Mobile Satellite Communications

Principles and Trends

Second Edition

Madhavendra Richharia

WILEY

MOBILE SATELLITE COMMUNICATIONS

MOBILE SATELLITE COMMUNICATIONS

PRINCIPLES AND TRENDS

Second Edition

Madhavendra Richharia
Knowledge Space Ltd., UK

WILEY

This edition first published 2014

Registered office
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom

For details of our global editorial offices, for customer services and for information about how to apply for permission to reuse the copyright material in this book please see our website at www.wiley.com.

Library of Congress Cataloging-in-Publication Data

Richharia, M. (Madhavendra)
Mobile satellite communications principles and trends / M. Richharia. – Second edition.
pages cm
Includes bibliographical references and index.
ISBN 978-1-119-99886-0 (cloth)
1. Artificial satellites in telecommunication. 2. Mobile communication systems. I. Title.
TK5104.R528 2011
621.3841′56–dc23
2013031745

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

ISBN: 9781119998860 (Hardback)
ePDF: 9781118810064
oBook: 9781118810170
epub: 9781118810149

Typeset in 10/12pt TimesLTStd by Laserwords Private Limited, Chennai, India

1 2014

To my parents

Contents

Preface

Much has happened in the arena of mobile satellite communications since the release of the first edition of the book in 2001. After the commercial failure of several operators in the early part of the first decade, the industry has revived and matured. A healthy growth has been forecasted for the next 10 years and the stage is set for the industry to meet this challenge. The long gestation period of 5–10 years in the preparation and uptake of mobile satellite service (MSS) products has been amply demonstrated over this period. We had introduced numerous products and technologies in the first edition. Such products introduced around the turn of the millennium are now being upgraded to the next generation and those introduced in the middle of the last decade have peaked. The next generation MSS systems are rolling out, as we speak.

Various technical, regulatory and commercial advancements have accrued since. Consider a few examples. The viability of non-geostationary orbit for MSS has been established, considerable progress has been made towards MSS radio interface standardization, satellite mobile broadcast technology has matured and standardized, the concept of satellite-terrestrial hybrid architecture has matured, the K_a MSS band has finally emerged as a viable MSS proposition, lower-end fixed satellite service (FSS) products now support mobility akin to MSS, dense spot-beam/high-power technology has matured and extraordinary advancements in terrestrial mobile communication technology have opened new technological vistas for the future.

Due to a rapid proliferation of systems and network architecture and the availability of a plethora of innovative applications, many telecommunication professionals and technical managers lose sight of an MSS perspective. A huge amount of technical and related information lies scattered in publications such as journals, magazines, conferences, expensive and lengthy specialist reference books, and the Internet. It is therefore difficult for individuals and companies to obtain coherent up-to-date technical information.

This book, written to bridge the information gap, compiles current system concepts, architecture and trends in a structured and easily understandable style. The book is also expected to serve as a reference source, as it adheres to technical concepts. Mathematics is limited to essentials, and where possible, equations are illustrated graphically; a comprehensive list of references has been included for the curious reader. Scientific principles have been amalgamated with business models for a balanced perspective of a commercial mobile satellite system. The treatment is unbiased and views expressed are the author's own; examples of commercial systems are chosen singularly on technical novelty and uniqueness.

Three chapters have been added to the previous edition and others are thoroughly revised. Operational matters, which were dealt in the system architecture chapter previously, have now been instated in a new chapter. A chapter dealing with radio interface standards and recommendations has been added to reflect the significant advances and exemplify amalgamation of new technologies into operational systems. A chapter dealing with mobile satellite broadcast technology has been included to capture the advances in this area; although the technology formally belongs to the broadcast regime this is a subject of interest to the MSS industry, since the technology and products are similar and there is a potential of including MSS for user-interactivity.

The book, comprising 14 chapters and an appendix, presents an in-depth review of concepts, practices and trends of the gamut of mobile satellite systems. Topics include: satellite constellations; propagation aspects peculiar to mobile communications including the emerging K_a band; applicable modulation and coding techniques and trends; design issues of hand-held units – including a review of biological effects of radio frequency radiation on humans; regenerative satellite transponder technology, intersatellite links and multi-beam antenna systems; business aspects of MSS, highlighting the intertwined relationship between technical and business aspects of the MSS; network optimization and vital elements of network operations; salient features of innovative operational systems; services with products similar to MSS such as satellite navigation receivers and mobile FSS terminals, highlighting an emerging inter-service paradigm; and innovative technologies such as multi-user detection, cognitive radio, multiple-input-multiple-output systems, and *ad-hoc* networks, amongst others, as potential solutions of the future. Useful data and mathematical formulas are included in an appendix.

The reader can visit the web site www.SatellitesAndYou.com for information and software pertaining to the book.

I hope the reader finds the book useful and enjoyable.

Madhavendra Richharia

About the Author

Madhavendra Richharia is a practising professional for well over three decades, with experience in diverse areas such as design and development of earth station products, satellite system design and operation, system planning, management, academics and research. He is currently a senior consultant and director of Knowledge Space Ltd, UK, prior to which he held senior technical positions at Inmarsat for over 15 years. Earlier, he was a faculty member at the University of Surrey. He has also contributed to satellite communication programmes of ISRO as an engineering scientist. He has published numerous technical papers on a variety of satellite communications topic, authored two books on satellite communication system design and co-authored a book covering concepts and technologies applicable to satellite systems for personal applications. The first edition of this book was well-received and continues to be recommended in many university courses. He was honored with the British Asian of the year award for technology in 2002 by the Guild of British Asians for his contributions to the field of satellite communications. He obtained his PhD degree from the University of Birmingham UK and Bachelor and Masters from the BHU. He is a member of the IEEE and the IET.

Acknowledgements

The author wishes to express gratitude to his former Inmarsat colleagues for encouragement to take up this project. Grateful thanks are extended to those individuals and organizations who have contributed to this compilation through their published work and by granting permission for their material to be included. The author thanks the reviewers for their valuable time and constructive critique which have enriched the quality of the book.

Every effort has been made to obtain permission from the copyright holders but in the event that any have been inadvertently overlooked they should contact the author in the first instance (via the publisher), who will endeavour to make appropriate copyright permission arrangements at the earliest opportunity.

The figures extracted from the ITU material have been reproduced with prior authorization of the Union as copyright holder. The sole responsibility for selecting extracts for reproduction lies with the beneficiary of this authorization alone and can in no way be attributed to the ITU. The complete volume(s) of the ITU material, from which the texts (figures) are extracted, can be obtained from:

International Telecommunication Union
Sales and Marketing Service
Place des Nations: CH- 1211 GENEVA 20 (Switzerland)
Telephone: +41 22 730 61 41 (English)/ +41 22 730 61 42 (French)/ +41 22 730 61 43 (Spanish)
Telex: 421 000 uit ch/Fax: +41 22 730 51 94
X.400: S = sales; P = itu; A = 400net; C = ch
E-mail: sales@itu.int
Web access: http://www.itu.int/publications

The author gratefully acknowledges the effort of the editorial and production staff of John Wiley & Sons, Ltd.

My son, Anshuman, deserves a very special word of appreciation for creating a number of graphics, and for his assistance in the preparation of the accompanying web site during his vacations from Imperial College, London. These figures have been acknowledged as (Graphics AR).

I thank my wife, Kalpana, for her invaluable support and understanding as I endeavoured to grasp every spare moment to complete the book. Finally, it is a pleasure to recall the enthusiastic support of my daughter, Meha throughout.

1

Introduction

1.1 Scope and Organization

This book aims to promote an in-depth understanding of mobile satellite communication system technology, a branch of telecommunications enriching the lives of millions of mobile and nomadic people globally. Its relevance to modern society is best demonstrated by a few real-life situations:

- Live satellite coverage from the Mount Everest summit camp brings a realism into people's homes never experienced before – a nation rejoices as the summit is conquered.
- A passenger ship cruising in high seas collides with a rock – the crew, in a desperate attempt for survival, send a satellite distress call; a rescue party arrives shortly.
- An old man's wizened face comes alive as he talks to his son from his thatched village home, tucked away, where time stands still.
- An executive, preparing for a meeting on a trans-oceanic flight, contacts head-quarters and receives vital documents within minutes.
- A distressed community in an aftermath of a natural disaster receives vital supplies aided by communications through satellite phones.

Mobile satellite service or MSS systems provide communications to terminals that may be in motion, or moved at will anywhere within the service area. Terminals may be mounted on an aeroplane, a ship or carried by individuals; alternatively, the terminal may be a portable communicator set up at a convenient location. The vital elements are unrestricted user mobility with minimum regulatory restrictions in the service area. Mobility is achieved by the use of a radio link for tetherless connectivity and incorporating network intelligence to manage mobility. Regulatory restrictions are minimized through appropriate spectrum selection and operating licenses. Service areas of an MSS can span a country, a region or indeed the world.

There was a steady and gradual growth in MSSs until early 1990, at which juncture demands began to accelerate as new services were introduced and public awareness heightened due to the success of terrestrial mobile systems. A further spurt in aggregate world-wide demand occurred when satellite phones were introduced in the third quarter of the 1990s, despite the business failure of some MSS operators who had primarily targeted

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

the hand-held sector. These failures were caused by differences in user expectations in terms of cost and quality, and operators' expectations in terms of the market size and rate of penetration dented further by advances in roaming arrangements of terrestrial operators that made the satellite service unattractive to the globetrotter. Since then, many systems have been reinstated – aided by financial support, enhanced public, governmental and commercial awareness about the strengths of satellite systems; with enhanced system capabilities; a variety of new applications; elegant user terminal (UT) designs; higher throughput and tighter integration with terrestrial systems.

As it happens in rapidly evolving technologies, collating cohesive information can be time consuming and difficult, as useful information lies buried and scattered in specialist reference books, journals, conference proceedings, trade journals, and so on. This book attempts to bridge the gap through a structured compilation of such knowledge to assist understanding of system architectures, their components, applications and trends, assisted by a comprehensive list of references.

This thoroughly revised edition supplements the concepts of the first edition with recent developments adding three chapters to cover new topics. The technology has evolved quite considerably since – consider a few examples:

- Auxiliary Terrestrial Component (ATC) technology has matured enabling seamless extension of satellite services into dense urban areas systems by terrestrial re-transmissions;
- Spacecraft antennas provide hundreds of spot beams enabling an order of magnitude higher EIRP (effective isotropic radiated power) with highly sensitive on-board receivers;
- User throughput has increased several-fold;
- MSS usage has extended to the K_a band;
- A number of new standards have emerged while others have evolved;
- There is much tighter integration with terrestrial systems and technologies;
- Various improvements have been made in the underpinning technologies such as air interface, radio resource management, UTs, etc.;
- Very small aperture technology (VSAT) – a fixed satellite service (FSS) technology with associated radio regulations (RR) have evolved to provide MSS-like solutions.

The International Telecommunications Union (ITU), a United Nations body that regulates world-wide allocation of radio spectrum, has for spectrum planning categorized radio services according to their broad application – broadcast satellite service (BSS), FSS or MSS, and so on. Personal satellite communications services such as voice, facsimile and multimedia services require a radically different system design with a commercial approach quite recent to the satellite community. The ITU has termed such services as Global Mobile Personal Communications Services (GMPCS). This book deals with system level technical issues of the MSS and GMPCS. We will refer to these services together as MSS, unless a specific distinction is essential.

This chapter introduces the subject beginning with a review of the evolution of mobile communications. A subsequent section presents basic concepts of MSS architecture to familiarize the reader with the topic early in the book. A plethora of telecommunications products often leaves users uncertain regarding the most suitable solution. To enable users and prospective operators to make informed decisions, salient features of satellite and terrestrial systems are compared, followed by an overview of applications typical of an MSS. Many satellite navigation, FSS and BSS products have entered the domain of mobile communications. The system features of such systems are summarized, and finally, emerging trends introduced.

The design and implementation aspects of satellite constellations were revisited extensively in the 1990s as an alternative to the well-established geostationary orbit. After introducing the basics of orbital mechanics for the benefit of readers unfamiliar with this rather specialized topic, Chapter 2 discusses the characteristics of various types of constellations and compares them for a number of well-known constellation designs.

Chapters 3 and 4 introduce the core technical concepts of mobile satellite communication systems. Some of these topics are applicable generally to satellite communication systems but the treatment here is slanted towards an MSS perspective.

The first part of Chapter 3 outlines the main features of ITU procedures for interference management with an example of spectrum forecast methodology. The second part covers MSS radio propagation characteristics and the associated system implications in the MSS frequency bands for each environment – land, maritime and aeronautical channels. The final part of the chapter addresses the fundamentals of radio link analysis.

Chapter 4 develops an understanding of the digital modulation and coding schemes prevalent and emerging in modern mobile satellite system, followed by an introduction to multiple access techniques.

Gateways provide a radio connection between ground and space segments. Chapter 5 highlights MSS-specific features of gateways, which are otherwise identical to medium earth stations of the FSS. However, the chapter mainly focusses on the characteristics and technology of mobile terminals, which have profound implications for the success of MSS operators. Considerable public interest has arisen regarding radio frequency radiation effects to humans; a section has been devoted to this topic where current state of understanding is summarized.

Satellites are undoubtedly the most vital node of an MSS. Chapter 6 highlights the main features of MSS satellite technology including on-board processing and reconfigurable satellite antennas.

Chapters 7 and 8 explore MSS in a system context by collating concepts discussed in previous chapters. Chapter 7 covers the core areas of network architecture. Topics include air interface concepts, system synthesis methodology taking into consideration external influences, constraints-tradeoffs; network concepts such as mobility management and intersatellite routing. A vast number of mobile satellite systems have been standardized and recommended in recent years necessitating the addition of Chapter 8; the chapter introduces various satellite air interface standards, ITU recommendations for satellite component of future mobile communication systems, and an interactive mobile satellite broadcast standard.

Chapter 9 addresses key operational aspects of MSS systems, covering topics such as subscriber and gateway commissioning, spectrum and EIRP management methods, traffic analysis, radio frequency monitoring, interference and quality-of-service management.

This book views the MSS in a commercial perspective, and in this context, business and technology are intricately entwined. MSSs have grown rapidly in the recent past – and yet, despite remarkable technical achievements, many MSS companies faced bankruptcy at the outset. Hence Chapter 10 changes the emphasis to MSS economics, illustrating its inter-relationship with technology. Topics include commercial aspects such as service distribution, billing, regulatory influences, traffic forecast from a commercial perspective and a case study from a maritime user's point of view.

A number of technically interesting systems have been proposed, each with some novel features and Chapter 11 discusses representative examples to demonstrate how concepts have been translated to practice. Some of these systems could not be realized for commercial reasons, but they are of immense technical interest. The choice of system examples are made purely on the basis of their technical variety and merit, with little bearing on their commercial performance or affiliations. Thus we cover geostationary satellite systems, LEO (low earth orbit), MEO (medium earth orbit) and hybrid-orbit systems, and a system based on auxiliary terrestrial transmission. Chapter 11 in conjunction with Chapter 8 represents the current state of the art in system architecture.

The technology for reception of satellite television broadcast on mobiles on modified direct broadcast satellite receivers is well established. However, the advent of satellite broadcast systems targeting reception of television and radio on personal and portable sets is relatively recent. Chapter 12 introduces the concepts and standards of direct-to-person radio, television and multimedia broadcasts. The chapter reviews various aspects of this transmission technology including system requirements and configuration, compliant space segment, transmission technology, receiver architecture and introduces salient aspects of the DVB-SH (Digital Video Broadcast-Satellite services to Handheld) standard, which is representative of the technology trends.

In Chapter 13 we discuss systems, which offer services akin to the MSS but are not formally a part of them. In this context, we begin the chapter with an introduction to the satellite component of Global Maritime Distress and Safety System (GMDSS), which constitutes an essential element to augment the safety of the shipping industry and provides communication support to individuals and aircrafts in distress. The embedded satellite navigation receiver is an established feature of modern mobile terminals providing a variety of network functions such as timing and frequency reference and applications like location-based information. The chapter introduces satellite navigation principles with examples of existing and planned navigation systems. It includes the principles of differential global positioning system (DGPS) and satellite based augmentation system (SBAS) that provide increased positional accuracy for a variety of mobile applications. In the past few years, FSS terminal sizes have shrunk to an extent that they are portable, and hence the distinction between fixed and mobile services has become blurred in applications where such terminals can be used, as for example on ships and aircraft. The chapter discusses the rationale, underlying issues and the technology of mobile very small aperture FSS UTs. A majority of upcoming and recent MSS systems are tightly integrated with terrestrial mobile systems implying that there is now a better appreciation of terrestrial cellular system by the satellite community and vice versa. In such an environment, the treatment of mobile satellite communication cannot be taken in isolation. We will therefore outline the salient physical layer aspects of terrestrial cellular systems in this chapter.

The final chapter discusses various emerging new techniques and concepts under investigation to promote the evolution of mobile satellite systems in the medium to long term. We introduce a few market projections to demonstrate the healthy growth of MSS together with a spectrum forecast, and present some capacity enhancement technologies, which offer the potential to enhance spectrum utilization efficiency. These include – multi-user detection technology, advanced frequency planning techniques, cross-layer radio resource optimization and cognitive radio technology. Next we address advanced system architectures under investigation – the role of MSS in 4G and heterogeneous *ad hoc* networks and hybrid network architectures. In the next part we introduce a few enabling concepts and technologies currently in the research domain. These include propagation in Q/V and higher frequency bands, aspects of modulation and coding, the concept of multiple input multiple output application to MSS and the application of software defined radio in future systems. It is anticipated that there will be a shortfall in capacity of existing aeronautical communication systems and hence there is intense interest in promoting the introduction of efficient communication technologies for aeronautical communication in the period beyond ~2015–2020. The role of satellite systems to meet the needs of the future aeronautical systems is addressed next. The chapter concludes on a speculative note presenting the concept of a utopian communication system where terrestrial, satellite and high altitude platform technologies are combined to provide a seamless global network.

Section A.1 provides coverage snapshots of a few representative non-geostationary satellite systems. Section A.2 provides a list of useful formulas for the system designer.

The book will be supported through the web site (www.satellitesandyou.com) with errata and updates.

1.2 Evolution of Mobile Telecommunications

For the purpose of this section, mobile communications systems are broadly categorized as terrestrial and satellite. In both, mobility is achieved by a RF link between the user and a relay station, which is connected to the fixed network that incorporates the mobility management functions. During the early phases of MSS, evolution of these two systems progressed independently. Terrestrial systems were best suited for urban environments, whereas satellites provided effective communications solutions for remote areas such as high seas, air corridors and remote land masses. By the beginning of 1990, MSS technology had matured to an extent that system planners began to evaluate benefits and techniques of integrating these two technologies, leading to the introduction of partially integrated systems by the end of the decade and full integration thereafter.

1.2.1 Terrestrial Systems

The potential of mobile communications was recognized from the outset of radio dating back to the late 1800s. Earliest use of mobile radio was for maritime navigation and safety.

Before and during World War II, mobile systems were generally confined to military users. Vacuum tubes and heavy batteries in transceivers (i.e. a transmitter-receiver unit) made them bulky, restricting their use to specialized applications. By the 1950s technology advancements enabled availability of man-pack very high frequency/frequency modulation (VHF/FM) radios, which extended the applicability of mobile systems to civil Private Mobile Radio (PMR). The growth of mobile radio was slow until the 1970s, because the problem of extending service range within the permitted frequency bands proved technically difficult. In general, mobile communication service remained an expensive communication medium. But even if the cost of mobile services had reduced significantly, the factor that hampered growth of mobile services was spectrum scarcity. A concept called 'cellular radio', which offered a solution to support large numbers of users in a limited spectrum, was proposed at Bell Laboratories in 1940s. However, enabling technologies to realize the concept would arrive much later, in the 1980s.

Mobile transceivers were permitted access to the public network of the United States by AT&T in 1946. The service was called Mobile Telephone Service (MTS), which operated in the 35 or 150 MHz band. The simple system was manual, where the user seized a vacant channel requesting an operator for a connection to it.

An improved MTS (IMTS) that offered duplex operation, automatic dialling and switching was launched in 1964. The system capacity was initially limited to 11 channels in the 152–158 MHz band. Due to a rise in demand, the capacity was enhanced in 1969 by 12 channels in the 454–459 MHz band. Each service zone was served by a single base station, which used a channel only once and hence the demand could not be satisfied; typically, around 550 users could be served in a zone. Other service limitations included bulky transceivers that had to be vehicle mounted, and the need for high-capacity batteries.

Introduction of the cellular concept provided the desired breakthrough in overcoming the spectrum limitations. The first cellular system known as the Advanced Mobile Phone System (AMPS) was developed in the USA by AT&T and Motorola, Inc. The analogue system was designed to operate in the 800 MHz band with a capacity of 666 paired channels.

However, the first *commercial* cellular system was deployed in Japan in 1979, followed by Nordic Mobile Telephone (NMT) system introduced in 1981 in Denmark, Finland, Norway and Sweden, the Total Access Communication System (TACS) in the United Kingdom and the AMPS system in 1983 in the USA. Yet other cellular systems were developed and introduced in other countries.

Rapid developments were made in radio, very large-scale integration (VLSI) and computer technologies, giving a huge reduction in costs through economies-of-scale and consequent increases in personal communications. Users have since been inundated with a variety of cellular phones, systems and services. Figure 1.1(a) illustrates the substantial world-wide penetration trend of the past decade for the world and by level of development. The number of mobile subscriptions increased from ~50 per 100 inhabitants in 2001 to 122.3 in 2011 in the developed world and ~18 to 85.7 across the world during the same period. Mobile broadband is a relatively new introduction. Figure 1.1(b) shows the penetration per 100 inhabitants of active mobile-broadband subscriptions for the period 2007–2011 for the world and by level of development. In this case, the penetration increased from ~18 to 51.3 per 100 inhabitants in the developed world and the corresponding worldwide growth increased from ~3 to 15.7 per 100 inhabitants.

A consequence of the runaway commercial success of cellular phones was the proliferation of a variety of systems incompatible to each other, causing inter-operability problems so that users could not operate their cellular phones outside a home territory. Concentrated efforts to harmonize evolution lead to second generation digital standards such as the Global System for Mobiles (GSMs) in Europe, which gave a cross-country roaming facility.

Despite high spectral efficiencies offered by cellular systems, capacity shortfall remained a pressing issue, and efforts to enhance capacity continued. In the United States, carrier spacing of AMPS system was reduced from 30 to 10 kHz, trebling the capacity. The second approach was to introduce digital voice compression with time-division multiple access, which resulted in a similar order of capacity enhancement. Spread spectrum modulation with code division multiple access (CDMA) was yet another approach. Its proponents claimed

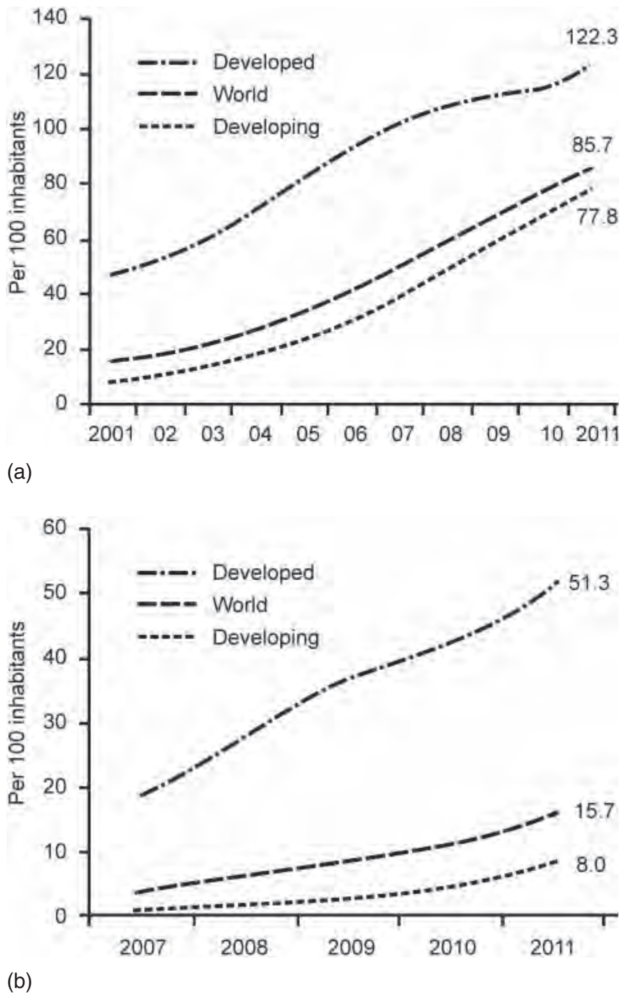


Figure 1.1 (a) Mobile-cellular subscription penetration during the period 2001–2011 for the world and by level of development. (b) Active mobile-broadband subscriptions penetration during the period 2007–2011 for the world and by level of development. (Both parts source: ITU World Telecommunication/ICT Indicators database. Reproduced with permission of ITU.)

a 10–20-fold capacity increase over the AMPS system. Subsequently personal communication system (PCS) systems were based around evolved GSM technology and a spread spectrum technology called IS-136 or D-AMPS. Satellite air interface standards based on GSM were introduced leading to a tighter integration of satellite and terrestrial systems.

Third generation cellular systems offering wideband were introduced in Europe and elsewhere to bring further homogeneity in mobile communications across the world, thereby offering benefits to users, manufacturers and operators. Satellite systems have since been integrated with the third generation terrestrial core network to offer seamless connectivity.

1.2.2 Satellite Systems

Table 1.1 summarizes some interesting milestones in the MSS evolution.

The earliest mobile communication experiments were conducted through National Association of Space Administration's (NASA) Applications Test Satellites ATS-5 and 6. A number of early MSS proposals never took off due to the perceived technical and financial risks. The potential of satellite systems to provide high reliability communications to ships was recognized by International Maritime Organization (IMO). The prevailing communication systems used high frequency (HF) band and proved unreliable under adverse propagation conditions, often requiring hours to establish communications from high seas when weather conditions were unfavourable. Ships on the high sea would occasionally be lost without trace. Satellite communications radio links are reliable under most conditions. Recognizing the advantages offered by a satellite medium, the IMO initiated the formation of an organization, then called, International Maritime Satellite Organization (Inmarsat) for provision of safety and commercial public correspondence services to ships for peaceful purpose. Inmarsat was founded in 1979 and the maritime communications services became available in 1982 using satellites leased from a number of satellite operators. The success of maritime services led to the introduction of services in land and aeronautical environments and generally vast enhancements in MSS technology. The size of terminals reduced progressively with technology evolution from heavy transportable terminals to desktop-sized personal telephones, briefcase-sized multimedia terminals, hand-held phones; high power satellites with spot beams were introduced while throughput continued to increase. The company was then renamed as International Mobile Satellite Organization to reflect its expansive portfolio, retaining the same acronym.

In the 1990s, a number of regional systems were introduced in the USA, Europe, Australia and Japan. For example, Omnitracs Inc. and EutelTRACS provided low bit rate services in the USA and Europe respectively; the American Mobile Satellite Consortium (AMSC) began voice and data services in the USA and Canada; and the Optus system was introduced for voice and data services in Australia.

Studies on non-geostationary orbits (NGSOs) were first conducted in 1960s when it was difficult to launch satellites to the GEO (geostationary earth orbit) and therefore lower orbits were considered appropriate. Subsequently, GEO became a favoured choice, due to advantages arising from static path geometry of geostationary satellites. Non-geostationary earth orbits (NGEOs) remained in use for specialized applications such as remote sensing, military reconnaissance, Earth resource survey, etc. The military were interested in the survivability and robustness offered by low orbit satellite systems due to their inherently distributed architecture (Chakraborty, 1989). Towards the end of 1980s a university group in the UK studied the feasibility of deploying a LEO satellite constellation for mobile communications with the conclusion that such systems were indeed feasible, compare favourably with geostationary satellite systems and could be implemented within a decade (Richharia *et al*., 1989). Within a year, quite independently, Motorola Inc. announced plans of a LEO satellite system to provide personal voice communication service via hand-held phones (Nelson, 1998). The announcement triggered a feverish activity in the satellite industry and within the next three years a number of organizations and companies announced similar plans, most of them intending to use non-geostationary constellations. Notable amongst them was a MEO system, which was conceived by Inmarsat but was to be implemented by a new private company

Table 1.1 Interesting events in the evolution of mobile satellite systems

Year	Milestone	Responsible
1957	Launch of first satellite	Former Soviet Union
1965	Launch of first geostationary satellite	Comsat, USA
1976	First demonstration of intersatellite link	LES-8 satellite, USA
1977	Mobile experiments using ATS-6	NASA, USA
1978	Introduction of global positioning system (GPS)	US government for the Department of Defense
1979	Inmarsat formation	International Maritime Organization
1980	IMO decides to deploy satellite communications for maritime safety.	UN
1982	Introduction of GLONASS	Former Soviet Union
1982	First civilian mobile satellite system introduced	Inmarsat
1987–1989	An architecture of LEO for mobile satellite communication investigated and proposed	University of Surrey
1990	First commercial satellite radio broadcast system filed	CD Radio Inc., USA
1990	First commercial non-GEO hand-held system announced	Motorola/Iridium
1990–1991	Commercial land and aeronautical mobile satellite services (MSSs) introduced	Inmarsat
1992	GSM system introduced	Europe
1992	Major changes to mobile satellite frequency allocation	WARC 1992
1993	Announcement of first commercial little-LEO satellite system (with secure finance)	Orbital Sciences Corporation – ORBCOMM system; USA
1994	First non-GEO FSS for personal communications announced	Teledesic Corporation; USA
1994–1996	Announcement of several regional 'super-geostationary' satellite systems	Agrani (Indian Consortium) ; APMT (China/Thailand); ACes; Thuraya, etc.
1996	Paging services introduced	Inmarsat
1997	Desktop-sized mobile terminals introduced	Inmarsat
1997	First non-geostationary little-LEO satellite system introduced	ORBCOMM
1997	Frequency allocation for non-GEO fixed system	WRC 1997
1997	Launch of first batch of LEO satellite system for voice communications (so called 'big-LEO')	Iridium
1997	Launch of first batch of non-geostationary satellite system for low bit rate data communications ('little-LEO')	ORBCOMM
1997	Mobile experiments using ACTS	NASA
1997	Navigation system: geostationary overlay capability available	Inmarsat

(*continued overleaf*)

Table 1.1 (*continued*)

Year	Milestone	Responsible
1997–1998	Start of world-wide spot beam operation for MSS	Inmarsat
1998	Introduction of first big-LEO satellite system	Iridium
1998	Introduction of dual-mode satellite-terrestrial handsets (i.e. combined satellite and terrestrial handset)	Iridium
1998	Safety Of Life At Sea (SOLAS) treaty introduced	UN
1998	Introduction of on-board processing satellites for MSS	Iridium
1999–2000	Serious financial difficulty experienced by new and proposed NGSO MSS systems	Iridium
	Introduction of Globalstar	Globalstar
2000–2005	Consolidation of new mobile satellite system operators despite financial losses	Various
2005	Introduction of wide-band portable land mobile communication system	Inmarsat
2006–2008	Extension of portable broadband system to mobile platforms	Inmarsat
2009–2010	Announcement of next generation systems	Inmarsat, Iridium, Globalstar, ORBCOMM
2009	LightSquared proposes ATC services in USA	LightSquared
2012	ATC service license suspended in USA due to interference issues to GPS	LightSquared
2014	Introduction of upper-end MSS broadband (up to ~50 Mbps) in K_a band	Inmarsat

ACTS, Advanced Communications Technologies and Services.

called ICO (Intermediate Circular Orbit) Global Systems Limited (referred in the remaining text as the ICO system); and a LEO system known as Globalstar.

Another approach was to use low-risk technology and smaller spacecraft for low bit rate niche applications such as messaging and machine-to-machine communication ensuring an early entry in to the market as proposed, for example by companies such as Orbital Science System.

In the mid-1990s, plans for hand-held satellite phone services using GEOs were announced. These systems planned to deploy several hundreds of spot beams with powerful transmitters to compensate for the relatively higher altitudes of a geostationary orbit. Interestingly, only a few years ago geostationary systems were discarded due to the perceived complexity of payload; intense research and development in the ensuing four to five years had matured the technology. Close on its heels, a number of PCSs for fixed services were announced – converging high bit rate service offerings of the MSS with the lower end FSS products. Most FSS systems intended to operate in 20–30 GHz FSS band and in a variety of orbits. Table 1.2 (Adapted from Evans, 1998) compares the salient technical features of various MSS non-geostationary satellite proposals for hand-held voice services with an equivalent regional geostationary satellite system of that period. The geostationary satellite system compensates for the additional path loss compared to MEO/LEO systems by deploying highly directive spot beams and transmitting high EIRP. Table 1.3 presents a comparison of LEO, MEO and GEO MSS systems. LEO systems have

Table 1.2 A comparison of main technical parameters of some existing and proposed MSS proposals of 1990s

System parameter	System			
	Iridium	Globalstar	ICO global communication (original proposal)	Typical regional system
Service (voice bit rate)	Hand-held telephony/ data (4.8 kbps)	Hand-held telephony/ data (up to 9 kbps, depending on channel conditions)	Hand-held telephony/ data (4.8 kbps)	Hand-held telephony/ data (2.4–4.8 kbps)
Type of orbit	Low earth polar orbit	Low earth inclined orbit	Medium earth orbit	Geostationary
Number of spot beams/ satellites	48/66	16/48	163/10	100–300/1
Nominal capacity per satellite	1 100	2 400	4 500	16 000
Service area	Global	Global	Global	Regional
Service link frequency (up/down) in gigahertz	1.616–1.6265/ 1.616–1.6265	1.62–1.63/ 2.48–2.5	1.98–2.01/ 2.17–2.2	(1.6/1.5) MSS band
Feeder link frequency band (up/down) in gigahertz	30/20	5.1/6.9	5.2/6.9	14/12
Gateway antenna G/T (dB/K)	24.5	28.5	26.6	37.0
Multiple access	FDM/TDMA	FDM/CDMA	FDM/TDMA	FDM/TDMA
RF bandwidth per channel (kHz)	31.5	1 250	25.2	5–10
Modulation	DQPSK	SS/QPSK	QPSK	QPSK
User terminal RF power (W)	0.45	0.5	0.625	0.5
User terminal G/T (dB/K)	−23	−22	−23.8	−23.8
Typical service link margin (dB)	16.5	11	10	10

QPSK, quadrature phase shift keying; SS, Spread Spectrum. (Adapted from Evans, 1998.)

Table 1.3 A comparison of system features using various types of orbit

	Geostationary orbit	Medium earth orbit	Low earth orbit
Number of satellites (world-wide coverage)	3–4	10–12	40–300
Regional coverage	Well suited	Specific orbital design necessary (depends on region)	Specific orbital design necessary (depends on region)
Coverage limitations	Within ±76° latitude	None	None
Approximate satellite lifetime – first generation non-GEO/second generation GEO (yr)	15	7–10	5–7
Operational cost and complexity	Low	Medium-high	High
Transmission delay (ms)	250	55–80	3.5–15

the lowest transmission delay, provide true global coverage but the space segment and the network is complex. MEO systems have intermediate transmission delay provide true global coverage with a moderately complex space segment and network. GEO systems exhibit the highest transmission delay with coverage limited to ∼±76° latitude; however, the network is simple and the system is supported by matured technology.

Most satellite system designers, noting the benefits of integrating satellite systems with terrestrial systems, proposed system architectures combining satellite and terrestrial system networks to various extents. Dual-mode satellite/cellular hand-held telephones were conceived.

As mentioned, the trend in space segment architecture diverged at the beginning of 1990 when a number of commercial non-geostationary satellite systems were proposed for hand-held voice and data communication services on the premises that hand-held services via geostationary satellites would require extremely complex spacecraft and suffer transmission delays. Their architecture varied widely in terms of orbit, complexity, transmission schemes, network routing and the addressed market.

Further, we observed that by mid-1990 there was a re-emergence of geostationary systems based on powerful satellites deploying several hundreds of spot beams.

International organizations, developing standards and formalizing concepts for third generation terrestrial systems, began considering role of satellite systems in future system architectures. It was well understood that cellular systems would prevail in populated areas where they are economically viable, leaving gaps in vast, sparsely populated areas best be covered by satellite systems. Thus, future systems would incorporate interfaces and interworking arrangements to support interoperability between terrestrial and satellite systems.

In parallel, an unabated growth in *satellite navigation* technology for mobile and personal use continued throughout the 1990s. The cost of GPS (global positioning system) receivers plummeted to levels affordable by the masses. A number of applications developed, combining navigation and communication capabilities. Notable progress was made

Table 1.4 Mobile satellite communications evolutionary trends

Phase	Year	Service	Satellite technology/ frequency band	Orbit
I	1980–1990	Analogue voice and 64 kbps data on 1 m dish terminals	Low power, single beam/L	GEO
II	1990–2000	Voice and low rate data on hand-held terminals; 64 kbps data on 0.8–1 m dish terminals	GEO: medium power, 1–8 spot beams/L NGEO: medium power, up to 48 spot beams, on-board processing/L and S	GEO and non-GEO
III	2000–2010	GEO: first generation hand-held; broadband up to 500 kbps on portable and mobile terminals NGEO: second generation hand-held, a wider application portfolio, first generation maritime and aeronautical products	GEO: high power, up to 200 spot beams, on-board processing/L NGEO: Enhanced first generation products and second generation proposals	GEO and non-GEO
IV	2010–2020	GEO: second generation hand-held Wider application portfolio, broadband up to 55 Mbps on large/portable K_a band terminals and 1 Mbps on small terminals NGEO: third generation hand-held; second generation products for maritime and aeronautical applications	GEO: K_a band high power satellites with multi-beam spot beams NGEO: second generation, high power, multi-spot beams, longer life time	GEO and NGEO

towards the introduction of *satellite radio service*, a service aimed to broadcast directly to receivers carried by individuals and such systems became available in the first decade the new millennium. The principles of these related systems are discussed in Chapter 13.

We may classify satellite system evolution on the basis of best available services and technology into a number of phases as shown in Table 1.4.

1.3 Satellite System Architecture

The main components of a mobile satellite system are shown in Figure 1.2.

The system offers communication services to a variety of mobile users within a predefined service area under the control of a network control centre (NCC). Users communicate with

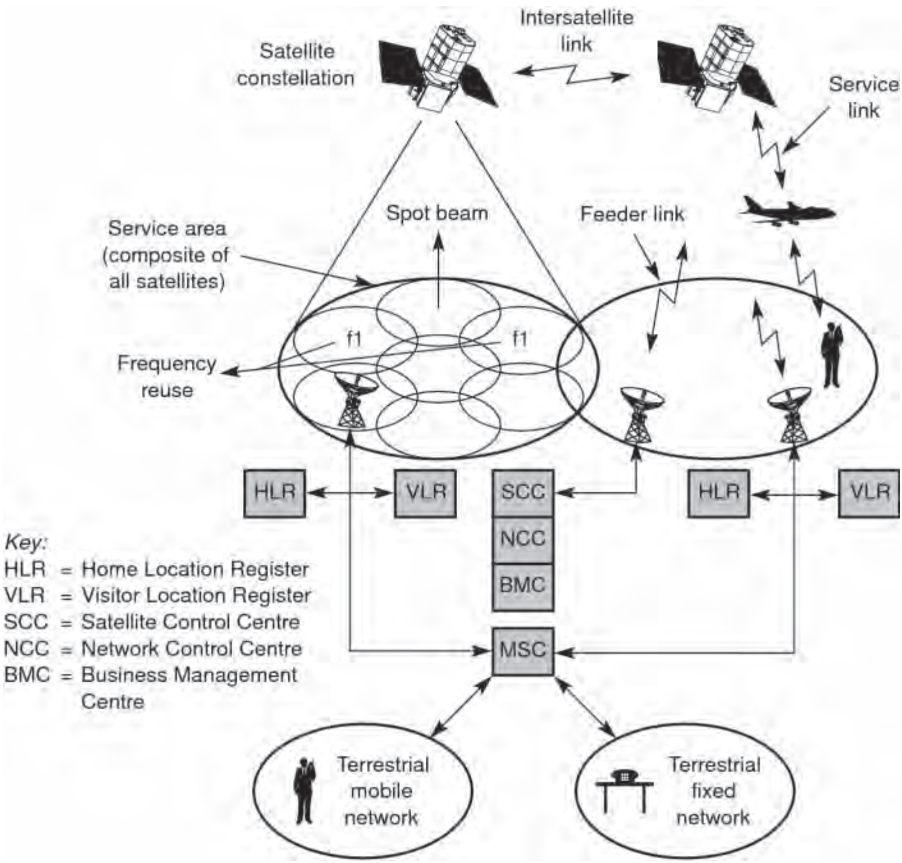


Figure 1.2 Main components of a mobile satellite system (MSC = Mobile Switching Centre)

other mobiles or with fixed users through the visible satellite. Users in the fixed network are accessed through large fixed stations called *gateways*, which must carry aggregated volumes of traffic, whereas mobiles are small portable units capable of supporting one to five channels (or the supported data rate). UTs' size and capability range from portable or handheld personal terminals, to units mounted on ships, aircraft and trucks capable of supporting several channels simultaneously. Services include paging, telephony and medium to broadband data. The fixed stations use very sensitive receivers and operate with stable radio links, whereas mobile receivers are small in size with limited sensitivity, and power received at mobiles undergoes wide fluctuations due to variations in a mobile's path profile. Hence the capacity of an MSS is governed by the power and bandwidth constraints of the service link.

Depending on the service area and application, the *space segment* consists of one or more GEO or NGEO satellites. Telemetry and control ground stations, used for monitoring and controlling satellites, constitute a part of the space segment. A satellite control centre (SCC) manages the functioning and maintenance of the satellites. To simplify the mobile terminals, complexity is shifted to the space segment and hence satellites tend to be large and complex with existing technologies typically capable of generating up to 12 kW DC power, EIRP in excess of 67 dB W in L-band with > 200 spot beams for efficient frequency reuse. A geostationary satellite remains almost fixed with respect to the Earth illuminating the service area with a *static* footprint that enables a simple network topology, whereas, NGEO space segments and network topology are complex due to non-stationary satellites.

In a centralized system, a NCC manages the traffic flow, call set-up and release, radio resources and signalling. On receiving a call request, the radio resource manager assigns the desired radio resource for the duration of the call. The home location register (HLR) and visitor location register (VLR) are responsible for mobility management. The mobile switching centre (MSC) manages call switching between the fixed network and the mobile network. A business management centre (BMC) is responsible for billing, new activations, customer support, and other business-related functions.

In an NGEO system, a call (or a data session) in progress must be re-routed to maintain connectivity as the connected satellite or spot beam moves out of visibility necessitating *handover(s)* to a rising satellite or spot beam.

A call may be routed to the final destination through intersatellite links, or via ground-satellite hops, or terrestrially. Figure 1.3(a) and (b) respectively illustrate source-destination connectivity in an NGEO satellite system deploying a terrestrial backbone and intersatellite links. The Iridium system deploys intersatellite links whereas the ICO system uses ground routing.

For non-interactive or store and forward services, a discontinuous coverage is acceptable. The message is stored in a satellite or ground station buffer and delivered to the destination via single or multiple satellite hops within a specified time limit. Figure 1.4(a) and (b) represents main entities of store and forward system with satellite and ground-based buffers, respectively. In Figure 1.4(a), the message M received at a satellite at $t = 0$, is transmitted to the destination at time t_2. In Figure 1.4(b), transmissions from a mobile van, received at a fixed site 'A' is stored and retransmitted when a satellite mutually visible with the destination N becomes available.

User data rate is commonly used to categorize MSS systems, as illustrated in Table 1.5 with example applications.

Although service requirements are the key enablers in developing the architecture of a MSS network, the outcome is often influenced by numerous constraints that include non-technical matters such as regional priorities and business goals, which lie beyond the scope of the book. Chapters 2 and 7 discuss the subject of network synthesis in detail.

1.3.1 Radio Frequency Environment

The constraints imposed by the service link impose significant limitations to the system capacity:

- *Forward link*
  - *Satellite EIRP and user's receiver sensitivity limitations*: The EIRP of a satellite, given as the product of available RF power and antenna gain, and the sensitivity of a UT set a bound to the UT throughput. The highest available EIRP cannot fully offset the link impairments and the low sensitivity of mobile terminals. Countermeasures include – transmissions in a robust format, increasing receiver sensitivity (traded

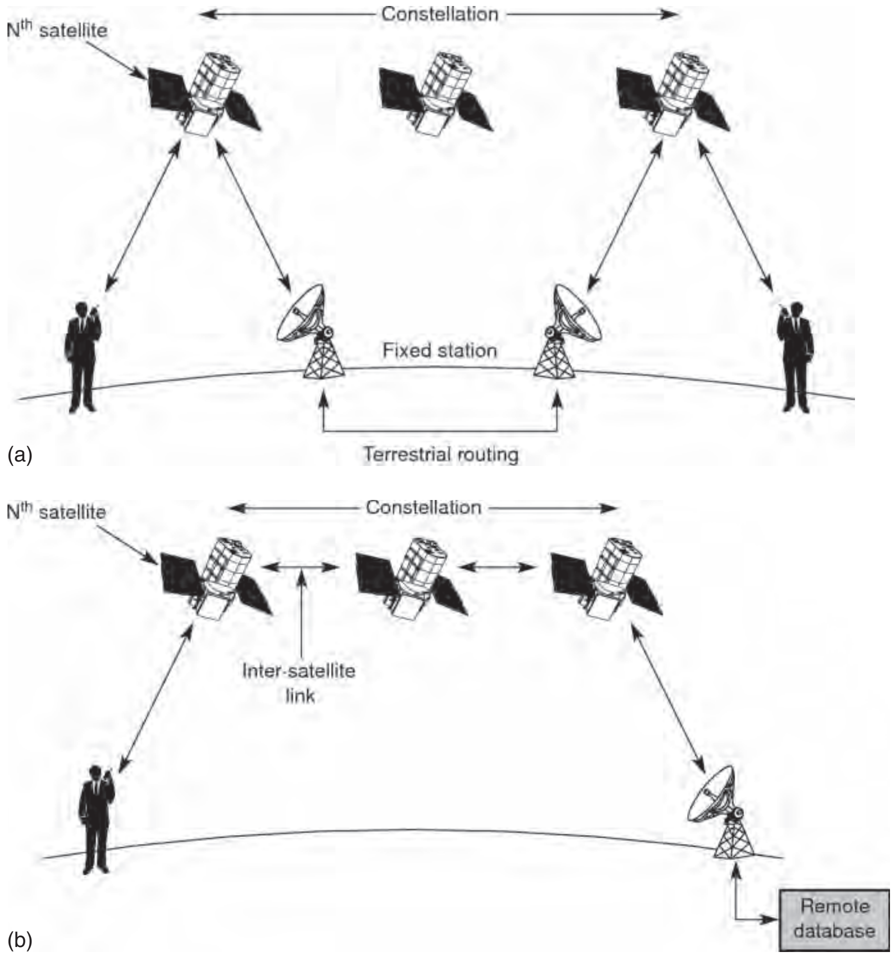


Figure 1.3 Connectivity architecture of non-geostationary satellite systems using (a) satellite-ground hops and (b) intersatellite links

with terminal weight, size and cost) and augmentation of satellite transmissions by terrestrial retransmissions.
  - *Propagation impairments*: Due to a non-stationary profile of the service link and limited multipath rejection capability of receiver antenna, the signal received at a UT is unstable necessitating a robust but rather inefficient radio link design (see Figure 1.5).
- *Return link*
  - *UT EIRP and satellite receiver sensitivity restrictions*: UT EIRP and satellite receiver sensitivity set a bound on the return link throughput (i.e. at a fixed station). The EIRP is constrained by antenna size restrictions due to space, weight and cost restrictions on portable or mobile UTs; for hand-held units EIRP is limited in compliance to radiation safety standards. Satellite receiver sensitivity is limited by the receive antenna gain of a satellite.

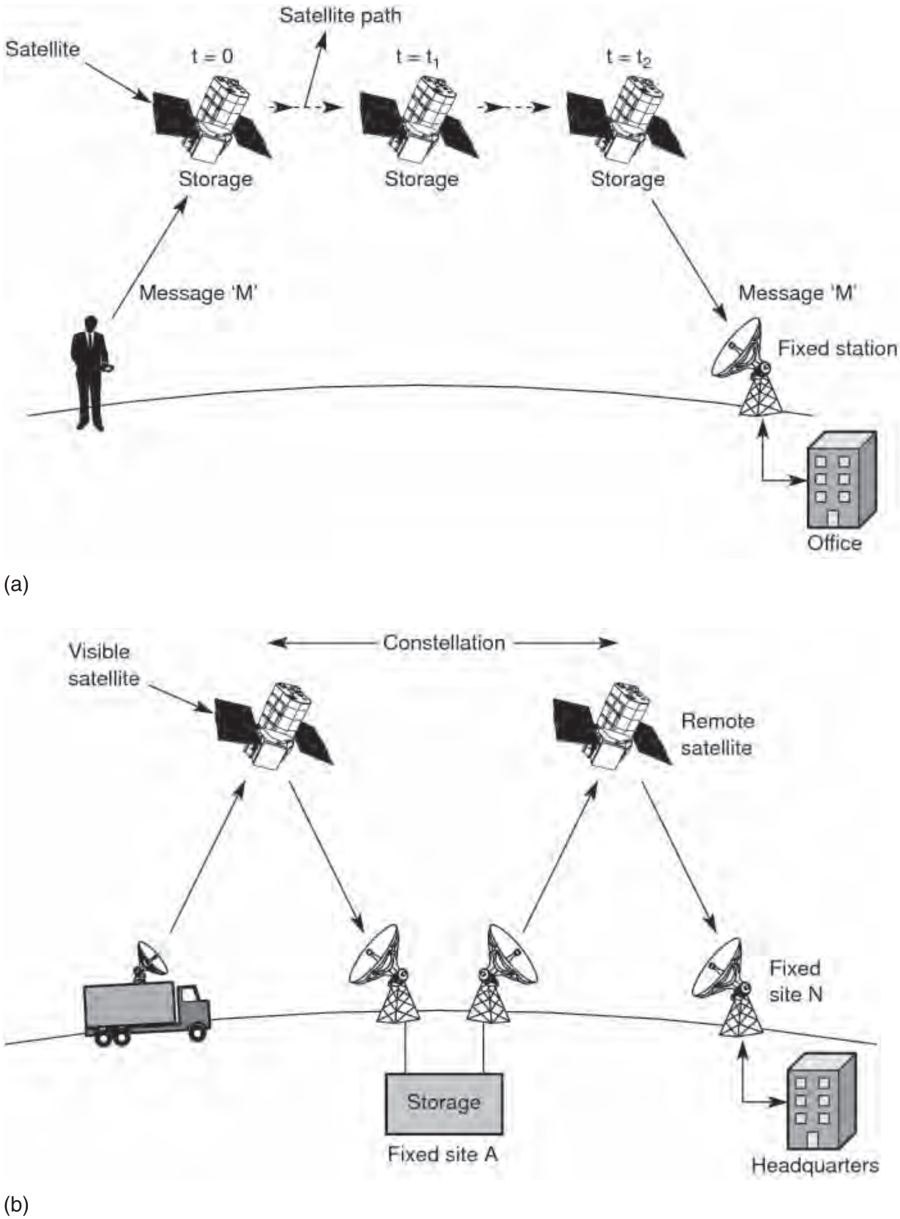


Figure 1.4 Main entities of store and forward system with (a) satellite buffer and (b) ground station buffer

Table 1.5 Categorization of mobile satellite services by throughput with example applications

Category	Throughput (forward link)	Typical applications	System example
Basic	A few bits per second to 1 kbps	Paging, message transfer, machine-machine communication, SCADA, remote monitoring	EutelTRACS (GEO); ORBCOMM (LEO)
Low bit rate	1–10 kbps	Voice, facsimile, e-mail, basic Internet	Globalstar (LEO); Iridium (LEO);
Medium bit rate	10–100 kbps	Voice, facsimile, e-mail, Internet, multimedia	Inmarsat (GEO), Thuraya (GEO)
Basic broadband	100–3 000 kbps	Internet, multimedia	Inmarsat (GEO), Thuraya (GEO), Iridium Next (LEO)
Broadband	3–10 Mbps	Terrestrial quality Internet and multimedia	–
Upper-end broadband	10–50 Mbps	Terrestrial quality fast Internet and multimedia	Inmarsat (GEO) (2014 start)

SCADA, supervisory control and data acquisition.

Figure 1.5(a) and (b) represents typical samples of L-band signals received on hand-held and stationary van-mounted terminals respectively demonstrating large receive signal variations (Jahn *et al.*, 1995). Referring to the figure we observe that signals received by the hand-held unit generally undergo peak-peak fluctuations in excess of 10 dB in 10 second segments, with a gradual variation of the mean level, while for the vehicle-mounted system the fluctuations at the most intense periods (∼ 50–100, ∼ 190–225 and ∼ 260–300 seconds) approach similar levels with stable but gradual change in the mean value in between.

1.3.2 Orbit

Orbits are categorized by altitude, inclination and eccentricity. The altitude determines a satellite's field of view (coverage) – higher altitude satellites can cover a larger area; Inclination influences the minimum-maximum latitudes covered on the Earth; the two extremes of the of orbital inclination are an equatorial orbit, which has an inclination of 0° and a polar orbit, which has an inclination of 90°. An equatorial orbit satellite would cover a belt around the equator, whereas a polar satellite would cover a belt orthogonal to the equator thus covering the full Earth over a period due to the Earth's west-east rotation beneath. Eccentricity of an orbit determines its shape. Satellites in a circular orbit travel at a uniform velocity to provide an unbiased temporal coverage, whereas satellites in an elliptical orbit travel at variable velocities and hence dwell longer over specific areas (see Chapter 2).

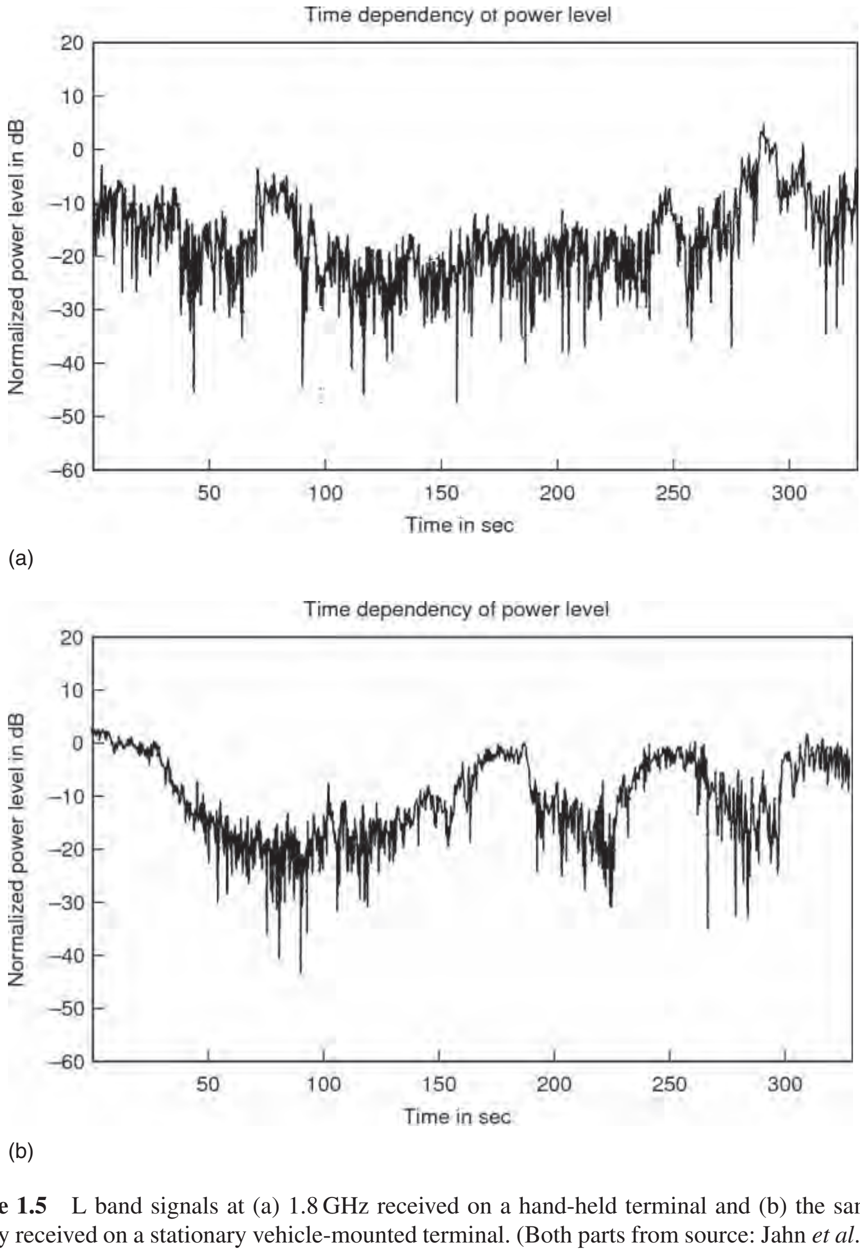


Figure 1.5 L band signals at (a) 1.8 GHz received on a hand-held terminal and (b) the same frequency received on a stationary vehicle-mounted terminal. (Both parts from source: Jahn *et al*., 1995. Reproduced with permission of Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

1.3.3 Tolerable Delay in Data Delivery

Delay tolerated by a system is application dependent. An end-end delay of more than ~ 400 ms is quite disturbing in a conversation, but delays of minutes and hours are acceptable for e-mail delivery. A non-real-time service can tolerate a break in communication link by data recovery techniques, whereas in an interactive service end-to-end connection must be maintained. The tolerable time delay influences several features of a mobile satellite communication system. For example, in a delay-tolerant system an NGEO satellite system can scale down the constellation size such that visibility statistics is restricted to the tolerable delay.

1.3.4 Handover

For a real-time geostationary system, ignoring RF link blockage due to obstructions, satellite location is static. The boundaries of spot beams are fuzzy, extending several tens of kilometres, which leads to a graceful degradation in signal quality for slow-moving vehicles. Thus, slow-moving mobiles need not handover a call to the next beam or satellite, as users spend a considerable period in this fuzzy zone where signals degrade gradually. However, handover is necessary for fast moving mobiles such as aircraft.

In non-geostationary satellite systems on the other hand, satellites are non-stationary and hence a user may have to communicate through different beams or satellites during a call, making handover a necessity.

Moreover, when intersatellite links are used handover between satellites becomes essential as path geometry is dynamic.

1.3.5 Mobility Management

The function of mobility management is to locate a called mobile, route the call efficiently and once established maintain the call, meeting the quality of service criteria.

In derivatives of the GSM system, each mobile is registered in a database, called home location register; if the mobile migrates outside the home territory, the mobile registers itself with the visiting system's VLR. The VLR conveys the location of each visitor to its HLR. Whenever a call is addressed to a mobile, the MSC interrogates the mobile's HLR to obtain the mobile's location and then establishes the call through an appropriate route. Figure 1.3(a) portrays a routing scheme in a NGSO system with satellite-ground hops, while Figure 1.3(b) represents an intersatellite link routing scheme. Various alternative routing schemes are possible and hence the routing strategy must be chosen carefully.

In an IP enabled network prior to a connection handover at the RF (physical layer) the IP address must also migrate to a new attachment. The Internet Engineering Task Force (IETF) have recommended mobility management schemes such as Mobile Internet Protocol version 6 (MIPv6), which have been adapted for satellite IP networks.

We have already introduced the concept of handover for maintaining RF link connectivity during a call in the previous section. Chapter 7 revisits the topic of mobility management in more detail.

1.3.6 Physical Environment

In a mobile satellite system, the physical medium in the vicinity of a mobile terminal and the receiver's antenna characteristics set a boundary on throughput. An area surrounded by obstructions would reduce the received signal power and hence the throughput; similarly, a reduction in antenna gain would reduce receiver sensitivity and hence throughput.

The behaviour of the received radio signals depends on the local surroundings (see Figure 1.5(a) and (b)) and receiver's antenna pattern. Low gain antennas tend to pick up multi-path signals from the surroundings, and furthermore, reduce receiver sensitivity. The size and hence gain of the mobile antennas depend on the available mounting space; for example a hand-held terminal would use a low-gain omnidirectional antenna whereas a ship-borne terminal may deploy a high gain 80 cm parabolic dish resulting in a more sensitive terminal and additionally, capable of transmitting higher power.

Figure 1.6(a–c) portrays various types of mobile terminals. Figure 1.6(a) is a mobile terminal, Figure 1.6(b) a portable wideband terminal and Figure 1.6(c) shows a satellite phone.

Impairments for narrow-band communication up to $\sim 100\,\text{kHz}$ are relatively benign in maritime and aeronautical channels at the L-band. Aeronautical channels tend to become dispersive beyond $\sim 100\,\text{kHz}$ when traversing over a quiet sea. Impairments are significantly high in land channels, tending to get worse in shadowed areas and with reduced antenna gain. Intermittent long and deep signal fades can break radio links causing discontinuity to real-time services while the rapid multipath fluctuations manifests as extraneous noise. Countermeasures include provisioning a higher link margin and robust modulation and coding.

The Doppler effect is introduced by relative motion between satellite and mobile; it affects aeronautical channels and NGEO satellite systems in particular. Countermeasure includes open and closed loop frequency correction arrangements.

Chapters 3 and 5 discuss these topics in detail.

1.3.7 Satellite Access

In an MSS environment thousands of users share satellite resources and therefore high satellite access efficiency is paramount. Demand assigned single channel per carrier (SCPC) frequency division multiple accessing (FDMA) or time division multiple accessing (TDMA) schemes, where a pool of channels is shared by all users on a per call or packet(s) basis, offer an effective solution. The channel pool can be managed by either a central or a distributed architecture. In a central architecture, a pool is managed centrally, whereas in a distributed architecture, separate pools are assigned to each participating fixed station for self-management.

In a CDMA scheme an RF channel is shared by all the users each using a unique code. This scheme offers advantage in terms of interference and multipath mitigation, and soft handover.

Data traffic tends to exhibit a variety of characteristics – ranging from sporadic bursts to continuous streams and therefore the accessing schemes are matched specifically to traffic characteristics. Common accessing schemes used for data communications include Aloha, slotted Aloha, Reservation Aloha and Time Division Multiple Access (see Chapter 4).

(a)

(b)

(c)

Figure 1.6 (a) A typical large ship-borne earth station configuration – antennas enclosed in radome above deck (left) and an application in progress below deck (right). (b) A typical portable terminal in use in a remote area. (Parts (a) and (b) source: www.Inmarsat.com. Reproduced with permission of Inmarsat.) (c) A tri-mode hand-held terminal supporting satellite, analogue and CDMA digital with 2.5–4.5 h talk time and 9–14 h standby time. (Source: Globalstar. Reproduced with permission of Globalstar.)

1.3.8 Spectrum Management

Frequencies are allocated by the International Telecommunication Union (ITU) and specified in the RR and managed by the local/regional regulatory regime taking into consideration engineering, commercial and political factors. From an operator's perspective, spectrum management includes selection of an appropriate frequency band, obtaining clearance from the regulatory authorities, and managing its usage efficiently.

At present, a majority of MSS systems operate in L (~1.5 GHz) and S (2 GHz) bands; and a few in the K_u band, with some beginning to use the K_a band. The L and S bands are suitable for communication-on-the-move because of relatively benign propagation attributes in these bands and mature technology of these bands. These bands are now congested and hence unsuitable to support for broadband ($> \sim 1$ mbps). Thus some operators prefer the K_a band for broadband applications on portable terminals with directive antenna.

Due to heavy usage of the MSS spectrum, interference management is an important consideration in the planning and operation of mobile satellite systems. A certain level of interference is budgeted in the radio link design to enable inter and intra system frequency sharing. To minimize the probability of unwarranted interference, operators follow a strict regime of spectrum monitoring and procedures to manage harmful interference. Techniques to maximize spectrum efficiency include spatial frequency reuse by spot beams, efficient radio transmission formats and judicious frequency planning. Chapter 9 discusses techniques for maximizing spectrum efficiency and addresses the issue of interference management.

1.3.9 Radio Link Reliability

Techniques for improving radio link reliability include use of robust modulation and forward error correcting codes (which govern link margin), fade countermeasures embedded in the system architecture, store and forward technique to support communication in deep fades and adaptive power or code rate control.

1.4 Business Plan

A crucial element of modern mobile satellite ventures is a viable and credible business plan with a sound market analysis, investment strategy and financial returns for raising finances and revenue. In recent years, we have observed user needs rather than technology influencing mobile satellite system products and system architecture. Prospective operators have tended initially to carry out extensive market research to select services, service areas and user expectations in terms of tolerance of transmission delay, signal fades, etc. System architecture is developed after acquiring a sound understanding of anticipated market requirements and user preference. Chapters 7 and 10 capture the system perspective essential in the development of business plans, demonstrating the intertwined relationship between technical and commercial aspects of MSS.

Invariably, all market forecasts of the 1990s projected a sharp growth in demands for satellite telephony, contrary to experiences of early entrants such as Iridium or Globalstar. Conclusions regarding the total number of systems required to support demands were less clear, with estimates varying between two and four. Clearly, accurate forecasting is a vital element for the success of an expensive MSS venture. More recently, the emphasis is shifting towards mobile wideband data, shadowing the trend towards wider bandwidth in terrestrial mobile and fixed networks. Chapters 9 and 10 present methods of short-term and long-term forecasting.

1.5 Regulatory Considerations

One of the first activities at the start of a venture is that of procuring an operating licence and frequency clearance from the regulatory authorities of the service area that may comprise a number of countries. This procedure can take up considerable time due to a variety of reasons such as bureaucracy, regulatory policies, political issues and involvement of national and international authorities. Other considerations include amongst others preparation of roaming arrangements with other operators, agreement of numbering schemes and selection of distribution partners. These and a range of associated topics are covered in Chapters 3 and 7.

1.6 Operational Considerations

The operation and maintenance of an MSS system is quite challenging, involving a wide range of complex activities such as the launching and maintenance of satellites, network management, commissioning of new terminals, billing, marketing, constant review of business plans and long-term planning. The scale of effort increases as the size of the constellation and network grows.

The starting phase of a system is critical because of a need to set up technical operations, commissioning, billing, customer relations, publicity, and a variety of related tasks, while maintaining cash flow and revenue projections. First generation non-geostationary systems were most susceptible to commercial risks due to uncertain markets, the introduction of untried system designs, the lack of operational experience and the pressure to introduce services on stringent deadlines.

Deployment of a new satellite constellation is time consuming, as it involves a number of launches spread over months. Moreover, satellites generally include a number of new technologies that make the system liable to a high number of early failures. The Iridium system, comprising 66 satellites, was deployed in about 12 months with about 10% of satellite failures caused by a variety of technical problems. Globalstar lost 25% of its constellation in a single launch failure.

Constellation maintenance requires regular monitoring of satellites, performing orbital manoeuvres and changes to spacecraft configuration, replacement of malfunctioning and ageing satellites, upgrading on-board software, and regular network health checks. LEOs deploy the highest number of satellites with the lowest satellite lifetime.

Measures to provide services in the eventuality of a satellite failure include deployment of in-orbit and ground spares; and/or incorporating failure-resistant features in the network architecture. If a region is served by more than one satellite, users can communicate from the operational satellite in case of failure of any one of them, thus providing coverage redundancy as depicted in Figure 1.7 for a geostationary system.

Non-geostationary satellite systems have an inherent resistance to failure due to their distributed architecture and dynamic footprints. The consequence of a failed satellite in such a network is a coverage gap that propagates around the satellite's coverage belt. Thus, the outage is distributed and time shared by subscribers within the affected belt. The gap can be filled by readjusting the position of other satellites and subsequently by the introduction of an in-orbit spare.

Network management involves among other tasks monitoring satellite health, RF transmissions and traffic flow through the network, maintaining quality of service, traffic trend analysis, commissioning of terminals. SCCs monitor telemetered signals from satellites continuously and take corrective action when necessary. Typical actions include the firing of thrusters to maintain a satellite's orbit, reconfiguring a payload in case of failure of a subsystem or in response to an operational requirement, etc. An operational system is susceptible

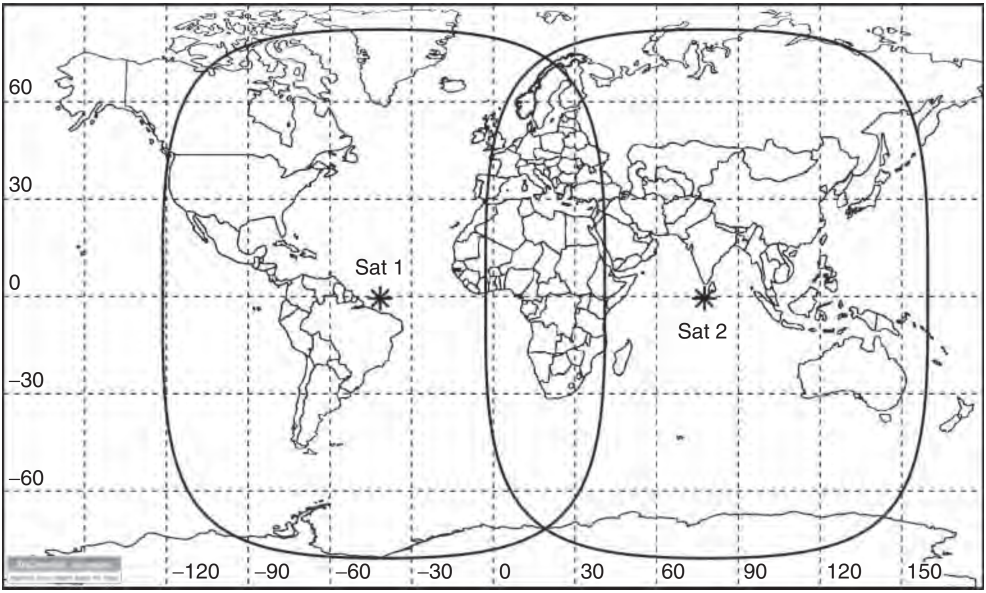


Figure 1.7 Coverage redundancy is available in the overlapping region

to a number of RF-associated problems such as inter and intra system interference and non-compliant transmissions from gateways. Satellite transmissions are therefore monitored to maintain RF integrity of the network and assist in routine RF-related tasks.

The network operator has to ensure that the user is satisfied with the quality of service, which can be measured as grade of service and signal bit errors and delay, in addition to matters related to commissioning, billing, after-sales service, etc. Sometimes failure of a critical subsystem, such as the network frequency management system, manifests itself indirectly as a loss in network traffic, and therefore monitoring of traffic flow through the network is essential. Forecasts used in the planning phase developed on theoretical assumptions and hypothesis can be refined in the operational phase by trending real data. Chapter 9 provides further details related to operational issues and network management.

The ITU has recognized personal mobile communications due to its potentially huge demands and unique technical requirements. The GMPCS is a specific category of MSS to support personal communications to individuals from hand-held personal communicators, much like the cellular telephone or pagers. A GMPCS can be categorized in various ways such as by orbit and service. Big LEO satellite systems deploy large, complex and powerful satellites in LEOs or MEOs capable of providing real-time communications, such as voice and facsimile to hand-held personal communicators; little LEO satellite systems deploy less complex and low-cost satellites operating in the lower part of the orbit designed to provide real-time or non-real-time low bit rate messaging services such as position reports, machine-machine communication and paging to pocket-sized terminals; sometimes the term ‘super’ GEO systems is used for modern GEO systems that deploy very large numbers of spot beams (> 100) enabling services to cellular telephone-sized terminals.

1.7 Mobile Systems – A Comparison

Satellite mobile systems are ideally suited for areas that are poorly served by terrestrial systems to complement terrestrial systems. The advantage of synergistic integration of terrestrial and satellite systems is well recognized. Most GMPCS systems support terrestrial services increasingly through an integrated network through dual-mode user units. Table 1.6 compares the main features of terrestrial and satellite mobile services.

Table 1.6 A comparison of the main characteristics of mobile services through satellite and terrestrial media

Satellite	Terrestrial
Wide area coverage is possible – typically thousands of kilometres	Relatively lower service area – typically hundreds of kilometres; some coverage breaks are possible
Roaming over a wide area is straightforward as coverage is seamless; roaming is limited by licensing issues; usually a single operator owns the space segment with one or more service distributors	Roaming over wide areas, encompassing several countries involves more than one operator/system requiring special system features and operational arrangements
Handsets resemble cell-phones; terminals deploying large antennas (0.3 – 1 m) can provide broad-band services (several tens of Mbps)	Terminals are small and attractively packaged with a wide variety of applications including broadband
Terminal costs are relatively high	Terminal costs are low
Call costs are relatively high	Call costs are low
MSS operate in aeronautical/land/maritime environments	Generally operate in land environment; limited coverage is possible in aeronautical and maritime environments
Service include voice and data with a throughput up to ~ 500 Kbps in L band and up to 50 Mbps in K_a band	Services include voice and data up to several megabits per second. higher throughputs are available in the fourth generation systems
Serve niche market – ships, aircrafts, trucks, international travellers and businessmen, cellular extension, tourism	Serve individuals for social and business needs in populated areas; coverage limited to coastal areas in maritime environments; air coverage is available in certain parts of the developed world
Frequency is reused at distances of hundreds of kilometres	Frequency is reused at distances of $\sim$100 m to $\sim$10 km
Handover between spot beams or satellites are not always essential for GEO but essential for NGEO systems	Handover is frequent and necessary in all cellular systems
Suited for wide area coverage and thin routes (e.g. traffic density < 0.1 Erlangs/km^2)	Suited for urban and suburban environment; uneconomic on thin routes (e.g. traffic density < 1 Erlang/km^2)

1.8 Example Applications

There are innumerable conventional and innovative applications of MSS. We have, therefore, categorized them generically as follows:

1. Maritime environment:
   - ship and cargo management;
   - distress;
   - social;
   - remote monitoring and control;
   - Internet access;
   - fleet broadcast;
   - tourism and leisure;
   - journalism;
   - business; and
   - others.
2. Land environment:
   - business;
   - remote monitoring and control;
   - personal;
   - fleet management;
   - tourism and leisure;
   - journalism;
   - government and aid agencies;
   - Internet access;
   - messaging; and
   - others.
3. Aeronautical environment:
   - cockpit communication;
   - passenger communications including facsimile/Internet access;
   - automatic dependence surveillance; and
   - others.

1.9 Practical Limitations

In this section, we highlight some of the existing practical problems related to the use of MSS and comment on their short- and long-term solution prospects. As illustrated in Table 1.7, many problems are likely to mitigate as technology evolves.

Voice call costs of satellite services have plummeted beginning from ~ \$8–10 per min in the early 1990s to less than 50 US cents; similarly, terminal cost and size to support voice communication has dropped from ~ \$25 000 at the outset to less than \$500 for hand-held units accompanied with phenomenal improvements in quality and ergonomics. Note that voice communication was only feasible through a large and expensive ship-borne terminal at the outset. Moreover, voice communication is increasingly being supported on shared data bearers to provide an increase in space resource efficiency.

Table 1.7 A summary of current limitations of MSS and their perspective

Current limitations	Comments
Expensive in terms of infrastructure, call and terminal costs	Terminal and call costs continue to reduce
Terminals are large compared to terrestrial systems	Hand-held units resemble mobile/cell-phones with a significant size reduction in recent years; terminal size is not crucial in many MSS applications (e.g. ship-borne, railway and aeronautical use)
User interface is complex	The limitation applies to specialist equipment such as ship-borne, airborne and large terminals; users see a simple interface similar to terrestrial systems
A general lack of awareness of the technology	Heightened awareness in recent years
Systems susceptible to local interference	Resolution of local terrestrial interference is indeed arduous
Routing arrangements can be complex and time consuming	These are one-off activities.
Concern about unauthorized bypass of a country's network	Practical solutions are feasible
Service limited to thin route	Terrestrial retransmissions allow extension to populated areas using an ancillary terrestrial component (ATC)
Service unreliable in areas susceptible to shadowing, e.g. urban and suburban locations	An ATC improves reliability

Harmful interference to MSS terminals from local transmissions has been a continuing problem in some parts of the world. Radio frequencies used for satellite communications are shared with terrestrial systems. Although the RR mandate protection of primary allocations (notably, L and S bands for the MSS) from secondary allocations, enforcement can be arduous, as it involves local operators and authorities. Sometimes transmitters in the vicinity, even if operating in a different frequency band, can saturate or overdrive the front-end of a satellite receiver, causing unacceptable degradation to signal quality. Harmonics of powerful terrestrial transmitters such as radar or television can enter receivers causing harmful interference. The probability of harmful interference increases in noisy radio environments such as busy ports. These types of problems can be minimized by practical measures such as communicating via another satellite so that the antenna points away from the interfering source, tuning the equipment to another channel, communicating through another gateway, or terminating and re-establishing a call. The operator can build a database of reported interference for alerting customers, better customer appreciation and to seek solutions formally.

Routing arrangements from the terrestrial segment to the space segment must be in place for the users to be connected efficiently. These arrangements are complex and time consuming for reasons of bureaucracy, politics, etc. The problem is compounded for an international operator, as the process has to be repeated, with each country having individual procedures and priorities, although these are one-off activities.

With the introduction of personal satellite communication services, concern has been raised regarding loss of business to local telecommunication operators because satellite systems can bypass them. Concern regarding security aspects has also been raised by countries vulnerable to antisocial activities. Operators therefore make agreements with local authorities to minimize revenue losses to local operators and build features into the system design to switch off transmissions in specified geographical regions or to permit communications only via specific gateways

1.10 Related Satellite Systems

Satellites provide a variety of mobile communication services, which are not MSS *per se* but complement or compete with the MSS offerings. Some of the more interesting ones are:

1. Mobile very small aperture terminal (M-VSAT) systems;
2. Satellite navigation systems; and
3. Direct audio/video broadcasts to personal and mobile terminals

1.10.1 M-VSAT Systems

VSATs are small FSS UTs for low-capacity/low-cost applications in use since 1970s. Advances in technology leading to high power K_u and K_a band multi-beam satellites and VLSI have enabled a reduction in size and cost of VSAT terminals such that they resemble broadband MSS terminal. Moreover, they provide higher throughput at lower cost/bit than the conventional L and S band MSS. The lower usage cost is due to the availability of larger bandwidth in the FSS bands and higher antenna gain for the same L band antenna dimension as antenna gain is proportional to $1/\lambda^2$ (λ = wavelength). However, since VSATs belong to the FSS, there are regulatory restrictions to their mobility because FSS terminals must by definition remain fixed at a specific location. Each move requires a further series of lengthy coordination meetings, making it impractical to move terminals freely.

The demand of low-cost broadband communication on ships and aircrafts prompted the development of mobile versions of VSATs. In recognition, M-VSAT operation was formally approved by the ITU in some regions in parts of the FSS band. Some countries allow unrestricted VSAT mobility within territories under their jurisdictions to encourage growth of the technology.

Similarly MSS systems have adapted VSAT technology in the K_u/K_a band with the added advantage of unrestricted global mobility. Thus M-VSAT technology has now extended to the MSS regime, giving the user a wide choice of technologies.

Let us briefly return to the issue of competition between these two services. Competition in core MSSs, such as communication from moving land-based terminals, hand-held services, etc., is restricted at these higher frequencies. For fixed-site applications that allow use of directive antennas, both classes of services are similar from a user's perspective. The main difference lies in economics, the regulatory advantage of MSS in terms of unrestricted connection to the public network, the ease of setting up MSS terminals at will and access to the Global Maritime Distress and Safety System (see chapter 13). VSAT networks have the advantage in terms of high throughput and lower cost for large volume data transfer between fixed locations.

We discuss the underlying technologies in Chapter 13.

1.10.2 Satellite Navigation Systems

Navigation is used for estimating the position of a vehicle on sea, in air or space and on land to ensure that the chosen route is followed accurately, both in the short and long term. Short-term navigation is required for making instantaneous changes in direction, speed and acceleration to avoid an obstacle, and long-term navigation is used for making a general correction to a route. A natural by-product of navigation system is the availability of user's position at the given instant.

Satellite navigation receivers have become a personal and mobile communications accessory due to a significant reduction in receiver costs, making them a part of regular gear carried by explorers, travellers, fleet managers, rally organizers, etc. They are embedded in a variety of personal appliances and although developed to aid navigation, the accurate position and time reference available are used in location-based and timing applications.

The GPS navigation system was introduced by the US military in 1978. A constellation of 18 GPS satellites can provide continuous world-wide two-dimensional coverage; increasing the constellation to 24 satellites gives three dimensional position fixes world-wide. GPS satellites are in 55° inclined circular orbits at an altitude of 20 200 km with an orbital period of 12 h, distributed in six orbital plane with four satellites in each plane; accuracy of fixes for military users is < 10.5 m, degrading to about 100 m for civilian users. Satellites transmit atomic clock-controlled timing signals together with their orbital parameters, which are used by receivers for range estimation. A GPS receiver estimates its location by measuring range from three (or four) most favourably positioned satellites simultaneously and solving three (or four) simultaneous equations. Another navigation system known as GLONASS (Global Navigation Satellite System) using a different type of transmission format was introduced by the former USSR at about the same time.

The Galileo navigation system, under developed in Europe, will be operational in the 2014–2015. Other nations such as China are also developing such systems. Thus numerous satellite navigation systems will be available in the near future.

Satellite-aided navigation systems can be categorized on the basis of their operating principles: Doppler signature, range determination, single satellite and multiple satellite transmissions. Chapter 13 summarizes the principles of operation of the most commonly used satellite navigation systems.

1.10.3 Direct Broadcasts to Individuals and Mobiles

Satellite radio broadcast systems for direct reception on portable and vehicle mounted terminals are in regular use in several parts of the world. These systems use terrestrial retransmissions and robust transmission format to provide reliable service throughout the service area, including cities, suburbs and highways. This is the concept used in the ATC based MSS systems.

Conventional direct-to-home satellite broadcasts are received routinely on mobiles with stabilized tracking antennas (e.g. KVH Industries Online, 2012). More recently, there is an interest in providing multimedia and television directly to individuals on small portable and vehicle-mounted receivers. Direct-to-home broadcasts are designed for fixed installations using dish antennas and are therefore unsuitable for small mobile receivers. The key technologies to support mobile reception include video compression, efficient transmission schemes to operate reliably in a harsh mobile propagation environment and synchronized terrestrial re-transmissions. Several standards have been proposed by the ITU and ETSI (European Telecommunication Standards Institute) to promote the evolution of these systems. Chapter 12 addresses these issues in detail.

1.11 Trends

1.11.1 General

The urge of people to remain in contact under all circumstances has been instrumental in the success of mobile communication technology. As people tend to expect from mobile communication systems services akin to those offered by fixed services, it is anticipated that trends in the fixed services are likely to be a precursor to those in the mobile systems. However, mobile wireless services give a lower end-user throughput than generally available from wired systems because of the difficulty in maintaining a high-quality mobile radio link, and at present, a rather limited available spectrum for these services. Nevertheless access to information databases while the user is on the move brings with it a variety of unique location-based applications related to the immediate surroundings.

In addition to the unceasing demands for voice communications, there is now an escalating demand for data traffic for both personal and business applications. Examples of such applications are internet access to large database, teleworkers transferring software, e-mails, image transfer, and so on. Similar demands are already placed on the mobile communication sector due to the vigorous growth in personal computers (PCs) such as the personal digital assistant (PDA). Thus, considerable effort is under way to provide high data rate mobile telecommunication services. This is evident in the evolutionary trend of the MSS, with data rates increasing from a few kbps in the first-generation to ~50 Mbps in the fifth-generation Inmarsat system. Interestingly, both FSSs and MSSs are vying for the personal broadband service.

A considerable headway has been made in convergence of the fixed component of mobile networks with the fixed network due to a growing number of commonalities between them. It is recognized that in dense urban areas terrestrial systems are better suited, but satellite systems offer unique advantages in providing wide-area mobile communications to ships, aircraft and land mobiles. Thus, from a user's viewpoint, there are advantages in developing a unified satellite and terrestrial network. All the satellite hand-held systems offer a facility to switch from one or more of the existing cellular standards to another. Considerable research activity is in progress to unify terrestrial and mobile systems into a universal network. Since integrated satellite-terrestrial networks offer seamless connectivity irrespective of users' location, synergistic solutions are a primary theme of modern MSS paradigm.

Satellite systems have grown significantly but due to their niche services, their penetration is significantly lower. The services offered by previous mobile satellite systems were targeted for specialized applications in aeronautical, maritime and land sectors on terminals that were heavy and expensive. However, modern MSS systems provide a cellular system-like service at comparable costs and broadband multimedia terminals are about the size of desktop-size telephones. Increase in the service link (L-band) EIRP of four generations of

Inmarsat, is representative of world-wide MSS growth trends since the 1980s. The EIRP of satellite has increased by about 34 dB, that is a factor of ∼2500; if we consider four such satellites for global coverage, the increase amounts to 40 dB – a factor of ∼10 000!

1.11.2 Market

Vast, sparsely populated areas throughout the world remain unserved by fixed or cellular systems, either because the service is uneconomic or due to a lack of infrastructure. Satellite systems are ideally suited to filling such coverage gaps.

Terminal costs and call costs have dropped considerably and awareness of the technology has increased significantly. Recent reports of MSS operators demonstrate an aggressive growth in MSS markets and market surveys conclude that the trend is likely to continue in the foreseeable future. All the major operators have announced plans to upgrade their technology aiming to provide next-generation services for well over a decade. Chapter 14 provides samples of recent market projections up to the year 2020 and beyond.

1.11.3 System Architecture

Until the late 1980s, most commercial mobile satellite systems used GEOs with bent-pipe transponders. In the 1980s, extensive studies were undertaken in the UK and by the European Space Agency (ESA) to investigate the feasibility of deploying on-board processing satellites in highly eccentric elliptical orbits for mobile communications in the European region. Studies also investigated the feasibility of deploying LEO satellites for mobile satellite systems.

The trend in space segment architecture diverged at the beginning of 1990 when a number of commercial non-geostationary satellite systems were proposed for hand-held voice and data communication services. The architecture proposed LEOs or MEOs, as hand-held services via geostationary satellites would require extremely complex spacecraft and suffer transmission delays. Their architecture varied widely in orbital choice, satellite complexity, transmission schemes, network routing and the market addressed.

By mid-1990 there was a re-emergence of geostationary systems based on powerful satellites deploying several hundreds of spot beams. A comparison of various types of systems and architectures is given in Chapters 2, 7 and 8.

Third generation mobile systems have been standardized on the premise that satellites form an integral part of the network. International forums that have been instrumental in the process include the ITU forum called the International Mobile Telecommunication-2000 and the ETSI, a European forum where the third generation mobile systems are called Universal Mobile Telecommunication Systems or UMTS. These bodies have now extended their remit to the fourth generation. Chapter 8 provides an extensive review of the prevailing standards. Figure 1.8 illustrates a generic approach used in the standardization where components of each system are connected through standard interfaces.

Examples of an early air interface standard includes Geo Mobile Radio (GMR) Standard that was derived jointly by the ETSI and the Telecommunication Industry Association (TIA) of the United States from the Third Generation Partnership Programme (3GPP) family of cellular digital standard. These systems allow seamless integration of satellite systems with GSM and 3G core networks. The standard has since evolved and branched. Examples of

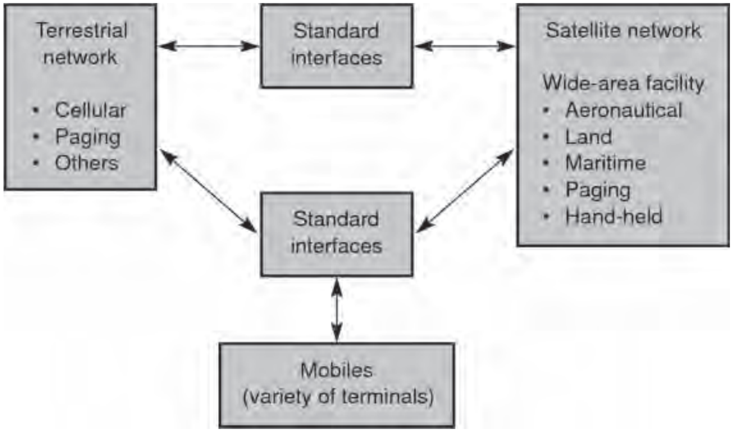


Figure 1.8 Concept of a Universal Mobile Communication System

application of these standards include Thuraya's satellite phone system based on GMR 1 (basic version) and Inmarsat phone (iSATphone) based on GMR 2 (an evolved version of GMR). GMR-1 3G technology is used by TerreStar, SkyTerra and Inmarsat's broadband system Broadband Global Area Network (BGAN). As MSS systems converge with VSAT at the upper end, it would appear to be advantageous attach a mobility component to VSAT standards – DVB RCS + M (digital video broadcast/return channel via satellite + mobility) is an example of such a standard. Work is now in progress to develop satellite architectures and standards for 4G systems.

All the major operators are beginning to introduce the next generation services. The general trends include a tighter integration with terrestrial mobile systems including the use of ATC; an increase in data throughput and provision of VSAT-like service through K_a band in direct competition with M-VSATs; and a deeper involvement in the development of MSS products and applications.

1.11.4 Spectrum

Rapid evolution of MSS fuelled by insatiable demands for broad-band services implies a greater demand on the spectrum. Since a majority of MSS systems operate in L and S bands, these bands are now congested despite the use of spectrum enhancement techniques such as spectrally efficient modulation-coding and advanced spot beam technology. There is a move towards implementing systems using MSS allocations in the K_a band for broadband applications with directive antennas. There is a growing interest in the possibility of higher bands such as V in the long term. Hand-held and other personal services are generally confined to L and S bands. Chapter 3 presents further details on spectrum-related matters.

1.11.5 Technology

To realize new services and applications and maximize spectral efficiency and revenue, several new technologies have been developed and are being refined. A few representative examples follow:

- **Satellites with regenerative transponders and on-board computing**: Regenerative transponders, by their ability to demodulate signals, can lower mobile terminal EIRP needs, reduce their size and cost and mitigate the effects of interference. The ability of regenerative transponders to decouple up- and down-links enables optimization of multiple accesses separately to maximize utilization of satellite resources. On-board processing can incorporate advanced space-based features such as signal routing and network functions. Software reconfigurable transponders minimize risk of obsolescence.
- **Intersatellite links**: Intersatellite links are space links between satellites for routing signals in space, thereby simplifying ground connectivity.
- **Spot-beam technology**: All modern mobile satellite systems deploy large numbers of spot beams to provide a dense spatial frequency ruse. Typically, several hundred spot beams are used for geostationary mobile satellite systems.
- **Multiple launch capability**: Large satellite constellations require a large number of reliable launches in quick succession; several launchers have been developed for multiple satellite delivery.
- **Mass satellite-production technique**: Traditionally, several years were spent in manufacturing a single communication satellite, clearly an unacceptable scenario for satellites that are part of large constellations. Satellite manufacturers have introduced mass-production techniques similar to those used in the automotive industry, reducing manufacturing time per satellite, in some cases, to a few weeks.
- **Software**: All aspects of satellite communications have benefited from the developments in software and workstation. Applications include computer-aided design, remote earth station operation, automated satellite control, constellation management, radio resource management, etc. Network control, data management and flow, as well as a majority of applications are now heavily dependent on software.
- **Advanced UT architecture and VLSI**: The challenge here is to produce low-cost, cellular-integrated satellite phones affordable by individuals. Phenomenal advances in VLSI, packaging, battery technology, and software have enabled introduction of affordable dual- and multi-mode hand-held portable transceivers and software configurable UTs.
- **Advancements** have been orchestrated in areas of modulation, channel coding, voice-coding, compression, access technology, system architecture, management of complex networks, and others.

Revision

1. What are the main components of a mobile satellite service? Outline the role of each component.

2. Table 1.2 gives a comparison of various technical parameters of several MSS systems. Explain the reason for different sensitivity requirement of the gateways (specified as G/T; a higher value implies a more sensitive receiver) in view of the similarity in G/T of the user terminal; Compare the total space segment capacity of each system assuming that three regional geostationary systems are required for world-wide coverage. Comment on your results.

3. Explain the difference between architectures of non-geostationary satellite systems, which provide a non-real-time communication service, and real-time communication services.

4. What is the rationale for using a low or medium earth orbit in preference to the geostationary orbit for the provision of a hand-held service?

5. The architecture of a satellite system is influenced by a number of technical considerations in addition to the service requirements. State these considerations. Briefly explain the role of each in system design.

6. Compare the characteristics of satellite and terrestrial mobile systems. Explain the reasons for a growing convergence between these systems.

7. What are the strengths and limitations of a mobile satellite service?

8. Briefly outline the factors likely to influence the evolution of mobile satellite systems.

References

Chakraborty, D. (1989) Survivable communication concept via multiple low earth-orbiting satellites. *IEEE Transactions on Aerospace and Electronic Systems*, **25**(6), 879–889.

Electronics and Communications Engineering Journal (1995) Technical demographics. December, 265–271.

Evans, J.V. (1998) Satellite systems for personal communications. *Proceedings of the IEEE*, **86** (7), 1325–1341.

Jahn, A., Buonomo, S., Sforza, M. and Lutz, E. (1995) Narrow and wide-band channel characterisation for land mobile satellite systems: experimental results at L band. International Mobile Satellite Conference, Ottawa–IMSC 1995, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communication Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 115–121.

KVH Industries (2012) Mobile Satellite TV, www.kvh.com/Leisure/RV-Systems/Television.aspx (accessed 20 August 2013).

Nelson, R.A. (1998) Iridium: From Concept to Reality. *Via Satellite*, Vol. XIII, Issue 9, pp. 62–70.

Richharia, M., Hansel, P., Bousquet, P.W. and O'Donnell, M. (1989) A feasibility study of a mobile communications network using a constellation of low earth orbit satellites. IEEE Global Telecommunications Conference, GLOBECOM'89, Dallas, TX, pp. 27–30.

2

Satellite Constellations

2.1 Introduction

The topic of satellite orbits has been extensively reviewed in the recent past due to the strong influence of orbit design on the space segment architecture of non-geostationary satellite systems. Satellite orbits were investigated for telecommunications at the dawn of the Space era. Early interests were confined to low earth orbits (LEOs) and medium earth orbits (MEOs) because of the limited capability of early launchers, however, geostationary satellite orbit (GSO) became the norm soon after availability of GSO launch vehicles. The interest in non-geostationary orbit (NGSO) satellite constellations revived in early 1990s, due to a combination of events – availability of advanced technology and industry's willingness to support high-risk commercial ventures motivated by a potentially lucrative hand-held market. Low and medium earth satellite systems, being closer to the Earth, could incur lower path loss and propagation delay, enabling the use of hand-held terminals while offering transmission delays approaching those of optical fibre and other terrestrial media. [Note: We have used GSO and geostationary earth orbit (GEO) interchangeably.]

System designers were faced with the challenge of optimizing NGSO networks and issues included the number of satellites, their orbital characteristics, efficient network routing schemes, cost, etc. System design takes into consideration the network in its entirety and is influenced by a variety of factors, such as service area, traffic distribution within it, acceptable space segment complexity, etc., though emphasis on each depends on the operator's preference. This is evident by the vast differences in orbital characteristics of various prevalent and proposed NGSO satellite constellations, ranging from LEOs and MEOs to elliptical, hybrid and GSOs.

In this chapter we introduce primarily the principles of satellite orbits essential in characterizing and optimizing satellite constellation in circular orbits. The chapter has been divided into two broad parts – the first part introduces the concepts of orbital mechanics and other related topics and the second addresses issues related to satellite constellations. To keep to the theme of the book the subject is treated at a system level. Optimization of satellite constellation is addressed solely from an orbital view point, that is, taking into consideration the coverage area, space environment and orbital characteristics. Optimization in a systems context is addressed in Chapter 7 after the reader has developed the necessary background.

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

The first part provides a brief introduction to orbital mechanics, which includes coordinate systems applicable to MSS (Mobile Satellite Service), the definition of orbital parameters and salient characteristics of various types of orbits. An understanding of the interrelationship between crucial parameters and their influence on orbit design is explained by a two-body model. The effects of perturbations caused by the Earth's non-uniform gravitational force and gravitational effects of the Sun and Moon are described and specific orbits related to these perturbations, namely the *Sun-synchronous* and *Molniya* orbits, are discussed. Commonly used map projections for illustrating a satellite's coverage area are introduced with examples of coverage maps of GSO and NGSO satellites. Subsequent sections discuss the solar eclipse, the Sun's interference, Doppler effects and the growing problem of orbital debris.

The second part of the chapter discusses various aspects of satellite constellation including theoretical bounds of constellation size, and reviews well-known techniques used in the optimization of polar, inclined orbit and hybrid constellations. The chapter concludes with a comparison of orbit-related performances of constellations deployed or planned.

2.2 Satellite Orbits

2.2.1 Orbital Mechanics Basics

This section provides a brief overview of orbital mechanics fundamentals essential in developing an understanding of satellite constellation.

Kepler's laws in conjunction with Newton's laws quantify satellite motion around the Earth. Kepler's laws of planetary motion define the shape of orbit and orbital period of each planet, while Newton's laws of motion and gravitation explain the reason for this behaviour. The motion of artificial satellites is governed by the same principles and hence they follow the same laws. The position of a satellite at any instant can be estimated precisely by the application of these principles and including the effects of all the forces acting on the satellite.

According to Kepler's law, a satellite orbits the Earth on a well-defined path that can be quantified by a set of orbital parameters. These parameters are used to define the position of a satellite uniquely in a three-dimensional space at any given instant. A number of coordinate systems are used in mobile satellite systems. These are often called *inertial coordinate systems* and the space is referred to as an *inertial space*, as each is referenced to stars.

The coordinate system used for definition of orbital parameters is called the geocentric-equatorial coordinate system shown in Figure 2.1.

The Earth's centre constitutes the origin and the equator is the fundamental X-Y plane. The positive direction of X-axis is towards the Vernal Equinox, which is the vector joining the geocentre to the direction of the Sun's position on the day of spring equinox. Its equivalent spherical coordinate system is called the *right ascension-declination coordinate system* and is commonly used by astronomers for defining the position of heavenly bodies such as the Sun and Moon. For satellite communications, a knowledge of the right ascension-declination system becomes necessary when positions of Sun and Moon are of interest, such as when determining occurrence of the Sun's eclipse on a satellite. Another coordinate system named the *celestial horizon coordinate system* or *topo-centric coordinate system* is used for pointing the antenna towards a satellite. The axes in such a system are called *elevation* and *azimuth*.

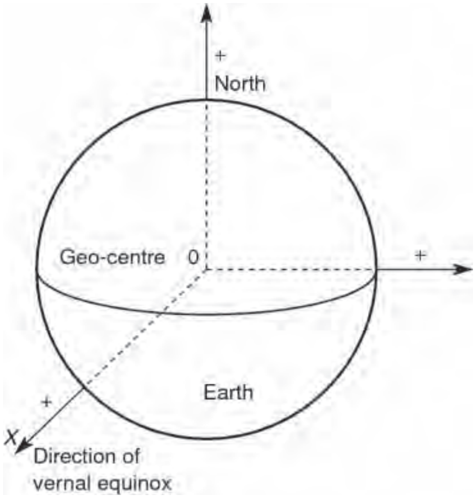


Figure 2.1 A geocentric-equatorial coordinate system

Elevation is the angle on the vertical circle from the local horizon to the position of the satellite, and azimuth is the angle of satellite measured in an eastward direction from true north on the local horizontal plane as shown in Figure 2.2.

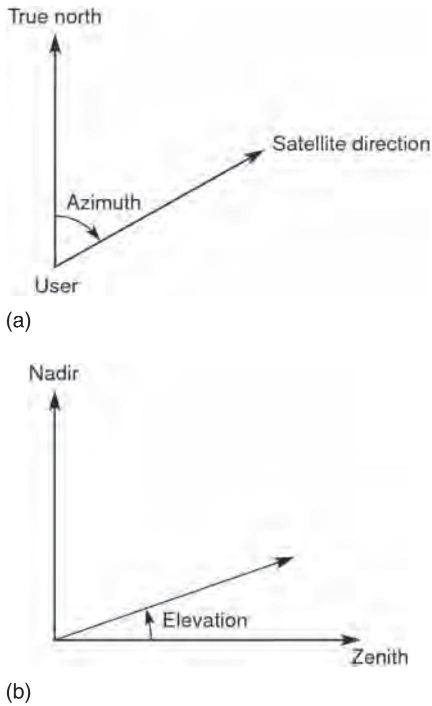


Figure 2.2 A celestial horizon coordinate system

The six orbital parameters, portrayed in Figure 2.3(a) and (b) are semi-major axis, eccentricity, inclination, right ascension of an ascending node, the argument of perigee and time elapsed from a reference point of the orbit.

1. The *semi-major axis, a*, describes the size of an elliptical or a circular orbit as shown in Figure 2.3(b). The term semi-major axis is appropriate for description of an elliptical orbit. The semi-major axis reduces to a radius for a circular orbit.
2. *Eccentricity*, *e*, represents the shape of an orbit. Eccentricity of a circular orbit is 1; the orbit takes on a more elongated shape as its magnitude is increased.
3. *Inclination*, *i*, describes the orientation of an orbit with respect to the equatorial plane as shown in Figure 2.3(a). It is the angle between the plane of the orbit and the equatorial plane. Referring to Figure 2.3(a) it can be observed that the orbit crosses the equatorial plane at two points called *ascending node*, where a satellite crosses the equator from the southern hemisphere to the northern and the *descending node*, where a satellite crosses the equator when moving from the Northern Hemisphere to the Southern.
4. *Right ascension*, Ω, shows the orientation of an orbit with respect to the X-axis of the coordinate system. It is the angle between the X-axis and the *line of nodes* – a line joining ascending and descending nodes.
5. The *argument of perigee*, ω, describes the orientation of an elliptical orbit's perigee with respect to the line of nodes measured as the angle between the line of nodes and the perigee.
6. *Time*, t_p is the time elapsed since a satellite has passed a reference point in an orbit, usually the perigee; where the reference time is known as an *Epoch*.

Applying Newton's Laws of Motion leads to the equation of motion of satellites in their orbit. A complete solution of the equation is quite complex, but a partial solution is adequate to illustrate the shape and size of an orbit. The simplified solution shows that the satellites follow a path of a conic section of the general form

$$\mathrm{r} = \mathrm{p}/(1 + e\cos(\theta)) \tag{2.1}$$

where r = the distance of any point on the trajectory from the geocentre, p = a geometrical constant – termed parameter of the conic, which determines the width of the conic at the focus, e = the eccentricity, which determines the type of conic section, θ = the angle between r and the point on the conic nearest the focus.

The trajectory of a satellite depends on the final velocity vector and the altitude at which the satellite is launched by the launch vehicle. Thus, for example a 900 km circular orbit can be achieved by launching a satellite to an altitude of 900 km so that it achieves a velocity of 7.4 km/s, that is the velocity required for a circular orbit at such an altitude, and tilting the satellite parallel to the Earth.

An orbit, in which a satellite moves in the same direction as the Earth, that is west to east, is called a *direct* orbit and when a satellite moves in an opposite direction to the Earth's motion, the orbit is a *retrograde* orbit. Satellites may also be categorized according to their inclination, altitude and eccentricity. An orbit with an inclination of 0° is called an *equatorial* orbit; when the inclination is 90° the satellite is said to be in a *polar* orbit; orbits having an inclination between 0° and 90° are called *inclined* orbits. Orbits may be categorized

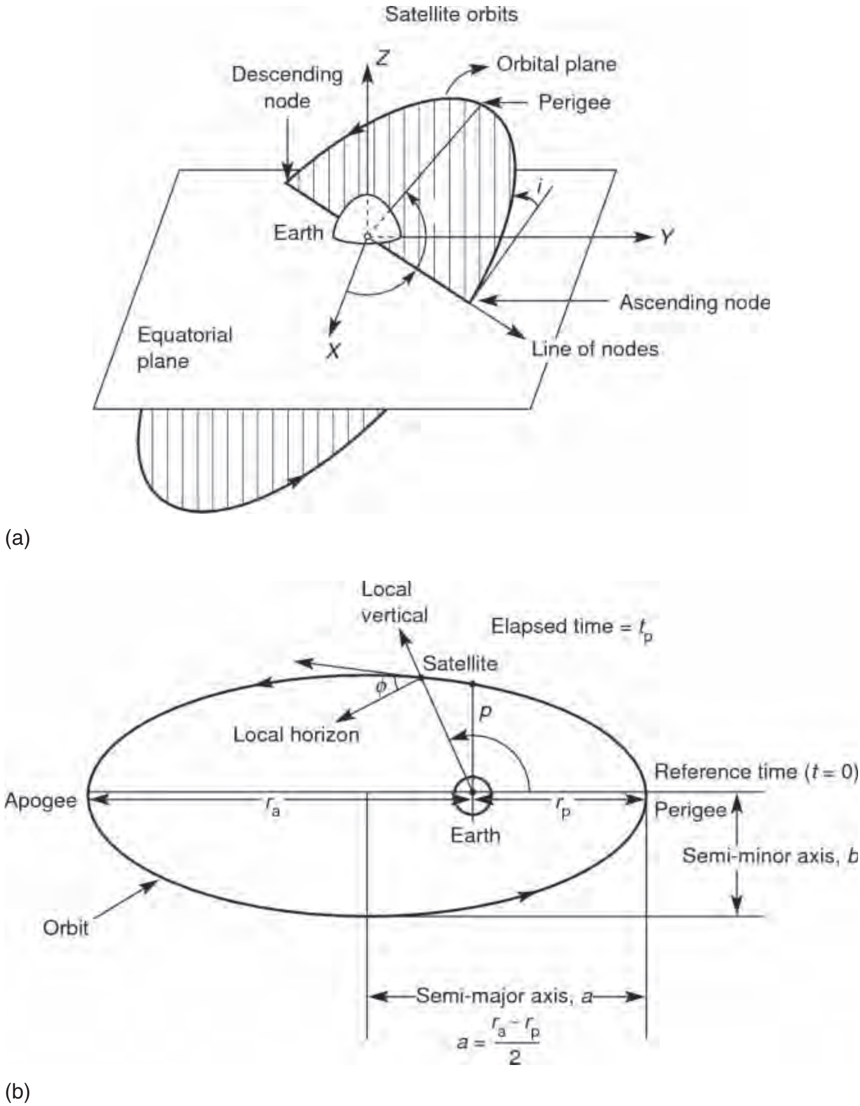


Figure 2.3 (a) A pictorial representation of orbital parameters; i = inclination; Ω = right ascension of an ascending node is the angle between the X-axis – the direction of the vernal equinox – and the ascending node; argument of perigee is the angle between the line of nodes and the perigee (b) Major parameters of an elliptical orbit (Both parts source: Richharia, 1999. Reproduced with permission of Palgrave-Mamillan.)

according to their altitude as LEO, MEO, also sometimes called the Intermediate Circular Orbit (ICO), and GSO or geostationary orbit (GEO). When eccentricity is 0, the orbit is circular and when eccentricity > 0, the orbit is elliptical. A high-eccentricity orbit is sometimes called a *highly elliptical orbit* (HEO).

For an elliptical orbit, the relationship between the period T of a satellite and the semi-major axis can be obtained by Kepler's Third Law

$$T^2 = 4\pi^2 a^3/\mu \tag{2.2}$$

For a circular orbit the equation reduces to,

$$T^2 = 4\pi^2 (R+h)^3/\mu \tag{2.3}$$

where R = the radius of the Earth, h = the satellite altitude and μ = gravitational parameter = 398 600.5 $km^3\ s^{-2}$.

Figure 2.4(a–c) shows the relationship between altitude and period of orbit. Also marked on the figure are the approximate practical bounds of low and medium altitude orbits.

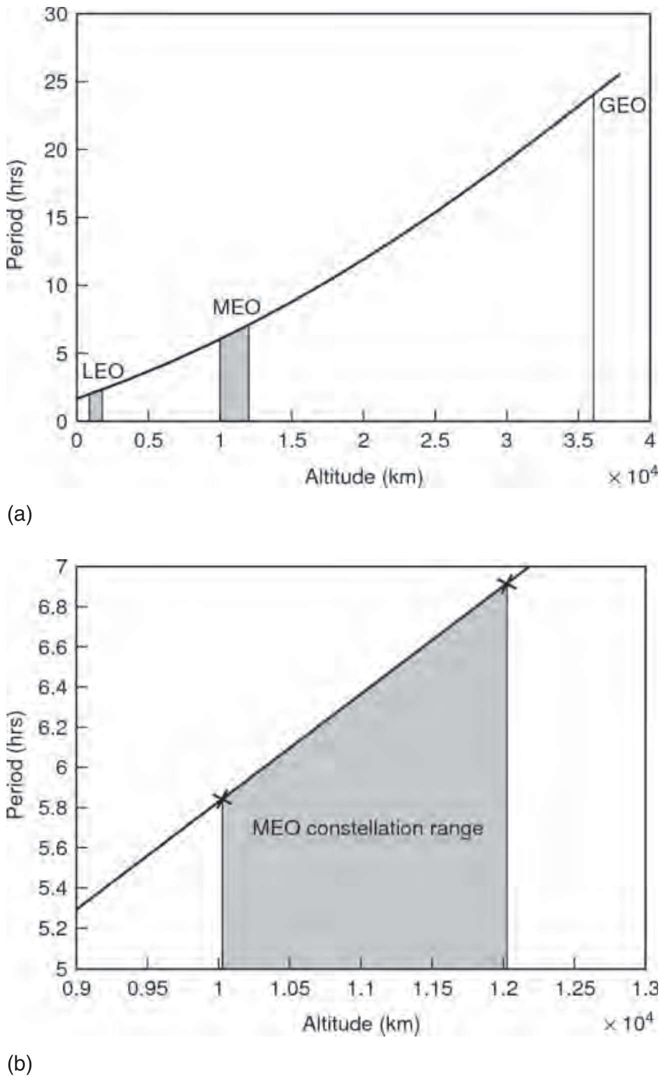
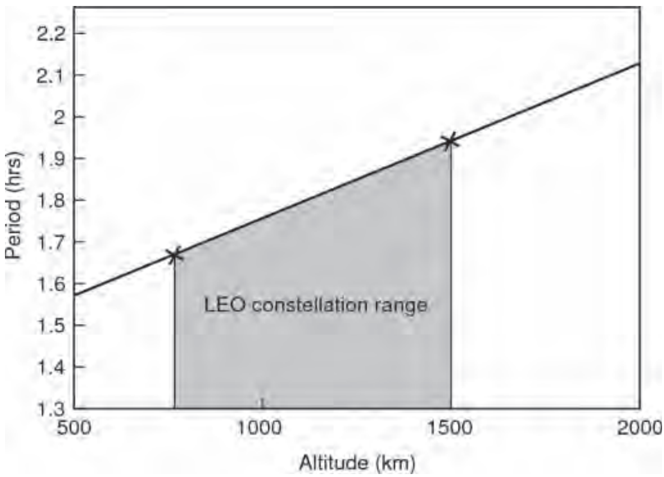


Figure 2.4 *(Continued)*

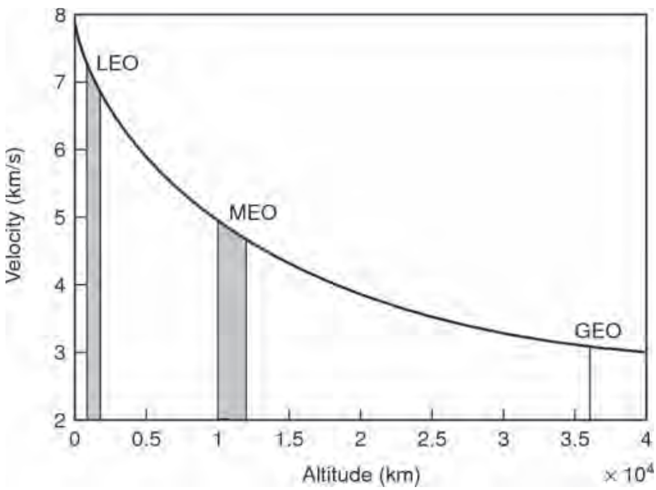


(c)

Figure 2.4 Relationships between (a) altitude and orbital period for altitudes up to the GEO, (b) altitude and orbital period for MEO, and (c) altitude and orbital period for LEO (All parts source: Graphics AR.)

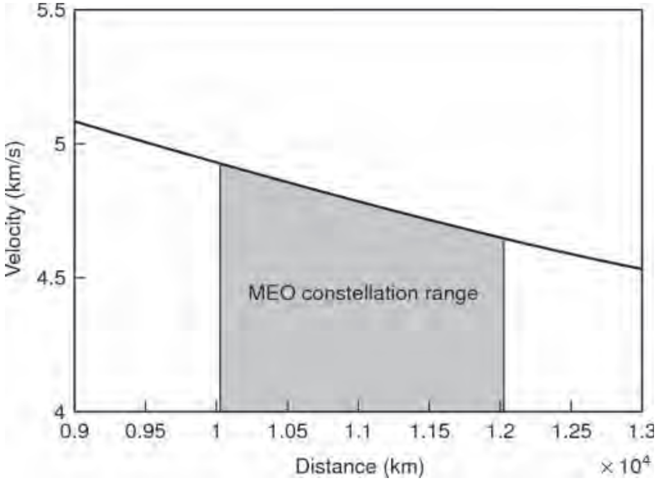
Figure 2.4(a) applies to altitudes up to the GSO, while Figure 2.4(b) and (c) pertain specifically to MEOs and LEOs. Similarly, Figure 2.5(a–c) demonstrates the relationship between altitude and orbital velocity for LEO, MEO and GEO satellites.

The region between low and medium altitudes contains a strong radiation belt that can cause damage to satellite components and software (see Section 2.2.7); it is therefore avoided as far as possible for commercial satellite systems. At low altitudes, satellite velocity decays due to friction from the Earth's atmosphere causing a loss in altitude until satellites re-enter

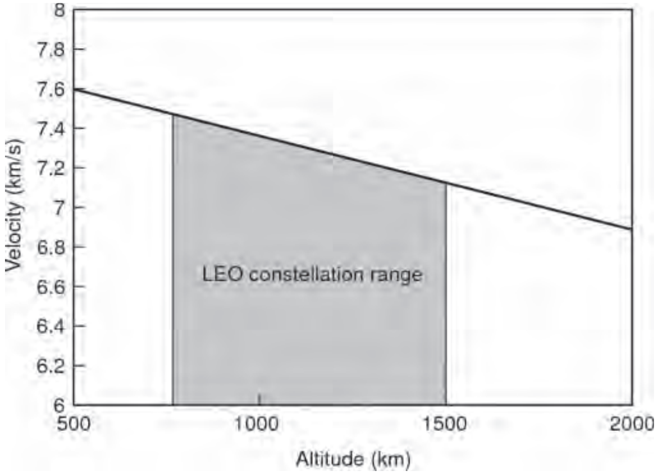


(a)

Figure 2.5 *(Continued)*



(b)



(c)

Figure 2.5 Relationships between (a) altitude and orbital velocity for altitudes up to the GEO, (b) altitude and orbital velocity for MEO, and (c) altitude and orbital velocity for LEO (All parts source: Graphics AR.)

the atmosphere and eventually a majority of them burn out due to frictional heat. Therefore, for practical purposes, altitudes below ∼180 km are not useful, except for military applications.

At an altitude of 35 786 km, the orbital period becomes 23 h, 56 min and 4.1 s, which is the time the Earth takes to rotate 360° around its axis. This period is called a *sidereal* day. When the inclination of the orbit is 0°, the relative motion of the satellite and Earth reduces to zero and therefore satellites in this orbit appear motionless to observers on the ground. This

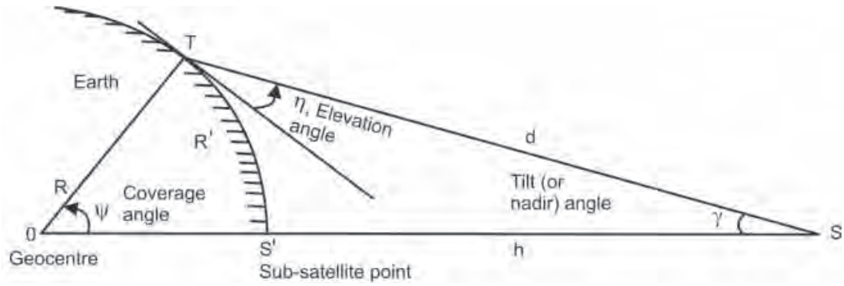


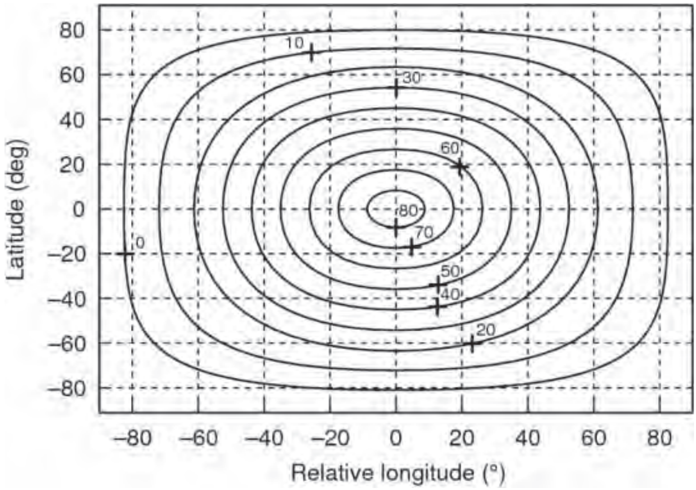
Figure 2.6 Orbital geometry of a geostationary satellite

type of orbit is called a *geostationary orbit* or geostationary satellite orbit (or, geostationary earth orbit), which has been used extensively for satellite communications as it offers various advantages.

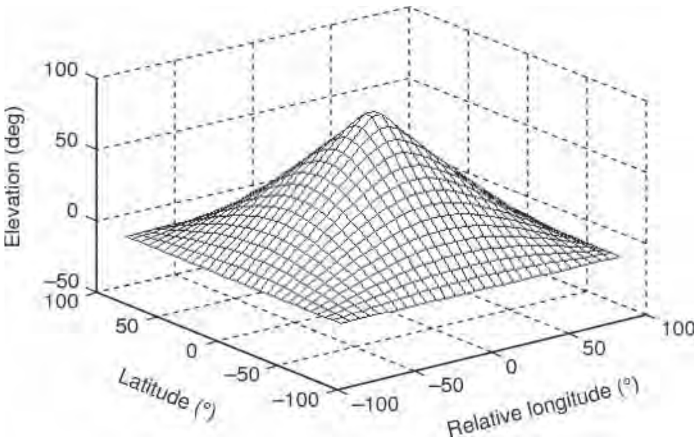
To effect communications it is necessary to determine the position of the satellite at any instant. The position of a satellite from a user's location can be estimated by applying the values of orbital parameters, obtainable from the satellite operator, to a set of vector identities. The knowledge of satellite position and geometry is also necessary to track a satellite in real time and, during the design phase to estimate radio link parameters such as satellite range and elevation angles to develop the radio link design. For GEO satellites, simple geometric relationships assuming a spherical Earth, as shown in Figure 2.6 are often adequate for radio link analysis.

Figure 2.7 (a) and (b) shows the contour and three-dimensional plot of satellite elevation as a function of latitude and longitude relative to the sub-satellite point of a geostationary satellite. Notice the variation in elevation angle as the latitude and longitude move further away from the sub-satellite. The furthest latitudes at which a satellite is visible (elevation $> 0°$) are $\sim \pm 81.3°$, when the observer is at satellite's sub-satellite longitude (i.e. relative longitude $= 0$). Similarly the furthest relative longitudes at the equator (latitude $= 0$) are $\sim \pm 81.3$. Figure 2.7(c) and (d) shows the contour and three-dimensional plot of azimuth as a function of latitude and longitude relative to the satellite sub-satellite in the four quadrants. Figure 2.8(a) and (b) illustrates variations in range to a geostationary satellite with latitude and relative longitude as a contour plot and in a three-dimensional representation, respectively. It emerges that range, and therefore path loss, increases as the observer moves away from the sub-satellite point with the highest loss at the edge-of-coverage (EOC) when the propagation conditions are least favourable.

The first-order behaviour of satellite motion can be interpreted by considering only the gravitational force of the Earth as this is dominant; however, a number of external forces act on a satellite causing its motion to deviate from the ideal. Therefore, satellite position has to be corrected periodically to ensure that it remains within tolerable limits. For example, relative positions between satellites of a *phased* constellation will change causing distortion of the service area, bearing in mind that in phased satellite constellations relative phase between satellites must be constant; similarly, the coverage contour of a geostationary



(a)



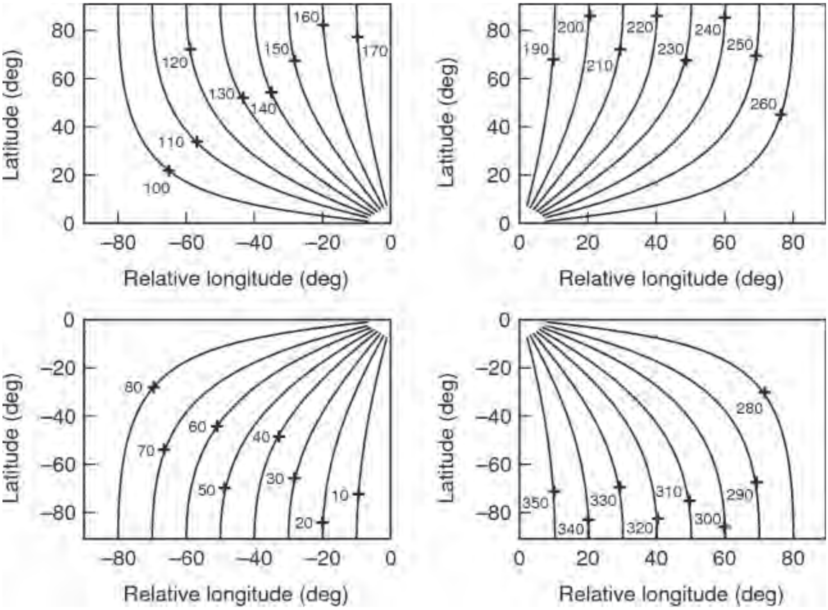
(b)

Figure 2.7 *(Continued)*

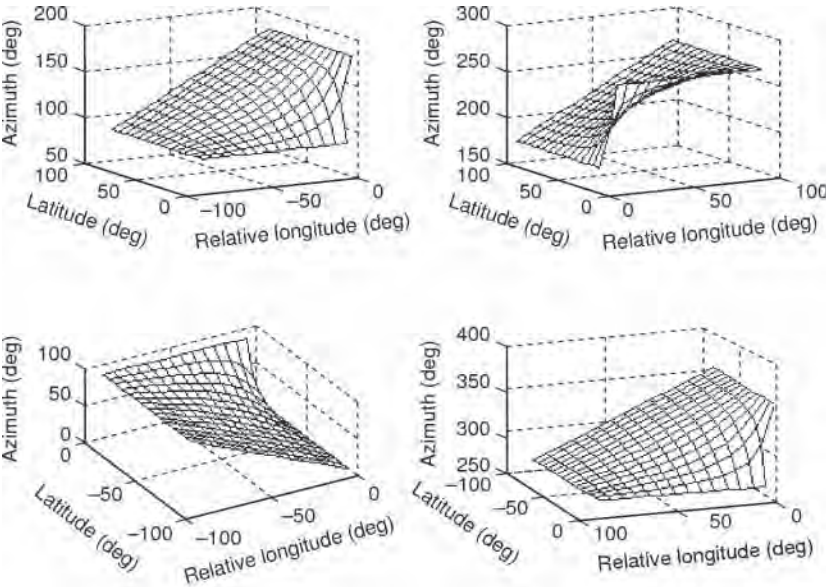
satellite will drift causing a loss of service in regions no longer within the footprint, and the satellite will eventually drift out of its allocated orbital location in breach of ITU (International Telecommunications Union) radio regulations.

The most significant perturbations are caused by:

- Non-uniform gravitation of the Earth;
- gravitational effects of the Sun and Moon;
- atmospheric drag; and
- solar radiation pressure.

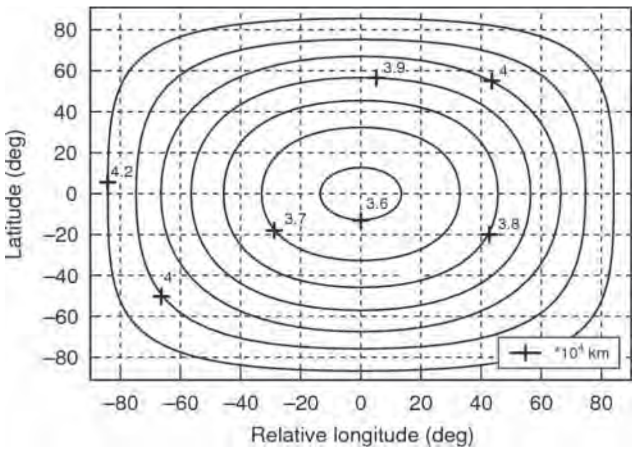


(c)

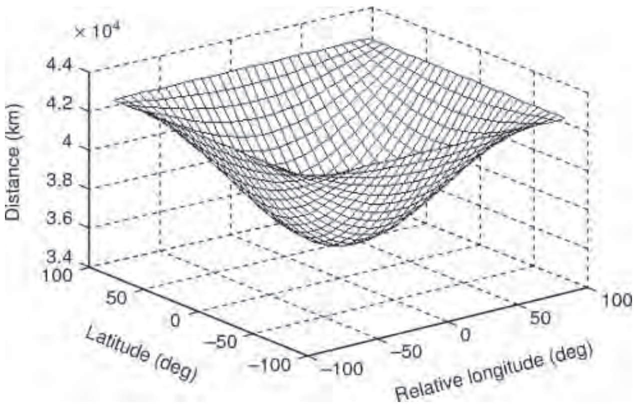


(d)

Figure 2.7 (a) Contour plot of elevation as a function of latitude and relative longitude (b) Three-dimensional plot of elevation as a function of latitude and relative longitude (c) Contour plots of azimuth angle as a function of latitude and relative longitude (d) Three-dimensional plots of azimuth angle as a function of latitude and relative longitude (All parts source: Graphics AR.)



(a)



(b)

Figure 2.8 (a) A contour plot of geostationary satellite range as a function of latitude and relative longitude and (b) a three-dimensional plot of geostationary satellite range as a function of latitude and relative longitude (Both parts source: Graphics AR.)

2.2.1.1 Non-Uniform Gravitation of the Earth

The gravitational force around the Earth varies due to the non-uniform distribution of the Earth's mass; furthermore the Earth's shape is slightly ellipsoidal as the polar radius is ∼21 km shorter than the equatorial radius. A cross-section of the Earth shows a semi-major axis approximately along the line 165°E and 345°E (15°W) and a semi-minor axis approximately along the line 75°E and 255°E (105°W). Therefore, gravitational force is no longer directed towards the geocentre, but towards the centre of the ellipsoid. Non-uniform gravitational fields cause the following effects:

- Precession of perigee in orbital plane.
- Precession of orbital plane around the Earth's north-south axis.
- Perturbing force in a direction along the orbit.

The first two effects are most noticeable in LEO and MEO satellites whereas the last mentioned mainly affects GEO satellites.

Precession of Perigee

Precession of perigee is represented pictorially in Figure 2.9.

The rate of change of argument of perigee, ω, of an elliptical orbit is given as,

$$\omega = 4.97(R/a)^{3.5}(5\cos^2(i) - 1)/(1 - e^2)^2 \ {}^\circ/\text{day} \tag{2.4}$$

where R = the mean equatorial radius (~ 6378 km), a = the semi-major axis, i = the inclination and e = the eccentricity.

Figure 2.10(a) illustrates a variation in rate of change of argument of perigee versus semi-major axis for a number of inclination for an eccentricity of 0.1° and Figure 2.10(b–d) shows a three-dimensional view of the variation for low, medium and geostationary satellites respectively.

Note that at the inclination of 63.4°, $\omega = 0$. This happens when the component of the equation containing cosine term in the right-hand side of the equation equals zero. A satellite in such an orbit retains its perigee over the same region of the earth. This characteristic can be usefully exploited for mobile satellite systems. Consider an eccentric orbit of 63.4° inclination with its apogee in the northern (or southern) hemisphere. The velocity of a satellite reduces with increasing altitude and hence reaches a minimum near the apogee of an elliptical orbit. By increasing the altitude sufficiently the satellite appears quasi-stationary at high elevation angles to an observer in the mid-to-high latitude of the northern (or southern) hemisphere. These features improve the radio link reliability of mobile satellite systems at mid/high latitudes (see Chapter 3). Figure 2.11 shows coverage of a satellite when the satellite is at the apogee of such an elliptical orbit.

Examples of satellite systems designed for this class of orbit are Archimedes, a European design concept for mobile and broadcasting applications, the ELLIPSO™ mobile system, a proposed US system for providing hand-held telephony service and the Sirius satellite radio broadcast system that services millions of users in North America. A derivative of this class of orbit, called the *Molniya* orbit after the Russian systems, was used in the high latitude

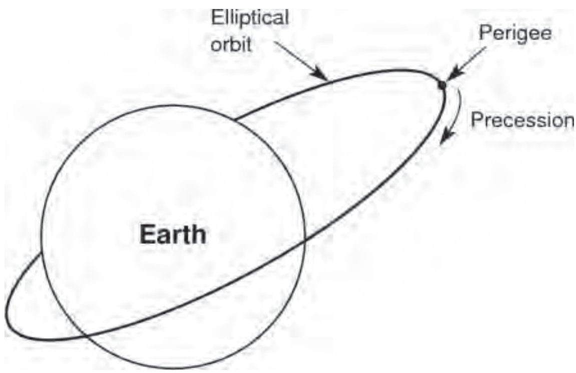
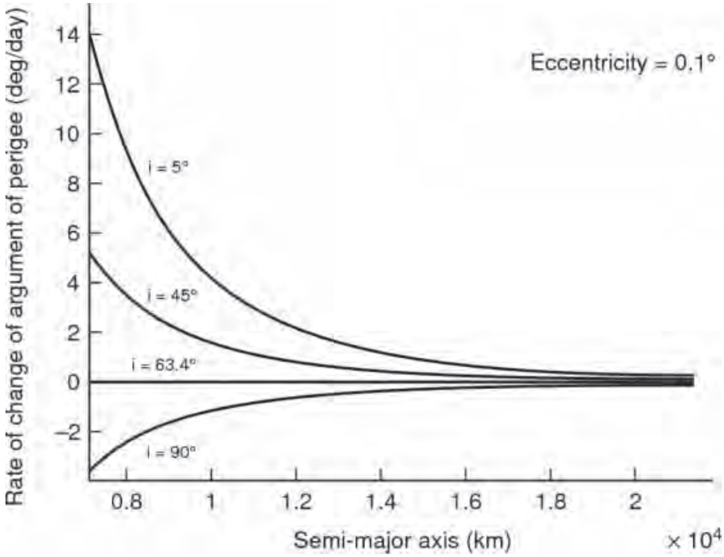
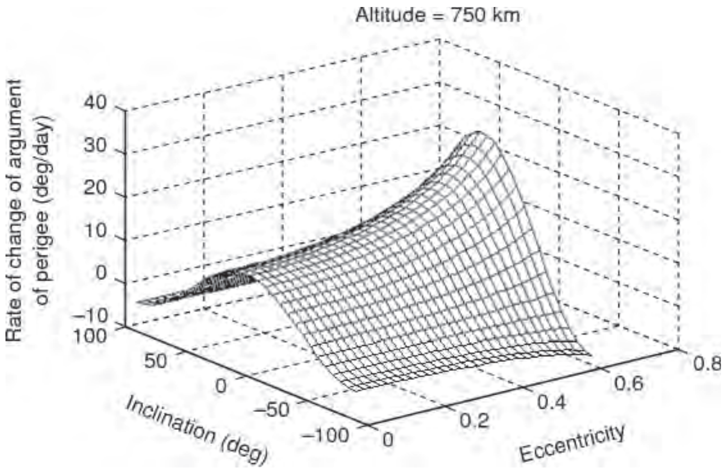


Figure 2.9 Precession of perigee



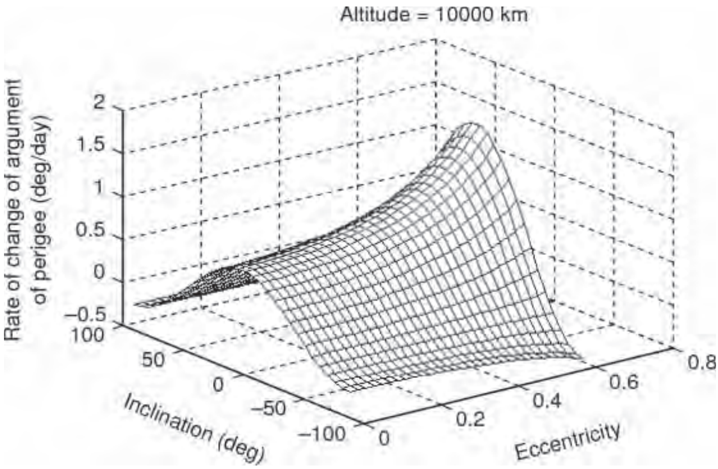
(a)



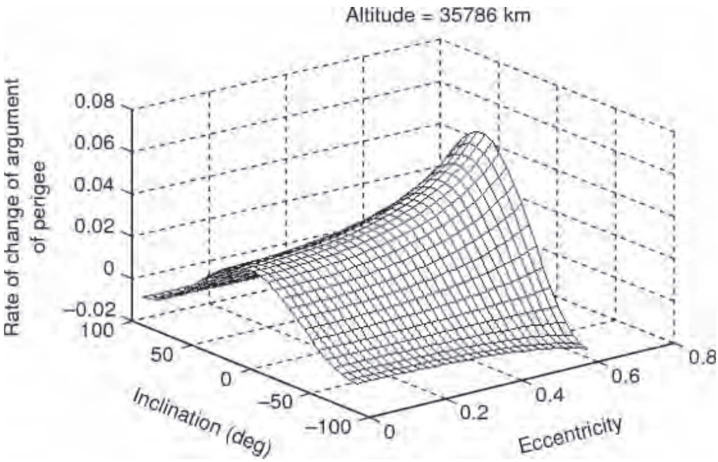
(b)

Figure 2.10 (a) Variation in rate of change argument of perigee versus semi-major axis for a number of inclination angles (b) A three-dimensional representation of variation in rate of change of argument of perigee versus inclination angle and eccentricity for 750 km altitude LEO (Both parts source: Graphics AR.)

regions of the Russian region for satellite communications. *Tundra* orbits belonging to this class of orbit with an apogee of ~46 300 km and a perigee of ~25 250 km have also been proposed for mobile satellite communications. Note that these orbits cannot provide true world-wide coverage on their own and, hence, are combined with other types of orbit when world-wide coverage is desired.



(c)



(d)

Figure 2.10 (c) A three-dimensional representation of variation in rate of change of argument of perigee versus inclination angle and eccentricity for 10000 km altitude MEO (d) A three-dimensional representation of variation in rate of change of argument of perigee versus inclination angle and eccentricity for a GEO (Both parts source: Graphics AR.)

Precession of Orbital Plane around the Earth's North-South Axis

Figure 2.12 illustrates precession of the orbital plane around north-south axis, essentially precession of the ascending node.

The rate of precession Ω is given as

$$\Omega = 9.95(R/a)^{3.5}(\cos(i)/(1 - e^2)^2)\ ^\circ/\text{day} \tag{2.5}$$

Rotation is in a direction opposite to the motion of the satellite.

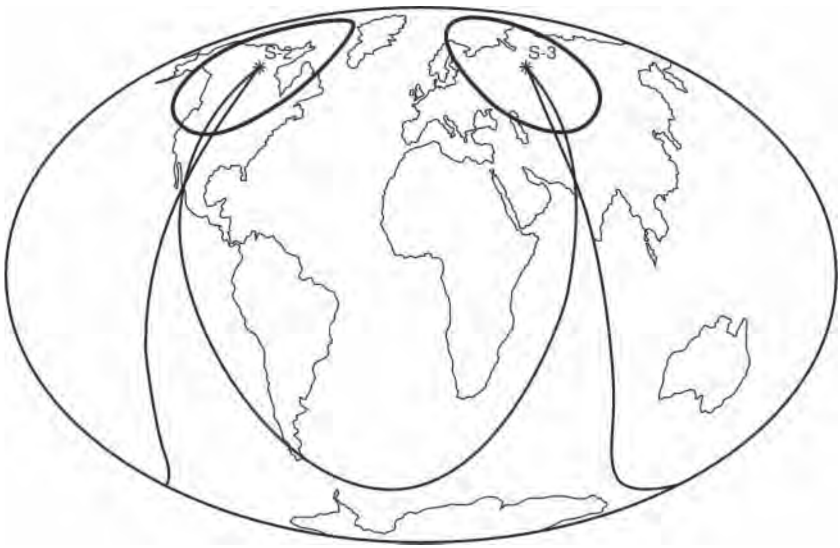


Figure 2.11 Coverage of a satellite from apogee of an elliptical orbit (Source: Graphics AR.)

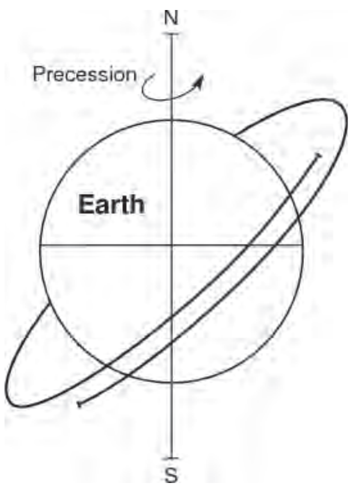
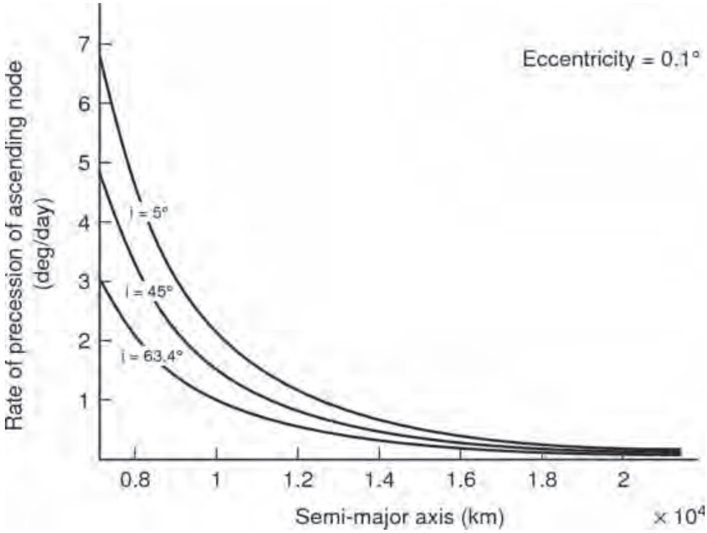


Figure 2.12 Precession of orbital plane

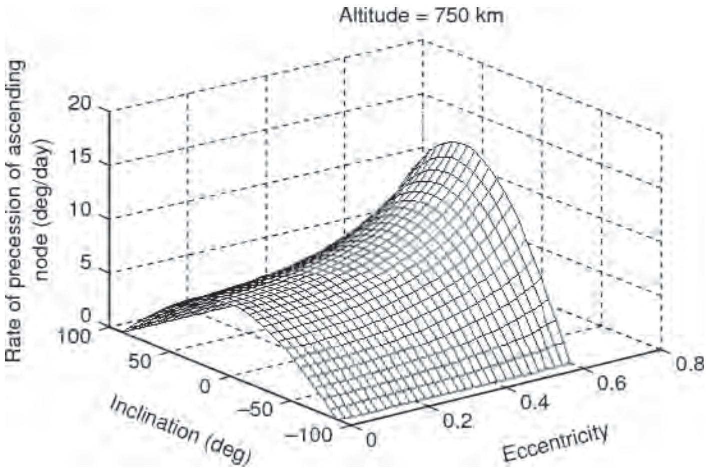
Figure 2.13(a) shows a variation in rate of precession of ascending node versus semi-major axis for a number of inclination and Figure 2.13(b) and (c) shows a three-dimensional view of the variation for low and medium orbit satellites respectively.

Precession of the orbital plane can, in fact, be utilized to give a useful class of orbit known as the *Sun-synchronous* orbit.

A satellite's orbit is referenced to the Earth and as the Earth moves around the Sun, the angle between the Sun and a satellite's orbital plane changes. If the orbital parameters, a, e and i, are adjusted such that the rate of precession of the orbit equals that of the Earth around the Sun, that is 0.986° per day, then the Sun–Earth vector remains invariant



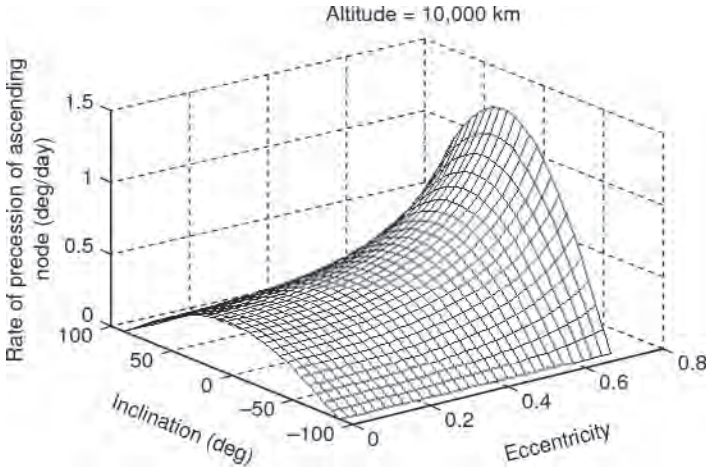
(a)



(b)

Figure 2.13 (a) Variation in rate of precession of ascending node versus semi-major axis for a number of inclinations (b) A three-dimensional view illustrating rate of precession of ascending node for a LEO (orbital altitude = 750 km) (Both parts source: Graphics AR.)

This is a Sun-synchronous orbit. Satellites in such an orbit always view the Earth under identical lighting, a useful attribute for earth resource survey and since satellite rise time at a location remains unchanged, it is easy for users to locate satellites and for earth resource. Furthermore, satellite tracking requirements at Earth stations can be simplified because a satellite follows an identical arc on each pass and if the orbital plane is made normal to the Sun, satellites never undergo eclipse. This orbital configuration simplifies satellite power supply requirements (see Chapter 6). Figure 2.14 illustrates an interesting perspective for the creation of a Sun-synchronous orbit using various combinations of a, e and i.



(c)

Figure 2.13 (c) A three-dimensional view illustrating rate of precession of ascending node for a MEO (orbital altitude = 10000 km) (Source: Graphics AR.)

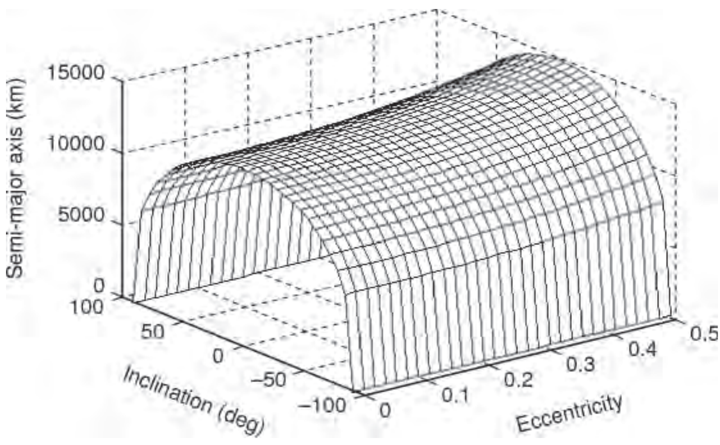


Figure 2.14 Solution space for achieving Sun-synchronous orbit with various combinations of inclination, eccentricity and semi-major axis (All parts source: Graphics AR.)

Perturbing Force along the Orbit

The non-uniformity in the Earth's gravitational field also causes a force vector along the satellite that varies around the Earth. The resultant force tends to cancel out in LEO or MEO because the position of satellites relative to the Earth keeps changing. However, the resultant force for geostationary satellites does not change because the position of a satellite relative to the Earth remains unchanged and hence satellites, depending on their orbital location, undergo acceleration towards one of the two stable points located approximately on the minor axis of the Earth's ellipsoid. Thus satellites, when left uncorrected, drift towards the nearest stable position of ~ 75°E or ~ 255°E longitude and oscillate around the point as they keep overshooting it due to inertia. Observations of free-drifting satellites indicate that these positions are closer to ~ 79°E or ~ 252.4°E (Morgan and Gordon, 1989).

In practice, geostationary satellites must remain within close limits of the designated location in compliance with Radio Regulations. Therefore, on-board thrusters are fired regularly to correct the drift and maintain the specified orbital location. This type of orbital manoeuvre is known as east–west station keeping.

2.2.1.2 Gravitational Effects of the Sun and Moon

For satellites in LEOs and MEOs, the gravitational effects of the Sun and Moon are negligible in comparison to the Earth's gravitational effects. However, satellites in GEOs are perturbed by the gravitational forces of the Sun and Moon. The main effect of the gravitational pull by the Moon is a change in the inclination of the orbit between ~ 0.48 and ~ 0.67 °/year with a period of about 18.6 years. The minimum occurred in the year 1997. The change is caused by variation in the inclination of the Moon itself. The yearly change in inclination due to the Sun is ~ 0.27°, which can be considered steady for practical purposes. The net effect of these two forces is to change the inclination of a geostationary satellite between ~ 0.75° and ~ 0.94° each year. These two forces act in the same direction on an average and hence the resultant is approximately the sum of these forces.

A component of the force due to the Earth's ellipsoidal shape is in a direction opposite to the force of the Sun and the Moon. These forces cancel each other at an inclination of about 7.5°. Consequently if the inclination of a geostationary satellite is left uncorrected then the inclination of the satellite oscillates around an inclination of 7.5° with a maximum of 15° and a period of 53 years. In practice, inclination is corrected regularly by firing thrusters on satellites in an orbital manoeuvre called *north–south station keeping*.

The north-south manoeuvre requires more fuel than the east-west station keeping. Since there is no regulatory limit on inclination, operators position the satellite at one extreme of tolerable inclination at the beginning of life and let the satellite drift to the other extreme before initiating the north–south manoeuvre in order to economize on fuel. For example, with an inclination drift of ~ 0.94 °/year and an operational limit of ±3°, a north–south correction becomes necessary only every ~6.4 years.

2.2.1.3 Atmospheric Drag

Atmospheric drag is caused by friction to a satellite's body, in the upper parts of the Earth's atmosphere. Therefore, satellites in LEO suffer the largest atmospheric drag. Below ~180 km, the friction causes excessive heat on a satellite's surface such that satellites burn out. From this perspective this altitude is considered as the lower limit of space. Atmospheric drag is directly related to the surface area and mass of a satellite and begins to become noticeable below ~750 km.

The orbital lifetime of a satellite is a complex function of the orbit, geometry and mass of the satellite and ionospheric conditions. Note the difference between the functional and orbital lifetimes of a satellite. The functional lifetime of a geostationary satellite may be of the order of 10–15 years, whereas its orbital lifetime can be thousands of years. On the contrary, a satellite in a 400 km orbit may have an orbital lifetime of a few months, whereas its functional lifetime is likely to be much longer.

2.2.1.4 Solar Radiation Pressure

When using large space structures it is necessary to consider the effect of solar radiation pressure. This is generally applicable to large geostationary satellites often deployed for MSSs. The net effect is an increase in the eccentricity together with disturbance along the north–south axis of the satellite that necessitates periodic corrections.

2.2.1.5 Miscellaneous Disturbances

A number of other types of forces act on a satellite, which must be compensated to maintain the desired orientation of satellite. These include effects of the Earth's magnetic field, meteorites, self-generated torque, and etc.

2.2.2 *Satellite Coverage*

2.2.2.1 Map Projections

The region illuminated by a satellite, where reliable communication is feasible, is called the *coverage area* or *footprint*. Within the coverage, service may only be offered to a selected region called the *service area* of the satellite. The shape and size of the coverage area depends on the satellite's antenna radiation pattern (see Chapter 6). The most elementary pattern is bell-shaped. The EOC area depends on the minimum permissible elevation of the satellite. The reliability of MSS radio link increases with an increase in elevation angle but the coverage area shrinks (see Figure 2.6). Minimum elevation angles of 5, 10 and 40° have been proposed, with 10° as a reasonable compromise.

To make better use of satellite resources and reduce the size and capacity of the mobile terminal, the coverage is divided into a number of small zones called *spot beams*. The size of spot beams is governed by on-board antenna technology and ranges from a few to several hundred at the L band with the existing technology. To represent the coverage area of a satellite, its antenna pattern is projected to a map of the Earth as iso-contours of antenna gain or elevation angle. Other types of map representations include iso-contours of effective isotropic radiated power (EIRP) and satellite receiver sensitivity given as antenna gain/system temperature (G/T).

A number of map projections are used, each suited to the application at hand. *Mercator projection* used commonly and is named after Gerardus Mercator, who published a world map on such a projection for ship navigation in 1569. In this type of projection, the longitudes are represented by vertical lines spaced equally and latitudes by horizontal lines, which are spaced closer near the equator than towards the pole. In such a map, coverage patterns get distorted. A number of plots with this projection appear in this book (e.g. see Figures 11.2a and 11.3a).

Satelli-centric or *hodocentric projection* is used when it is necessary to preserve the shape of the antenna pattern. This projection is a view of the Earth from a satellite. See Figure 11.2(b) for an example.

Polar projection is used by radio amateurs as it is quite simple for plotting ground tracks. In a *rectangular projection* the X-axis is represented as longitude and the Y-axis as latitude.

The coverage contour of a satellite in GEO is stationary and hence visualization and design of the service area is simple as the relative position of any location within the service area is invariant. In non-geostationary systems, the path geometry varies with time, which requires a statistical description of visibility statistics from any given point. Section 2.3.8 summarizes visibility statistics of some non-geostationary satellite systems (see appendix A for coverage snapshot of various NGEO constellations).

Consider the salient coverage features of a few satellite orbits used in an MSS.

Polar orbits provide unbiased coverage to all geographic regions – a single satellite can view the entire Earth over a period of time. Satellites in a polar orbit undergo a large number of eclipses and require regular station-keeping to maintain their orbital position. These orbits are well-suited for applications, such as world-wide MSSs and global Earth resource survey. Satellite constellations in a polar orbit favour polar regions because all the satellites in the constellation converge towards the pole.

Coverage provided by inclined orbits between 0 and 90° depends on the inclination of the orbit. Figure A.1 (in Appendix) shows the ground trace of ICO satellite constellations that are in a 45° inclined MEO.

GEO is the most commonly used orbit in satellite communications because of its numerous advantages. Figure 11.3(a) illustrates a 5° elevation contour of an Inmarsat system, which deploys four geostationary satellites for its global mobile and distress communication services. Some of the main features of GEOs are summarized in Table 2.2 (see Section 2.2.8).

2.2.3 Space Environment

Space environment affects the choice of altitude of an orbit as shown in Figure 2.15, which depicts the radiation environment around the Earth.

The Earth's magnetic fields deflections get trapped around the Earth's radiation belts called the Van Allen radiation belts, named after the American scientist who discovered them. From Figure 2.15, we note that the radiation intensity varies gradually with altitude, increasing to the highest intensity of 10 000 counts between ~ 2000–4000 km and ~ 13 000–20 500 km. The intensity also varies around the Earth, being highest above the equator and lowest above the poles. Regions in the south Atlantic between the middle region of South America and the lower tip of Africa exhibit much higher radiation levels than other regions at the same latitude. This effect, called the South Atlantic anomaly, affects satellites mostly in LEO.

High ion density causes adverse effects on electronic devices. Ionization causes an increase in background noise, increases the rate of degradation of solar cells and causes single-event failures, which result in either temporary or permanent failure of electronic components. Single-event failures increase as the level of integration of electronic components in onboard electronic circuits increases.

Taking atmospheric drag as the lower limit and the first Van Allen radiation belt as the upper limit, the approximate altitude of LEO is ~750 and ~1500 km. The altitude of MEO usually lies in the range ~10 000 to ~12 000 km, the window between the first and second Van Allen belts.

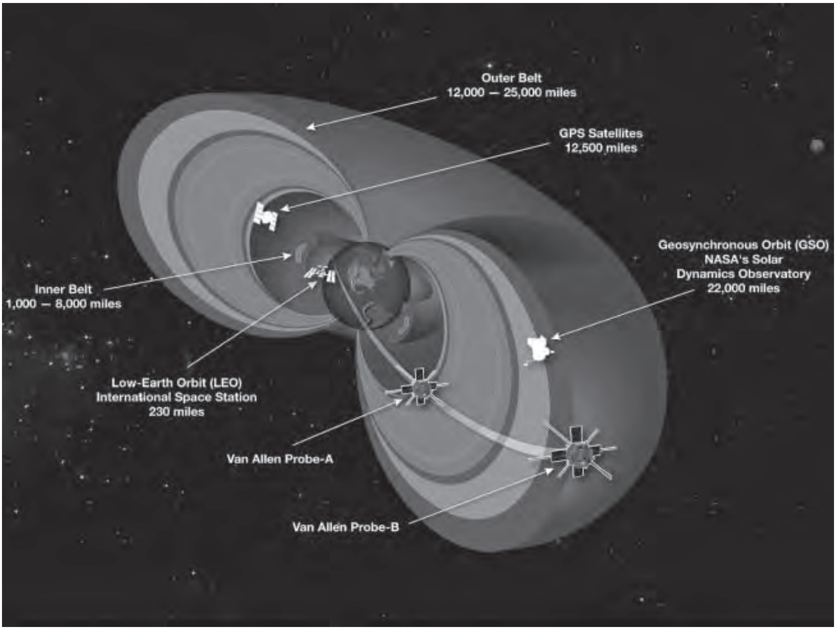


Figure 2.15 A cut-away model of the radiation belts with the twin Van Allen probes flying through them. The radiation belts are two doughnut-shaped regions encircling the Earth, where high-energy particles, mostly electrons and ions, are trapped by Earth's magnetic field. [Note: NASA's twin spacecrafts known as Van Allen Probes launched in August 2012 are credited to have revealed a third radiation belt, which exhibits a transient behaviour (NASA, 2013)] (Source: NASA, 2013. Reproduced with permission of NASA.)

2.2.4 Eclipse on Satellites

Satellites derive their electrical power from sunlight. Just as we occasionally observe the solar eclipse on the Earth, so also do satellites in their respective orbits; the number and duration of eclipses depends on the altitude, inclination and eccentricity of the orbit. When the Sun is in an eclipse, satellites are powered by storage batteries that are subsequently charged when the Sun reappears. Regular charge/discharge cycles reduce the lifetime of batteries. Eclipses also cause a load on a satellite's thermal control system as the system must react to rapid changes in temperature when satellites enter and leave an eclipse. As mentioned before, a Sun-synchronous orbit oriented at right-angles to the Sun-orbital plane vector eliminates eclipse occurrence and therefore offers an advantage in this respect.

To understand the principle, consider the geometry of an Earth-induced solar eclipse on a geostationary satellite. Figure 2.16(a) shows the angular variation of the ecliptic plane (i.e. plane of the Earth around the Sun) with respect to the Earth's equatorial plane.

We note that at certain points the equatorial and ecliptic planes coincide and therefore the Earth lies in the same line as the Sun and the geostationary arc. These points correspond to the Spring and Autumn Equinoxes, respectively. The Earth's shadow subtends an angle of about 17.4° at the altitude of the GEO and the Sun is eclipsed on days when the shadow intercepts the GEO during intervals when a satellite lies behind the Earth.

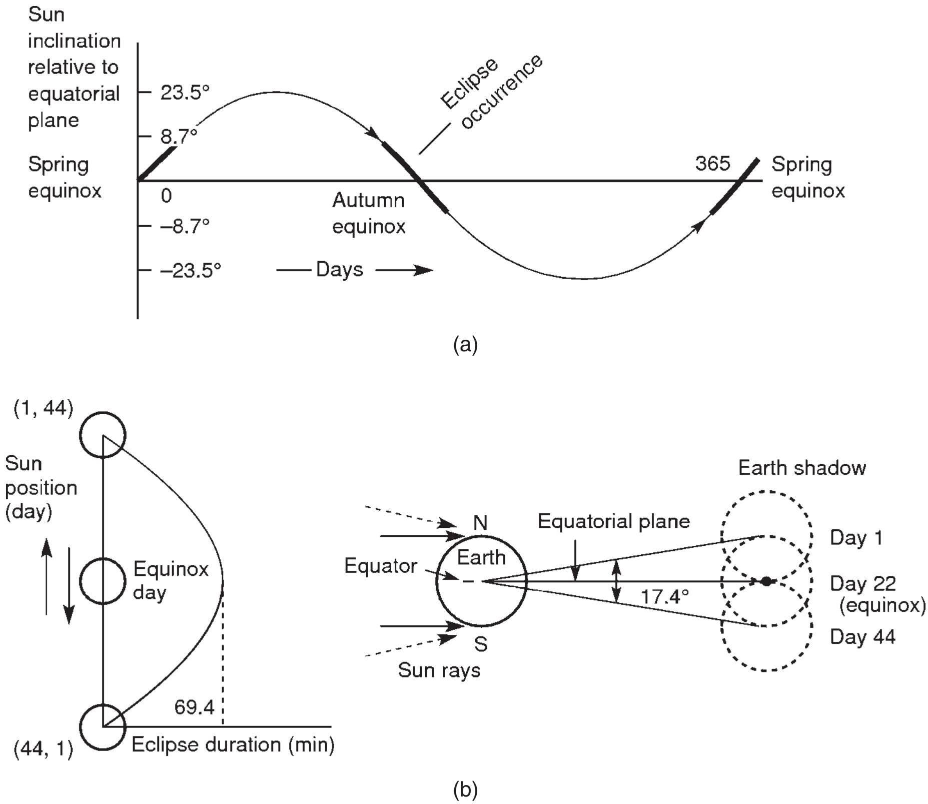


Figure 2.16 (a) Motion of the ecliptic plane with respect to the Earth's equatorial plane over a year (b) Eclipse duration in minutes on a geostationary satellite (left) and the Earth's shadow on geostationary arc on days 1, 22 and 44 of the eclipse (right) (Both parts source: Richharia, 1999. Reproduced with permission of Palgrave-Mamillan.)

Figure 2.16(b) shows the Earth's shadow on the geostationary arc progressively from the first day of the eclipse when the shadow begins to graze the GEO to the last day when it moves past the orbit. The duration of the eclipse increases progressively, maximizing on the day of equinox, and then subsides as the shadow moves away from the geostationary arc. On the day of the equinox, the Earth and the GEO are on the same line and hence the duration of the eclipse is highest, reaching around 69.4 min, which is essentially the time it takes the satellite to traverse 17.4°. The Sun is shown moving relative to a fixed Earth and directions represent the apparent motion of the Sun relative to the equator around the Spring and Autumn Equinoxes (see Figure 2.16b: left).

The peak eclipse occurs at the sub-satellite midnight. Therefore, by selecting the satellite to the west of the coverage region, the onset of the eclipse can be delayed past the region's midnight when the traffic carried by a satellite is usually low and hence the spacecraft battery's capacity requirement is reduced. Figure 2.17 shows the time of occurrence of the eclipse for various longitudes relative to satellite longitude. For example, setting the satellite 15° to the west would set back the eclipse to 01:00 a.m.

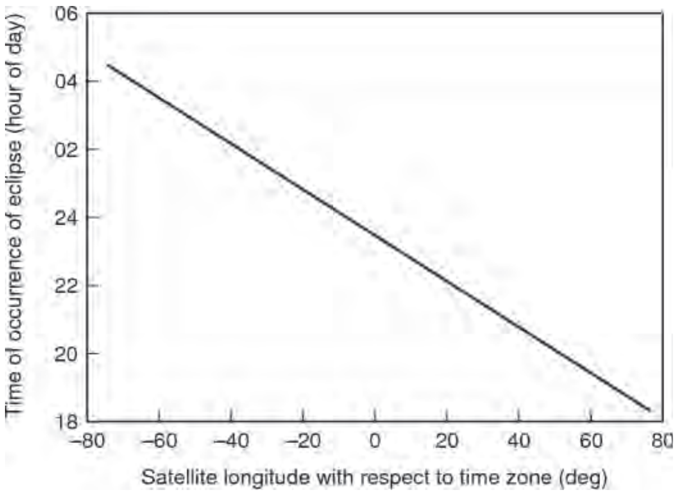


Figure 2.17 Time of occurrence of eclipse for various longitudes relative to a satellite longitude for a GEO (Source: Graphics AR.)

LEO satellites can undergo several thousand eclipses in a year; the number depends on the relationship between the Sun-Earth vector, altitude and the inclination of the orbital plane. Some eclipses could last for a considerable part of the orbital period. For example, a satellite in an equatorial orbit at an altitude of 780 km can remain in the Earth's shadow for 35% of the orbital period. Such eclipses can occur over 14 times per day during the equinox and the total eclipse period could last over 8 h in a day. For a MEO under similar conditions, the maximum eclipse duration would be about 12.5% of the orbital period and the total duration would be of the order of about 3 h in a day, with a maximum of about four eclipses per day. Furthermore, the eclipse can occur at any time. Figure 2.18 presents the maximum fraction

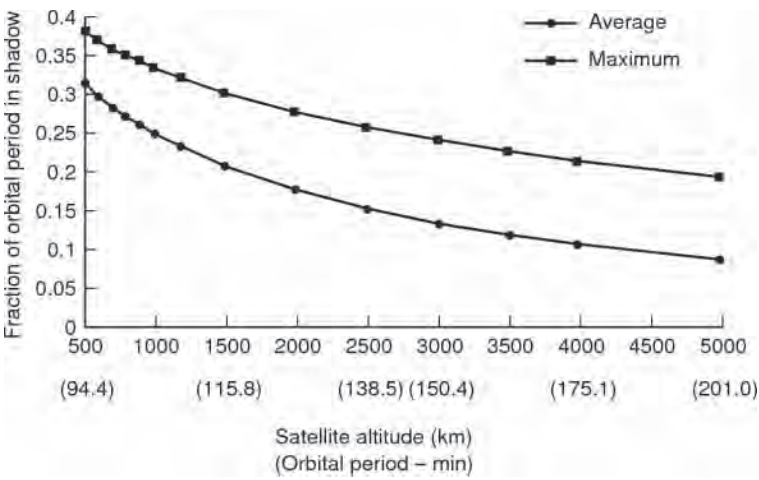


Figure 2.18 Maximum and average fraction of the orbital period that a satellite stays in the Earth's shadow as a function of orbital altitude (Source: Graphics AR.)

and average fraction that low Earth satellites spend in the Earth's shadow (data from Gavish and Kalvenes, 1998).

The Sun can also be eclipsed by the Moon, when the Moon's shadow falls on a satellite. However, occurrence of the Moon (lunar) eclipse is less predictable because of time-variant geometric relationships and may occur at any time on a geostationary satellite, unlike the Earth eclipse that is easily predictable and tends to occur late at night when the traffic is low.

2.2.5 *The Sun's Interference*

The Sun is a strong radiator of electromagnetic radiation. Therefore, whenever the Sun appears behind a satellite, it receives strong radiation that can cause disruption to service. Generally, the effect is quite small in mobile terminals because receiver noise is significantly greater than the Sun's noise. However, a disruption in communication can occur in gateways that are far more sensitive. Solar interference occurs within ± 22 days around the equinox when the Sun's declination angle equals the angle between an Earth station and the equator. Figure 2.19(a) and (b) respectively show the maximum number of days when interference occurs and the duration of the Sun's transit on the day when the passage is the highest for various antenna beamwidths.

2.2.6 *Doppler Effect*

In a mobile communication system, the frequency shift caused by the Doppler effect must be compensated to minimize the effect of frequency error on demodulation of signal. Whenever a moving object transmits a wave, the received frequency changes by an amount

$$\Delta f_d = \pm v_r f_t / c \tag{2.6}$$

where v_r = relative radial velocity between the observer and the transmitter, c = velocity of light and f_t = transmission frequency.

The sign of Doppler frequency shift is positive when the satellite is approaching the observer. The Doppler shift is related to the radial velocity, which can be predetermined and this feature can be utilized at the receiver for frequency compensation. For a LEO satellite, an observer typically sees an 'S-curve' with an increase in frequency as the satellite rises and a reduction as the satellite moves away. Figure 2.20 shows the Doppler curve of an L band signal received at a point near Bhopal in central India for LEO and MEO satellite passes. In passing, we note that Doppler shift measurements at a location can be used to estimate the position of the receiver, a principle used in some navigation systems.

2.2.7 *Orbital Debris*

There are two types of debris that can affect satellites – natural and man-made. Natural objects most likely to affect satellites are isolated meteorites or showers, such as the well-known Leonid showers that were in the news in 1966 and 1999–2000, as their activities peaked. Satellite operators take precautions to minimize the impact of such showers on spacecraft electronics. There is a growing concern regarding the number of man-made objects that are cluttering space as new instances of debris collision or debris avoidance manoeuvre with operational spacecraft get reported – for example the Iridium33-Cosmos

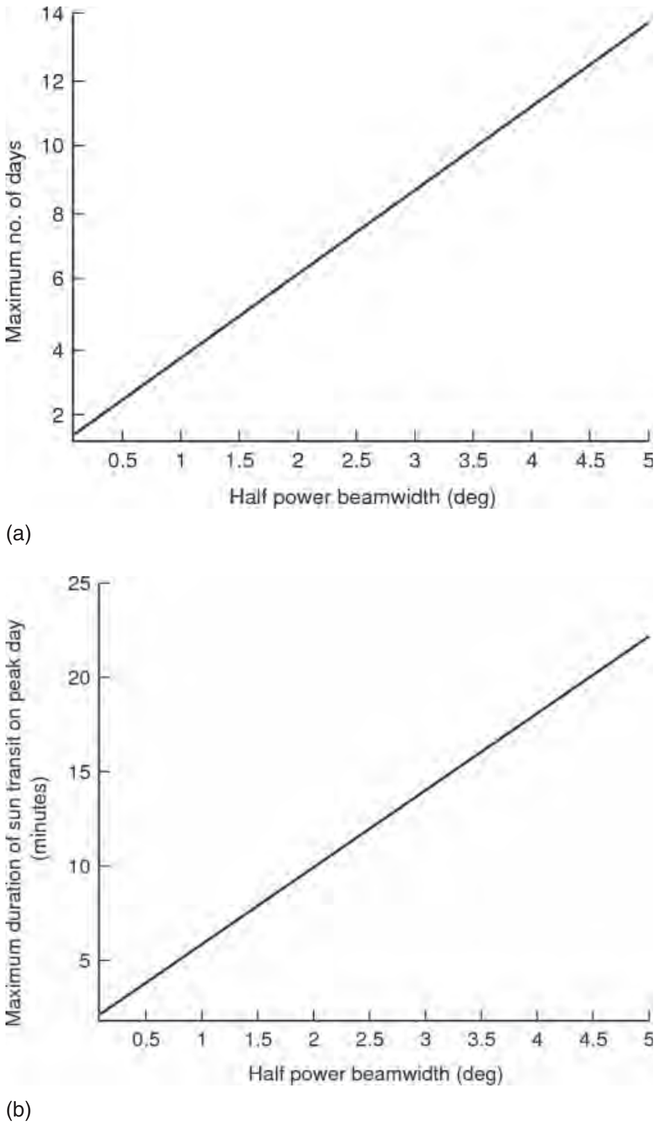


Figure 2.19 (a) Maximum number of days when Sun interference occurs in a GEO satellite for various antenna beamwidth and (b) duration of Sun transit when sun passage is the highest for various antenna beamwidth (Both parts source: Graphics AR.)

2251 collision in 2009 at an altitude of 790 km. Hundreds of thousands of objects, from sub-millimetre in size to several metres, are estimated to be orbiting the Earth and several hundred new objects are catalogued each year. Figure 2.21(a) shows the number of catalogued objects in Earth orbit up to the year 2010, categorized as fragmentation of space hardware, defunct spacecraft, mission related debris and defunct rocket bodies (NASA, 2010). Catalogued objects are typically > 10 cm s such that they can be tracked

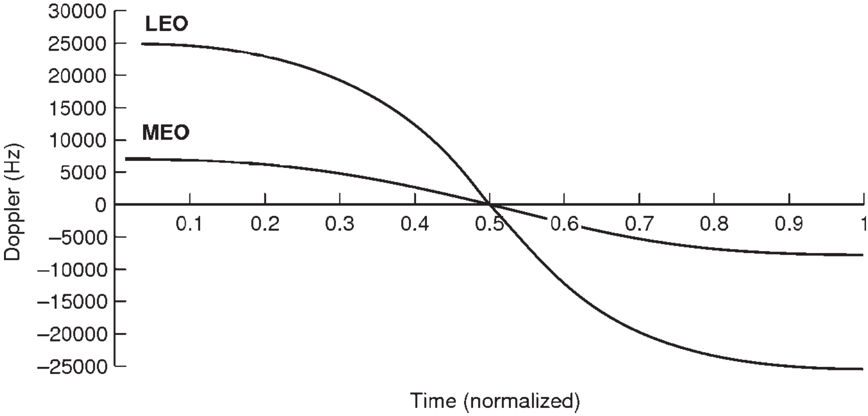
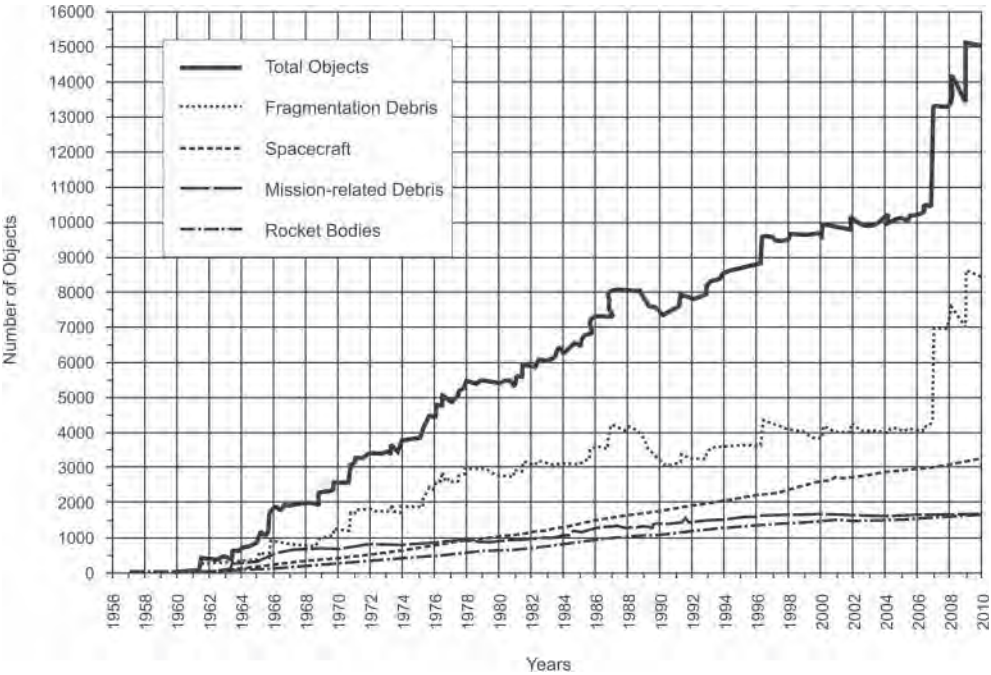


Figure 2.20 Doppler frequency shift as viewed from a location in central India (77.5°E and 23°N) for a LEO satellite: altitude = 1400 km, i = 52°, ascending node = 60°E and a MEO satellite: altitude = 10 350 km, inclination = 45°, ascending node = 60°E (Source: Graphics AR.)



(a)

Figure 2.21 (a) Number of catalogued objects in Earth orbit up to the year 2010 (Source: NASA, 2010. Reproduced with permission of NASA.)

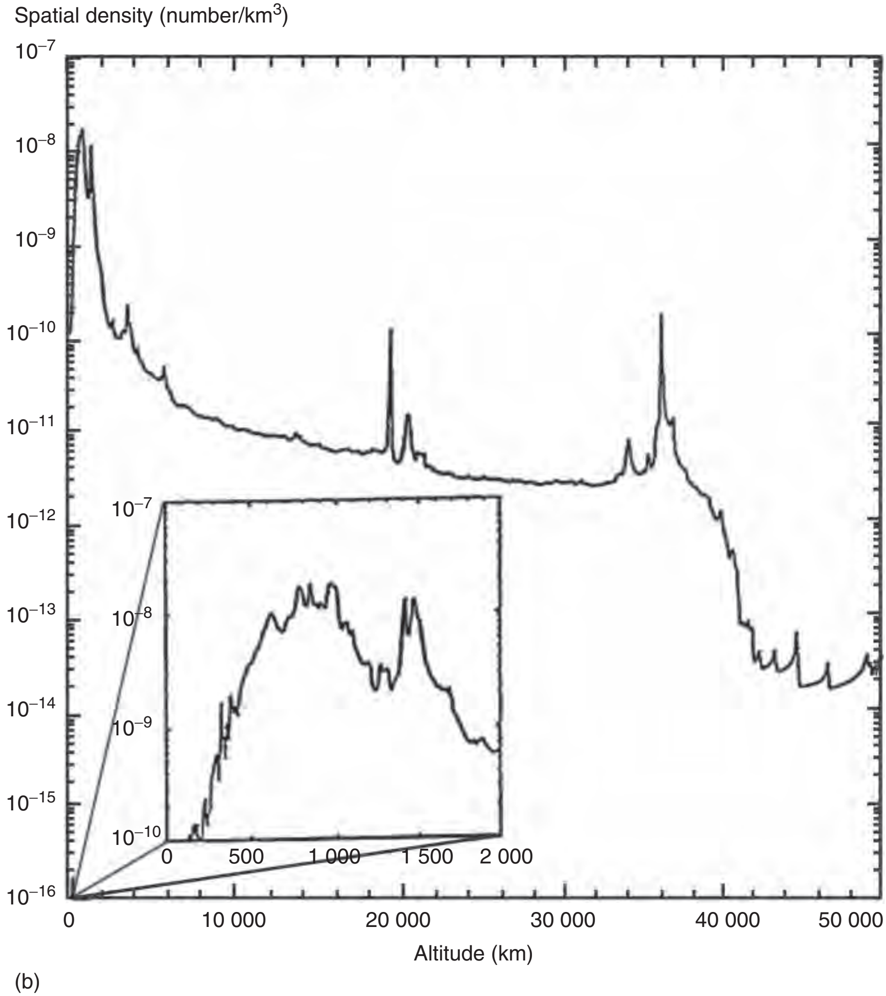


Figure 2.21 (b) representation of altitude profile of debris population until the 21 August 1997, given as number of objects per km^3 (Source: UN, 1999. Reproduced with permission of UN.)

(explained later in this section). Beginning at near zero at the outset of modern space technology in the year 1957, the number increased to ~15 000 by 2010. The growth is generally linear, apart from the discrete jumps of 2007 and 2009, which occurred respectively due to break up of a Chinese satellite Fengyun 1-C in 2007 and the Iridium-Cosmos collision in 2009. Figure 2.21(b) shows the spatial density of catalogued objects as at 21 August 1997 (UN, 1999). Notice the high spatial density at altitudes where a majority of satellite systems operate – in LEO at 750 and 1500 km; at MEO altitude of ~ 20 000 km close to the area where a number of navigation satellite systems operate; and at the GEO altitude of ~ 36 000 km.

It is recognized that proliferation of orbit debris must be minimized, if not eliminated otherwise space will become unusable (UN, 1999). As yet there is no radio regulation for controlling orbital debris, but there is an ITU recommendation to de-orbit a satellite from the GEO at the end of its lifetime; moreover, other practical and advanced measures such as space clean-up options are either in use or under consideration. De-orbiting a satellite requires fuel that can be used to maintain the satellite in orbit thereby earning revenue, and therefore some operators may be reluctant to follow a recommendation unless it is mandatory. There is no regulation as yet regarding tolerance in the orbital parameter of the non-geostationary satellite; to minimize the risk of collisions, regulations will therefore be necessary.

There have been several analytical studies for investigating the evolution of current orbital debris in order to predict the probability of collision hazard in the future and thus facilitate formulation of suitable regulation. Studies conducted by Kessler and Burton (1978) demonstrated that a self-sustaining chain reaction of orbital collision may be initiated when the spatial density of orbital debris reaches a threshold, causing severe restriction to the use of space. At present, atmospheric drag is the most practical way of removing debris from the Earth's orbit and hence below ~750 km, where atmospheric drag effects become dominant, debris decay and eventually get burnt in the Earth's atmosphere (Petro and Talent, 1989). However, above this altitude debris continues to orbit the Earth for hundreds of years unless a suitable and cost-effective debris-removal technology such as retrieval is applied.

Studies of space debris and its impact on future space endeavours address a number of inter-related topics:

- Techniques to estimate numbers and distribution of space debris.
- Effects of space debris on present and future space operation.
- Modelling of debris distribution in short term of days and months, medium term of up to 10 years and long term covering periods extending to a 100 years.
- Risk assessment.
- Mitigation methods.

Details of these topics are outside the scope of the book and hence we will merely introduce the topics due to potential long-term effects of space debris on satellite communications (UN, 1999).

Ground-based measurements utilize radars or optical systems such as telescopes and *space-based* measurements utilize sensors and satellites developed specifically for estimating the effects and statistics of debris. Radars are used for measuring debris in LEO orbits and optical measurements are used for HEO orbits, as many objects continue to be illuminated by the Sun, even at night, in these orbits against a dark background, and accuracy of optical systems in comparison to radar tends to be better at such altitudes due to strong radiation intensity available from solar rays.

Space-based measurements assess the impact of debris on satellite hardware brought back to the Earth by analysing the surfaces exposed to the space environment. Such measurements are generally limited to LEO for cost reasons. Examples include Long-Duration Exposure Facility (LDEF) of the USA, European Retrievable Carrier experiment and Salyut and Mir missions of the Russian federation. Debris monitoring in space by specialized satellites is also used. Examples include infrared astronomical satellites that point at the celestial sphere for object siting and the impact ionization detector Geostationary Orbit Impact Detector (GORID) on-board the Russian telecommunication satellite Express 12.

A catalogue is a record of debris characteristics such as orbit, radar cross-section, mass, and so on, that have been derived from measurements or records. Catalogues are maintained by the US and Russian Federation. They are used for various practical purposes such as maintenance of historical records, inputs to modelling of debris behaviour and management of operational activities such as collision avoidance. The minimum size of catalogued debris is 5–10 cm and statistical methods are used to estimate debris population of smaller objects. Whether a debris environment affects a space system depends on factors such as its projected area, altitude and inclination. Whereas large debris of > 10 cm are known to be potential threats, smaller debris do not cause mission failure but cause damage to surface and subsystems. Nevertheless precautions are necessary to minimize risk of impacts to humans in human-related space operations. Space debris also tends to contaminate observation windows of astronomers.

A number of mathematical models have been developed to predict the distribution and evolution of orbital debris. Such models provide a mathematical description of the distribution, movement and the physical characteristics of debris for risk assessment, avoidance manoeuvres and long term assessment. Deterministic models provide a concise description of each object; statistical models treat the debris as an ensemble to describe their behaviour statistically; and hybrid models combine the relevant features of each. Considerations include the debris *source* such as launch or satellite breakups and *sink* such as removal characteristics of atmospheric drag below 750 km (McInnes, 1993). For example, the observed spatial debris density data available in early 1990s was projected by spline interpolation (Klinkrad and Jehn, 1992). Figure 2.21(b) shows an updated version of such data (UN, 1999). In this particular effort, debris evolution was approximated as a hydrodynamic process allowing the derivation of an analytical expression, which described collision evolution of debris under various scenarios. The most interesting scenario represented debris evolution with a slowly increasing rate of debris deposition to represent future debris deposits. Results demonstrated that catastrophic growth through collisions is likely to occur at an altitude of 1000 km where the debris population is at its highest, although there is a bias towards lower altitude due to the increased mean collision speed. For the case when the deposition rate is constant, the catastrophic growth occurs at an altitude of ∼ 1000 km after about 150 years (see Figure 2.22) and for the case when the rate of deposition and collision is increased by 1% per year due to the growing volume of satellites, the situation can occur at the same altitude by 2100 and if deposition and collision are increased by 10%, the catastrophic growth is brought forward to 2050 (Figure 2.23). It is observed that the dates may be postponed by depositing spent satellites at local minima, such as within the high radiation zone of the Van Allen belt. The results provide an insight into the altitudes and approximate period at which problems may occur and its sensitivity to debris deposition. Another element of interest is the probability of occurrence of collision to an operational satellite. Clearly at each altitude, the collision probability will increase in the same proportion as the growth in orbital debris, rising exponentially in the orbital belts of catastrophic collision.

Limitations of such models include sparse data volume that affects validation; lack of reliable mechanisms to predict orbital failure; uncertainty in future developments such as change in space traffic or introduction of innovative debris-retrieval systems; and limitations of the mathematical model itself. Hence, these models must be updated regularly. Table 2.1

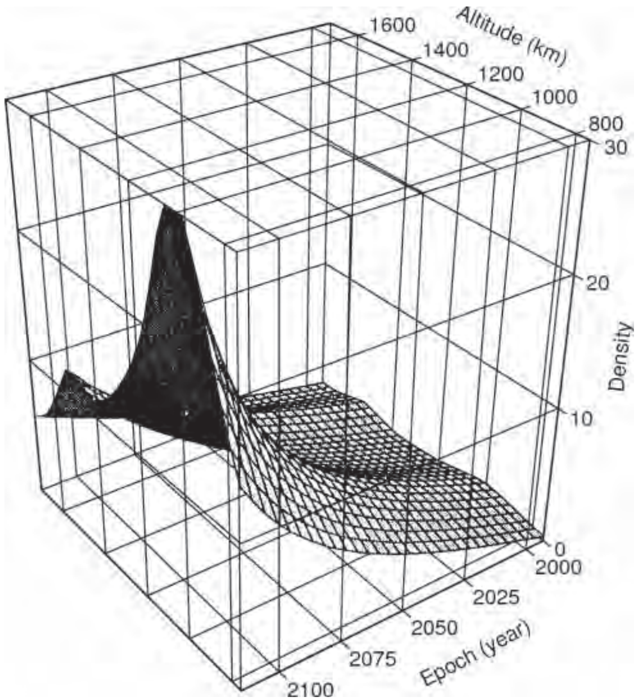


Figure 2.22 Growth in density of orbital debris when deposition and collision growth are constant (Source: Klinkrad and Jehn, 1992. Reproduced with permission of ESA.)

summarizes various aspects of short- and long-term models developed over the years. Due to variations in assumptions the results do not always agree and hence an envelope encapsulating all the predictions is a useful indicator. The general trend emerging from these models can be summarized as follows:

- Debris population will grow at an increasing rate because of an acceleration in collision probability as the number of space-borne objects increases.
- A majority of existing orbital debris is due to fragmentation from explosions and these may remain the biggest debris source in the future.
- An exponential growth may occur sometime in the future due to multiplicative impact of collision fragments.
- Even if instances of collisions are reduced or eliminated, collisions will still occur due to a growth in the number of space-borne objects.
- Space debris mitigation techniques will reduce the probability of collision; examples of such techniques in use include: measures to limit mission-related debris, prevention of accidental explosions, ejection of GEO satellite out of this orbit at the end of a mission and deorbiting LEO satellites to lower altitude for burnout.

The emerging risk trend is as follows:

- The probability of collision of operational satellites with man-made debris poses little risk at present but the risk may grow if no action is taken;

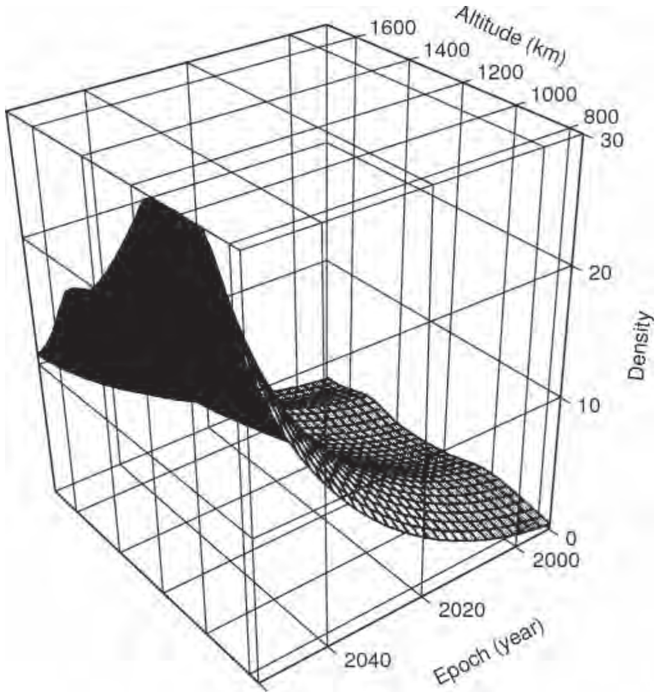


Figure 2.23 Growth in density of orbital debris when deposition and collision growth are 10% (Source: Klinkrad and Jehn, 1992. Reproduced with permission of ESA.)

- The probability of collision in LEO is, at least, a magnitude higher than in MEO;
- Probability of collision with large objects of the order of >10 cms is considerably lower than from smaller debris;
- Probability of collision of a catalogued object with a GEO satellite was estimated at 10^{-5} towards the end of 1990s (UN, 1999);
- Spacecraft designers already take countermeasures such as shielding of critical spacecraft parts.

2.2.8 Summary of Orbital Characteristics

We have categorized satellites according to inclination (e.g. equatorial, polar) eccentricity (i.e. circular or elliptical) and altitude (e.g. LEO, MEO or GSO). Equatorial, inclined or polar orbits favour equatorial, mid-latitude and polar regions, respectively. Eccentricity increases the dwell time over specific regions. Satellites dwell longer over the apogee and it is therefore possible to take advantage of this prolonged dwell. Both path loss and propagation delay reduce as *altitude* is lowered, each is a useful attribute for provision of low-delay services to small terminals. Reducing the altitude reduces its coverage area, thereby increasing the number of satellites in a constellation and hence increasing the complexity of the network. Main features of LEO, MEO and GEO are summarized in Table 2.2 (Richharia, 1999).

Table 2.1 Models to predict short and long term evolution of orbital debris.

Model name	Source	Period	Minimum debris size	Model basis	Altitude
CHAIN	NASA	Long term	1 cm	Analytical model	LEO
CHAINEE	ESA	Long term	1 cm	Analytical model	LEO
EVOLVE	NASA	Short and long term	1 mm	Quasi-deterministic population propagation techniques	LEO
IDES	DERA	Short and long term	0.01 mm	Semi-deterministic model using historical and future traffic models	LEO
LUCA	Technical university of Braun-schweig	Long term	1 mm	Semi-deterministic model; detailed traffic model; uses several source model options for explosions, collisions and leaks	LEO/MEO
MASTER	ESA	Short term	0.1 mm	Semi-deterministic environment model based on 3-D discretization of spatial densities and transient velocities	LEO/GEO
Nazarenko	Russian science academy	Short and long term	0.6 mm	Semi-analytic, stochastic model; takes account of averaged debris sources and atmospheric drag	LEO
ORDEM96	NASA	Short term	1 µm	Semi-empirical, based on remote and *in situ* observations (used for International space station operations, etc.)	LEO
SDM/STAT	ESA/CNUCE	Short and long term	–	Semi-deterministic model takes into account traffic model and source model options for explosions, collisions, etc.	LEO/GEO

(Data source: UN, 1999.)

2.3 Satellite Constellations

In this section the concepts developed in the preceding section are applied to the problem of optimization of satellite constellations from an orbital viewpoint. We will consider orbital altitude, space environment and geographical coverage as the optimization criteria for minimizing the number of satellites for *real-time* and *non-real-time* applications. Although orbital characteristics of constellations have a significant influence in the design

Table 2.2 A comparison of orbital features of GEO, HEO, MEO and LEO satellite systems

Orbit	Advantages	Limitations
Geostationary orbit	Well developed and proven technology	Coverage unreliable between ±76° and ±81° latitude, and unavailable beyond ±81° latitudes
	Signal strength stable due to constant ground-satellite range during calls	Poor service link reliability at mid-to-high latitudes, particularly for land mobile satellite services
	Interference effects are easy to predict due to stable and simple geometric relationships	Large propagation delay (~ 240 ms one-way) – affects voice and time-sensitive data protocols
	Low Doppler	Large path loss
	Supports hand-held and broadband service	Spectrum efficiency lower than LEO and MEO
	Coverage is available to most populated areas of world (but see the limitations)	High launch cost
	Only three satellites can provide near world-wide coverage	In-orbit back-up satellite increases system cost disproportionately
	A single satellite is adequate for regional coverage	
Highly elliptical orbit (HEO)	Reliable service links possible at mid-to-high latitudes	Inefficient for global coverage on its own
	Lower launch cost than GEO	Propagation delay can be higher than GEO system
	Distributed space segment architecture allows partial service in case of a satellite failure; 1 : 1 satellite redundancy is not required	Doppler effect quite significant
		Hand-over is essential
		Satellites tend to pass Van Allen radiation belts regularly to the detriment of electronic components
Medium earth orbit (MEO)	Can provide true global coverage	Large number of satellites necessary (10–12)
	Lower path loss than GEO;	Receive signal strength is variable due to variability in range and elevation angle
	Medium propagation delay (~ 55–80 ms)	Doppler effect significant
	Enables efficient use of spectrum	Complex network architecture: e.g. handover, intersatellite links, dynamic satellite resource management, routing, etc.

(*continued overleaf*)

Table 2.2 (*continued*)

Orbit	Advantages	Limitations
	Distributed space segment architecture allows partial service in case of a satellite failure; 1 : 1 satellite redundancy is not required	Tends to increase orbital debris because of need of large number of satellites per system
		Relatively long time necessary for constellation deployment
		Space segment maintenance is complex due to large number of satellites and distributed network architecture – higher number of satellite replacement than GEO but less than LEO
Low earth orbit (LEO)	Can provide true global coverage	Large number of satellites necessary
	Lowest path loss	Signal strength varies rapidly due to variability in range and elevation angle;
	Lowest propagation delay ($\sim$3.5–15 ms one-way from transmitter to receiver) comparable to optical fibre system	Doppler effects is significant (highest)
	Efficient use of spectrum	Most complex space segment and network architecture: high rates of hand-over, intersatellite links, dynamic satellite resource management, complex routing, etc.
	Distributed space segment architecture allows partial service in case of a satellite failure; 1 : 1 satellite redundancy is not required	Possibility of very large number of eclipses with a large number of charge/discharge cycles
	Possibility of including position determination as a value-added service	Rate of depositing orbital debris is the highest but it is easier to remove debris from LEO
		Space segment maintenance is involved – satellite replacement rate is more than in GEO or MEO

(Source: Richharia, 1999. Reproduced with permission of Palgrave-Macmillan.)

of the overall mobile system architecture, a number of practical considerations influence the optimization of the space segment. These may be, for example communication service, network issues, traffic distribution, technological risk and project schedule. Such considerations have resulted in a wide range of constellation architecture, depending on the emphasis imposed by each operator. Optimization of constellation in entirety is discussed in Chapter 7, after the reader has become familiar with the relevant concepts.

2.3.1 Considerations in Constellation Design

Considerable research has been directed to the problem of minimizing the number of satellites for continuous single or multiple visibilities. The purpose has been varied – world-wide communications, communication capacity expansion, global distress communications services, Earth resource survey, scientific data gathering, robust and survivable communications, military reconnaissance, navigation and mobile communications.

In a *random constellation*, all or some constellation parameters such as altitudes, inclination, interorbital plane separation, intersatellite phase are chosen at random whereas in a phased constellation, the phases have a well-defined relationship with one another. Random constellations, although simpler to maintain, are inefficient in terms of coverage property and tend to crowd the celestial sphere randomly. Such a constellation has been used in the past by the former Soviet Union for store and forward communications. Only *phased constellations* are considered here because they are economic, have a more reliable coverage property and do not clutter space. They can be categorized by inclination, eccentricity and altitude. We will summarize the main features of a few interesting and diverse constellation designs (Richharia, 1999).

Constellations using a circular orbit have been developed where orbital planes intersect at a common point, for example a polar constellation. Here such constellations are called type 1 (Luders, 1961; Beste, 1978; Adams and Rider, 1987). Another type of orbit design, which we call type 2 here, optimizes constellations in inclined circular to distribute the satellites more uniformly on the celestial sphere (Walker, 1973; Mozhaev, 1972, 1973; Ballard, 1980). In both types of design, satellites are in circular orbit at a common altitude, that is their time period is the same.

Constellations can also be optimized by controlling eccentricity of the orbits, which allows coverage to favour specific areas. Coverage from circular orbits is unbiased whereas eccentric orbits tend to be biased towards the field of view around the apogee.

It is also possible to deploy hybrid constellations, consisting of a combination of circular and elliptical orbits. They are used to tailor coverage towards specific regions of the Earth – an example of is the ELLIPSO system that was developed to illuminate higher Northern latitudes and up to 55° of Southern latitude, a region that represents the most populated regions of the world (see Section A.1).

In all cases, altitude is an important consideration as the number of satellites and vital communications parameters depend on the altitude. Since the field-of-view increases, number of satellites decrease with an increase in altitude (see Figure 2.26).

With minor exceptions, to date Walker's constellations, which use inclined orbits, have been the most efficient for unbiased single visibility up to a constellation size of 15 satellites, deployed at medium and high altitude. Examples of constellations based on this approach include Odyssey, Globalstar, MAGSS-14 (Medium Altitude Global Mobile Satellite System) (Benedicto *et al.*, 1992), Deligo (Meenan *et al.*, 1995), GIPSE (Global Integrated Personal Satellite Multimedia Environment) (Sammut *et al.*, 1997), and others.

J. Draim, an American aerospace engineer, is reported to have developed a constellation that can provide continuous global coverage with only four satellites instead of five as demonstrated by Walker (Logsdon, 1995). The satellites are placed in a 31.5° inclined orbit with a period of 26.5 h or greater and can see each other continuously; a US launch from Cape Canaveral would require very little fuel for orbital plane change.

Polar constellations based on the work of Beste, Adams and Rider, and Walker, have preferred for LEO to date. Some propriety hybrid constellation designs have been developed for specific applications.

Figure 2.6 shows the geometry of a single geostationary satellite, useful in understanding a constellation's optimization. Optimization is achieved by setting a minimum permissible elevation angle, η, thereby fixing the altitude and then adjusting orbital parameters to minimize the number of satellites. Another approach is to select a constellation pattern of known characteristics and minimize the largest value of great circle range R' for all observation points and instants of time by trial and error.

A useful constellation performance measure is its comparison with theoretical bounds that may be estimated in a number of ways; the Earth can be divided into a number of non-overlapping equilateral spherical triangles; this way the coverage is distributed uniformly around the world (Ballard, 1980). The bounds can also be approximated by dividing the Earth into a number of hexagons and enclosing circles around them (Beste, 1978):

$$N \sim 2.42/(1 - \cos\psi) \tag{2.7}$$

where $\psi =$ great circle range for which the stationary bound is required and $N =$ the number of stationary satellites.

The coverage efficiency of a constellation, consisting of N satellites can be given as:

$$N\Omega/4\pi \tag{2.8}$$

where $N\Omega =$ Total solid angle of the constellation, $\Omega =$ Solid angle bounded by a single satellite $= 2\pi\ (1 - \cos\psi)$ Steradians and $4\pi =$ solid angle of a sphere.

The best coverage efficiency of 1.0 transpires when two satellites are located diametrically opposite at a distance of infinity, theoretically the minimum number of satellites for worldwide coverage.

When comparing the uniformity of coverage of two constellations, the system designer is faced with the problem of quantifying the degree of uniformity of candidate schemes. The coverage uniformity of constellations can be measured as the temporal uniformity index u_T and the spatial uniformity index u_S, which respectively indicate how unbiased the coverage is over an orbital period and geographical area; an ideal constellation should give a value of 1 for each.

2.3.2 Polar Constellations

Polar constellations have been studied by several researchers (Luders, 1961; Beste, 1978; Adams and Rider, 1987). As an illustration, consider the approach of Beste who founded some of his development on Luder's earlier work.

Beste optimized single-visibility coverage for the complete world as well as for regions extending from the pole to any arbitrarily chosen latitude. He adopted type 1 constellation, that is, satellites are placed in orbital planes that intersect at a common point such as over the pole. Plane separation and satellite spacing in each plane are adjusted to minimize the total number of satellites. Figure 2.24 shows the geometry used in optimization of a polar constellation for full Earth coverage.

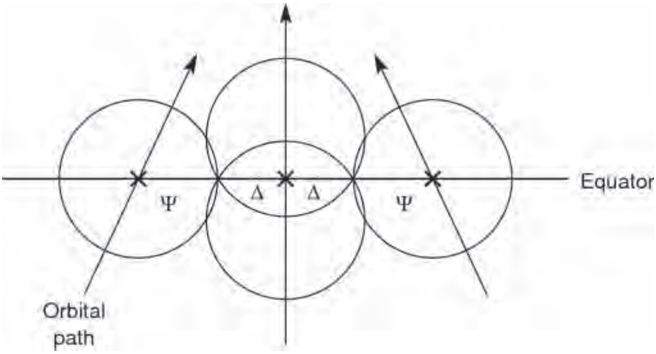


Figure 2.24 Geometry for optimization of a polar constellation

Satellites in adjacent planes move in the same direction, offset relative to each other by half intra-orbit satellite separations, that is half of (2π/m) or π/m, where m is the number of planes. The arrangement staggers satellites at the equator giving a relatively uniform coverage. The separation between adjacent planes, where satellites move in the same direction, is ($\psi + \Delta$). The relative geometry of such orbital planes remains constant because the satellites move in phase; however, there is one set of adjacent planes where satellites move in opposite directions. The relative geometry between adjacent planes is not constant and hence these planes are placed closer together at 2Δ, such that the equator is covered by the common area between adjacent satellites. The constellation is defined by the following parameter set:

ψ = the coverage circle;
n = the number of orbital planes;
m = the satellites per plane;
$\psi + \Delta$ = the plane separation between adjacent planes containing satellites moving in the same direction;
2Δ = the separation between planes at the boundary where satellites move in opposite directions;

where

$$\Delta = \cos^{-1}[\cos\ \psi/(\cos\ \pi/m)]. \tag{2.9}$$

The optimization is based on minimizing the number of satellites at the equator, where the coverage requirement is most stringent due to the widest plane separation. The following condition must be satisfied for this orbital region:

$$(n-1)\psi + (n+1)\Delta = \pi \tag{2.10}$$

A trial and error method is used to solve Equation 2.10. The method can be extended for multiple visibilities based on the same iterative search process so as to obtain multiple visibilities at the worst point. The constellation geometry is similar to the single-coverage case.

Using the same definition of intra and interorbital spacing as used by Beste (Adams and Rider, 1987) provide optimum solutions using the following set of constellation definitions:

$$P = (2/3)k(\pi/\psi) \tag{2.11}$$

$$Q = (2/\sqrt{3})\, j(\pi/\psi) \tag{2.12}$$

$$N = (4\sqrt{3}/9)n(\pi/\psi)^2 \tag{2.13}$$

where P = the number of orbital planes, Q = the number of satellites in each orbital plane, N = the total number of satellites, ψ = the coverage angle, k = the multiple coverage factor in different planes, j = the multiple coverage factor in the same plane and n = the multiple coverage factor of the constellation.

For single visibility, k = j = n = 1.

2.3.3 Inclined Orbit Constellations

A family of constellations originally proposed by Walker (1973) provides efficient unbiased regional or world-wide coverage using circular orbits at the same altitude and inclination, distributed uniformly on the celestial sphere. Regional coverage can be provided by partial deployment of the constellation, according to the region of interest. Adjacent orbital planes are separated equally around an arbitrary reference plane that can be the Earth's equator. Within each orbit, neighbouring satellites are equally separated. The initial position of a satellite is proportional to the right ascension angle of the orbital plane – the right ascension angle being measured on the reference plane. Orbital perturbations are not included in optimization and hence in order to preserve constellation geometry with time, station keeping becomes necessary. A slight overdesign of the constellation can cancel station keeping errors. In the original work, Walker specified inclination as δ and hence called the coverage patterns 'delta patterns'. The orbital traces of satellites of a constellation on the celestial sphere resemble petals of a flower and therefore Ballard (1980), who formalized and extended Walker's work, called such constellations 'Rosettes'. Identical results have also been independently obtained by Mozhaev (1972, 1973). Ballard's work is used here for definition.

The constellation is designated as (N, P, m) where N = the total number of satellites in the constellation, P = the number of orbital planes and m = the harmonic factor, which determines the initial distribution of satellites and the rate at which patterns precess over the celestial sphere. When m is an integer, the satellite constellation consists of 1 satellite per plane, when it is an unreduced integer S/Q, the constellation consists of a Q satellite in each orbital plane. For example, a (5, 5, 1) constellation consists of a total of five satellites in five planes with one satellite in each plane; a (12, 3, ¼) constellation consists of a total of 12 satellites in three planes with four satellites in each plane.

The parameters of the rosette constellation geometry are illustrated in Figure 2.25. The parameters of our interest here are;

α_i = the right ascension angle of the ith orbit plane,
β_i = the inclination angle of the ith orbit, the same for each orbital plane,

γ_i = the initial phase angle of the ith satellite in its orbital plane at t = 0, measured from the point of right ascension,

χ = Phase angle at an arbitrary time t,

H = satellite altitude,

R_E= the Earth radius.

$$\alpha_i = 2\pi i/P, i = 0 \text{ to } N - 1 \tag{2.14a}$$

$$\beta_i = \beta \tag{2.14b}$$

$$\gamma_i = m\alpha_i = mQ(2\pi i/N); \tag{2.14c}$$

where m = (0 to N−1)/Q, N = total number of satellites = PQ, P = number of orbit planes and Q = Number of satellites per plane.

$$\chi = 2\pi t/T, \text{ where } T = \text{orbital period} \tag{2.14d}$$

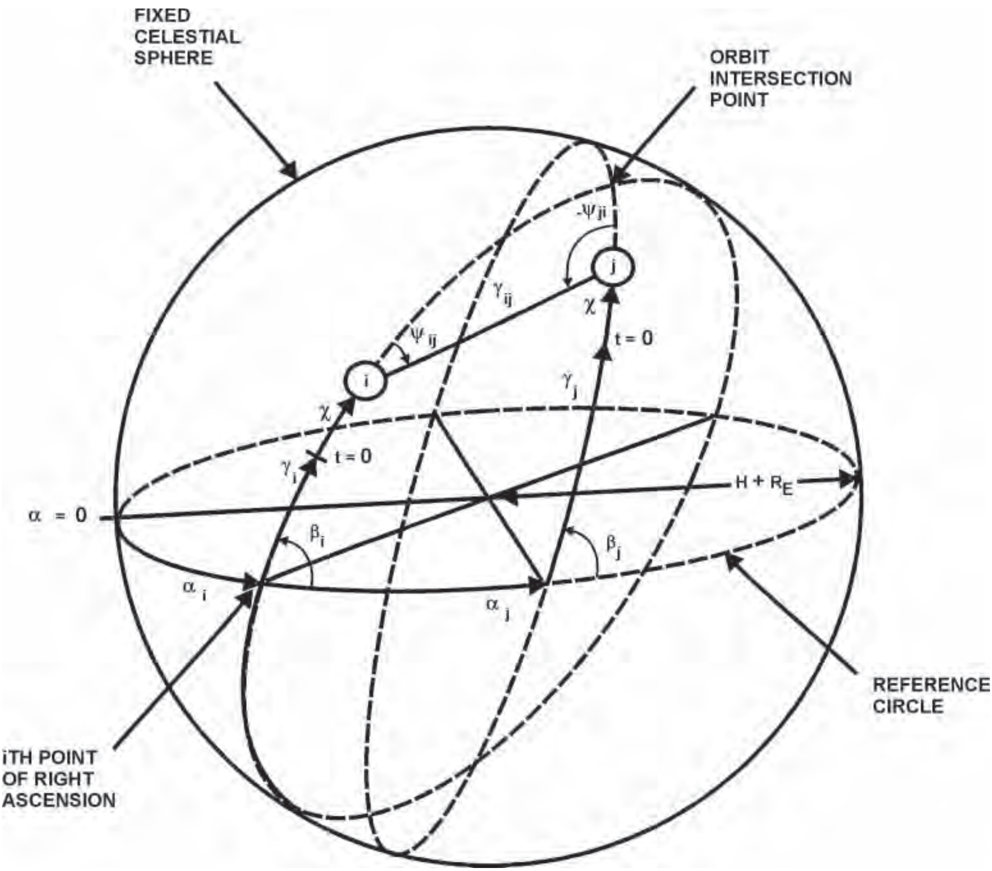


Figure 2.25 Constellation geometry for the Walker constellation. (Source: Ballard, 1980. © 1980 IEEE. Reproduced with permission.)

The great circle range ψ, shown in Figure 2.24 (see also Figure 2.6), is used as a measure of optimization. For a chosen set of constellation parameters, the highest value of ψ is minimized at all observation points for all instants of time. The worst observation point on the Earth is the centre of a spherical triangle formed by joining the sub-satellite points of three satellites. If the distance of the point to any of the sub-satellite point is R_{ijk}, and a constellation provides useable coverage to $\psi < R_{ijk}$ then the centre point of the spherical point is left uncovered, thereby failing the full coverage test. On the other hand, if $\psi \geq R_{ijk}$ the visibility at the worst point increases to three. A measure of goodness of coverage of a constellation is, therefore, the highest value of R_{ijk}, $R_{ijk}(\max)$ taken over all instants of time over the orbital time period T; a constellation with a lower value of $R_{ijk}(\max)$ being better suited from communications view point, because this would ensure that in such a constellation satellites are closer to an observer located at the worst point. A total of $2N-4$ non-overlapping triangles are formed with N satellites, all of which must be examined at each instant of time over the orbital period. Furthermore, the process must be repeated for all values of inclination to obtain the minimum value of $R_{ijk}(\max)$. A search of such a magnitude can best be performed by the use of a computer. The only effect of Earth's rotation is that location of R_{ijk} on the Earth varies with time and hence it is not necessary to include earth's rotation in the optimization.

Uniformity of coverage over an orbital period and geography is measured by temporal uniformity index, u_T, and spatial uniformity index, u_s. An ideal constellation should give a value of 1 for each. These uniformity indexes are defined as follows:

$$u_T = R_{ijk}(\max)_w / R_{ijk}(\max)_b \tag{2.15}$$

where, $R_{ijk}(\max)_w$ and $R_{ijk}(\max)_b$ are respective values of $R_{ijk}(\max)$ at the worst and the best phase angles

$$u_s = \text{the maximum of } [R_{ijk}(\max)/R_{ijk}(\min)] \text{ over all phase angle} \tag{2.16}$$

where, $R_{ijk}(\max)$ and $R_{ijk}(\min)$ are the maximum and minimum value of R_{ijk}, respectively at the same phase angle.

Ballard derived results of best single visibility using exhaustive searches for $N = 5-15$, summarized in Figure 2.26(a) and (b). The average temporal and spatial uniformity indexes of these constellations were calculated as 1.06 and 1.25 respectively. The minimum number of satellites required to guarantee single world-wide visibility is the constellation (5, 5, 1), that is one satellite in each of five planes at an altitude of 4.232 eru (~26 992 km). As expected, the total number of satellites N for world coverage reduces as the altitude is increased.

We can increase the reliability of the radio link by increasing the minimum look angle. This can be achieved by choosing a higher altitude at the expense of an increase in path loss.

From a practical viewpoint, there are some advantages in choosing the orbital time period to be the integer ratio of the Earth's rotation period; the ground track of such orbits follows fixed repetitive patterns on the ground. This makes algorithms for searching and tracking of satellites, and handover between satellites, simpler due to the simple predictability of the satellite position from the ground. It is also possible to deploy a constellation in manageable stages, building up from the areas where coverage is most required, or, to deploy partial constellations for regional coverage. There are an infinite number of possibilities depending

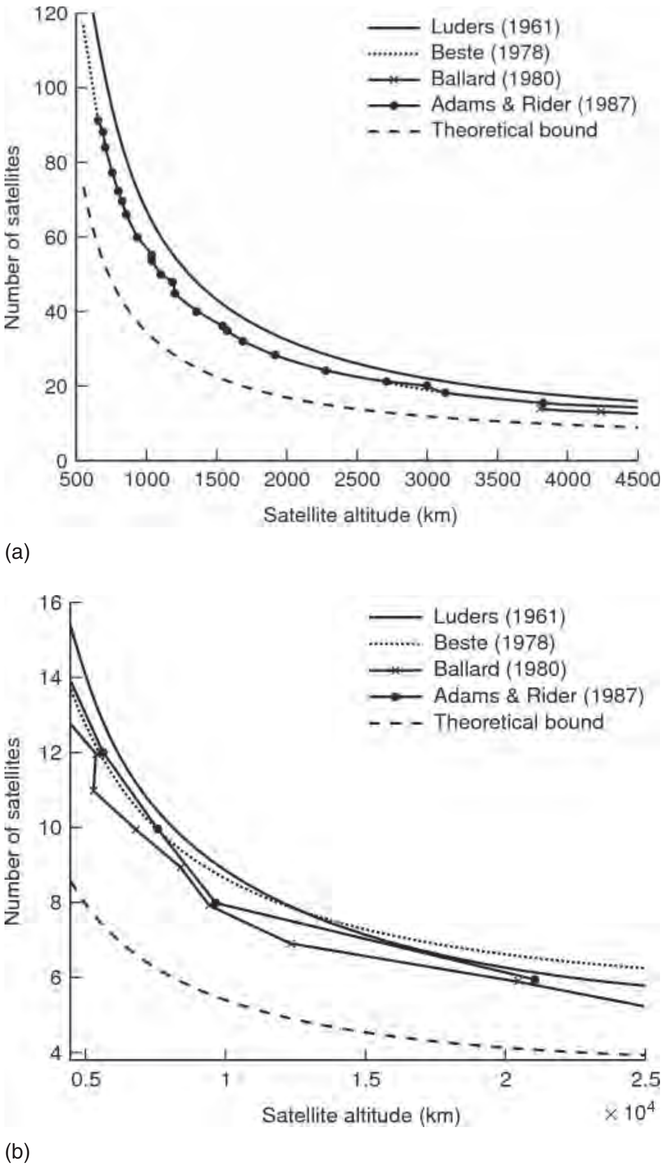


Figure 2.26 Optimized constellation derived by various methods for (a) altitudes up to 4500 km and (b) altitude between 5000 and 25 000 km (Both parts source: Graphics AR.)

on the orientation between the constellation's reference longitude relative to the equator, and the constellation parameters (N, P, m).

An interesting example of a quasi-stationary constellation, known as 'Loopus' (Loops in Orbit Occupied Permanently by Unstationary Satellites) has been proposed by Donald (1984). The constellation parameters are chosen such that satellites follow the same track, forming small loops on the celestial sphere; changeover between satellites takes place at the loop intersection, that is, just when a satellite is leaving the loop and another entering it. Satellites appear to remain within the small loop in the sky, much as a geostationary satellite. Loops can be created over high latitudes, enabling a high elevation angle service to polar and high latitude regions.

The discussion above dealt with single visibility of satellites; a system designer may consider multiple satellite visibility for achieving a path diversity advantage and/or to offer navigation capability. For the Rosette constellation, it has been shown that for world-wide V satellite visibility, at least $2V + 3$ satellites are required. The definition of R_{ijk} in optimization in this case changes to the distance where V satellites are visible at the worst observation point.

Figure 2.26 summarizes solutions derived by various methods including the theoretical bound as a reference. Figure 2.26(a) covers the LEOs, while Figure 2.26(b) covers the MEOs. Note that Ballard's solution offers the highest efficiency (i.e. beyond ~3800 km). Iridium constellation is based on Adams and Rider solution, and Globalstar constellation uses Ballard's approach.

2.3.4 *Hybrid Constellations*

It is possible to combine various types of orbits for full Earth coverage. Such orbits may have different orbital periods, inclination and eccentricity. For example, in one such configuration circular orbit would cover the equatorial region and elliptical orbit the higher latitude regions. Chapter 11 discusses an interesting hybrid constellation design that maximizes coverage of populated areas of the world including high latitude regions around Canada, proposed for the ELLIPSO system. Other possibilities would be to use GEO for covering the equatorial region and either inclined orbits or LEO to cover the polar region.

An example of this type of hybrid orbital design is the proposed WEST broadband system. The system combined a MEO constellation called JOCOS (Juggler Orbit COnStellation) comprising nine satellites with a GEO satellite (Pennoni and Bella, 1995). In this design, satellites are placed such that they follow the same sub-satellite path. By dividing the Earth into three parts, 120° apart, and placing satellites in an 8 h orbit, the design is capable of serving most of the inhabited regions.

2.3.5 *Regional Coverage*

When regional coverage is required, it is necessary to ensure that all satellites in the constellation pass over the service area. Partial deployment of Walker constellations can provide a continuous ground track of regions of interest. For example, equatorial regions can be covered by deploying a sequence of satellites in an equatorial orbit such that their footprints follow each other.

Constellations using a HEO inclined at 63.4° have been used for covering high latitudes as satellites in such orbits dwell over high latitudes for a considerable time of the orbit. Such a system has been in use to cover high-latitude regions of Russia. They were proposed for mobile and broadcasting applications in Europe and for mobile communication in the USA and Canada. More recently, such a constellation has been deployed by Sirius satellite system in the American region for radio broadcasting.

2.3.6 Constellations for Non-Real-Time Systems

Constellation design for systems that do not require real-time coverage, for example messaging/paging systems, is less stringent because gaps in coverage can be tolerated, provided that the regions of interest have at least one satellite visible within ($t_a - t_d$) seconds, where t_a is the specified end-to-end delay and t_d is the delay in message transfer, which is the sum of connection, processing and propagation delays. The ORBCOMM system uses this approach during the constellation deployment phase or during periods when constellation gets depleted.

2.3.7 Use of Spot Beams

The optimization techniques discussed above ensure a minimum elevation angle visibility within the coverage area. However, from link design considerations it becomes necessary to partition the coverage area into small segments covered by spot beams. Spot beams increase the satellite antenna gain and EIRP (effective isotropic radiated power), receiver sensitivity so as to satisfy the quality-of-service and improve spectrum reusability; the penalty is the complexity in payload, an increase in the weight of a satellite and an increase in the number of handovers between beams, though they are low when the spot beams are Earth-fixed (see Chapter 7).

2.3.8 Availability Considerations for Non-Geostationary Satellites

A fundamental requirement of an orbital design is to guarantee a minimum elevation angle visibility from all locations within the coverage area. A system should provide the desired visibility of one or more satellites from the coverage area, depending on the type of service, and other considerations such as whether radio link diversity is used. In a real-time service, communication is lost when the radio link breaks and hence it is important to maintain at least one satellite within visibility. In a store and forward system the system stores information until such time as a satellite becomes available, when the stored data is transmitted in a burst.

The reliability of the service link depends on the radio path profile, or more generally, the skyline. The higher the elevation of a satellite, the less likely is the probability of path obstruction, but the number of satellites increases to ensure coverage availability right up to the EOC. Therefore propagation degradation caused by the skyline reduces progressively in the following order: city, suburbs and open areas (see Chapter 3).

Link availability can be increased if communication can be established through separate paths, either from the same satellite (e.g., through separable multipath signal reception) or through separate satellites; signal quality can be improved by combining signals from each path. When path diversity is established through different satellites, there are other potential advantages. It is possible to reduce congestion on a heavily loaded satellite by diverting traffic through the less-congested satellite; improve handover reliability by using robust and seamless ‘soft’ handover algorithms; and signal quality can be improved by combining signals from each path. Table 2.3 shows an example of reduction in link margin for one, two and three satellite visibilities from a measurement campaign at the L band during the design phase of the ICO system (Hart, Goerke and Jahn, 1995). The required link margin in a rural

Table 2.3 Diversity versus link margin

Number of visible satellites	Link margin (dB)	
	Rural environment	Urban environment
1	16.5	23.5
2	7	8
3	5	8

(Source: Hart *et al*., 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa cosponsored by Communication Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, 469–75.)

environment reduces from 16.5 to 5 dB when the satellite visibility increases from a single to three satellites and the margin reduces from 23.5 dB to 8 dB in the urban environment when the visibility increase from 1 to 2 satellite. The diversity gain depends on elevation angles of visible satellites and azimuth separation of satellites. Generally, higher elevation satellites are likely to offer a clearer line of sight as shown in Figure 2.27, and wider azimuth

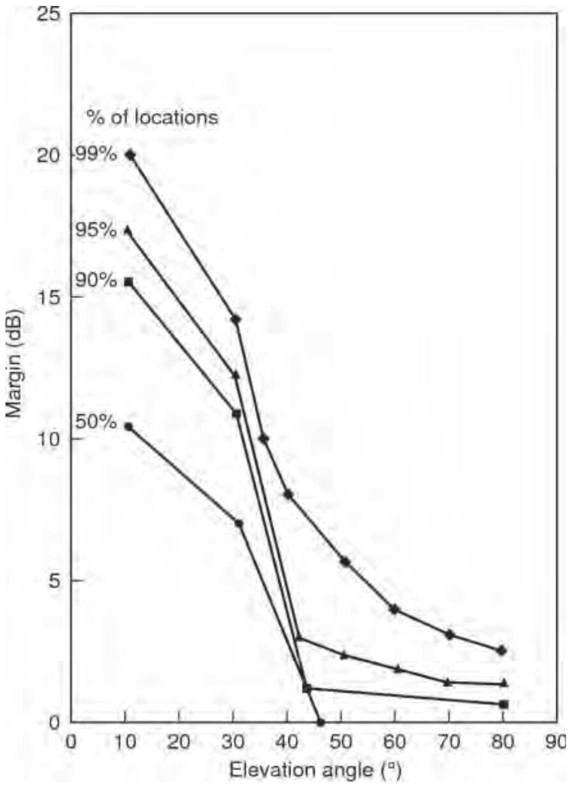


Figure 2.27 Elevation angle dependence of fade margin for various percentages of locations (Source: ESA, 1988. Reproduced with permission of ESA.)

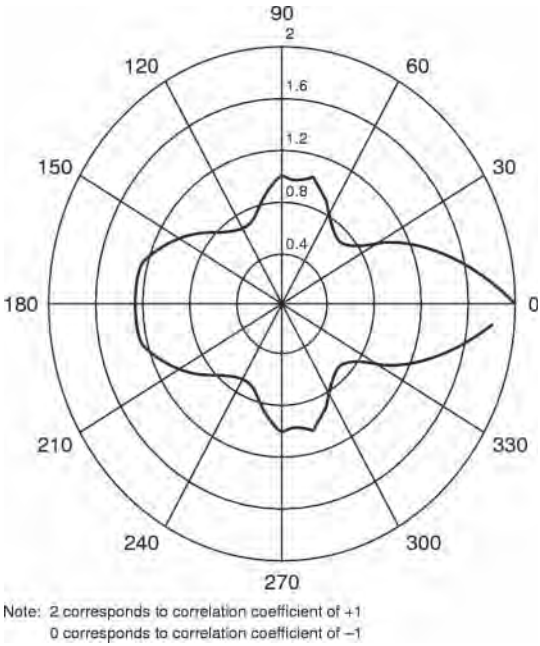


Figure 2.28 Azimuth angle correlation of shadowing in an urban environment (Source: Hart *et al.*, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

separation reduces the probability of simultaneous fading from visible satellites; however, the extent of improvement is strongly influenced by the local environment (see Chapter 3 and notice the link behavior between 2 and 3 satellite visibility in the urban environment). Figure 2.28 shows an estimated correlation in fading with azimuth in an urban environment (Hart, Goerke and Jahn, 1995). For example, the figure shows a high correlation ($\sim 0.7-1$) in the azimuth range of $\pm 15°$ from the boresight and low correlation (~ 0.4) between 30° and 60°.

Look angles of satellites in non-GEO are time and location dependent and therefore for design and planning purposes expressed statistically and generally obtained through computer simulation because of non-linear interaction of a number of system variables. Statistics should be collected for a large number of locations and time for a better statistical representation. Some useful statistics for a given location are summarized later in this section. It should be emphasized that when considering radio link performance, geometric visibility alone cannot be taken in isolation as the reliability of the radio link is severely affected by the local environment (see Chapter 3 for propagation effects). Statistics used in system design include:

- probability and cumulative distribution of look angles from representative locations;
- average, minimum and maximum elevation and azimuth angle of the highest visible satellite from the worst location;

- probability and cumulative distribution of look angles for a given percentage of locations;
- maximum period when a satellite is unavailable from the worst location;
- probability of a call of a given duration completing successfully;
- sensitivity of the quality-of-service such as delay to incomplete constellation (e.g. caused by satellite failure or at the start of operation);
- elevation angle and azimuth separation between two satellites – these parameters provide a measure of satellite diversity improvement.

Consider, as an example, the performance analysis of a few LEO and MEO constellations in a hypothetical urban environment with the following assumptions (Sandrin and Haschart, 1993):

- orbital characteristics specified in Table 2.4;
- no service when line of sight is absent, that is diffraction, light shadowing, multipath were not considered;
- environment representative of a typical urban US environment;
- handover to the next satellite is perfect and based on a dynamic check of channel availability;
- call maintained as long as at least one satellite has an unobstructed view to the user;
- call occurrence at random times – user either initiating or receiving the call;
- 5° minimum look angle to the satellite.

A number of interesting conclusions were derived:

- Both ICO constellations provide a high probability of completing calls.
- LEO-54 provides a better probability of completion for short calls of 7 min or less in one case.
- LEO-54 provides a better probability of call completion than LEO-35.
- The system is very sensitive to the local environment; a 45° rotation of city intersection causes a dramatic effect on probability.
- Satellites with higher minimum elevation angle have a better availability.

Table 2.4 Main characteristics of orbits used for developing system statistics

Constellation	Number of satellites	Number of planes	Satellite/plane	Inclination (°)	Orbit altitude (km)	Orbital period (min)	Minimum elevation at 30° latitude	Approximate view time for 5° elevation (min)
LEO-35	35	5	7	90	1 584.9	117.9	10	21
LEO-54	54	9	6	55	1 800	122.7	20	26.5
ICO-12	12	3	4	50.7	10 355	359	20	144.8
ICO-15	15	3	5	53.5	11 622.2	400.6	30	164.5

(Adapted from Sandrin and Haschart, 1993.)

- The conditioned call initiation is expected to be beneficial only for incomplete constellations when there is a large gap in the visibility of satellites.
- The sky profile has a marked influence on performance even in a very specific environment; for example the availability of satellites in a polar orbit on a street intersection with a clear north–south is better an east-west road with obstructions on either side.

Visibility statistics are an effective way of comparing constellations as they give an insight into the effectiveness of a constellation performance. On comparison it emerges that there is no unique solution – some such as Globalstar chose LEO to optimize coverage in specific

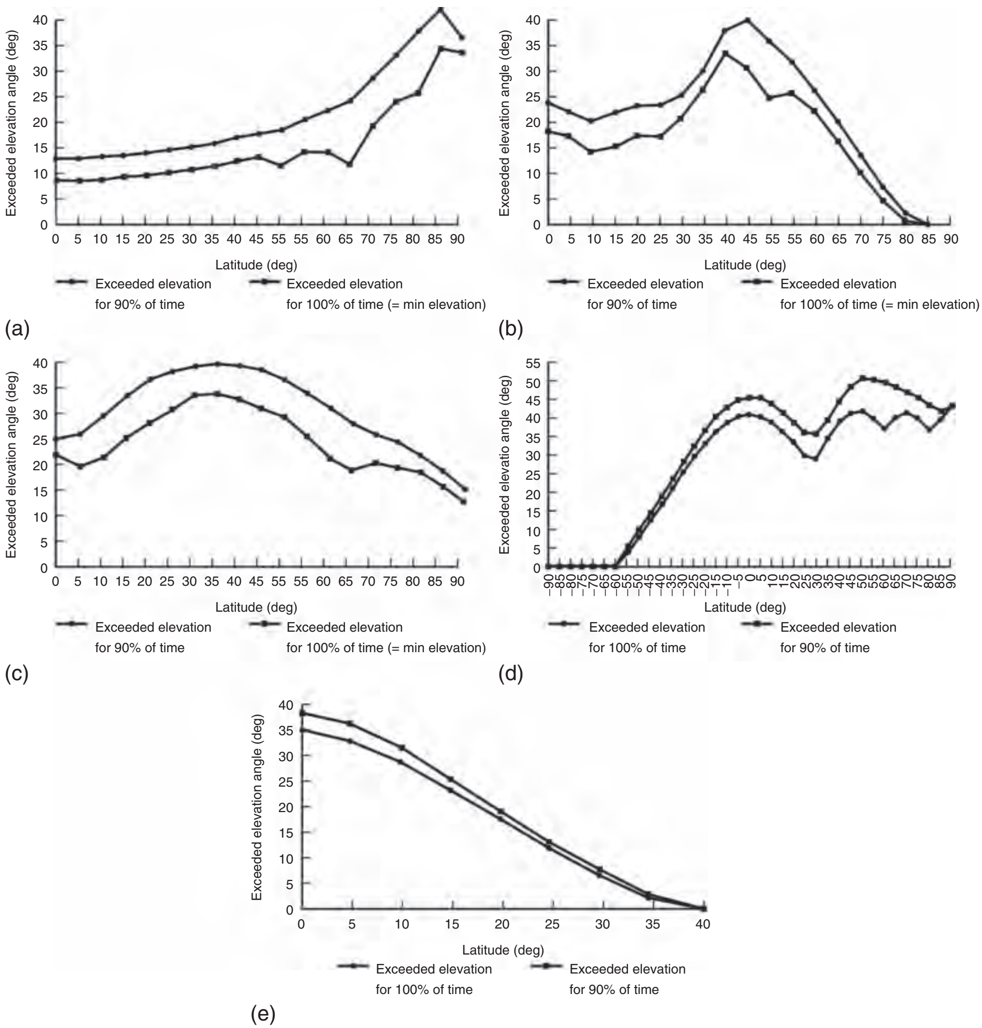


Figure 2.29 Visibility statistics in terms of percentage of time elevation is exceeded for 90 and 100% of time for (a) Iridium, (b) Globalstar, (c) ICO, (d) ELLIPSO, and (e) ECCO (All parts source: Krewell and Maral, 1998. Reproduced with permission of John Wiley & Sons, Ltd.)

regions; others such as Iridium chose LEO to provide an unbiased world-wide coverage; yet others preferred non-continuous coverage to offer non-real-time service. The choice depends on the service area, service, that is real-time, store and forward, etc., target link reliability, system cost, available expertise, and others.

Visibility statistics of various constellations have been compared by various authors. Here we summarize the findings of Krewel and Maral (1998). Figure 2.29(a–e) demonstrates elevation angles exceeded for 90 and 100% of time versus latitude for five types of constellations – Iridium (LEO/polar), Globalstar (LEO/inclined), ICO (Medium/inclined), ELLIPSO (hybrid – elliptical/circular) and ECCO (LEO/equatorial). Some interesting features are as follows:

- **Iridium**: Has the best visibility performance above latitude of 65°; below this latitude the elevation angle varies between 8° and 15° (Figure 2.29a). Although the system has a good dual-satellite diversity performance beyond latitude of around 50°, the system has not been designed to use diversity (see Chapter 11).
- **Globalstar**: Offers the best visibility in mid-latitude between 35 and 50° (both the hemispheres) where elevation angle ranges between 20 and 32° (Figure 2.29b). The system offers good path diversity and has been designed to exploit this advantage.
- **ICO**: Visibility angle of 18° is exceeded up to latitude of 80° with the highest elevation angle visibility at latitude of 35° (Figure 2.29c). The system offers good path diversity with a feature to exploit diversity on demand.
- **ELLIPSO**: Has been optimised for the northern latitudes, with minimum elevation angle > 27° at latitude of 28.5° and 41° at latitude of 50° (Figure 2.29d). The system offers the highest probability of dual and triple visibility in the latitude range of 20–40°.
- **ECCO**: (Equatorial Constellation Communications Organization) Is an equatorial constellation offering the highest elevation angle at the equator and a minimum elevation angle of 8° within latitude of ±30° (see Figure 2.29e). Satellite diversity is possible for 50% of the time.

Revision

1. Define and explain the significance of orbital parameters in an MSS context.
2. Figure 2.7(c) represents contour plots of azimuth angle as a function of latitude and longitude relative to the sub-satellite point of a geostationary satellite. Estimate the azimuth of a geostationary satellite located at 15°E when the user is located at: (i) 55°E/20°N; (ii) 75°E/80°S; (iii) 35°W/40°N and (iv) 55°W/40°S.
3. What are the various types of perturbations on satellites in Earth orbits? How do such extraneous forces affect satellite motion?
4. Outline the principle, characteristics and advantages of a Sun-synchronous orbit.
5. Determine the inclination of a Sun-synchronous circular orbit of 750 km altitude (Earth radius = 6378 km).
6. How are the parameters of an eccentric orbit adjusted to provide service to high latitude region above 81°N? Suggest limitations of this approach.

7. What is the rate of change of argument of perigee of an elliptical orbit with the following characteristics: Eccentricity = 0.15; Semi-major axis = 10 000 km; Inclination = 45°? (Earth radius = 6378 km). Plot a graph of argument of perigee versus inclination ranging between 5° and 90°. Comment on the results.
8. Outline the concept of solar eclipse at the geostationary orbit? What is the impact of such eclipses on a spacecraft? Suggest an orbital configuration with an illustration in which solar eclipses can be eliminated.
9. What are the types of constellations used for mobile satellite services? Outline the benefits and limitations of each.
10. Which constellations would you select to provide mobile satellite service for: (i) a seamless global coverage; (ii) an equatorial coverage within ±30° latitude and (iii) a regional coverage to serve equatorial and mid-latitude regions? Justify your choice.
11. What are the differences in the radio link design of geostationary and non-geostationary satellites from an orbital perspective? Suggest a technique to improve reliability of non-geostationary radio links.

References

Adams, W.S. and Rider, L. (1987) Circular polar constellations providing continuous single or multiple coverage above a specified latitude, *Journal of the Astronautical Sciences*, **35** (2), 155–192.

Ballard, A.H. (1980) Rosette constellations of earth satellites. *IEEE Transactions on Aerospace and Electronic Systems*, **AES–16** (5), 656–673.

Benedicto, J., Fortuny, J. and Rastrilla, P. (1992) MAGASS-14: A medium-altitude global mobile satellite system for personal communications at L-band, ESA Journal, 92, **16** (2), 117–133.

Beste, D.C. (1978) Design of satellite constellation for optimal continuous coverage. *IEEE Transactions on Aerospace and Electronic Systems*, **AES–14** (3), 466–473.

Donald, P. (1984) Loopus opens a new dimension in satellite communications. *International Journal of Satellite Communications*, **2**, 242–250.

ESA (1988) Archimedes Study Report.

Gavish, B. and Kalvenes, J. (1998) The impact of satellite altitude on the performance of LEOS based communication systems. *Wireless Networks*, **4**, 199–213.

Hart, N., Goerke, T. and Jahn, A. (1995) A discussion on mobile satellite systems and the myths of CDMA and diversity revealed. International Mobile Satellite Conference, Ottawa, IMSC'95, The Fourth International Mobile Satellite Conference Ottawa, Canada, Co-sponsored by Communication Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 469–475.

Haviland, R.P. and House, C.M. (1965) *Handbook of Satellites and Space Vehicles*, Van Nostrand Reinhold, New York.

Kessler, D.J. and Burton, G.C. (1978) Collision frequency of artificial satellites: the creation of a debris belt. *Journal of Geophysical Research*, **83** (A6), 2637–2646.

Klinkrad, H. and Jehn, R. (1992) The space-debris environment of the Earth. *ESA Journal*, **16** (1), 1–11.

Krewel, W. and Maral, G. (1998) Single and multiple satellite visibility statistics of first-generation non-geo constellations for personal communications. *International Journal of Satellite Communications*, **16**, 105–125.

Logsdon, T. (1995) *Mobile Communication Satellites*, McGraw-Hill, p. 160.

Luders, R.D. (1961) Satellite networks for continuous zonal coverage. *American Rocket Society Journal*, **31**, 179–184.

McInnes, C.R. (1993) An analytical model for the catastrophic production of orbital debris. *ESA Journal*, **17**, 293–305.

Meenan, C., Sammut, A., Tafazolli, R. and Evans, B.G. (1995) DELIGO: repetitive ground track 100% dual-diversity LEO satellite constellation for S-PCN. *IEE Electronics Letters*, **31**, 1407–1408.

Morgan, W.L. and Gordon, G.D. (1989) *Communications Satellite Handbook*, John Wiley & Sons, Inc., New York.

Mozhaev, G.V. (1972) The problem of continuous earth coverage and kinematically regular satellite networks, I'. *Cosmic Research*, **10** (6), 729–882(UDC 629.191) translation in CSCRA7 (Consultants Bureau, New York), vol. 10, no. 6, 729–882.

Mozhaev, G.V. (1973) The problem of continuous earth coverage and kinematically regular satellite networks, II. *Cosmic Research*, **11** (1), 1–152 (UDC 629.191), translation in CSCRA7 (Consultants Bureau, New York), vol. 11, no. 1, 1–152.

NASA (2010) Orbital Debris Quarterly News, Vol. **14** (1), January.

NASA, (2013) http://science.nasa.gov/science-news/science-at-nasa/2013/28feb_thirdbelt/ (Accessed July 2013)

Pennoni, G. and Bella, L. (1995) JOCOS: a triply geosynchronous orbit for global communications: an application example. IEE Conference, publication no. 403/2, pp. 646–652.

Petro, A.J. and Talent, D.L. (1989) Removal of orbital debris. *Progress in Astronautics and Aeronautics*, **121**, 169–182.

Richharia, M. (1999) *Satellite Communication Systems*, 2nd edn, Macmillan Press Ltd, Basingstoke and London.

Sammut, A., Mertzanis, I., Tafazolli, R. and Evans, B.G. (1997) GIPSE: a global integrated personal satellite multimedia environment. 4th European Conference on Satellite Communication (ECSC-4), Rome, Italy, November 18–20, pp. 11–15.

Sandrin, W.A. and Haschart, D.V. (1993) Availability prediction for hand-held communications system using non-geo-synchronous satellites. *COMSAT Technical Review*, **23** (2), 237–269.

UN (1999) *Technical Report on Space Debris*, United Nations Publication, New York, Sales No. E.99.I.17, ISBN: 92-1-100813-1.

Walker, J.G. (1973) Continuous whole earth coverage by circular orbit satellites. Presented at the IEE Satellite Systems for Mobile Communications Conference, Paper 95, March, pp. 13–15.

3

Radio Link

3.1 Introduction

This chapter provides an overview of concepts and techniques related to a satellite radio link. The satellite–mobile radio link is one of the most critical components of a Mobile Satellite Service (MSS), as it gives mobility and concurrently sets bounds on throughput. Keeping to the theme of the book, topics are discussed at a system level to enable the reader to grasp relationships between system components.

The chapter introduces spectrum-related issues including interference estimation and management methods, international spectrum allocation procedures and trends. An understanding of radio wave propagation is of fundamental importance in the design and understanding of the mobile radio link, as propagation impairments set binds to throughput. Beginning with a general introduction to radio wave propagation, the chapter reviews some of the prevalent propagation models applicable to land mobile, maritime and aeronautical links, followed by a review of radio link analysis highlighting factors that affect mobile radio links.

3.2 Spectrum Issues

Radio waves are information carriers and a fundamental, albeit limited, natural resource of any radio communication system. Consequently, useful parts of the radio system, particularly the MSS bands, are currently in great demand and precious. Commonly used MSS ranges are the L (1.5–1.6 GHz) and S (2.0–2.5 GHz) bands; however, higher frequency ranges such as the K_u band (~ 14–15 GHz) and K_a band (20/30 GHz) are also in use.

In an ideal world, an operator would select a frequency band that offered an optimum compromise in terms of propagation degradation, user equipment size and cost attractiveness, manufacture and launch of satellites, technology, and so on. Such a freedom would result in a certain degree of chaos, as all the operators would quite probably converge to a few favourable bands, thereby causing interference to each other as well as to other radio services operating in the band. To avoid such a scenario, RF usage is regulated by the International Telecommunications Union (ITU). The ITU assigns spectra to all types of radio services and specifies procedures for their management on a global basis, taking into consideration the

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

demands of each service by region and globally, radio wave propagation characteristics and technical status, while respecting the needs and concerns of each country. Within this framework, each country manages a spectrum through its own regulatory body such as the FCC in the USA or the DTI in the UK. Thus, an operator's freedom in the selection of spectra is constrained to specific ranges and within them, those bands with technical and commercial advantages become contentious, leading to a need for recognized regulatory procedures for spectrum coordination between operators. Such procedures, developed by the ITU, require a new operator to notify the ITU with appropriate technical details, announcement of the intention by the ITU and coordination of the applicant with incumbent operators who may be affected by the new system.

The ITU has divided the world into three regions for the purpose of spectrum allocations, as shown in Figure 3.1.

These regions are broadly:

- **Region 1**: Europe, Africa, the Middle East and the European regions of the former USSR.
- **Region 2**: The Americas.
- **Region 3**: The remainder of Asia, plus Australasia.

Frequency allocations are listed in Article 8 of the Radio Regulations (ITU, 2008, or the most recent). A number of categories of allocations are provided. These are primary, secondary, footnote, planned, exclusive, and shared. Exclusive primary frequency allocations are preferred for the MSS. In such allocations, a service has an exclusive right of operation, provided mutually satisfactory coordination between operators of the service has been

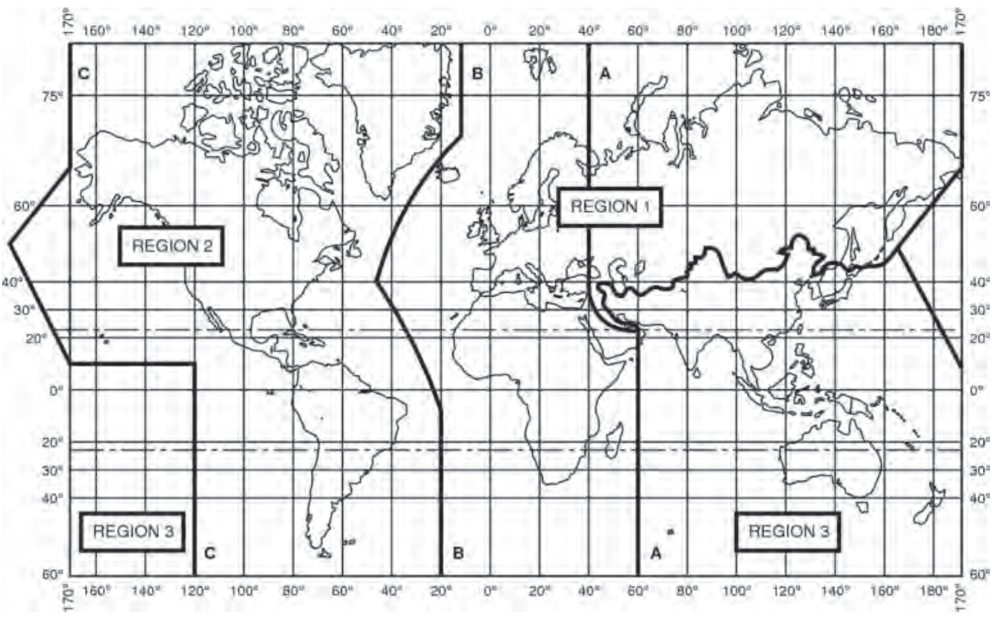


Figure 3.1 ITU regions of the world used for frequency allocation (Source: RR Article 8. Reproduced with permission of ITU.)

achieved. Secondary allocations are of limited use for the MSS, because a service with such an allocation is not guaranteed interference protection and neither is it permitted to cause interference to services with primary status; in all, this has an adverse impact on user mobility. The problem can be mitigated by countermeasures such as use of robust modulation, coding and accessing schemes, for example spread spectrum modulation with code division multiple access (CDMA) or use of directive antennas in order to minimize interference to adjacent satellite systems.

At present the radio spectrum used in satellite communication extends from a few hundred megahertz to around 30 GHz with an interest in extending the upper range as demand for spectrum increases. Because of the heavy and growing dependence of the modern society on radio systems, some parts of the radio spectrum are in great demand. As mentioned, MSS allocations in the L (1.5–1.6 GHz) and S (2.0–2.5 GHz) bands are in heavy use. Broadband mobile services at bit rates comparable to those of the fixed satellite services (FSSs) cannot be supported on these limited allocations. There is a wide allocation to MSS in K_a band hence this band is used for broadband MSS applications in land portable, maritime and aeronautical applications. Moreover, the extra high frequency (EHF) bands have also begun to receive attention from the research community (see Chapter 14). Several World Radio Conferences (WRCs) addressed MSS-related issues, resulting in the merging of land, maritime and aeronautical allocations at the L band, allocation of spectrum for non-geostationary satellite systems, satellite components of third generation mobile systems (IMT-2000, International Mobile Telecommunication-2000) and direct-to-person radio and multimedia broadcast systems. Table 3.1 summarizes the main MSS link allocations, but this list is indicative, not

Table 3.1 A sample of MSS service-link frequencies allocated by the ITU

Up-link (MHz) (Earth-Space)	Down-link (MHz) (Space-Earth)	Comments
1626.5–1660.5 except 1645.5–1646.5	1525–1559 except 1544–1545	Generic MSS band; primary allocation; used by a majority of MSS; in great demand now; previously the band was segmented into land, maritime and aeronautical bands
1645.5–1646.5	1544–1545	Distress and safety band
1610–1626.5	2483.5–2500	Primary allocations, allocated in WARC-92 (World Administrative Radio Conference); used by big-LEO systems but allocation can be used by any type of MSS. In the USA the band is reserved for big-LEO MSS operators who cover the world (excluding polar regions) for 75% time and provide telephony service to the USA; a footnote permits use of 1613.8–1626.5 MHz in Space-Earth direction (used by Iridium)
2670–2690	2500–2520	Effective from 1 January 2005
1980–2010	2170–2200	Satellite component of IMT-2000 world-wide
1970–2010	2160–2200	Satellite component of IMT-2000 in USA

(Adapted from ITU, 2008.)

exhaustive. The reader should refer to Article 8 of the most recent issue of the *Radio Regulations* for the latest allocations.

3.2.1 Spectrum Sharing Methods

Intersystem spectrum sharing must be compliant with a formal ITU procedure for which ITU has laid out specific guidelines, including standardized calculation methodology to avoid conflicting interpretation, while intrasystem interference is managed by the operator who will try to maximize the capacity following techniques best suited for its operational philosophy. For example, an operator may trade off interference against fade margin to achieve increased network capacity in favour of link reliability; essentially a trade-off between revenue and customer satisfaction. At present, typical intrasystem interference margins used in practice for demand assigned (DA) single channel per carrier (SCPC) multiple access schemes are 0.5–1.0 dB and intersystem margins are 6% of total noise, per interference entry. In the limiting case, MSS systems become interference limited, similar to their terrestrial counterpart rather than thermal noise limited (as are most satellite systems). We will see in Chapter 4 that this is particularly true for CDMA schemes. Degradation caused by an interfering carrier can be more severe if interference is coherent or intelligible. This can occur, for example, when its modulation characteristics are identical to those of the wanted carrier; gateways are more susceptible in this respect, because of their high sensitivity.

3.2.1.1 Interference Calculations

Figure 3.2(a) and (b) shows general models for interference calculations.

Figure 3.2(a) shows a model for estimating interference caused to a mobile or received on a satellite. The offset angle θ between the wanted and interfering source determines the discrimination between the wanted and the interfering signals depending on the receiver antenna gain at this offset; for a feeder link, the mobile station is replaced by a fixed earth station.

Figure 3.2(b) illustrates a model for estimating the interference between spot beams of the same satellite; note that, in this case, isolation between the wanted and interfering carrier is provided by the spacecraft antenna pattern. A carrier transmitted at an effective isotropic radiated power (EIRP) of P dBW towards the centre of coverage of spot beam 2 is received as an interference at −18 dB lower at a location M positioned at the edge of spot beam 1 resulting in carrier to interference ratio (C/I) of 15 dB when the frequency is reused in spot beam 1. The interfering carrier power level in the forward direction (space to Earth) is given as

$$\mathrm{I} = \mathrm{I_e}(\theta)\,\mathrm{P_i I_s}(\varphi)(\mathrm{lp})(\Delta \mathrm{c}) \tag{3.1}$$

where $\mathrm{I_e}(\theta)$ = the ground station antenna gain towards interfering source, $\mathrm{P_i}$ = the satellite transmitter output of interfering carrier, $\mathrm{I_s}(\varphi)$ = the satellite antenna gain towards interfering source, lp = the path loss and Δc = the fraction of interfering carrier power captured within the wanted receiver's bandwidth.

In the return direction, satellite and ground station roles are reversed. The return service link interference analysis exhibits a unique feature because the locations of both wanted and

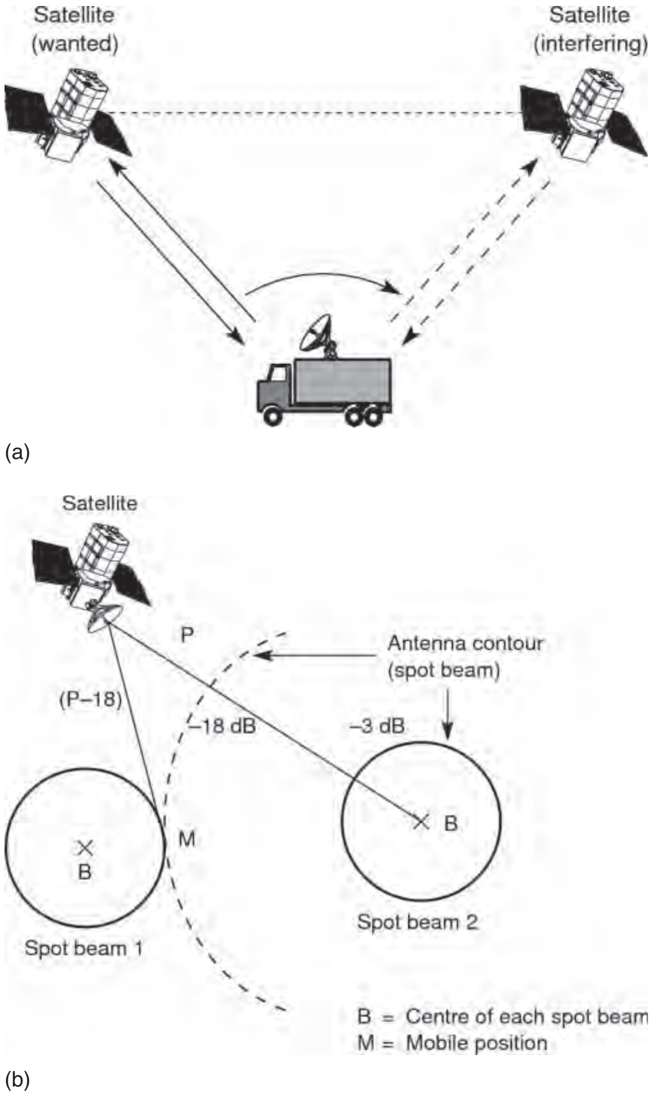


Figure 3.2 (a) A generic interference model of a mobile satellite system. (b) Interference model of a mobile satellite system in a spot beam environment

interfering mobile earth stations are unknown. The worst locations of each are useful for planning purposes but in practice, this assumption is pessimistic as the probability of such an occurrence is low and temporal. This aspect is addressed later in the section.

Note that the main isolation is derived through discrimination of either the satellite antenna or the ground antenna or both in combination with Δc, which can be reduced by staggering the central frequency of the interferer to reduce the interference power captured by the receiver demodulator.

Interference calculations for the forward direction can be summarized as follows:

1. Calculate the offset angle θ.
2. Estimate ground station antenna gain towards the interfering satellite.
3. Estimate satellite antenna gain towards the ground station.
4. Calculate power (I) of each interfering carrier received within the pre-detection bandwidth of the wanted receiver.
5. Calculate wanted carrier power (C).
6. Calculate C/I_t, where I_t is the sum of the interfering power from all the sources. As an approximate rule, 20% of the total noise is budgeted for interference, which includes inter and intrasystem interference.

Carrier levels are estimated using standard link calculations (see Section 3.4). Note that carrier levels at mobile receivers fluctuate randomly and hence C/I_t is a statistical quantity. Offset angles for geostationary satellites may be estimated using standard identities in the ITU recommendations. For non-geostationary orbit (NGSO) constellations, satellite movement causes randomness in C/I_t and therefore its statistics are estimated by computer simulation, or where possible, analytically. To assist intersystem coordination, frequency planning and antenna manufacturers, satellite and mobile station antenna side-lobe patterns for a given application are specified generically as a mask that sets bounds on side-lobe performance (angular offset from boresight versus isolation). For hand-held and portable antennas, omni-directional or near-omni-directional antennas are most cost-effective. Unfortunately, that implies little antenna discrimination and therefore the discrimination must be derived from satellite antennas. This is one reason why MSS systems for hand-held services use a large number of narrow spot beams to provide spectrum utilisation efficiencies necessary for supporting these high-volume services. However, note that the primary reason for deploying narrow spot beam is to close the radio link, which is made possible through highly directive spot beam antenna.

Contrary to FSS, where ground stations remain fixed, in MSS systems the mobile Earth stations are permitted to have unrestricted mobility within the service area. The traditional approach for intersystem coordination and interference analysis is to ensure that the (C/I) ratio remains within the limits when the mobile is situated at the point that causes or receives the highest levels of interference. For intrasystem interference management, this stringent specification can be relaxed due to the following considerations:

- The probability of simultaneous transmissions from worst locations is low, especially if the locations happen to be in remote or inaccessible areas.
- MSS systems usually have voice-activated transmissions, which reduces average interfering power.
- MSS systems use power-controlled transmissions, which tend to reduce the average interfering power.
- The probability of occurrence of simultaneous fading when the link margin of the wanted carrier is exhausted in conjunction with the worst interference geometry is low and hence some fade margin is likely to be available to accommodate increased interference.
- In MSS systems that cover several time zones, beams in differing time zones exhibit different time-of-day loading pattern (note: satellites exhibit a cyclic loading that peak during business hours). Thus the probability of co-channel interference in such beams is reduced.

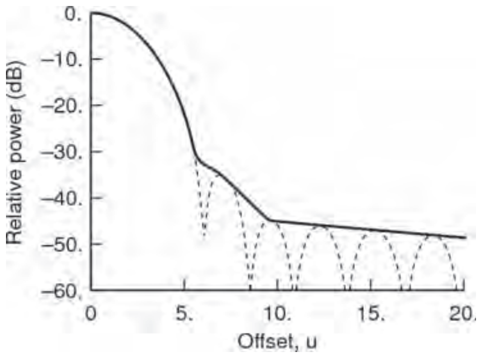


Figure 3.3 An example of a spot beam pattern and its mask. (Source: Vatalaro *et al.*, 1995. © 1995 IEEE. Reproduced with permission.)

It is therefore conceivable to derive C/I statistically as a function of percentage of locations and accept slight degradation to a small percentage.

The problem of interference calculations becomes complex in non-geostationary (NGSO) satellite systems, due to a dynamic path profile that causes temporal variations in offset angles, path loss, propagation conditions and interference environment as satellites rise and set. Moreover, handover between satellites or beams causes discontinuity in interference levels, due to abrupt changes in path profile and transmission frequency. Interference in such scenarios is calculated through computer simulations and defined statistically.

The simulation model for interference analysis in the service link of an NGEO system comprises the following components:

- **constellation definition**: which comprises orbital parameters of each satellite of the constellation with appropriate phase relationships between satellites;
- **ground projection of antenna pattern including the side-lobe performance of each satellite**: Figure 3.3 (Vatalaro *et al.*, 1995) illustrates an example of a spot beam pattern and its mask, assuming a Bessel function model of satellite antenna pattern. A Bessel function of the first kind and order of 2 with a 20 dB edge taper was used for this diagram;
- **regions** of interest;
- **communication traffic model**: Probability of traffic generation and its geographic distribution provide a realistic estimate of interference; for example probability of a service targeted for land being used in oceans is low. Furthermore traffic volume is correlated to the time of day and hence interference is time dependent. Furthermore, traffic generation is strongly correlated to the time of day and hence level of interference is also time dependent; thus reuse between otherwise interfering areas may well be possible due to staggering of busy hours of spot beams (see Chapter 7). Intersystem coordination is, however, done assuming a worst interference scenario to avoid unnecessary controversy;
- **mobile Earth station antenna pattern mask**: including its side-lobe performance, which is usually provided by the operators as an antenna radiation pattern envelope;
- **carrier parameters**: defined in terms of frequency, EIRP, voice activity, modulation scheme, pre-detection bandwidth and C/I tolerance;
- **propagation model**: characterising fading and power control. Section 3.3 discusses a number of propagation models;

- **simulation granularity**: in terms of time increments and number of Earth points, which together with run-time and simulation method sets bounds on statistical accuracy of the simulation results; a trade-off can be applied between accuracy and run-time for a given computer hardware;
- **statistical data processing**: results obtained must be presented statistically, such as cumulative distribution of C/I in terms of geographical location, to enable system-level performance evaluation;
- **graphical representation** of results is essential for comprehension and analysis, as trends in large volumes of data are easier to assimilate visually.

Figure 3.4 represents a pictorial view of a model that captures the essential units for interference calculations of a non-geostationary system.

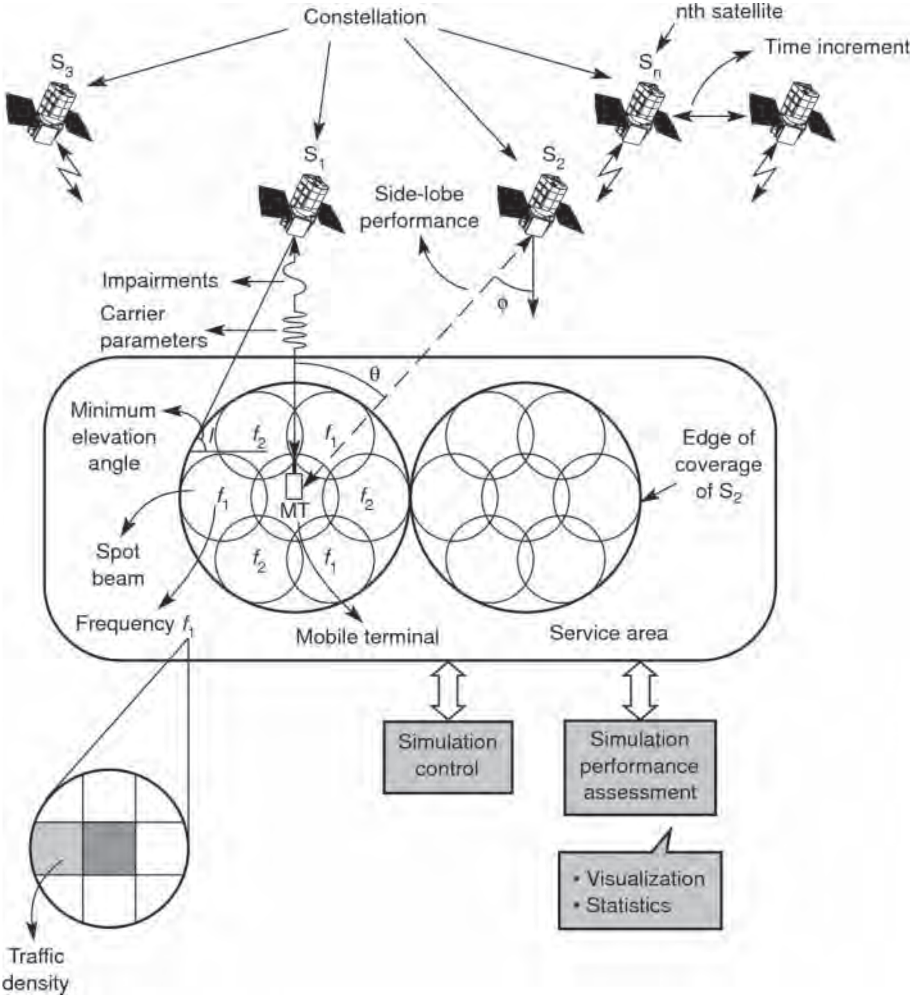


Figure 3.4 Entities of a simulation model for interference calculations

Figure 3.5 shows an example of a simulation comparing four types of constellations similar to Iridium (LEO1), Globalstar (LEO2), Odyssey (MEO) and a four-satellite geostationary orbit (GEO) constellation under simplified assumptions.

In each case, 49 spot beams were used comprising seven clusters, each with seven cells. A frequency pool was divided into seven subsets and each subset was used in the corresponding cell of other sets (see Figure 3.6).

Each satellite uses the same set without attempting to minimize interference at the intersection of the orbital plane; the return link was evaluated with mobiles transmitting at the same power, that is, without any power control. Simulation was carried on uniformly distributed grid points over the Earth within the latitude $\pm 70°$, assuming uniform traffic density. Performance was estimated for each grid point at each simulation instant and averaged, weighting appropriately for traffic density and service area (i.e. weight = 0 above ± 70 and 1 below these latitude ranges). Finally, the results of each instant were averaged over the constellation period. The outage probability was defined as the probability of exceeding the C/I threshold. The high outage probability for non-geostationary systems was attributed to the simplistic frequency planning arrangement used. The simulation was extended to study various strategies in frequency planning to achieve better reliability (see Chapter 9). The model was further refined to include Ricean fading.

The spectrum allocated by the ITU has to be shared equitably between operators. In fact, a regulatory authority may have to apportion a spectrum within its jurisdiction to more than one operator. Thus, the spectrum sharing has to be managed both at country and international levels, where more than one jurisdiction is affected. The international MSS L band spectrum is managed through regular meetings between operators, where each operator's requirements and usage are reviewed and mutually agreed interim sharing arrangements are

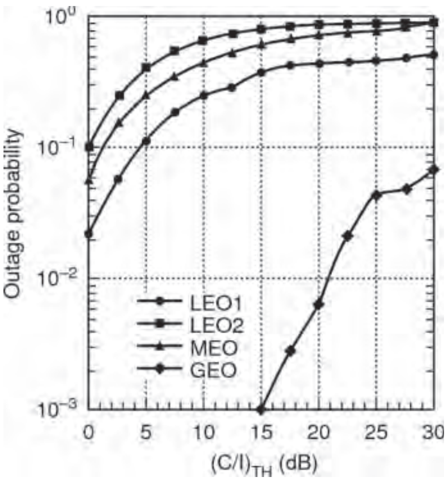


Figure 3.5 Example of a simulation comparing threshold carrier to noise ratio of outage probability of four types of constellations similar to Iridium (LEO1), Globalstar (LEO2), Odyssey (MEO) and a four-satellite GSO constellation under simplifying assumptions. (Source: Vatalaro *et al.*, 1995. © 1995 IEEE. Reproduced with permission.)

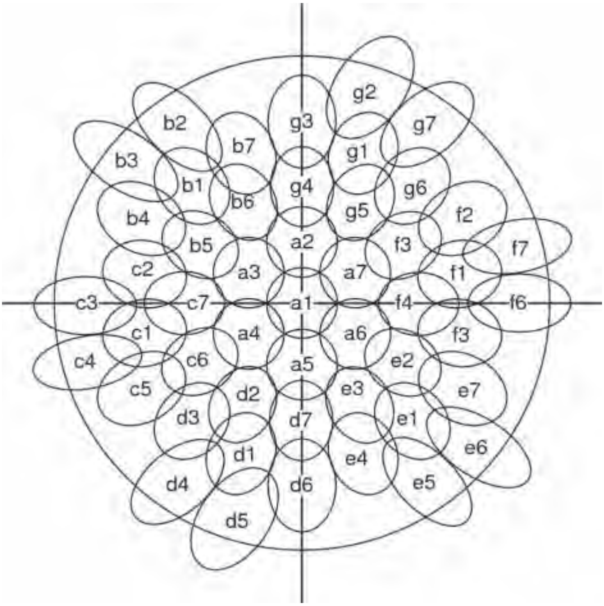


Figure 3.6 Spot beam pattern and cluster subdivisions used in the simulation. (Source: Vatalaro *et al.*, 1995. © 1995 IEEE. Reproduced with permission.)

coordinated. This way sharing arrangements are revised regularly on the basis of requirements. The process offers control over spectrum hoarding, where an operator claims rights over the spectrum without actual usage.

The method by which a regulatory authority divides spectra between operators within its jurisdiction varies. The regulatory authority may use its own licensing policy, such as first-come-first-served, an open bid, spectrum auction, etc. The authority has to negotiate the spectrum on behalf of the operators through the ITU when the demand is not covered by existing frequency allocation, as was the case, for example in the early 1990s when the licence for non-geostationary satellite systems was sought in the USA.

3.2.2 Spectrum Forecast Methodology

Regulators require long-term spectrum forecasts for planning growth and decline of a service generically and operators target their forecast to specific markets for both short- and long-term planning. Spectrum forecasts are also necessary by each operator for operational planning and readjustment of sharing arrangements between operators. For operators with existing services, short-term forecasts can be based on trend analysis including the effects of known events. Chapter 9 outlines further details regarding such methodology. Such forecasts tend to deviate from reality in the longer term because of extraneous influences such as competition, technological obsolescence or breakthrough. There is usually no historic data at the start of service. In these circumstances, a model is developed from existing information and influences. Chapter 7 discusses a methodology for developing forecasts in these latter situations.

Figure 3.7 illustrates a flow chart of a generic methodology, which may be used for deriving long-term MSS spectrum forecasts for a chosen service – a country, region or the world. The model can be modified and tailored on the basis of available or estimated data.

Spectrum requirements are categorized by the environment because of differences in traffic requirements and propagation conditions. For each environment and service, the target population is estimated; traffic penetration within the population gives an estimate of market size, which is weighed for loss to competition, other service offerings such as very small aperture terminal (VSAT) or terrestrial mobile systems, and others. The extent of penetration is influenced by the existing infrastructure, affordability of the target population, their attitude towards acceptance of new technology, exposure/experience with similar technologies and social trends such as the way people communicate during work or during leisure, Internet penetration, etc. It is then necessary to estimate traffic generated by each terminal in the busiest hour, the holding time of each call for circuit-mode services and the average

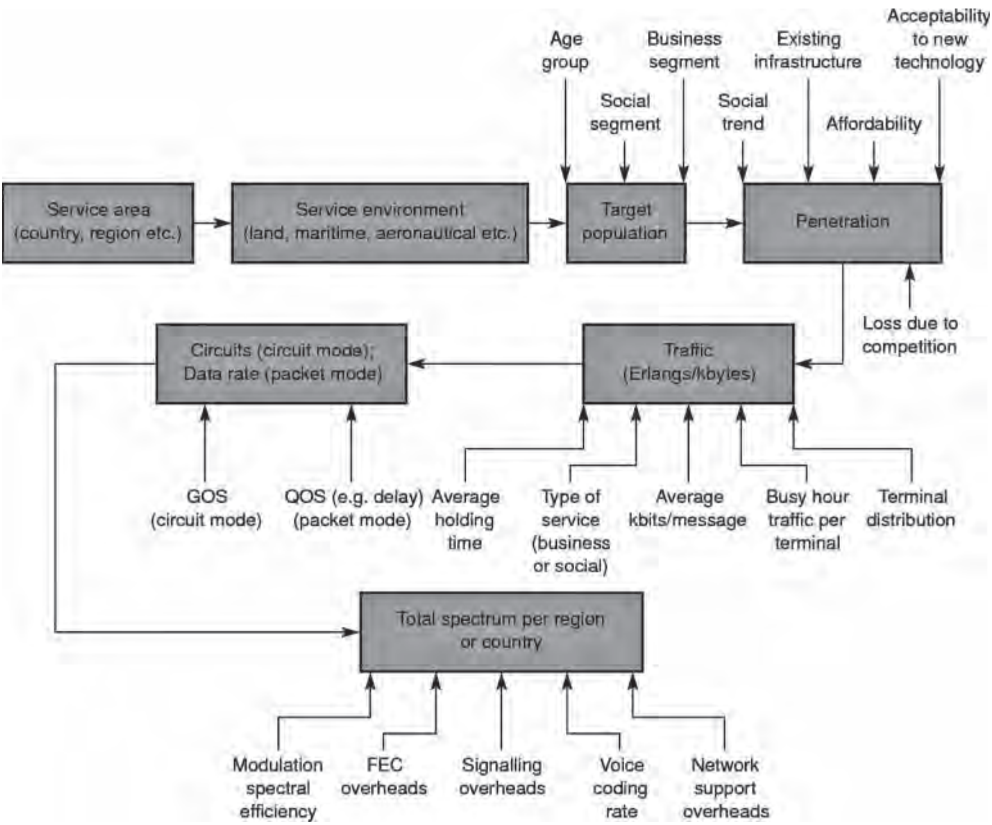


Figure 3.7 A schematic of methodology for estimating long term spectrum forecast in a given region

message length for packet-mode services. The busiest hour may be staggered if the service carries a mix of social and business traffic due to differences in their usage characteristics. The methodology is useful for early planning of a service, but its accuracy is sensitive to assumptions.

The total traffic carried for the circuit-mode traffic is

$$E = T_a * N[\text{Erlangs}] \tag{3.2}$$

where T_a = average traffic per terminal during busy hour and N = number of terminals. The total traffic for the packet mode is

$$P = [M + C]N \text{ Kbytes} \tag{3.3}$$

where M = average message length per user (Kbyte) during busy hour and C = coding overheads (Kbyte).

The total spectrum for each service is estimated individually and summed, taking into consideration the modulation efficiency, coding overheads, grade of service (for circuit mode) and permissible delay with packet retransmissions (packet mode), including network overheads, such as signalling, network test and support channels.

3.3 Propagation Characteristics

In an MSS system, feeder link propagation conditions are stable and much easier to manage than in the service link. Feeder links use large fixed stations, whereas service links operate to small mobile earth stations exhibiting dynamically variable path geometry. Generally, the propagation environment is more hostile in a land mobile environment than in maritime and aeronautical environments. The problem of fading in the land mobile environment is exacerbated by a need to use small and hence low-gain, wide beam-width antennas due to size and economic constraint resulting in low received power and signal fluctuations due to multiple signal entry. It is for this reason that land mobile communication systems only became viable in the 1990s even though the market potential of this service was already known for some considerable time; it required almost two decades for the industry to develop the technology to offset these inherent limitations.

Propagation effects are frequency dependent. At present, feeder link frequencies range from 4–30 GHz, whereas service links are mostly in the L and S bands with the use of the K_a band at an early stage and a growing interest in higher frequency bands.

Propagation characteristics are also affected by the orbital characteristics of the constellation. Additional variables to be considered for non-geostationary satellites are time-dependent range and elevation angle. General propagation characteristics, which apply to both the feeder and service links, irrespective of the local environment but caused by the troposphere and ionosphere are summarized in the next section, followed by review of the propagation models applied specifically to land, maritime and aeronautical environments. The magnitudes of various degradations are quoted only to demonstrate their likely impact on radio links. The reader may observe that land mobile channels have been dealt with in more detail than others. This is due to the variability of land mobile communication environments, which has necessitated a large number of studies throughout the world resulting in a vast number of measurement campaigns as well as theoretical and empirical modelling methods.

3.3.1 General Propagation Characteristics

Signals travelling between a satellite and the ground are affected by the intervening medium causing changes to signal level, polarization and noise contamination. The effects are frequency-dependent and caused mainly by the troposphere and the ionosphere. The troposphere consists of the first few tens of kilometres of the atmosphere where clouds, rain and fogs form, and the ionosphere is the ionized region that extends between ~80 and 1000 km around the Earth. The main sources of degradation in the troposphere are gaseous absorption in the atmosphere, absorption and scattering due to fog, cloud and rain, signal fluctuations due to atmospheric turbulence and depolarization due to rain. In the ionosphere, signals undergo changes in polarization and occasionally suffer rapid signal fluctuations known as scintillation. In addition to these impairments, the received signals become contaminated by noise from extraterrestrial and man-made sources.

Tropospheric effects on radio-wave propagation are well documented, therefore only a brief summary is presented here. Since K_a band MSS systems are emerging, some results for the K_a band service link are included in Section 3.3.2.

The most noteworthy tropospheric effects are:

- gaseous absorption;
- attenuation by hydrometers;
- scintillation;
- depolarization.

Gaseous absorption increases with frequency and peaks around 22.2 GHz due to water vapour absorption and close to 60 GHz due to oxygen (see ITU-R Rec. 390-4). The absorption depends on temperature, pressure and humidity of the atmosphere as well as the elevation angle of the ground–satellite path. Absorption reduces with a reduction in the humidity and an increase in elevation angle. For example, in the frequency range of 1–18 GHz, one way gaseous absorption for 100% humidity at the zenith varies approximately in the range (calculated for vertical polarization using United States standard atmosphere for July, at 45° N latitude) ~0.03–0.5 dB, which increases to ~0.35–5.7 dB at an elevation of 5°. These values become critical when propagation margins are low, as in MSS links and at higher frequencies. The ITU's Radio Telecommunication Sector (ITU-R – formerly the CCIR) model for estimating gaseous absorption through the atmosphere is made up of the following four steps (ITU, 1990):

- Determine frequency-dependent specific attenuation coefficients in decibels per kilometre for oxygen and water vapour.
- Determine equivalent height for oxygen and water vapour (dependent on location latitude).
- Adjust these values to account for surface temperature.
- Determine total gaseous attenuation.

Attenuation by hydrometers refers to attenuation caused by water particles existing in the atmosphere, such as fog, cloud, rain and ice, out of which rain produces the most significant attenuation by scattering and absorption. Over the past decades, considerable effort has been spent throughout the world for the development of rain attenuation prediction methods (e.g. ITU-R Rec. 564-3; Ippolito, 1986; Crane, 1980). Prediction techniques have been developed for the FSS, which also apply to the feeder links of the MSS. Representative link margins for 99.50–99.95% link reliability are respectively of the order of 3–20 dB at 20 GHz and 6–30 dB at 30 GHz for the continental climate region of the USA. However, tropospheric effects at the L band are negligible compared to the shadowing and multipath loss.

Attenuation A_{rain} of radio waves propagating through rain extending length L of the path is given as

$$A_{rain} = \int_0^L \alpha d\alpha \tag{3.4}$$

where α = specific attenuation of rain (dB/km).
α can be estimated theoretically, but in practice, empirical statistical methods are used due to randomness associated with atmospheric conditions. The specific attenuation is estimated as

$$\alpha = \mathrm{aR^b} \tag{3.5}$$

where a and b are frequency-dependent constants and R = the surface rain rate at the location of interest.

The most reliable estimates are obtained by measurements in regions of interest. Measured data are presented as cumulative distribution of fade levels. Prediction models are based on these results and extended to other regions. Generally, prediction techniques use Equations 3.4 and 3.5 as follows:

$$A_{rain} = \mathrm{aR^bL(R)} \tag{3.6}$$

where L(R) = effective length of rain.

The difference in prediction method lies in the method used for estimating L(R). The most accurate method is to use a model that applies specifically to the region of interest, but this is not always possible due to insufficient data. The ITU-R recommends a model that can be applied to any climatic region of the world (ITU-R Rec. 618-10). The inputs to the model are point rainfall rate for the location of interest, height of the location above mean sea level, satellite elevation angle, latitude of the location and frequency of operation. Referring to Figure 3.8, steps involved in calculating A_{rain} by this method are as follows:

- Calculate the effective height (h_R) of the location.
- Calculate the radio path length, Ls up to the rain altitude.
- Calculate the horizontal projection of the radio path, L_G.
- Obtain 1 minute of integrated rain intensity, $R_{0.01}$, for 0.01% of an average year; if local data is not available then values given in the recommendation can be used. These values have been derived from world climate data.
- Obtain the reduction factor, $r_{0.01}$, for $R_{0.01} < 100$ mm/h.
- Calculate the specific attenuation, γ_R, using tabulated frequency-dependent coefficients.
- Calculate the predicted attenuation for 0.01% of time using the equation:

$$\mathrm{A0.01} = \gamma_R L_s r_{0.01} \,\mathrm{dB}. \tag{3.7}$$

- If necessary, obtain attenuation for another percentage of time using the available empirical relationship.

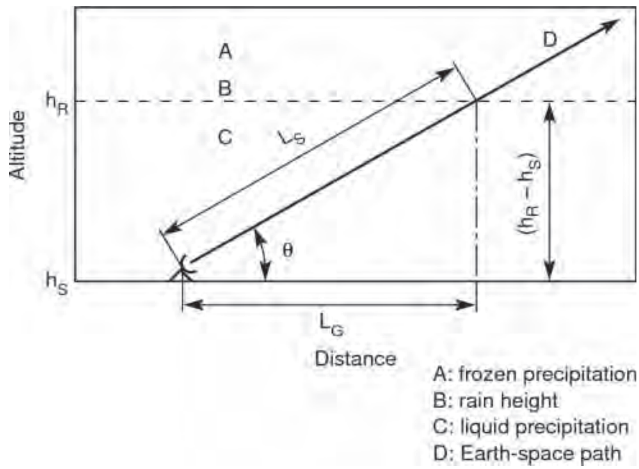


Figure 3.8 Geometry for estimating attenuation due to rain (Source: Figure 1 of ITU-R, 1992-1. Reproduced with permission of ITU.)

ITU-R Recommendation 840 provides a method for estimating attenuation due to water clouds or fog, of known liquid content. It has been observed that attenuation due to clouds up to 30 GHz is relatively insignificant for fixed links, but at higher frequencies, attenuation from clouds of high water content (e.g. cumulonimbus) becomes significant, for example at 100 GHz the attenuation can reach 4–5 dB, increasing up to 8 dB at 150 GHz.

When a new frequency band is considered reliable rain attenuation data are not available. This was the case when the K_u and K_a bands were introduced and is currently the situation for bands above 30 GHz (e.g. Q and V bands). One approach is to scale data from a frequency for which data are well characterized. A number of empirical scaling models have been developed in ITU-R Report 721 (ITU, 1986).

To meet the specified link reliability throughout a year, it is essential to convert the annual p% attenuation statistics to the worst month of the year, as certain months are the wettest (e.g. monsoon season in the Indian subcontinent). Techniques for this type of scaling are well documented, for example ITU-R Recommendation 581 provides a method of converting the worst-case statistics to annual statistics. However, it has been observed that there is a notable difference in year-to-year statistics; variations in excess of 20% rms are possible.

Site diversity offers a solution for mitigating effects of heavy attenuation occurring at the K_a band and above. It has been observed that intense rain cells typically have dimensions of a few kilometres. Thus, if sites are spaced several kilometres apart, it is unlikely that severe fading occurs at both sites simultaneously. By selecting or combining signals from two spaced sites, it is therefore possible to reduce demands on spacecraft power and improve link reliability. Such a scheme is designed in the feeder link of the Iridium and the (proposed) Teledesic systems.

Diversity performance is characterized by two measures – a diversity improvement factor and diversity gain. The diversity improvement factor is the ratio of single site time percentage for a given attenuation to that obtained after application of diversity for the same attenuation. Diversity gain is the difference in decibels between attenuation on a single site before

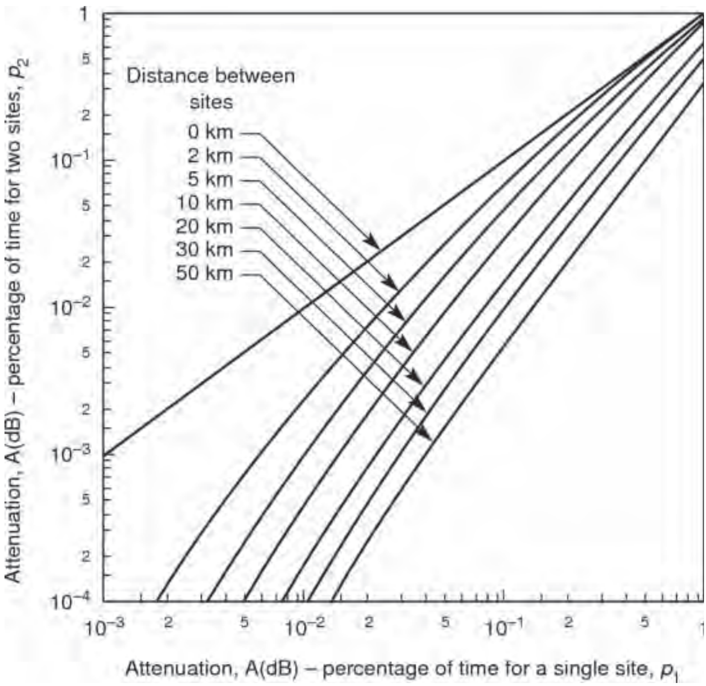


Figure 3.9 Improvements through two-site diversity (Source: ITU-R, 1992-1. Reproduced with permission of ITU.)

and after application of diversity for the same time percentage. Empirical methods for estimating both types of measures are available in the literature (ITU-R Rec. 618-2). Figure 3.9 represents two-site diversity improvements as percentage of time with and without diversity for the same attenuation (ITU-R Rec. 618-2). Notice the improvement in reliability as the site separation increases – for example the 10^{-2}% time reliability of a single site improve to 2×10^{-3} and 10^{-4}% with a site separation of 2 and 30 km respectively.

When a dual-polarized radio wave travels through rain or ice, a part of the power from one polarization gets coupled to the orthogonal component as noise due to the anisotropic behaviour of the medium causing impairments to the orthogonal component. Degradation is measured as cross-polar discrimination (XPD) or cross-polar isolation (XPI), given as

$$\mathrm{XPD} = 20 \log |E_{11}| / |E_{12}| \tag{3.8}$$

where E_{11} = the received co-polarized electric field strength and E_{12} = the electric field strength coupled to orthogonal polarization.

XPI is defined as

$$\mathrm{XPI} = 20 \log |E_{11}| / |E_{21}| \tag{3.9}$$

where E_{21} = the cross-polar component received from the other polarization.

For practical purposes, XPD and XPI are assumed equal.

It can be shown that XPD degrades with decrease in frequency at a given co-polar attenuation; and increase in co-polar attenuation at a given frequency. It has also been observed that generally XPD of a vertically polarized wave is better than that which horizontal polarization and circularly polarized waves have; ~ 10 dB lower XPD than horizontally polarized waves for the same co-polar attenuation. Depolarization caused by ice occurs without accompanying co-polar attenuation; however, its magnitude is about 25 dB and therefore not very significant.

In the absence of measurements, prediction models recommended by ITU-R can be used. Empirical relationships show dependency on frequency, polarization, elevation angle and the canting angle of rain. Measurements at 19 and 28 GHz taken in the USA show XPD of 16–18 (Horizontal-Vertical polarization at 38.6°. elevation) and 12–13 dB (vertical polarization 38.6–46° elevation) respectively for 0.01% of the time.

Small-scale refractive index variations of troposphere (or ionosphere) cause signals to arrive at the receiver via different paths causing rapid level variations, called scintillation, due to random phases and power of the multipath signal components. Tropospheric scintillation depends on the season, the local climate, frequency and elevation angle. Its magnitude increases with frequency and a reduction in elevation angle. Degradation caused by tropospheric scintillation becomes noticeable above 10 GHz and is significant in the K_a band and above. Scintillation can be accompanied with rain. The interdependency of rain fades and scintillation is at present not well understood; hence they are treated as statistically independent. Models for estimating the distribution of scintillation are available in the literature (e.g. ITU-R, 2012). A technique for laboratory simulation of tropospheric scintillation has also been reported by this author and used for evaluation of a tracking system. Typical values of scintillation in K_a band are reported for a mid-latitude/elevation location to be of the order of 0.2–0.3 dB peak-to-peak level in winter, 1 dB in the clear conditions of summer and 2–6 dB in some types of cloud while fade rates range from 0.5 to over 10 Hz.

The K_a band (20–30 GHz) has received considerable attention recently particularly for the service link because the band is relatively less crowded and high antenna gain is possible through small apertures, attributes attractive for wideband personal and mobile very small aperture terminal (MVSAT) communication services. Deep signal fades due to rain and to a certain extent tropospheric scintillation pose the severest problems in this band. The problem gets exacerbated by motion of mobiles. A large number of propagation measurements have been conducted since 1970, mostly applicable to fixed geostationary satellite systems. Due to continuous changes in elevation angle of non-geostationary satellites, these results are not directly applicable to systems deploying non-geostationary satellite constellations. Moreover, only a limited amount of work has been conducted to establish the effect of motion on rain fade statistics (see Chapter 14).

For radio link design of fixed earth stations operating through non-geostationary satellites, it is necessary to combine attenuation due to path loss, rain, scintillation, gaseous absorption, shadowing and multipath as a function of elevation angle. Feeder links use large antennas and are sited to avoid blockage, therefore shadowing loss and multipath variations are minimal. Path loss, rain fade, tropospheric scintillation and gaseous absorption for the specific site and constellation are estimated as a function of elevation angle, using appropriate models such as discussed earlier; cumulative path loss distribution obtained, and standard link budget analysis used for sizing Earth stations and satellites. Figure 3.10(a) illustrates the

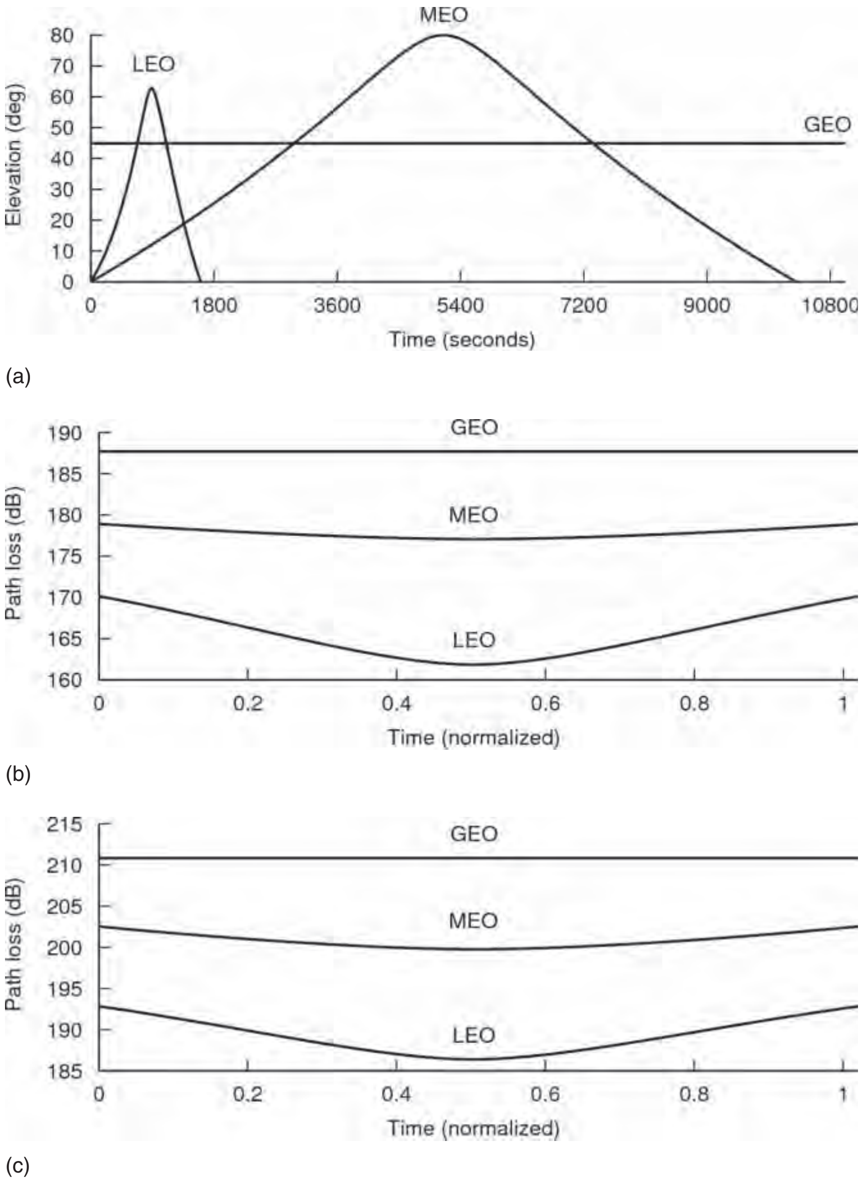


Figure 3.10 (a) Elevation angle variation for LEO, MEO, GEO satellites from Bhopal, a location in central India (77.5°E and 23°N). (LEO parameters: altitude = 1400 km, inclination = 52°, ascending node = 60°E; MEO parameters: altitude = 10 350 km, inclination = 45°, ascending node = 60°E.)
(b) A comparison of path loss at 1.5 GHz for LEO, MEO and GEO satellites from London, UK; Orbital parameters of LEO and MEO are identical to those listed in Figure 3.10(a).
(c) A comparison of path loss at 20 GHz for LEO, MEO and GEO satellites from London, UK; Orbital parameters of LEO and MEO satellite are identical to those in Figure 3.10(a) (All parts source: Graphics AR.)

elevation angle variation from Bhopal, a city in central India for a low earth orbit (LEO) and a medium earth orbit (MEO) satellite pass and a geostationary Earth orbit (GEO). Figure 3.10(b) and (c) demonstrates respectively the path loss at 1.5 and 20 GHz for the same LEO, and MEO satellites and a GEO satellite from London, UK. Note that path loss decreases as the satellite rises until the satellite is closest to the ground station, increasing again as the satellite sets; this observation causes no surprise, as it essentially implies that the worst conditions occur at the lowest elevation angle.

Figure 3.11 illustrates variations in elevation angle, path loss and available link margin during a single pass of a LEO satellite link similar to that in the Iridium system, from White Sands in the USA (Ippolito and Russel, 1993).

It is thus evident that the link margin is a function of elevation angle, for example when path loss reduces from 69 dB at the edge of coverage (EOC) to 59 dB, the margin increases from 5 to 15 dB in a system designed with a worst-case (EOC) margin of 5 dB. It follows that link reliability depends on the minimum operational elevation angle; this rather obvious conclusion has considerable influence on the size and cost of a non-geostationary satellite constellation, as observed in Chapter 2. Figure 3.12 represents variation in 20 GHz gaseous attenuation and margin for rain effects for a single pass of the LEO satellite. Annualized outage probability for the link using the ITU-R global attenuation model is captured in Figure 3.13; note wide variations in outage probability for the pass; outage probability in the middle 3 min of the pass is 0.01% and increases to 1% at ± 5.5 min in the 14.7 min pass. A corollary to these observations is the need of a large dynamic range in earth station receivers due to large variations of signal level.

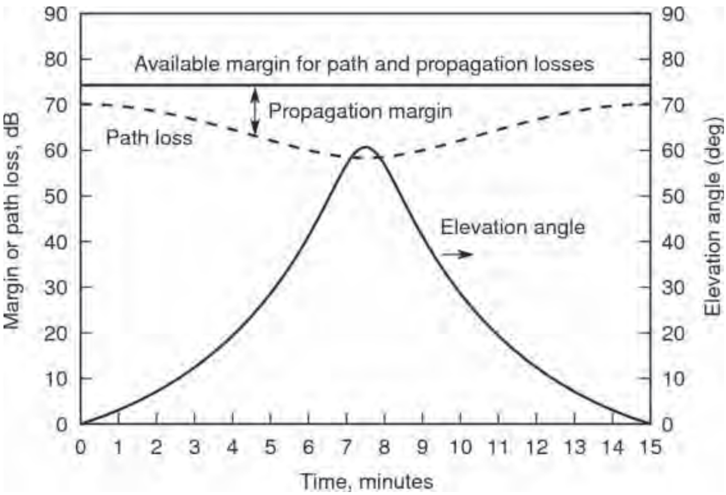


Figure 3.11 Elevation angle, path loss and available link margin for a 20 GHz link during a single pass of a LEO satellite from White Sands in the USA (ascending node = 100°W, altitude = 756.3 km). (Source: Ippolito and Russel, 1993. © 1993 IEEE. Reproduced with permission.)

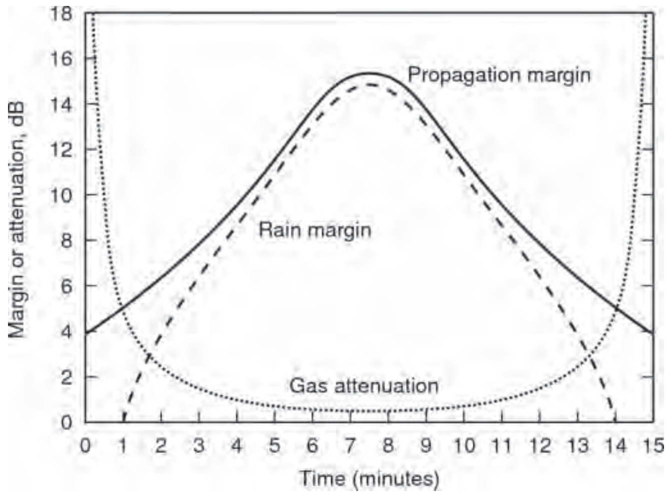


Figure 3.12 Variation in 20 GHz gaseous attenuation and margin available for rain effects for a single pass of a LEO satellite. (Source: Ippolito and Russel, 1993. © 1993 IEEE. Reproduced with permission.)

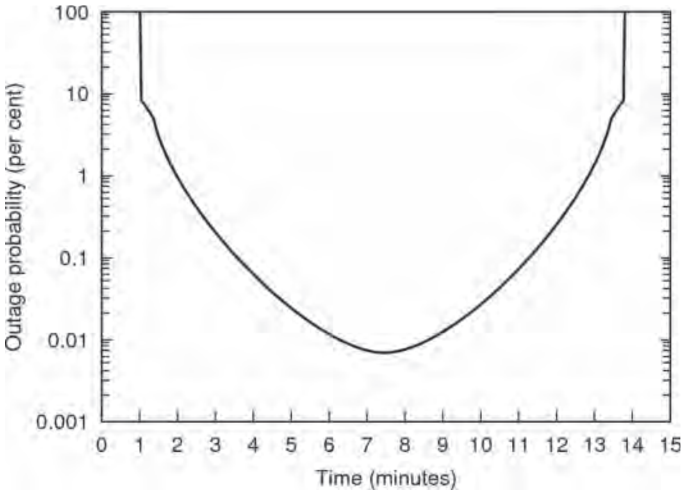


Figure 3.13 Annual 20 GHz outage probability for a single pass of a LEO satellite (ascending node = 100°W, altitude = 756.3 km). (Source: Ippolito and Russel 1993. © 1993 IEEE. Reproduced with permission.)

3.3.1.1 Ionospheric Effects

The ionosphere affects radio waves in a number of ways. It causes rotation of radio wave polarization, propagation delay, refraction, variations in angle of arrival, absorption, dispersion and scintillation. Table 3.2 provides the worst-case estimated values of these effects at 0.5, 1, 3 and 10 GHz.

Table 3.2 Worst case estimated one-way values of rotation of radio wave polarization, propagation delay, refraction, variation in angle of arrival, absorption, dispersion and scintillation for 0.5, 1 and 3 and 10 GHz at 30° elevation (Assumptions: Total Electron Content = 10^{18} electrons/m^3, a high daytime value at low latitudes with high solar activity)

Effect	Frequency (f) dependence	Frequency (GHz)			
		0.5	1	3	10
Faraday rotation (°)	$1/f^2$	432	108	12	1.1
Propagation delay (μs)	$1/f^2$	1	0.25	0.028	0.0025
Refraction (s)	$1/f^2$	< 144	< 36	< 4.2	< 0.36
Variation in angle of arrival (rms sec)	$1/f^2$	48	12	1.32	0.12
Absorption (auroral and polar cap) (dB)	$\sim 1/f^2$	0.2	0.05	6×10^{-3}	5×10^{-4}
Absorption (mid-latitude) (dB)	$1/f^2$	< 0.04	< 0.01	< 0.001	$< 10^{-4}$
Dispersion (ps/Hz)	$1/f^2$	0.0032	0.0004	1.5×10^{-5}	4×10^{-7}
Scintillation (dB peak to peak) (observed near equator, early local night time, at equinox under conditions of maximum sunspot number)	$1/f^3$	–	> 20	~ 10	~ 4

(Adapted from ITU-R report 884-2.)

For mobile satellite communications operating in 1–2 GHz band, polarisation rotation, also known as the Faraday Effect, and scintillation are dominant. The Faraday Effect is caused by the interaction of electromagnetic waves with Earth's magnetic field. Circular polarized waves are unaffected by the Faraday Effect and therefore MSS links use circularly polarized waves, as this eliminates the need for polarization tracking, whereas feeder links can use linear (or circular) polarization as they can incorporate polarization tracking without significant cost impact. The effect is generally predictable and hence can be compensated by rotating the polarization of transmitted waves (or receiver antenna) in an opposite sense.

Peak-to-peak signal-level fluctuations of up to 20 dB are known to occur at ~ 1.5 GHz. Hence, it is not surprising that scintillation can cause occasional outage to MSS links at the L and S band, for several hours in a year, but fortunately, the onset of severe scintillation events is late in the evening when traffic volumes are quite low. Personal communication systems that often operate close to threshold are particularly susceptible. Feeder earth stations in the equatorial regions are known to suffer occasional outage due to scintillation. We observe that navigation systems such as the global positioning system (GPS) and Global Navigation Satellite System (GLONASS) also operate in the L band and therefore are susceptible.

Ionospheric irregularities occur due to certain solar, geomagnetic and upper atmospheric conditions. Hence ionospheric scintillation is affected by sunspot activity, position of the Sun relative to location and hence time of day, latitude of location, season and magnetic activity. Table 3.3 summarizes the main features of scintillation.

Table 3.3 Main features of scintillation

Parameter	Region			
	Equatorial	Mid-latitude	Auroral	Polar
Dominant cause of scintillation	Plasma bubbles	Occurrence of ionospheric spread, F	Related to auroral and geomagnetic activity	Related to auroral and geomagnetic activity
Severity	Largest intensity	Very quiet to moderate	Moderately active to very active	Intensity directly related to sunspot numbers
Diurnal behaviour	Night – maximum Day – minimum	Night – maximum Day – sporadic	Night – maximum	Night – maximum
Seasonal behaviour	Longitude dependent; for example Ghana: maximum – November/March, minimum – solstices Peru: Maximum – October to March Minimum – May to July	Maximum – spring/summer Minimum – winter	Function of longitude sector	Function of longitude sector
Effect of solar cycle	Activity/intensity strongly dependent on number of sun spots	Limited data	Activity/intensity strongly dependent on number of sun spots	Activity/intensity strongly dependent on number of sun spots
Magnetic activity (measured in terms of K_p)	Depends on longitude (increasing or decreasing with the measure K_p)	Independent of K_p	Activity increases with K_p	Activity increases slightly with K_p

(Adapted from ITU, 1992a.)

The severity of scintillation is defined in a number of ways. A commonly used measure is known as the S_4 scintillation index, defined as the standard deviation of received power divided by the mean value of the received signal power

$$S_4{}^2 = \sigma/\mu \tag{3.10}$$

where σ denotes ensemble average and μ is the mean signal power.

The fading period is another useful system parameter. Its value is variable from event to event and ranges in the region of 1–10 s in the gigahertz range of interest. Longer periods lasting tens of seconds have also been observed at very high frequency (VHF) and ultra high frequency (UHF).

Due to large regional variability, the best estimate of scintillation is obtained from measured local data. In the absence of measurements, an applicable model can be used (for example see ITU, 1992). Some useful conclusions of ITU-R recommendations are summarized as follows:

- Measurements taken at frequency f can be scaled to another frequency according to the dependence, $f^{-1.5}$.
- Instantaneous fluctuation of a scintillation event can be approximated by the Nakagami probability density function, which can be used for estimating cumulative fade distribution.
- Power spectral density of scintillation events varies widely due to variations in drift velocity of refractive index irregularities; power spectra density slopes of f^{-1} to f^{-6} have been reported; a value of f^{-3} can be used as an approximation in absence of real data.
- Scintillation magnitude $s_4{}^2$ varies as 1/cos(i) up to i ~ 70° where i is the zenith angle, and for lower values of i, variation is between cos(i) and $(\cos(i))^{1/2}$
- Seasonal and longitudinal dependence of S_4 can be approximated as,

$$S_4 \ \alpha \ \exp(-\beta/W) \tag{3.11}$$

  where β is seasonal and longitudinal dependent parameter and W is a location and day of year dependent constant.
- Figure 3.14 (ITU, 1992a) illustrates geomagnetic latitude, longitude and local time dependence of scintillation at 1.6 GHz.
- The cumulative distribution of fading statistics for 1 dB peak-peak fluctuation as a function of monthly sunspot numbers, taken from a 4-GHz measurements campaign for of a number of locations is illustrated in Figure 3.15 (ITU Rec 531-2). In an equatorial region (Hong-Kong) 1 dB peak to peak variation can occur for up to 5% of time in periods of high solar activity.
- The probability of simultaneous occurrence of ionospheric scintillation and rain fading in equatorial regions is relatively high especially in years of high sunspot activity resulting in differences to statistics. The occurrence probability of such events has to be considered in the design of high reliability radio links, such as for critical safety applications.

3.3.2 Land Mobile Channel

A service link provides mobility to a user, but at a cost! Motion of mobile terminals causes variability in signal path geometry, resulting in environment and velocity-dependent signal

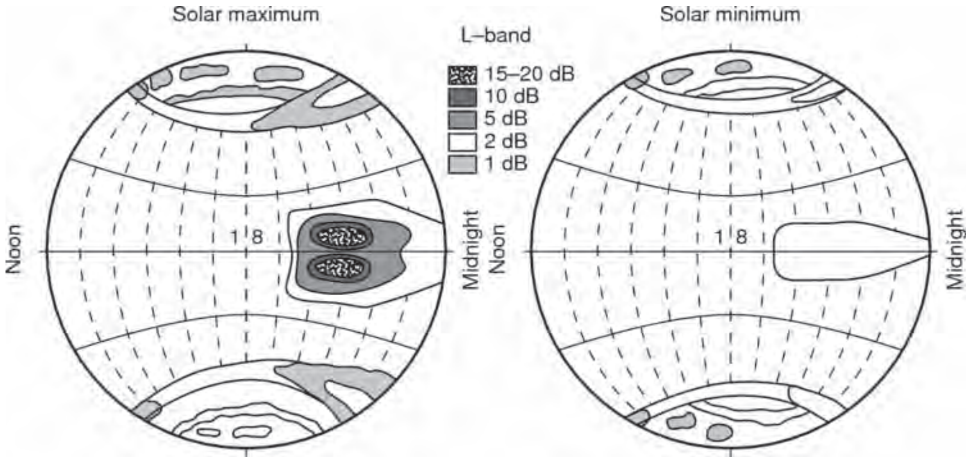


Figure 3.14 Scintillation occurrences from observations at 1.6 GHz showing geomagnetic latitude, longitude and local time dependence of scintillation (Source: ITU, 1992a. Reproduced with permission of ITU.)

impairments at the receivers. Additionally, reduction of antenna size (and hence wider beam width), essential for user mobility, causes reception of multipath signal emanating from nearby objects. The rate of signal fluctuations due to shadowing and multipath depends on the user's velocity; coupled to this is a velocity-dependent Doppler component; satellite motion introduces a further Doppler component. Generally, propagation conditions are most demanding for the land mobile channels.

In narrowband applications at data rates less than 100 kb/s, the signal fades simultaneously across the full bandwidth; this is called *flat fading*. However, when signals are sufficiently wideband, such as in high-fidelity sound broadcasting or wideband CDMA links, signals can suffer frequency selective fades, which affect only parts of the signal bandwidth causing degradation by inter-symbol interference and thereby a reduction in throughput. Furthermore, when terminals operate near their threshold, some otherwise relatively minor effects become noticeable.

Figure 3.16 depicts the general trend in throughput at the L band for various sizes and speeds of user terminal in a land environment. Throughput reduces as the size of user antenna reduces and the travelling speed increases. The throughput in the L band is bound by limited availability of spectrum; for this reason K_a band has become attractive for high-throughput/low-mobility systems with the additional advantage of a relatively smaller antenna dish, although this band is severely affected by rain fades. It is worth noting here that this band is in regular use for MVSATs, including those installed on ships and air crafts.

Figure 3.17 illustrates a signal path geometry to characterize propagation behaviour in a land mobile channel.

Received signals, already affected by the ionosphere and troposphere, depending on the environment around the receiver, may comprise various components, as shown in Figure 3.17: a direct path; diffracted components, specular components caused by reflections

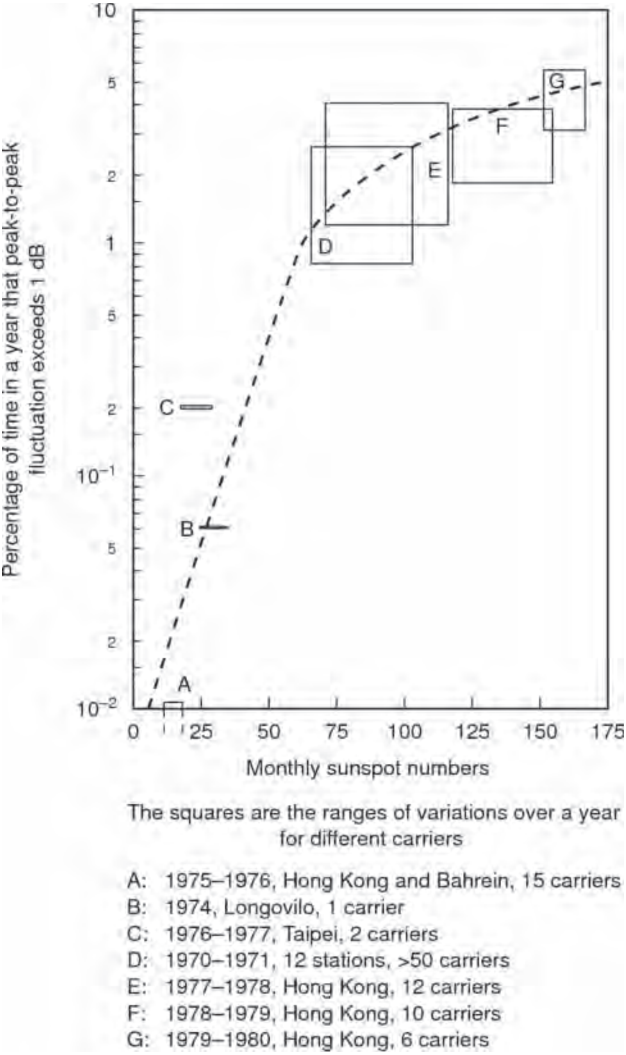


Figure 3.15 Cumulative distribution of fading statistics taken from a 4-GHz measurements campaign (Source: ITU, 1992a. Reproduced with permission of ITU.)

from metallic or smooth objects such as the body of a car or a smooth sea surface; and diffused components caused by reflection and scattering of objects around a mobile. The magnitude of each component is strongly influenced by the environment around a mobile. In an urban environment, the direct path may be shadowed for a considerable period by obstructions such as buildings and therefore the diffracted or scattered components will dominate.

The resultant signal r(t) at a receiver can be represented as

$$r(t) = \sigma(x)a(t) + s(t) + d(t) \tag{3.12}$$

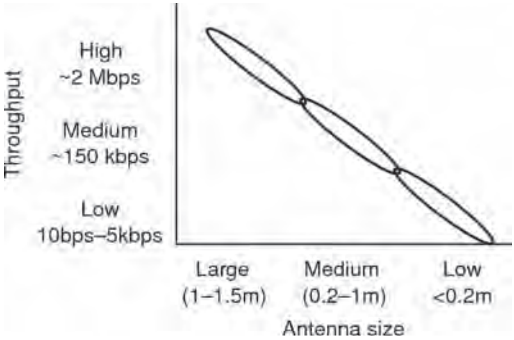


Figure 3.16 Throughput versus antenna size

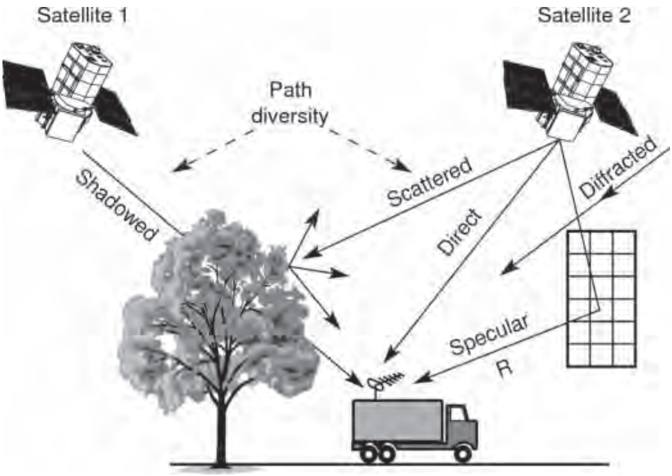


Figure 3.17 Geometry of an MSS communication channel showing various signal components – direct, shadowed, scattered, specular components and path diversity

where $\sigma(x)$ is an environment, elevation and RF-dependent attenuation factor, $a(t) =$ magnitude of direct path, $s(t) =$ magnitude of specular components and $d(t) =$ the magnitude of diffused components.

The magnitude and behaviour of each component depend on a number of factors, which include terminal antenna characteristics, speed of travel, user cooperation, elevation angle, type of environment and RF bandwidth.

The following considerations are necessary to characterize the propagation of MSS radio links:

- tropospheric and ionospheric effects, which were discussed in Section 3.3.1.1;
- receiver antenna characteristics;
- environment dependence;
- elevation angle dependence;
- frequency dependence;
- path loss variation in non-geostationary satellite systems.

Common descriptors for characterisation include fade margin, level crossing rate (LCR), power spectral density, phase characteristics, and others. A number of approaches are used in characterizing land mobile channel:

- **Empirical models** are based on applying regression fits to the measured data. It is possible to derive simple algebraic equations relating channel impairment such as fade level to experimental variables chosen by the experimenter for the application at hand. The impairment can be determined as a function of elevation angle, type of environment, antenna characteristics, frequency, season, foliage content, and so on.
- **Statistical models** are based on applying numerical analysis to measured data to obtain statistical descriptions of signals such as probability distribution function and cumulative probability distribution, which are used to derive useful system parameters such as the fade margin for a given percentage of time, LCR, power spectral density, etc. This type of model is sometimes also based on heuristic reasoning, based on understanding and simplification of the propagation mechanism.
- **Deterministic models** are based on an accurate characterization of the environment and propagation mechanism, applying ray or electromagnetic wave theory in conjunction with models of receiver antenna pattern characteristics, vehicle velocity, elevation angle, and so on. A received signal is the sum of individual wave components.

A model derived from measured data can provide an accurate statistical estimate of the system parameters for the area where measurements were undertaken. Due to the vast variability in landscape, the results should be extrapolated to a new area cautiously. Thus, characterization of 'global' areas (e.g. Europe) requires a vast database covering a huge number of locations and numerous variables such as elevation angle and environment description.

The mathematical approach avoids the need for complex and expensive experiments in situations, such as characterization of signal in specific environments and at frequencies for which data are not available, and it is best performed by computer simulations. Examples of such instances include: characterization within buildings, urban areas or at EHF where measurements may be impractical; investigations on diversity; characterization of unusual scenarios such as aeronautical propagation over sea routes; site selection of fixed terminal in a dense urban area; wideband characterization, etc.

3.3.2.1 Environment Dependence

The first part of the section addresses the statistical approach to propagation modelling with specific examples of prevalent models; this is followed by methods for mathematical generation of data.

Equation 3.12 defines the signal arriving at a mobile antenna. Each component of the equation varies randomly, and therefore can be characterized statistically. Depending on the physical interpretation of the propagation mechanism, a number of variants of statistical models have been proposed; the difference in models lies in the interpretation of the shadowing mechanism on direct and scattered paths, as illustrated in the examples that follow.

- When direct line of sight is available, the probability distribution of the signal amplitude is considered Ricean, with the assumption that the component a(t) is constant; in-phase and quadrature components of the diffused signals are independent of each other and normally distributed with zero mean. In the absence of a direct component, the signal is assumed to

be Rayleigh distributed with a log-normal mean; scattered signals are also assumed to be shadowed. Ricean distribution is characterized by the Ricean factor, defined as the ratio of direct signal to multipath ratio; and the Rayleigh function is characterized by a mean and standard deviation.
- In a variation on the previous approach, only the direct path is assumed to be shadowed.
- Yet another approach is to assume scattered signals to be a composite of clear and shadowed components.
- In the environment-dependent approach, the probability distribution is taken as a composite of a number of probability distribution. A Ricean distribution is assumed for the fraction of time when a signal is unshadowed and a Rayleigh distribution with log-normally distributed mean for the period when the signal is shadowed. From a practical viewpoint, the inclusion of environment dependence provides a powerful technique for quantifying system performance in a mix of environments.
- A further refinement of the environment-dependent approach is to represent channel behaviour as a combination of a Markov process and a statistical model, where the state of a channel is described by a Markov process and the signal variations within each state are modelled by a statistical model appropriate for the environment. The Markov process allows the signal to assume one of the M states with a probability, which only depends on its previous state. A number of authors have used this approach; their solutions differ in the number of states of M and the chosen probability distribution of signal within each state. We discuss this further in a following section.

Table 3.4 summarizes the characteristics of several prevalent narrowband models.

These statistical descriptions are strictly valid under simplifying mathematical assumptions, such as applicability of the central limit theorem; therefore, the description does not apply to a specific situation. When a particular area is of prime interest, data stored from a previous measurement campaign provides the most accurate characterization. In absence of measured data the simulation approach provides an alternative.

Both primary and secondary statistics are useful in the analysis and design of satellite systems. Primary statistics refer to parameters that deal with long term behaviour of the signal, and secondary statistics deal with short time parameters of the signal. The primary statistical descriptor 'cumulative fade distribution' relates fade level and its probability of occurrence; and is used for estimating link margin for a given link reliability. LCR, fade duration at a given fade level, phase characterization including Doppler statistics are useful secondary statistics for system design and evaluation. Their magnitude depends on the probability distribution of the signal envelope and the velocity of the mobile.

In the Markov chain modelling method, two matrices are used in defining the process – a $1 \times M$ state probability array whose ith element defines the probability of a system being in a state i and a $M \times M$ transition probability array whose element defines the probability of transitioning from one state (say, i) to another (say, j). Models have been proposed, ranging from two to four states. Probability distributions within each state depend on the environment and preference. The overall probability distribution function is given as the summation of $p_i w_i$ where p_i is the probability distribution function of signals in state i and w_i is the probability of the signal existing in state i. Figure 3.18(a) shows a two-state Markovian chain as proposed by Lutz *et al.*, (1996). The channel is assumed to alternate between

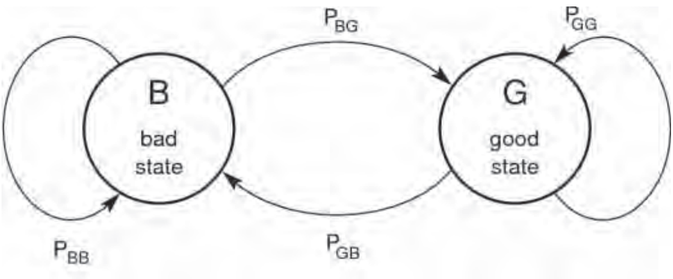
Table 3.4 Characteristics of some prevalent models

Model	Complex channel process, $r = r_x + jr_y$
Loo (1985)	$r = Se^{j\phi_o} + Re^{j\phi}$ S: log-normal R: Rayleigh distributed ϕ_o, ϕ : uniform
Rice-log-normal (Corazza and Vatalaro, 1994)	$r = RSe^{j\theta}$ S: log-normal R: Ricean θ: uniform
Generalized Rice-log-normal (Vatalaro, 1995)	$r = RSe^{j\theta} + x_1 + jy_1$ S: log-normal R: Ricean distributed $x_{1,}\ y_1$: Gaussian
Hwang *et al*. (1997)	$r = A_c S_1 e^{j\phi} + RS_2 e^{j(\phi+\theta)}$ S_1,S_2: log-normal R: Rayleigh distributed A_c: constant
Poca-lognormal (Babalis and Capsalis, 1998)	$r = RSe^{j\theta}$ S: log-normal R: Poca-log-normal distributed θ: uniform Note: Poca distribution is similar to Rayleigh distribution; here, the number scattering components is assumed finite instead of infinite as in Rayleigh distribution
Pätzold *et al*. (1998)	$r = Se^{j(2\pi\ f_{pt}+\theta_p)} + x_1 + jy_1$ S: log-normal x_1, y_1: Gaussian distributed f_p and θ_p : constant

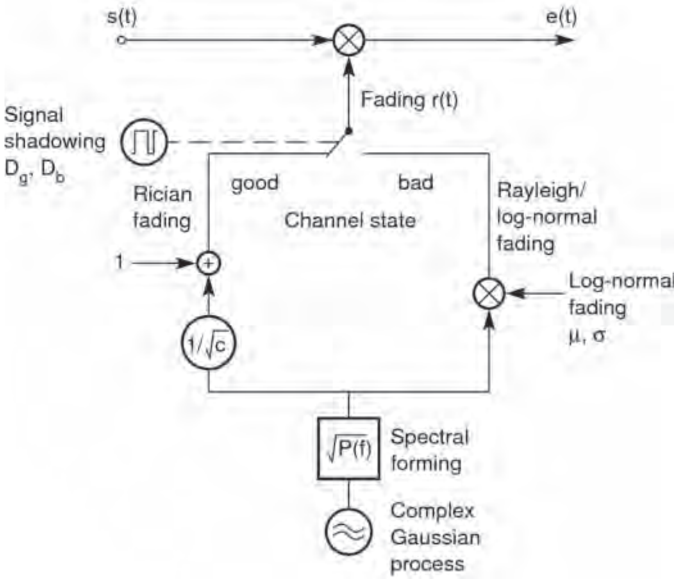
(Adapted from Karaliopoulos and Pavlidou, 1999.)

a Ricean and Rayleigh/log-normal statistical state. The transition probabilities for various environments were determined through measurements.

A simulator of a land mobile channel based on the principle is presented in Figure 3.18(b) (Lutz, Papke and Plöchinger, 1986). Fading statistics are switched between Ricean and Rayleigh distributions for mean periods of D_g and D_b, representing good and bad channel states respectively; for hand-held communicators, D_g and D_b can be measured in metres. The magnitude of D_g and D_b depends on environment and vehicle speed, and switching between the two states follows a two-state Markov model. The channel is shadowed on an average



(a)



(b)

Figure 3.18 (a) A two-state Markovian chain representing environment dependence in a land mobile communication channel representing environment dependence; P_{BB} = probability of returning to a bad channel state; P_{BG} = probability of transitioning from a bad state to a good state; P_{GB} = probability of transitioning from a good state to a bad state; P_{GG} = probability of returning to a good state. (Source: Lutz *et al.*, 1996. Reproduced with permission of John Wiley and Sons, Ltd.).
(b) A narrowband propagation channel simulator based on the environment-dependent approach. (Source: Lutz *et al.*, 1991. © 1991 IEEE. Reproduced with permission.)

for $D_b/(D_b + D_g)$. Table 3.5 summarizes observations in highway and city environments of Europe for a number of satellite elevations for L band transmissions (Lutz, 1998). Note the progressive improvement in channel conditions with elevation angle – increase in mean duration of the Ricean channel (good state), improvement in Rice factor and a reduction in fade levels.

Trees, utility poles and overpasses are known to affect propagation. Other physical objects and factors affecting propagation include terrain, orientation of the satellite with respect to direction of travel and seasonal dependence.

Table 3.5 Characteristics of L band land mobile channels in Europe

Environment	City			Highway		
Satellite elevation	10	30	50	10	30	50
Mean duration (m): Ricean	7	25	50	100	1000	10 000
Mean duration (m): Rayleigh	70	50	35	50	30	20
Ricean factor (dB)	5	9	10	10	14	18
Time share: shadowing	0.9	0.7	0.4	0.3	0.2	0
Mean attenuation: shadowing	12	12	15	9	10	14
Standard deviation of received power: shadowing (dB)	4	4	4	4	4	4

(Source: Lutz, 1998.)

Other factors known to influence propagation include satellite azimuth with respect to direction of travel, foliage density and vehicles in the vicinity. Features lying within the main lobe of the antenna pattern of the receiver influence the propagation characteristics primarily.

Attenuation caused by a variety of trees such as white pine, holly, burr oak and pin oak performed at 870 MHz for elevation angles between 20° and 45° show a mean value of 12 dB resulting in attenuation coefficients from 0.8 dB to 1.3 dB/m. The authors used Fresnel diffraction theory to model the attenuation from a single tree (Vogel and Hong, 1988). Utility poles often erected beside roads cause chirp-like signal fluctuations, due to interference between direct and scattered signals. The authors present a theoretical model for estimating the electric field, treating the pole as a scatterer. Overpasses near motorway junctions or crossings of major roads, shadow the signals completely. Measurements taken in Australia show typical shadowing of > 15 dB lasting for 0.5–1.0 s, occurring for 0.5–1% of the time.

Figure 3.19 illustrates cumulative distribution of UHF and L band fade in a clear environment without obstructions, typical of a desert (Vogel and Goldhirsh, 1990). Note that the signal level exceeds the mean signal level by about 1.5 dB for 1% of time.

Figure 3.20 compares L band fading of two environments – one dominated by multipath and the other by shadowing. It can be noted that for 1% level multipath fade is 6 dB, whereas for the shadowed environment the 1% fade level is about 20.5 dB. Similar results have been reported by several experimenters. In one case the transmitting source mounted on a helicopter (helo) was behind and in the other to the right-hand side (RHS) of the data collection vehicle.

Propagation characteristics demonstrate a seasonal dependence in environments dominated by trees, due to variations in foliage density and flowers during the year. Figure 3.21 (Butt, Evans and Richharia, 1992) reports cumulative fade distribution of the same environment for two seasons – spring 1992, when deciduous trees were without leaves, and summer 1991 when there was 100% blossom, and leaves high in water content. As expected, the spring distribution gives the lower fade level. The levels are only marginally lower at 60° elevation at 1% and identical at > 5% probability – this may be attributed to the high elevation angle when the path is clear of the trees, whereas at 80° elevation the difference between the spring and summer is about 9 dB at 1% and about 4.2 dB at 5% probability.

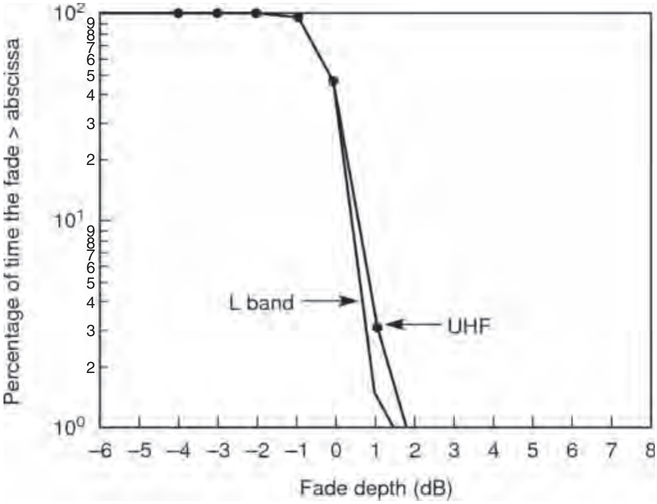


Figure 3.19 Cumulative distribution of UHF and L band fade in a clear environment without obstructions. (Source: Vogel and Goldhirsh, 1990. © AIAA 1990. Reproduced with permission of the American Institute of Aeronautics and Astronautics, Inc.)

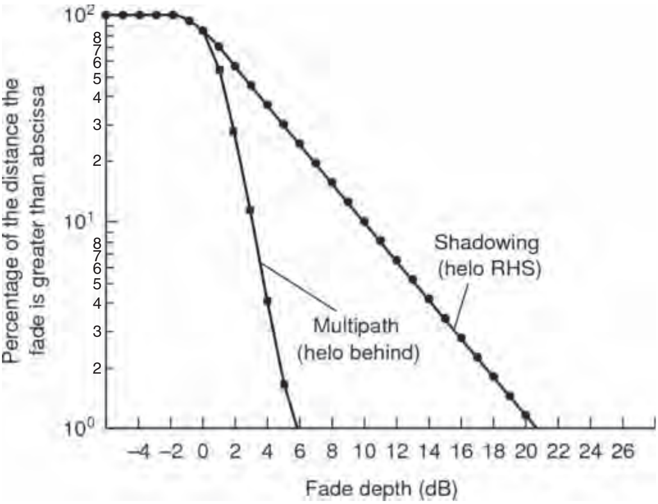


Figure 3.20 A comparison of L band fading in two environments – one dominated by multipath and the other by shadowing. (Source: Vogel and Goldhirsh, 1990. © AIAA 1990. Reproduced with permission of the American Institute of Aeronautics and Astronautics, Inc.)

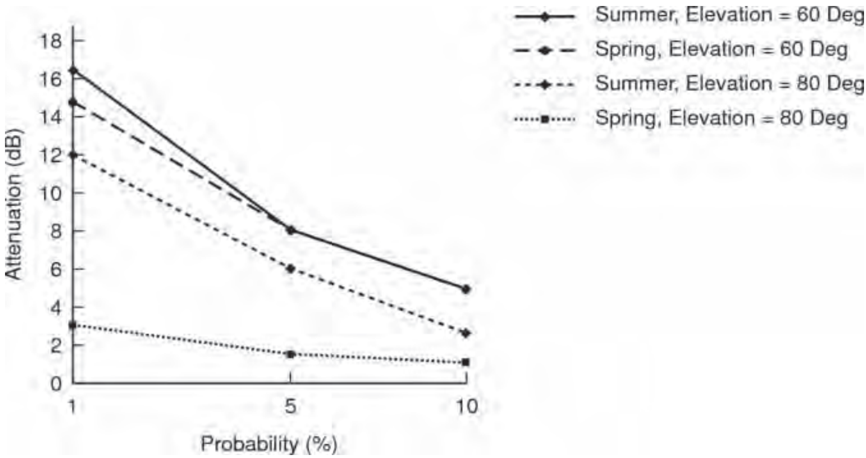


Figure 3.21 Cumulative fade distribution in the same environment for two seasons. (Source: Butt *et al*., 1992. Reproduced with permission of IET.)

Figure 3.22 demonstrates that the direction of travel on a dual carriage way and its orientation with respect to the satellite affect the fade level notably. For example, 5 dB fade depth occurs for less than 3% of the route in the southerly direction of travel whereas the same fade is experienced for about 7% of the route when travelling North. The reason of the difference is attributed to the path geometry of the environment as illustrated in the figure.

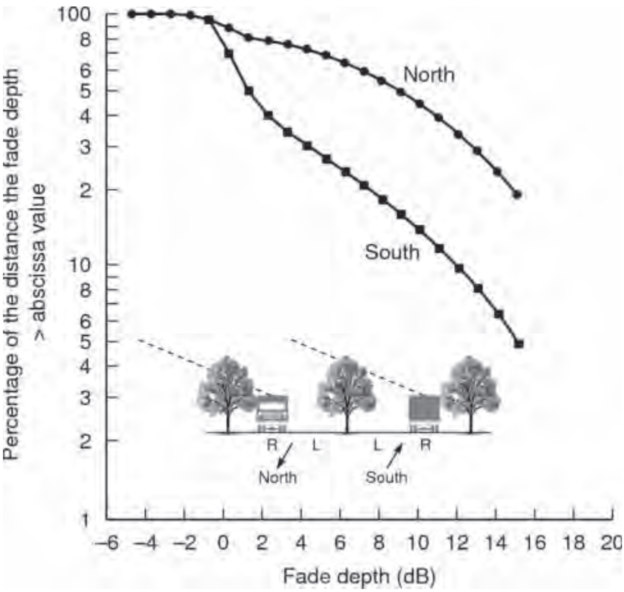


Figure 3.22 Effect of direction of travel in a dual carriage way and its orientation with respect to satellite. (Source: Vogel and Goldhirsh, 1990. © AIAA 1990. Reproduced with permission of the American Institute of Aeronautics and Astronautics, Inc.)

An example of an empirical model for estimating L-band fade at a given probability level in the range 1–20% is the empirical roadside shadowing (ERS) model derived from 640-km measurements in various types of road in the USA in a predominantly tree environment with some utility roadside poles.

The fade depth F exceeded for P% of distance covered is represented as

$$F = -M \ln(P) + B \text{ dB} \tag{3.13}$$

where $20\% > P > 1\%$, $M = (3.44 + 0.0975\ \theta - 0.002\ \theta^2)$, $B = (-0.443\ \theta + 34.76)$ and θ = elevation of satellite in degree.

Polarization of an electromagnetic wave changes on reflection, with complete polarization reversal when angle of incidence exceeds the Brewster angle. Measurements show that at higher fade levels, XPI reduces to unacceptably low values, due to multipath scattering. Measurements taken in Australia on orthogonal polarization indicated that XPI reduced from 18 dB for 0 dB fade to 0 dB at 12 dB fade level. Further data are necessary to validate such a conclusion, because measurements on each polarisation were taken at different times (Vogel and Goldhirsh, 1990). From a system perspective, this would imply that frequency reuse on opposite polarization can be difficult using conventional schemes; however, the use of spread spectrum scheme and multiuser technique could enable such a scheme (see Chapter 14).

An example of fade duration statistics at fade threshold levels of 2–8 dB in 1 dB steps for a typical US rural road, consisting of trees and utility poles as main obstructions, is given in Figure 3.23. To make the results independent of the vehicle, they have been given in terms of distance travelled in metres.

The statistics followed a log-normal distribution. The non-fade duration for the same case followed an exponential variation (Vogel and Goldhirsh, 1990). The empirical representation

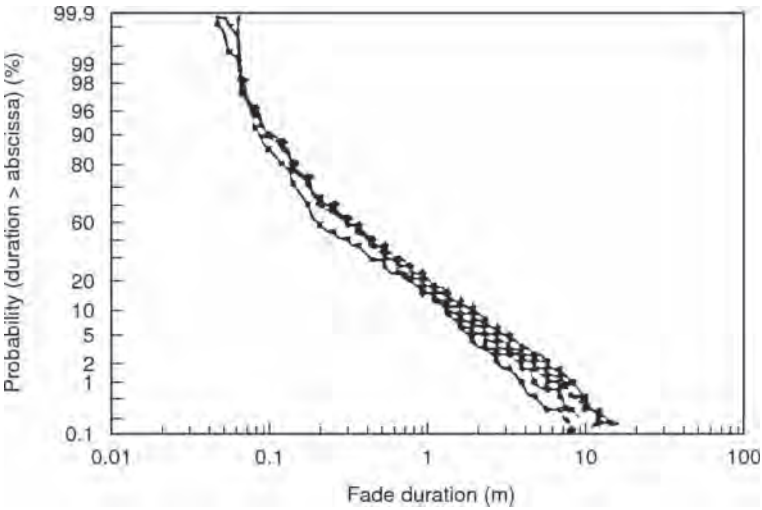


Figure 3.23 Fade duration statistics in the range 2–8 dB for a typical US rural road. (Source: Vogel and Goldhirsh, 1990. © AIAA 1990. Reproduced with permission of the American Institute of Aeronautics and Astronautics, Inc.)

of fading was derived as:

$$P\,(d_n > x) = b\,x^{-c} \tag{3.14}$$

where $P(d_n > x)$ = Percentage probability of the non-fade distance d_n exceeding x; b and c are constants in the range 12–13 and 0.55–0.84 respectively.

Equation 3.15 specifies the diversity improvement empirically possible for the environment. Diversity improvement D_{im} is defined here as the ratio of the single antenna probability to the joint probability at a specified fade margin and antenna separation. A diversity improvement of 2, for example implies that a single antenna system will fail at twice the number of locations for the chosen threshold. Figure 3.24 demonstrates the diversity improvement achievable when antennas are kept far enough apart.

$$D_{im}(dF) = 1 + [0.2\ \ln(d) + 0.23]\,F \tag{3.15}$$

where d = the antenna separation (metres) and F = the fade margin (dB).

ITU-R recommendation 681 (ITU, 2009) provides a number of techniques that predict parameters needed in planning land mobile satellite communication systems. Prediction methods are recommended for various aspects of shadowing, multipath and diversity. The scope of these techniques are summarized below.

Shadowing models address roadside tree and buildings environments with additional considerations for hand-held terminals. A technique to model building blockage using street masking functions has been suggested.

The *roadside tree-shadowing* model is an empirical model to predict fade statistics for shadowing density of 55–75% on basis of measurements at 870 MHz, 1.6 GHz and 20 GHz.

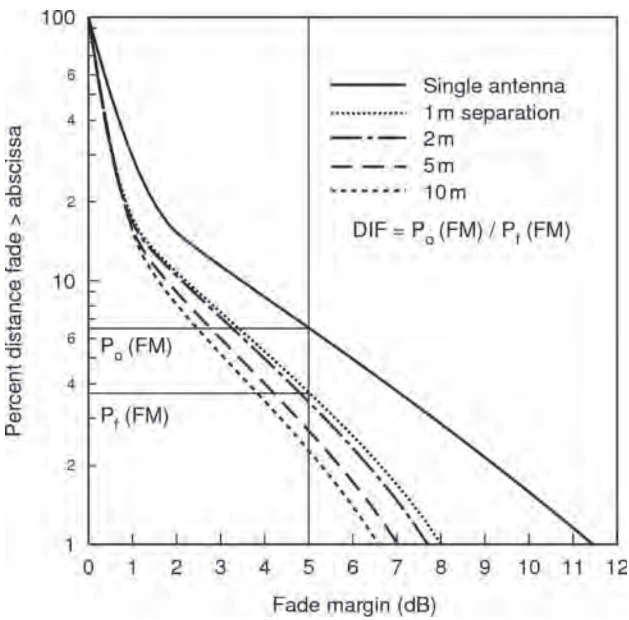


Figure 3.24 Illustration of diversity improvement. (Source: Vogel and Goldhirsh, 1990. © AIAA 1990. Reproduced with permission of the American Institute of Aeronautics and Astronautics, Inc.)

The extent of tree cover is specified by the percentage of optical shadowing due to trees at an elevation angle of 45° towards a stationary satellite. An extension of the model covers satellites in NGSO by combining statistics at each elevation angle of a satellite pass.

The *roadside building-shadowing* model is based on the assumptions that roadside building heights in an urban area are Rayleigh distributed with a provision to include Fresnel clearance to determine blocking probability.

Another approach estimates building blockage mathematically through a 'street masking function' to quantify the range of azimuth and elevation angle blocked for any given scenario. Such functions are suitable for simplified scenarios to obtain approximate statistics quickly and are derived through photogrammetric measurements or by ray-tracing.

It has been observed that the antenna pattern is modified by user's head when using a handheld terminal. It is anticipated that hadheld users will generally cooperate by moving to a favourable location during a call both with respect to head and the local surroundings. However, since this is not feasible for LEO systems, the fading probability is estimated by using the fading model proposed for non-geostationary satellites, taking into consideration the modified antenna pattern in conjunction an averaged azimuth.

Propagation data related to signal loss for reception within buildings and vehicles are available in Recommendation ITU-R P.679.

Multipath prediction models are based on empirical fits to measured data for clear line-of-sight conditions, mountainous surroundings and roadside tree environment. An omni-azimuth pattern antenna with gain variations of less than 3 dB between 15° and 75° elevation and a reduction of at least 10 dB below the horizon was used for the measurements.

The models mentioned here apply to local surroundings; in a real land mobile system the mobile encounters a variety of conditions. The *mixed-state model* is a three-state narrowband statistical model (as already described earlier in this section) – representing clear, lightly shadowed and blocked states. Procedures are suggested for predicting statistics for a single satellite link including the probability for each state and probability distribution of the signal in each state. The probability density distributions of the signal for the three states are modelled respectively as Nakagami–Rice, Loo and Rayleigh.

The *physical-statistical wideband model* proposed for mixed propagation conditions is a linear transversal filter whose output is the sum of delayed, attenuated and phase shifted versions of the input signal as described later in the section (Equation 3.33). The model comprises wide band transmission in the frequency range 1–2 GHz from a satellite located at an elevation ε, and an azimuth φ and a mobile receiver. The model is valid for bandwidth up to 100 MHz. It is based on deterministic and stochastic parameters derived from measurements from a synthetic environment as described elsewhere in this chapter.

The ITU recommendation describes two cases of *satellite diversity*. In the uncorrelated case the signals from diversity satellites are uncorrelated; and in the correlated case a certain degree of correlation is present.

We will now briefly review some representative second order statistics that influence transmission format and mobile terminal design – the LCR and Doppler. The LCR N_l is defined as (see Figure 3.25)

$$N_l = N_t/T_t \tag{3.16}$$

where N_t = total number of positive crossings and T_t = total observation time.

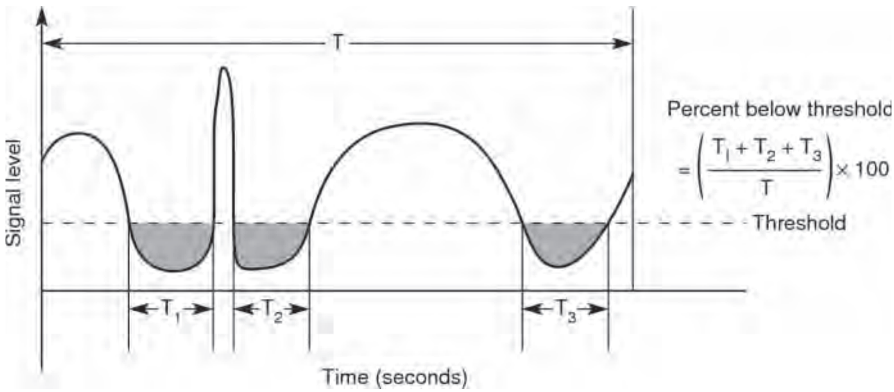


Figure 3.25 Representation of level crossing rate

Analytical expressions of LCR have been derived in the literature (Loo, 1985).

Average fade duration T_a is the average duration of a fade at a depth r (Lee, 1986),

$$T_a = F(r)/N_l \tag{3.17}$$

where F (r) = cumulative distribution of fade for a fade of r and N_l = LCR.

The Doppler spectrum of a mobile receiver using an omni-directional antenna in a two-dimensional space can be given as (Clarke, 1968),

$$S(f) = \frac{1}{\pi f_m \sqrt{1 - \left(\frac{f-fc}{f_m}\right)^2}} \tag{3.18}$$

where f_m = maximum Doppler frequency.

A two-dimensional model can represent a terrestrial system, but is not adequate for a satellite system that is better represented by a three-dimensional system. Using the geometry illustrated in Figure 3.26, Aulin derives signal auto-correlation that can then be readily Fourier transformed to give Doppler frequency in a three-dimensional space (Aulin, 1979).

Angles α_n, β_n, magnitude c_n and phase of wave components are random and statistically independent; ϕ_n and α_n are uniformly distributed between 0 and 2π; the angle β_n has a probability distribution of $p(\beta)$ that is zero outside $(-\beta_m, \beta_m)$ and $\beta_m \leq \pi/2$. Figure 3.27 illustrates the power spectral density for two definitions of $p(\beta)$ and includes results of two-dimensional model for the case of $\beta_m = 45$ (Kanatas, Kanderakis and Constantinou, 1995). Definitions of $p(\beta)$ are (Aulin, 1979) and (Parsons, 1992) respectively;

$$\begin{aligned} p(\beta) &= \cos\beta/2\sin(\beta), |\beta| \leq \pi/2 \\ &= 0 \text{ elsewhere} \end{aligned} \tag{3.19}$$

$$\begin{aligned} p(\beta) &= \pi/4\beta_m \cos((\pi/2).\beta/\beta_m), |\beta| \leq \pi/2 \\ &= 0 \text{ elsewhere} \end{aligned} \tag{3.20}$$

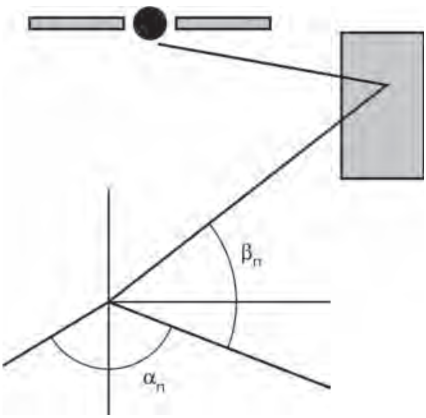


Figure 3.26 Geometry for estimating Doppler

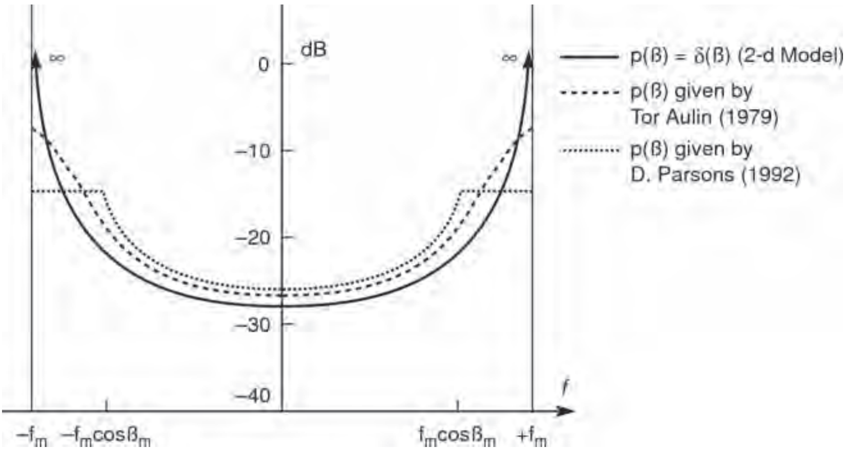


Figure 3.27 Theoretical Doppler spectrum using a two-dimensional model and two cases of three-dimensional models; f_m = maximum Doppler frequency. (Source: Kanatas, Kanderakis and Constantinou, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

The maximum Doppler frequency of the carrier is given as,

$$f_m = u/\lambda \tag{3.21}$$

where u = velocity of the mobile and λ is the wavelength of the RF.

Doppler components associated with a land vehicular user terminal are attributed to:

- *Satellite motion*: satellites move relative to the Earth, causing a Doppler shift; the relative motion from the ground can be determined when the position of the observer and orbital

parameters are known – for a geostationary satellite the Doppler is of the order of a few hundred hertz in the L band, depending on the relative positions (e.g. ± 250 Hz over 24 h);
- *Vehicular motion*: the magnitude depends on the vehicle-satellite velocity vector;
- *Multipath signals*: depending on the characteristics of the local surroundings, multipath signals arrive from random directions with random amplitude resulting in Doppler that is spread over $\pm f_m$ with the magnitude shaped by the receiving antenna's radiation pattern.

Figure 3.28 illustrates spectral signature of a 10-s signal sample recorded on a vehicle travelling at variable speeds on a motorway near London while receiving unmodulated transmissions from Inmarsat's AOR-E satellite at an elevation of 29.8° (Gambaruto, Richharia and Trachtman, 2008). The I and Q samples of the carrier were recorded on a vector signal analyser, the environment variables of each scenario was captured by means of a digital camera and recorded audio commentary; and the position and velocity of the vehicle on a GPS recorder. The spectrum was estimated by Fourier transforming 10-s sections of I/Q signal samples. Using the satellite's orbital parameters and the vehicle position the Doppler of the line-of-sight signal was estimated accurately from the instantaneous path geometry; the same data was derived by Fourier analysis of the received signal and observed to be in close agreement. Satellite Doppler and Doppler rate was estimated by Inmarsat's satellite control centre from orbital data and was observed as negligible over the sample period. Superimposed on the main Doppler is a Doppler spread caused by scattered signals that arrive through various paths, hence, angle of arrival and magnitude. The spread was captured by analysis of 10-s samples, where the path geometry was likely to remain quite invariant (Figure 3.28). Ideally the timescale of analysis should be instantaneous but for practical

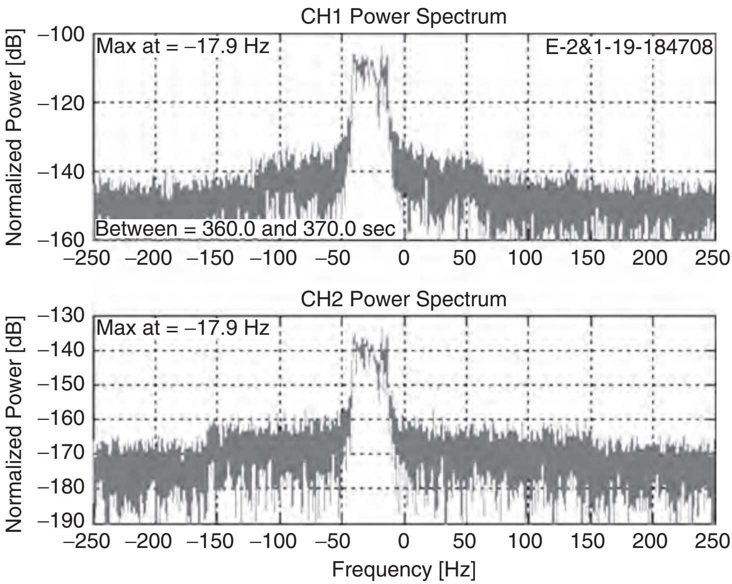


Figure 3.28 Example of Doppler spread of the main signal represented by the central high level component and the raised hump caused by multipath signals. (Source: Gambaruto, *et al.*, 2008. Reproduced with permission of John Wiley & Sons Ltd.)

reasons 10-s samples were analysed, which corresponds to typical averaging time used in land-mobile receivers (such as the Broadband Global Area Network (BGAN)).

Referring to Figure 3.28, channel 1 and channel 2 correspond to data received on receiver with two-element antenna and one-element antenna respectively. The two-element antenna is a low-gain tracking antenna that provides a receiver G/T of −15.5 dB/K at 5° elevation and 1525 MHz; and one-element antenna is a non-tracking antenna pointed towards the zenith that gives a receiver G/T of −17.1 dB/K at 5° elevation and 1525 MHz. The one-element antenna has a near omni-directional pattern in the azimuth plane and a 3 dB beam-width of about 140° (± 70°) in the elevation plane.

The centre spectrum spread of the figure is due to the Doppler variation of the direct signal. The multipath components can be observed around the direct signal for both the two-element antenna receiver (channel 1: CH1) and the one-element antenna receiver (channel 2: CH2). Thermal noise at the edges of the two-element antenna is about 10 dB below the multi-path Doppler spectrum and about 5 dB below the one-element antenna.

The multi-path spectrum is spread within ± 75 Hz of the centre frequency received on two-element antenna receiver and ± 155 Hz around the centre for one-element antenna receiver. The maximum Doppler for a speed of 110 km/h at the test frequency of 1546.195 is 155 Hz, which tallies with the two extremes of the spread of channel two (azimuth-omni-antenna). Further, note that the power spectral density of the spread for the azimuth-omni-antenna channel is reasonably uniform. Thus for this sample of the run the hypothesis of uniform scattering around the vehicle is reasonably valid with scatterers exhibiting a similar coefficient. The upper figure illustrates a narrower spectral spread as the spectrum is shaped by the spatial filtering of the azimuthally directional antenna. The qualitative conclusions from these figures indicate that multipath signals are received uniformly from all the directions for the give experimental conditions.

3.3.2.2 K_u Band and Above Land Mobile Channels

The digital video broadcasting with provision of a return channel was developed in the first instance for services to small fixed user terminals at frequencies above 10 GHz for applications like digital broadcast, video on demand, broadband Internet service – the addition of mobility to these standards extends these services to mobile platforms including railway trains, aircrafts and (see Chapter 8). This new standard, known as digital video broadcasting second generation/return channel by satellite + mobility (DVBS2/RCS+M), has received considerable attention in regions with well-developed transport infrastructure and high broadband penetration.

Although road vehicles and railways are both land platforms, they are distinguished in the propagation literature as the land mobile satellite channel (LMSC) and the railway satellite channel (RSC) because of differences in their operational environments.

Radio wave propagation becomes more ray-like at K_u band and above in comparison to the L and S bands; and since the attenuation due to foliage and other obstructions is more severe, the transition between good and bad states is distinct (Butt, Evans and Richharia, 1992). Therefore, the Markov chain model is suited for performance evaluation in these frequency bands as suggested by several authors (e.g. Scalise, Ernst and Harles, 2008). The transitional probabilities of an N-state Markov chain model correspond to N grades of channel states.

The railway channel, for a considerable part, has a clear line-of-sight to the satellite similar to a typical motorway channel, particularly above $\sim 20^\circ$ elevation. The difference with the road path geometry arises due to the presence of metallic structures and cables along the railway tracks that lead to periodic spikes in the received signal; in addition, railway channels undergo long durations of blockage while traversing tunnels or when stationed at railway platforms. The general approach has been to deal with such prolonged periods of blockage by the use of 'gap fillers' when communication is handed over to a terrestrial system (see Chapters 7 and 8). The open areas are treated statistically using, for example the Markov chain approach, with a deterministic element superimposed to deal with the impairments caused by power supply structures (whose positions on a railway route are known) and blockages caused by tunnels, etc., again, whose positions are known. The overall characterisation is obtained by combining these features on a given route. The PDF (Power Density Function) of the received signal power S can then be defined as (Lutz, Papke and Plöchinger, 1986):

$$p_s(S) = \sum_{k=1}^{n} V_k \cdot p_{S,k}(S) \tag{3.22}$$

where n is the number of states, V_k is the probability of being in state k, obtained from the state transition Matrix $M = [p_{ij}]$ and $p_{S,k}$ is the PDF associated to the fast fading and multipath phenomena of state k.

Table 3.6 illustrates the parameter of a three-state Markovian channel at Ku band obtained by measurements in the European region (Scalise, Ernst and Harles, 2008). The authors modelled the probability distribution of the received signal for the LOS (Line of Sight) segment as Ricean (represented by Ricean Factor c in the table), the shadowed portion as Suzuki/Lognormal-Rayleigh distribution (represented by mean, μ and standard deviation, σ); and in the blocked segment the signal fell below the measurement threshold.

Figure 3.29 shows a typical arrangement of power arches along a railway track.

To estimate the loss due to power arches the authors adapted a knife-edge diffraction model; the attenuation A_s is defined as the ratio of the received field with and without the obstacle.

$$A_s = \frac{E_1}{E_0} = \left[\frac{1+j}{2} \cdot \int_{v_0}^{\infty} e^{-j\frac{\pi}{2}v^2} dv\right] .G(\alpha) \tag{3.23}$$

Where

$$v_o = h\sqrt{\frac{2}{\lambda}\frac{(a+b)}{a.b}} \tag{3.24}$$

Table 3.6 Example parameters of a land mobile satellite channel in K_u band measured in Europe; μ = mean, σ = standard deviation, c = Ricean factor

Environment	LOS (%)	Shadowed (%)	Blocked (%)	μ (dB)	σ (dB)	c (dB)
Highway	90	7	3	−8	1.5	17
Rural	78	16	6	−7	2	17
Urban	60	10	30	−7	2	17

(Adapted from Scalise *et al*., 2008.)

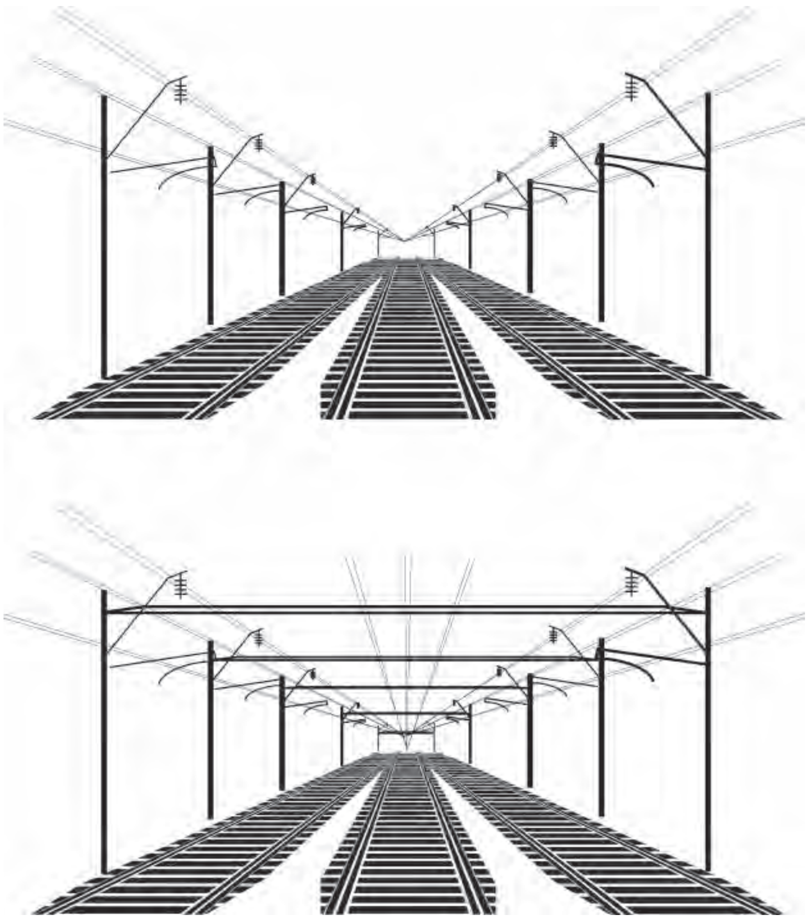


Figure 3.29 Typical arrangements of power arches on a railway track

λ = wavelength; a = train antenna to edge-of-arch distance, b = edge-of-arch to satellite distance, h = height of the edge above train antenna, measured from the point where the LOS path intersects the edge. $G(\alpha)$ = Train antenna gain towards the knife edge

Using values obtained from the main Italian railways company and a double-edged diffraction model to account for the two edges of the trellis (a = 2.5 m, distance between knife edges = 0.4 m), it was estimated that the diffraction loss became noticeable when h lies in the range $\pm$ 0.5 m with the worst case value of $\sim$ 18 dB in a V-shaped double dip spike. At 53 m distance between arches, the event would last for about 6 ms, every 600 ms at a train speed of 300 kmph. The estimates obtained from the model were found to be in good agreement with measured results at 1.5 GHz. Note that Equation 3.23 has to be applied for both the edges so that an additional component E_2 to account for the second edge must be included as follows,

$$A_s = \frac{E_1 + E_2}{E_0} \tag{3.25}$$

At the K_u band tropospheric effects, particularly rain fades, can be severe in additional to the impairments caused by mobility. The effects of rain fading on fixed links are well-documented, and therefore constitute the starting point in formulating the joint impact (Arapoglou, Liolis and Panagopoulos, 2012). It has been shown that stochastic properties of rain rate received on a fixed gauge and a moving rain gauge are similar, which implies that statistical properties of rain attenuation observed on a mobile terminal is similar to those on fixed terminal but conditioned by mobility. The factor for converting fixed rain statistics in to mobile rain statistics is given as (Matricciani, 1995)

$$\xi = \frac{v_{\mathrm{R}}}{|v_{\mathrm{M}} - v_{\mathrm{R}} \cos \varphi|} \tag{3.26}$$

where

v_{M} = speed of mobile in kmph,
v_{R} = speed of rain cell in kmph (advection or front speed), respectively,
φ = angle between mobile and rain cell velocity vectors

The complementary cumulative distribution function (CCDF) of the mobile rain attenuation random attenuation process $a_M(t)$ (dB) is specified as:

$$\mathrm{PM} = \Pr\{a_{\mathrm{M}}(t) \geq A_{\mathrm{th}}\} \tag{3.27}$$

where A_{th} is a chosen threshold.

Similarly, the CCDF of the fixed rain attenuation random process $a_F(t)$ (dB) is given as

$$P_{\mathrm{F}} = \Pr\{a_{\mathrm{F}}(t) \geq A_{\mathrm{th}}\} \tag{3.28}$$

Finally, P_M and P_F are related as,

$$P_{\mathrm{M}} = \xi . P_{\mathrm{F}} \tag{3.29}$$

where, $0 < \xi < \frac{1}{P_F}$ to ensure $P_M < 1$.

Using established techniques to estimate P_F, Equation 3.29 provides an approximate estimate of the CCFD; refinement and validation is in the research domain.

The probability distribution of the received signal on a fixed link is widely agreed to be log-normal and therefore the assumption is that the mobile link would also follow the same distribution, although this has to be validated by extensive measurements. With this assumption the probability distribution of the mobile rain attenuation process is given as (Arapoglou, Liolis and Panagopoulos, 2012),

$$P_a(a_M) = \frac{1}{\sqrt{2\pi}.\sigma_M.a_M} \exp\left[-\frac{(\ln a_M - \ln m_m)^2}{2\sigma_M^2}\right] \tag{3.30a}$$

The CCFD is given as

$$P_M = \frac{1}{2}\mathrm{erfc}\left[-\frac{\ln A_{th} - \ln m_m}{\sqrt{2}\sigma_M}\right] \tag{3.30b}$$

The log-normal statistical parameters $\{m_M, \sigma_M\}$ are related to the fixed log-normal distribution through Equation 3.29 and erfc(.) is the complementary error function.

In practice, propagation data cannot be collected for all conditions and are often lacking when developing new concepts including use of untried frequency bands. In such cases, simulation methods involving mathematical modelling offer an economic and viable alternative.

A signal is received at a mobile via a number of paths, as shown in Figure 3.17, so it should be possible to construct the received signal at a mobile by vector summation of all major components. A number of authors have attempted such an approach to LMSC modelling (e.g. Dooren *et al.*, 1993; Sforza, Dibernardo and Cioni, 1993; Matsudo *et al.*, 1993; Richharia *et al.*, 2005). As an illustration, the approach adapted by Dooren *et al.*, is elaborated.

The method, demonstrated for an urban environment, is based on the *Uniform Theory of Diffraction*, which includes effects of non-perfect conductivity of obstacles and their surface roughness. The physical environment in the vicinity of mobiles is defined by approximating buildings as block-shaped obstacles and cylindrical shapes as a combination of polygonal cylinders or elliptic cylinders. Data for such a detailed definition of a location can be obtained from digital databases. The field strength at a mobile receiver is calculated by summation of contributions from LOS waves, reflected waves, edge- or corner-diffracted waves, reflected-diffracted waves, diffracted-reflected waves and double-diffracted waves. Spatial filtering of the receiver antenna is included through either theoretical or measured antenna weight functions in amplitude, phase and polarization. A ray tracing procedure is applied at each observation point and each relevant parameter such as the complex received field strength, type of ray (LOS, etc.), direction of propagation and the absolute path length are stored, from which system parameters, such as fading statistics, Doppler shift and time-delay spread can be calculated by applying statistical methods. The wave relationship is defined as

$$E_o = E_i \mathbf{C} A(s) e^{-j\,k\,s} \tag{3.31}$$

where E_o and E_i are the outgoing and incident waves respectively; the coefficient $\mathbf{C}$ describes the physical interaction of a wave and an object and depends on the property of the material, direction of travel of incident/outgoing wave, wavelength and shape of the obstacle edges and surfaces; the factor A(s) describes divergence of outgoing wave, s is the distance between the observation point and the point of interaction and k is the wave number for free space.

The total received field at the output of the antenna is given as:

$$E^{obs} = \varepsilon^{los} G^{los} E^{los} e^{los}_{pol} + \sum_{l} \varepsilon_l^{refl} G_l^{refl} E_l^{refl} e_{pol}^{refl} + \text{other types of rays} + \sum_{m} \varepsilon_m^{dd} G_m^{dd} E_m^{dd} e_{pol}^{dd} \tag{3.32}$$

where ε accounts for blockage by obstacles, the superscripts *los*, *refl*, *dd* respectively denote line-of-sight, reflected, double-diffracted components of electric field E or antenna voltage gain pattern G or polarization e_{pol}, as applicable; G applies to the angle between the applicable ray and antenna boresight and e_{pol} accounts for antenna polarization discrimination.

A simulation package developed on a similar principle, applicable in the range of 1–60 GHz, has been developed by Sforza, Dibernardo and Cioni (1993). The user can emulate a full description of a communication channel for any environment through a user-friendly interface. The core of the package comprises the ray tracer and geometric theory of diffraction (GTD) solver with supporting functions performed by an *Urban Area Modeller*, an electromagnetic mesher and a post-processor unit, as illustrated in Figure 3.30.

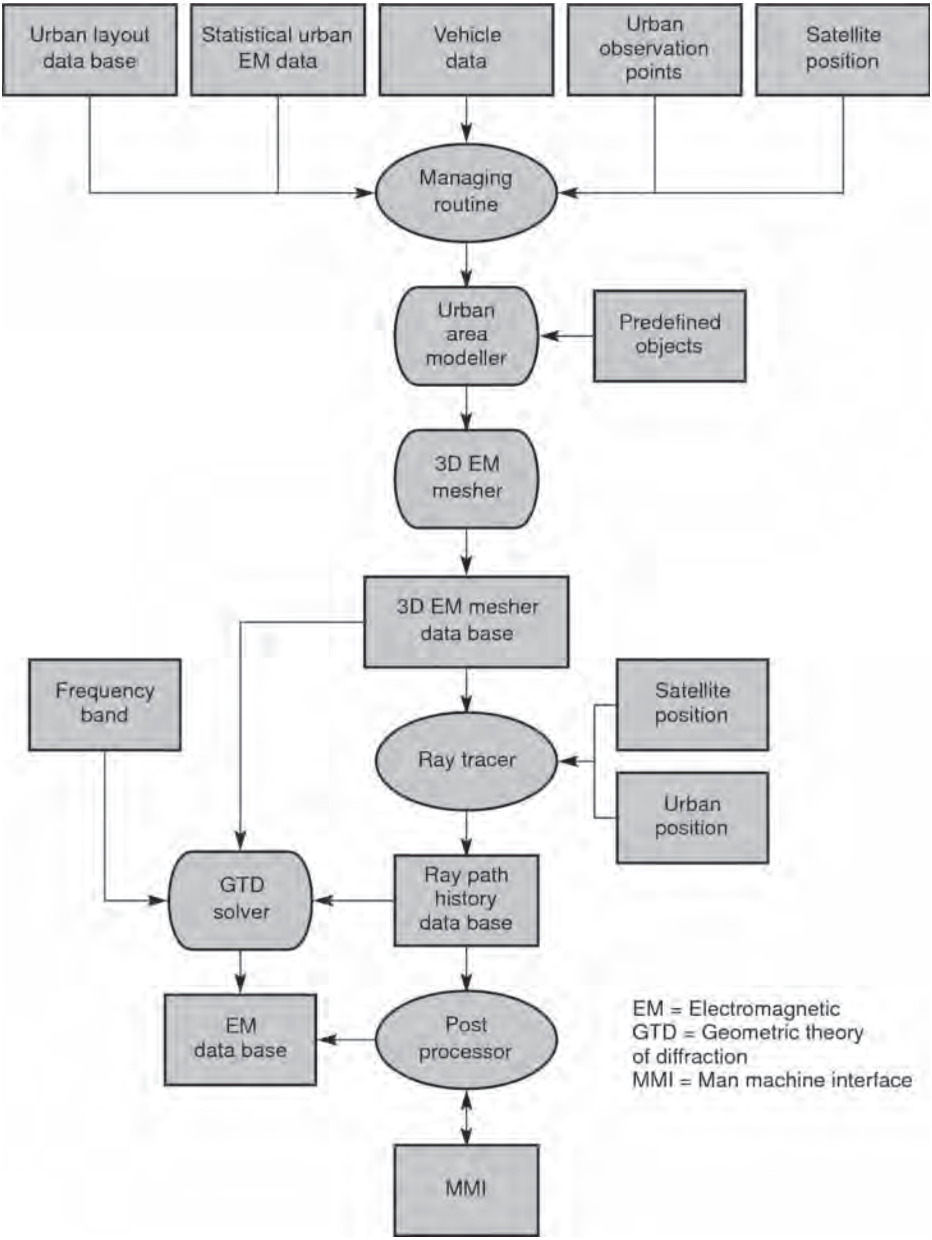


Figure 3.30 The main constituents of a simulation package for predicting propagation characteristics of an urban LMSS (Land Mobile Satellite Service) channel. (Source: Sforza *et al.*, 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

The inputs include satellite position, urban area description, vehicle speed and path with number of electromagnetic interactions for inclusion; up to second-order electromagnetic interactions can be included. In addition to the GSO, a capability to simulate LEO, MEO and HEO has been included. The GTD solver can include the effects of leaf absorption either through user-entered weight functions or empirical relations. The urban layout can be characterized by features such as buildings, tunnels, overpasses, street lamps, phone booths and parked cars, by using boxes, cylinders, spheres, and so on. Trees are characterized by spherical leafage and a cylindrical trunk. The layout of cities can be introduced as digital terrain data-base through input files. The electromagnetic mesher interfaces the urban layout modeller with the electromagnetic solver by verifying and reformatting input data, and introducing each object's electromagnetic property. The post-processor unit extracts a range of parameters such as the time series of the received field and antenna pattern weightings, and performs narrowband or wideband statistics. Narrowband analysis includes probability and cumulative distribution functions, average fade duration, LCRs, fade/connection distributions and time shares of fades/connections. Wideband analysis includes Doppler and power delay profiles, delay spread and coherence bandwidth. Figure 3.31(a) shows a typical ray tracer output and Figure 3.31(b) illustrates a time series of the test case. A simulation package offering such features is useful to a land mobile satellite (LMS) system designer for characterization a specific location, without the need to perform expensive and time-consuming field measurements.

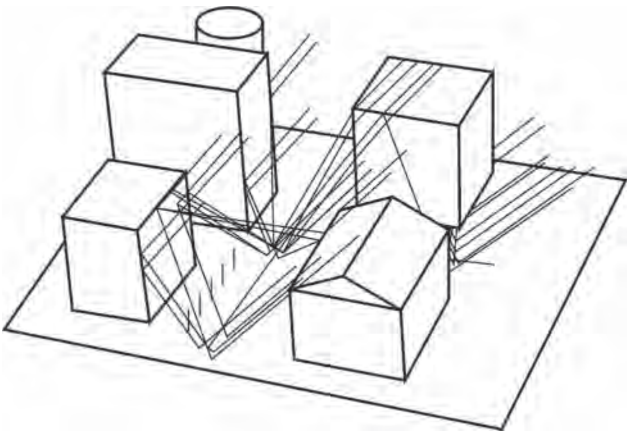
Vogel and Hong (1988) suggest a model based on simple physical and geometrical description of an environment developed to predict effects of elevation angle, distance of scatterer and bandwidth of a land mobile channel. Under simplifying assumptions of a single isotropic scatterer and a receiving antenna gain independent of azimuth, the model was used to duplicate measured data and predict the effects of elevation angle, scatter distance and bandwidth. Furthermore, the work demonstrated that modelling attenuation, caused by roadside trees as Fresnel diffraction, gave results of correct order of magnitude, despite violations of such an assumption.

3.3.2.3 Wideband

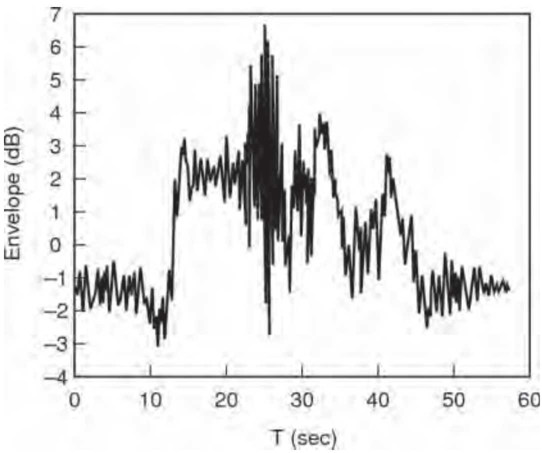
All components of narrowband transmissions undergo simultaneous variations level with time and position. On the contrary, broadband signals undergo environment-dependent frequency selective fading. The frequency dependence of fading is measured as correlation bandwidth, sometimes called coherence bandwidth, defined as the frequency separation, where correlation in frequency response is more than an arbitrary fraction K for which values of 0.5 or 0.9 have been used. Although frequency-selective fading can cause an increase in bit error rate (BER) due to inter-symbol interference, it allows frequency diversity to be designed into a system, as discussed later (see Section 4.2.5, as an example).

Figure 3.32(a) shows a two-dimensional plot illustrating various components useful in characterization of a wideband channel, while Figure 3.32(b) presents a snapshot of power delay profile of a satellite channel at an elevation of 25° received on a handheld receiver in a suburban European environment.

Typical measured values of delay spread, using an aircraft as the source, were observed to be less than 600 ns (coherence bandwidth = 0.265 MHz) on a mobile travelling on a European highway, and 100–500 ns (coherence bandwidth: 1.59–0.32 MHz) on hand-held



(a)

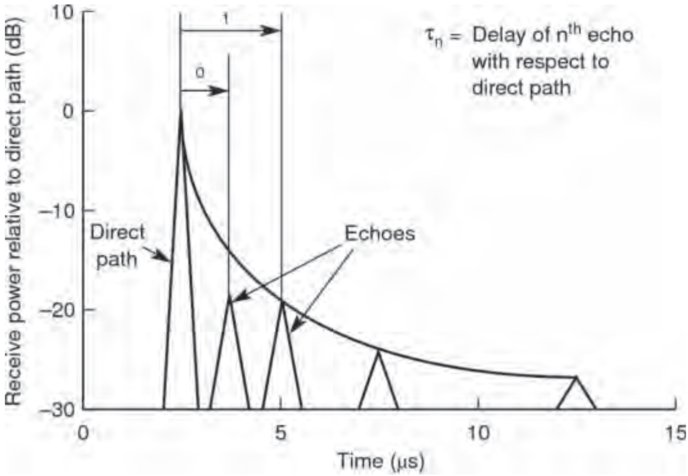


(b)

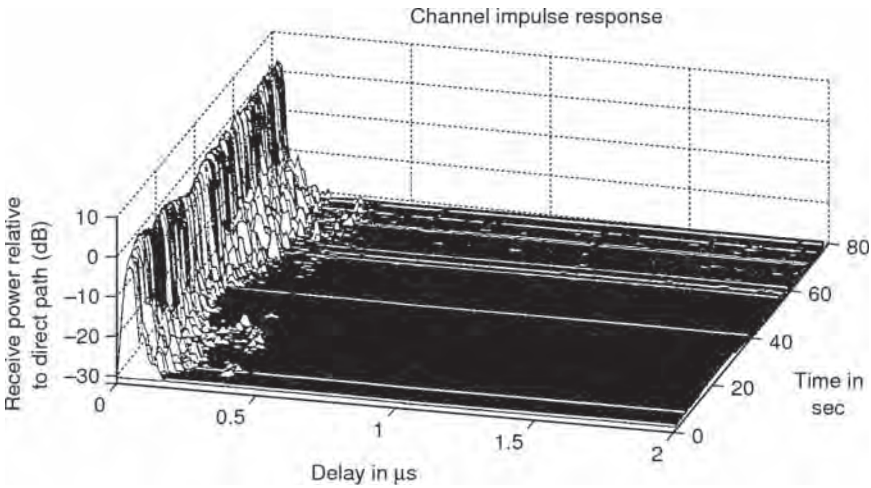
Figure 3.31 (a) Ray model of a test case (Sforza, Dibernardo and Cioni, 1993). (b) Time series of the signal strength for the test case (Sforza, Dibernardo and Cioni, 1993). (Both parts source: Sforza *et al*., 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

receivers, neglecting echoes of less than −25 dB (Lutz, 1998). Measured delay spreads in two Japanese cities, Shinjuku Tokyo and Sapporo, were reported to be respectively 0.04 s to 0.18 μs (coherence bandwidth = 3.97 to 0.88 MHz) and 0.05–0.19 μs (coherence bandwidth = 3.18 to 0.837 MHz) (Ikegami *et al*., 1993).

Considerable experimental work has been done in wideband characterization of terrestrial land mobile systems, which indicates that the channel is approximated as quasi-wide sense stationary uncorrelated scattering (QWSSUS). The received signal of such channels consists of echoes from independent point scatterers. As multipath mechanisms on the slant path are



(a)



(b)

Figure 3.32 (a) Components for characterizing a wide-band channel. (b) A representative power delay profile on a satellite channel at 25° elevation in a suburban environment as received on a handheld receiver. (Part b source: Jahn *et al.*, 1995-2. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

identical to those on horizontal terrestrial paths, the terrestrial model can also be applied to satellite channels. The complex-valued impulse response of QWSSUS channels can be represented as a tapped delay line with a complex valued impulse response

$$\sum_{k=0}^{k=\infty} a_k \delta(t - t_k) e^{j\theta_k} \tag{3.33}$$

The propagation medium can be characterized by a set of path amplitude (a_k), path arrival time (t_k) and phase (θ_{k}) the direct path being represented by amplitude a_0, delay t_0 and phase (θ_0). The components can be measured with a wide-band probe that uses the correlation properties of a pseudo-random sequence (see Section 4.2.6). Impulse response of the channel can be obtained from such a measured power delay profile as follows,

$$\mathrm{p(t,}\,\tau) = |\mathrm{h(t,}\,\tau)|^2 \tag{3.34}$$

Best fit equations to measured data provide environment-specific characterization for system studies.

For an exponential power delay profile the coherence bandwidth is given as (correlation ≥ 0.5),

$$\mathrm{B_c} = (1/2)(\pi)(\delta) \tag{3.35}$$

The net effect of the multipath dispersion is an irreducible lower limit on the BER of a digital transmission and a limit in minimum distortion in an frequency modulation (FM) system.

3.3.2.4 Elevation Angle Dependence

Radio waves are electromagnetic in nature, as light waves follow a similar (but not identical) elevation angle dependence to solar rays, that is depth of shadowing is governed by type and extent of obstruction(s), and hence, reduces as elevation angle is increased. Degradation of signal quality at low elevation angle is of particular significance to high latitude regions where GSO satellites appear at low elevation. This difficulty prompted investigations into other types of orbit that could provide higher elevation angle coverage at mid and high latitudes. For the European region, a HEO was considered to be a viable candidate by some researchers.

Characterisation of propagation channels operating with LEO and MEO mobile satellite systems is of considerable interest. Elevation angle changes continuously, even for a stationary observer, necessitating models that characterize elevation angle dependence. ITU-R recommendation 681 presents a prediction technique for elevation angle dependency of fade (ITU, 2009; see previous section).

Corazza and Vatalaro (1994) propose a statistical model based on a combination of Ricean and log-normal probability distributions applicable to LEO and MEO in various environments. Shadowing is assumed to affect both direct and diffused components and its applicability was extended to various types of environment and a wide range of elevation angles by altering the mean, standard deviation and Ricean factor extracted from the European Space Agency's (ESA's) L band database. Figure 3.33 demonstrates the close correlation of the estimated cumulative distribution functions with the measurements in the elevation angle range 20–80° in a rural tree-shadowed environment.

The applicability of the model in system evaluation is demonstrated in Figures 3.34 and 3.35, which present respectively BER for Differential binary phase shift keying (BPSK) at various elevation angles, and coverage percentage at various elevation angles for Iridium (66 satellites), Globalstar (48 satellites) and Odyssey (12 satellites) constellations (excluding the polar region).

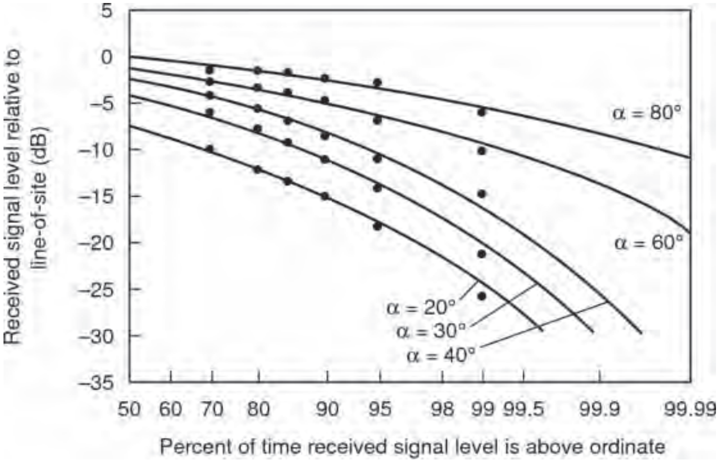


Figure 3.33 Comparison of measured (circles) and estimated fade (continuous line), as a function of elevation angle, α, in a European tree-shadowed environment. (Source: Corazza and Vatalaro, 1994. © 1994 IEEE. Reproduced with permission.)

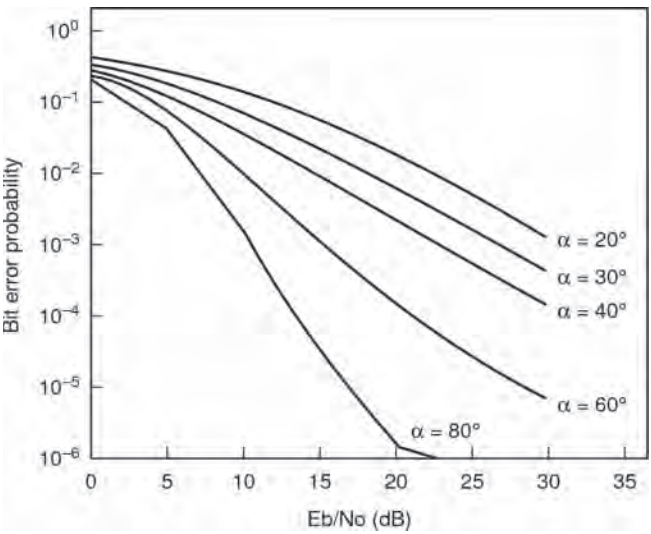


Figure 3.34 Bit error probability as a function of energy per bit by noise power density (E_b/N_o) for differential BPSK at various elevation angles (α). (Source: Corazza and Vatalaro, 1994. © 1994 IEEE. Reproduced with permission.)

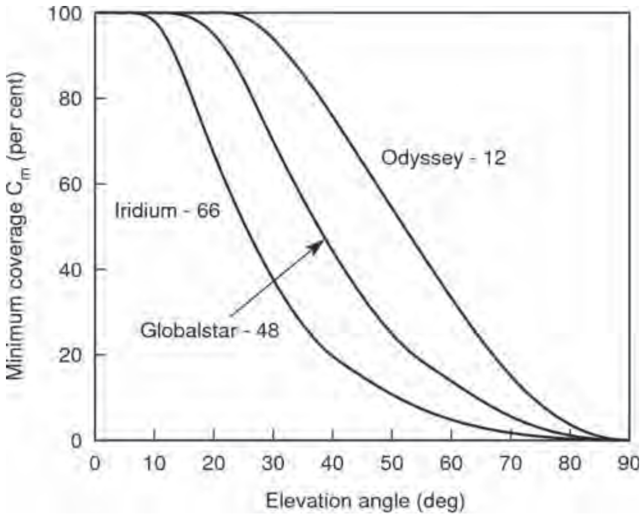


Figure 3.35 Percentage of worldwide coverage, excluding the polar region, as a function of elevation angle for three constellations – Globalstar, Iridium and Odyssey. (Source: Corazza and Vatalaro, 1994. © 1994 IEEE. Reproduced with permission.)

Sforza, Dibernardo and Cioni (1993) elaborated the ERS model (also known as ERM) (Goldhirsh and Vogel, 1992) using ESA's L and S band land mobile propagation database. The range of elevation angles extended from 20–80° and the percentage of optical shadowing ranged from 35–85% with roadside tree varieties being similar to the ERS model. The approach was to derive empirical expressions of fading probability by two different forms of curve-fitting as follows:

$$\mathrm{F}(\mathrm{Pr}, \theta) = -\mathrm{A}(\theta)\ \ln(\mathrm{Pr}) + \mathrm{B}(\theta) \tag{3.36}$$

$$\mathrm{F}(\mathrm{Pr}, \theta) = \alpha(\mathrm{Pr})\theta^2 + \beta(\mathrm{Pr})\theta + \gamma(\mathrm{Pr}) \tag{3.37}$$

where Pr = the percentage of distance where fade is exceeded, θ = the elevation angle and $\mathrm{A}(\theta)$ and $\mathrm{B}(\theta)$ are elevation angle-dependent parameters and coefficients α, β and γ depend on Pr.

Parametric curves of the modified empirical roadside shadowing (MERS) model are plotted with experimental data in Figure 3.36; the results show an rms error lying within 0.5 dB. The model was used for studying diversity advantage of a multi-satellite constellation with the conclusion that a MEO system can offer improved link reliability over a geostationary system.

Jahn and Lutz (1994) present results of a propagation measurement campaign conducted on behalf of ESA for narrowband and wideband characterization of LEO, MEO and HEO constellations for hand-held and car-mounted terminals in a variety of environments, and elevation angles in 0–80° range at 1.82 GHz, using a light aircraft. For the narrowband experiment, an unmodulated signal was transmitted, while wideband measurements used 30 MHz wide pseudo-random transmissions.

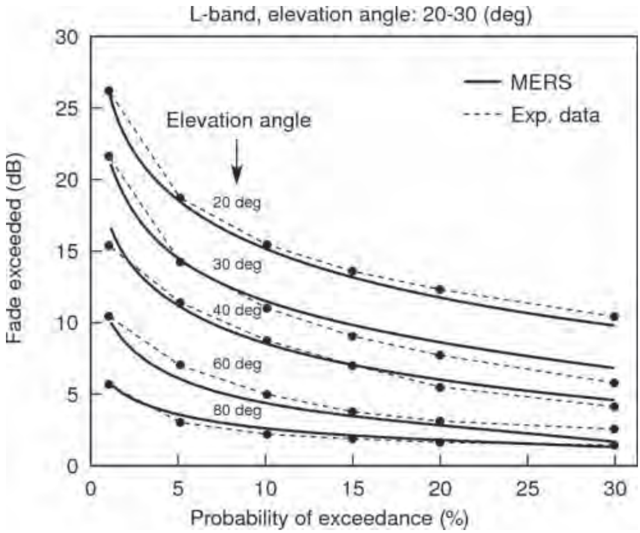


Figure 3.36 A comparison of the MERS model with experimental data. (Source: Sforza *et al*., 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

The conclusions of the work for the narrowband measurements can be summarized as follows:

- For elevation angles between 10 and 30°, shadowing is dominant, ranging up to 20–30 dB.
- Car roof-mounted antennas provide better channels than hand-held antennas while stationary and when driving.
- Hand-held antenna signal performance degrades because of shadowing from head and line-of-sight path, due to its lower height compared to car roof-mounted antenna.
- The cumulative fade distribution of hand-held channels in urban and suburban environments for 95 and 98% of time are compared in Figure 3.37, demonstrating a need for around 10 dB link margin at elevation angles above about 55° for the suburban region (95%), rising to over 15 dB below about 40°. This would make a system for interactive service unreliable with current technology without suitable countermeasures such as diversity.

The wide-band results can be summarized as follows:

- Echoes appear with delays smaller than 600 ns, equivalent to a distance of 200 m from the objects causing the echo, and are attenuated by 10–30 dB.
- Power of echoes with small delays decrease exponentially with delay.
- Figure 3.38 presents a delay spread for the hand-held terminal for urban, suburban and highway versus elevation angle. Delay spread decreases as elevation angle is increased. Generally, the delay spread lies in the range 0.5–2 μs.

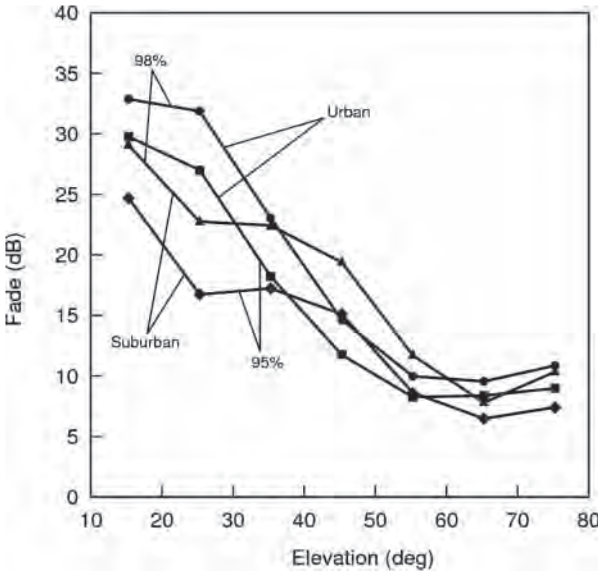


Figure 3.37 Fade depths for 95 and 98% time as a function of elevation angle in urban and suburban environments for a hand held terminal using right hand polarized drooping dipole antenna. (Source: Jahn *et al*., 1995. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

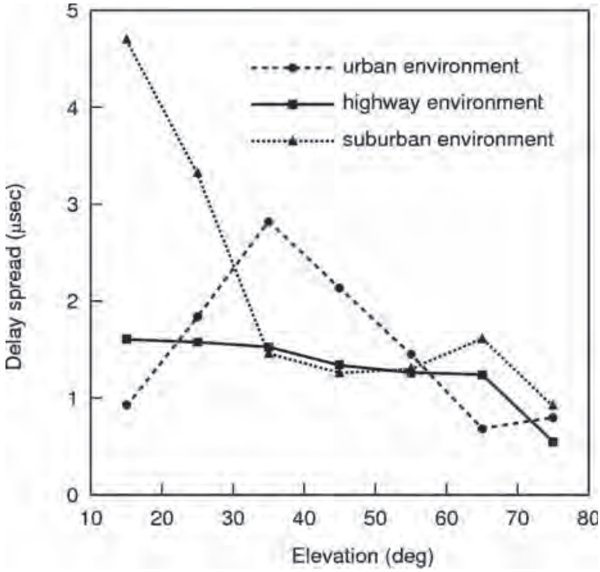


Figure 3.38 Delay spread as a function of elevation angle on a hand-held terminal in three environments (Source: Jahn *et al*., 1995. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

3.3.2.5 K_a Band Propagation

Due to a need for large amounts of spectrum for mobile communications there has been significant interest in utilizing the K_a band in the service link as the band offers large bandwidth and additionally offers the advantage of a lower sized directive antenna. Radio signals in this band undergo considerable degradation in the troposphere and are affected more adversely by the local environment than in the L band, because diffraction advantage is less and penetration loss is more severe. This would, for example result in notable shadowing loss from small branches and leaves that pose only minor problems in the L band.

At present, the K_a band propagation database for MSS is limited because the interest in this band for MSS is recent. Empirical models have been developed by scaling data from L band to K_a band and measurement campaigns have been conducted for similar purposes. Table 3.7 (Butt, Evans and Richharia, 1992) lists examples of multiband (L, S and K_u bands) fade levels measured in southern England for 60, 70 and 80° elevation in three types of environment over a simulated satellite path illustrating trends in attenuation with an increase in frequency and variations in the environment and elevation angle. At an elevation of 60° and 5% probability of occurrence the fade levels are 8, 9 and 19.5 dB for the L band (1297.8 MHz), the S band (2450 MHz in summer and 2320 MHz in spring) and the Ku band (10 368 GHz) respectively. The propagation conditions are relatively benign in open areas and high elevation angles where the K_u band attenuation is 2.8 and 1.7 dB at 70 and 80° respectively.

Consider a representative K_a band measurement campaign in Europe, using 18.7 GHz propagation transmissions of Italsat F1 (Murr, Arbesser-Rastburg and Buonomo, 1995). Measurements were conducted in a number of European environments, open, rural, tree-shadowed, suburban, urban and mixed constituting four types of environment – and in the elevation angle range 30–35°, maintaining an orientation of 0, 45 and 90° with respect to the satellite. The test gear comprised a vehicle-mounted high gain 2.4° beamwidth antenna

Table 3.7 A sample of fade level as a function of frequency in various environments of southern England during summer 1991 (polarization: circular; frequencies: L band = 1297.8 MHz, S band = 2450 MHz in summer and 2320 MHz in spring, K_u band = 10.368 GHz; antenna beamwidth: L and S bands = 64° and K_u band = 80°)

Frequency	Probability (%)	Suburban (dB)		Wooded (dB)		Open (dB)	
		60°	80°	60°	80°	70°	80°
L band	1	16.5	12.0	18.5	8	2.8	2.0
	5	8.0	6.0	11.0	4.5	2.2	1.2
	10	5.0	2.5	7.5	3.0	1.8	0.9
S band	1	18.5	16.0	22.5	–	3.5	2.0
	5	9.0	8.5	13.0	–	2.6	1.6
	10	6.0	6.0	8.8	–	2.0	1.4
K_u band	1	27.5	26	28	24	4.0	2.5
	5	19.5	18.5	22.5	15.5	2.8	1.7
	10	13.0	13.0	18.5	10.5	2.3	1.4

(Adapted from Butt *et al*., 1992.)

system employing a dual-gyro stabilization system, Doppler correction and a fast phase lock loop (PLL) enabled receiver that provided an overall measurement accuracy of better than 1 dB. Environments chosen included the Netherlands (open, suburban), France (tree-shadowed, open and mixed), Germany (urban, suburban and tree-shadowed) and Austria (mixed, suburban and tree-shadowed). To minimize ambiguity from tropospheric effects, measurements were conducted under clear sky conditions or under similar weather conditions. Statistical analyses included cumulative distribution function, probability distribution function, LCR, cumulative connection time and timeshare of fade. As expected, statistics showed dependence on environment, orientation of mobile with respect to satellite, type of tree leaves, height and diameter of trees. Table 3.8 provides a summary of statistics for 80,

Table 3.8 A summary of K_a band statistics in a European environment

Environment	Location	Cumulative probability (%)	Power level with respect to line of sight		
			Azimuth 0°	Azimuth 45°	Azimuth 90°
Suburban		95%			
	Netherlands		−3.3	−10	−24.7
	Germany		−1	–	−24.7
	Austria		–	−24.7	–
		90%			
	Netherlands		−1.7	−5.7	−21
	Germany		−0.6	–	−19.3
	Austria		–	−15	–
		80%			
	Netherlands		−1	−2	−14.6
	Germany		−0.3	–	−7
	Austria		–	−0.5	–
Tree shadowed		95%			
	France		−12	−20	−20
	Germany		−1.7	–	−32.3
		90%			
	France		−10	−17	−17
	Germany		−0.6	−34	−28.7
		80%			
	France		−7.3	−13.9	−13.3
	Germany		−0.3	−29	−24.7
Urban		95%			
	Germany		−1.3	–	–
		90%			
	Germany		−1	−34.8	−33.3
		80%			
	Germany		−0.3	−32	−29.3

(Source: Murr *et al*., 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, co-sponsored by Communication Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

90 and 95% probability levels. Notice the strong influence of the azimuth in all the results. At 95% probability and 0° azimuth, the worst and best fades for suburban areas are 3.3 and 1 dB respectively; the corresponding levels at the same locations worsen to 24.7 dB at 90° offset. A similar trend is observed across all the locations. The worst fade levels of 1.3 dB in urban areas appear uncharacteristically low at 0° azimuth indicating a direct line-of-sight path; however, the fade worsens to 33.3 dB at 90° offset.

Results of a K_a band experiment conducted over National Association of Space Administration's (NASA's) Advanced Communication Technology Satellite (ACTS) demonstrate that for both the K band (19.914 GHz) and K_a band (29.634 GHz) fading and multipath characteristics are similar (Pinck and Rice, 1995). Data were gathered for lightly shadowed suburban (infrequent partial blockage), moderately shadowed (occasional full blockage) suburban and heavily shadowed suburban (frequent full blockage).

A representative set of results of one run taken in Pasadena, California is illustrated in Table 3.9. The similarity in results at the two frequencies is remarkable. The K band measurements were taken in the forward link gathered at the mobile and the K_a band measurements were taken in the return direction, that is, collected at the fixed site.

3.3.2.6 Multiple Input Multiple Output Channels

Multiple-input multiple-output (MIMO) technology provides attractive capacity gain compared to single-input single-output (SISO) systems in terrestrial systems without extra spectrum and power overheads! Therefore there is considerable interest in application of the technology to satellite systems; but since its implementation to satellite systems poses various challenges due to differences in transmission techniques, its applicability to satellite systems is being researched (e.g. Liolis *et al.*, 2009; King, 2007). In this scheme signals are transmitted/received via more than one antenna – for example two transmit antennas and two receive antennas in a 2×2 MIMO arrangement (see Chapter 14). The parameters of interest for a MIMO channel include – spatial and temporal correlation between MIMO channels, LOS shadowing effect, XPD, dependence on operating frequency, elevation angle and operational scenario. We outline a small sample of the reported work.

Table 3.9 A measurements sample at K band (19.914 MHz) and K_a band (29.634 MHz) in various categories of suburban settings of Pasadena, California using beacons from ACTS

Category	1% fade level (dB)		3% Fade level (dB)		5% Fade level (dB)	
	K band	K_a band	K band	K_a band	K band	K_a band
Lightly shadowed suburban	8	9	1	1	1	1
Moderately shadowed suburban	27	26	17.5	17.5	12.5	12.5
Heavily shadowed suburban	$\gg 30$	$\gg 30$	$\gg 30$	$\gg 30$	> 30	> 30

(Source: Pinck and Rice, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, co-sponsored by Communication Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

Liolis *et al*. (2009) address statistical modelling of MIMO channels for satellite digital multimedia broadcasting (SDMB) applications using dual polarisation diversity where the same information is transmitted on two separate orthogonally polarized antennas. Due to scarcity of measured MIMO propagation data, relevant SISO data of LMS system and terrestrial MIMO systems are extrapolated to give a statistical model of an L/S band correlated dual circularly-polarized 2×2 statistical LMS MIMO channel with a provision to refine the model on basis of measurements. The model is utilized to assess LMS channel capacity statistics, and in particular, to characterize the performance of the DVB-SH broadcast standard (see Chapter 12). The large scale component representing shadowing is assumed to be log-normal with parameters (α,φ) and small scale fading component is Rayleigh distributed with parameter (*MP*) of Loo's model (Loo, 1985). The 2×2 MIMO channel is modelled by the matrix,

$$\mathbf{H} = \overline{\mathbf{H}} + \widetilde{\mathbf{H}} \tag{3.38}$$

$$\mathbf{H} = \begin{bmatrix} \overline{h}_{11} & \overline{h}_{12} \\ \overline{h}_{21} & \overline{h}_{22} \end{bmatrix} + \begin{bmatrix} \widetilde{h}_{11} & \widetilde{h}_{12} \\ \widetilde{h}_{21} & \widetilde{h}_{22} \end{bmatrix} \tag{3.39}$$

where

$\overline{h}_{ij}$, $(i,j = 1,2)$ represents the amplitude of the large scale component,
$\tilde{h}_{ij}$ $(i,j = 1,2)$ represents the small scale component.

The overall amplitude of fading, **H**, follows the Loo distribution. The parameter triplet (α, φ, *MP*) is estimated from tabulated experimental dataset as detailed in the literature (Burzigotti *et al*., 2008).

A ray-tracing method outlined in the preceding section was used to develop a time series of signals received in an MIMO receiver (King, 2007). A simplified version of the ray-tracing algorithm was used to quicken the simulation. The model included impairments typical of a land mobile channel, that is; blockage, shadowing, rooftop diffraction and multipath. The accuracy of such a model depends on closeness of the scattering objects with the chosen environment. Urban and highway environments in Munich, Germany were selected, since published measurement data was available for validation. A time series of the received signal is obtained by moving the MIMO receiver along the route as rays are traced from the MIMO source to the receiver at each point based on the geometry of the path and scattering objects around the mobile. A reflection coefficient is assigned randomly between 0 and 1 with a phase between 0 and 2π for each scatterer in a cluster. The magnitude and phase of the reflection coefficient are both uniformly distributed and their values are chosen to be identical for each scatterer of a cluster. Simulated and measured probability levels at each fade depth were observed to be in agreement.

3.3.2.7 Propagation Measurement Techniques

Narrowband Measurement

The main components of any mobile propagation measurement experiment are a transmission source, a mobile receiver, a representative environment and data analysis. A satellite source is preferred when experimental conditions are satisfied. A number of experiments have been conducted using L band transmission from various geostationary satellites around the world. However, when measurements are required for typical conditions, such as characterization of elevation dependence, geostationary satellites are not the best choice, as elevation angle variations are very gradual over geographical span available for a measurement campaign. Platforms such as aircraft and helicopter are used as alternative sources as they provide a flexible control of environment and look angles. Similarly, there are not many wideband sources with the required signal characteristics for wideband measurements in popular MSS bands. However, GPS satellites do provide wideband transmission and hence have been considered as a potential source at the L band.

With the launch of LEO satellite systems, several L/S band satellite sources have become available for an accurate characterization of narrow-band NGEO channels.

Propagation receivers use standard coherent detection with a wide dynamic range similar to those used for propagation measurements of FSSs. The hardware is 'ruggedized' to accommodate mechanical vibrations during vehicular motion. Antennas are chosen to be as close to the real system as is possible, due to the sensitivity of fading characteristics to antenna patterns. Data are gathered in a digital recorder and analysed by statistical data processing software to give mean, standard deviation, cumulative distribution of fade, best-fit probability distribution or elevation dependence, and so on. Care has to be exercised in setting a stable reference signal due to signal drift with time, and further to recognize the reference during data processing. One technique is to obtain a stable reference signal at the start of each measurement run.

Wideband Measurement

Wideband measurement methods for satellite channels have been derived from terrestrial measurement systems (Cox and Leek, 1972). One of the most reliable transmission methods is the use of a pseudorandom (PN) sequence transmissions and correlation receiver. Each multipath component, including the direct path, appears as a correlation peak at the receiver output; the shape of the correlation profile depends on the environmental features. Taking the direct path as a reference, a signal path delay profile may appear as shown in Figure 3.32. The spatial resolution of the multipath depends on the chip rate of the pseudo-random sequence as (chip rate/velocity of light); for example a 30 Mbps chip rate gives a spatial resolution of 10 m.

3.3.3 Maritime Channel

Satellite services for maritime applications were the first to be introduced due to a pressing need for reliable communication on the high seas, aided by a relatively lower technology demand as maritime propagation conditions are quite benign and large terminals can be used on ships. The consolidated conclusions from investigations on maritime channels are that signal fades depend on elevation angle, sea conditions including wave height, slope and wind, receiver antenna characteristics (beamwidth, side lobe and axial ratio), height of antenna above the sea and the structure of the ship. Other influencing factors include antenna pointing accuracy and polarization mismatch.

The signal at the receiver comprises a summation of direct, specular and diffused components as already discussed (Equation 3.12). The magnitude of the specular component reduces with an increase in wave height and as satellite elevation angle and the transmission frequency increase; diffuse components dominate under rough conditions; fade depth depends on amplitude and phase difference between the direct and indirect waves. The amplitude of the diffused component tends to follows a Rayleigh distribution with a uniform phase distribution. The received signal amplitude tends to exhibit a Nakagami–Rice distribution (ITU-R report 1007).

To compare fading statistics in a consistently, the condition of the sea is characterized by wave height quantified by 'a sea-state number' (ITU-R report 884-2; Hogben and Lumb, 1967). For example, sea-state numbers 0 and 1 represent calm seas with significant wave heights of < 0.15 m, sea-state number 4 represents a moderate sea with significant wave heights of 2–4 m and sea-state number 8 corresponds to an extremely high-sea condition with wave heights > 9.5 m. Statistics of the sea states are available for the entire world (Long, 1975); averages over all regions of the world are tabulated in Table 3.10.

The diffused component of the signal depends on the slope distribution of the sea-surface facets, the effects of which are RF dependent. Sea slopes range between 0.04 and 0.07 for rough sea conditions.

When specular reflection coefficient of the sea, rms height of the sea waves, RF, elevation angle and antenna gain pattern are known, it is possible to estimate the magnitude of specular components theoretically. Figure 3.39 (ITU-R 884-2) illustrates the calculated values of coherent (specular) components of a 1.5 GHz circularly polarized wave received through an omni-directional antenna at elevation angles of 5, 10 and 15° as a function of wave height. Note the reduction in magnitude of the specular component as elevation angle increases and wave height are increased.

Similarly, with the knowledge of average scattering cross-section per unit area of sea surface, scattering angles, receiver antenna pattern and height of antenna above sea level, the magnitude of diffused components can also be estimated. Figure 3.40 (ITU-R 884-2) illustrates theoretical values of carrier to multipath ratio versus surface rms slope for a circularly polarized wave at a frequency of 1.5 GHz received through a 12 dB gain antenna at elevation angles of 5, 10 and 15°, assuming worst-case sea surface. Observe a reduction in

Table 3.10 Average sea state for the world (Long, 1975)

% occurrence	Height of waves	Sea condition
22	0–0.9	Calm
23	0.9–1.2	Smooth
20.5	1.2–2.1	Slight
15.5	2.1–3.6	Moderate
9.5	3.6–6	Rough
9.0	> 6	Very rough-high

(Adapted from ITU Report 884-2.)

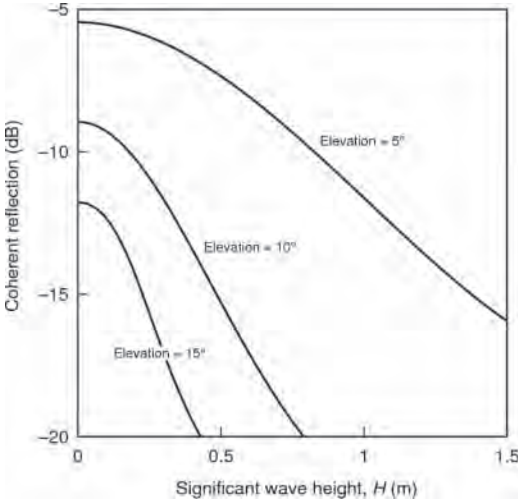


Figure 3.39 Magnitude of coherent reflection relative to direct signal as a function of significant wave height; frequency = 1.5 GHz, polarization: circular; antenna: omni-directional. (Source: CCIR (ITU-R) report 884-2. Reproduced with permission of ITU.)

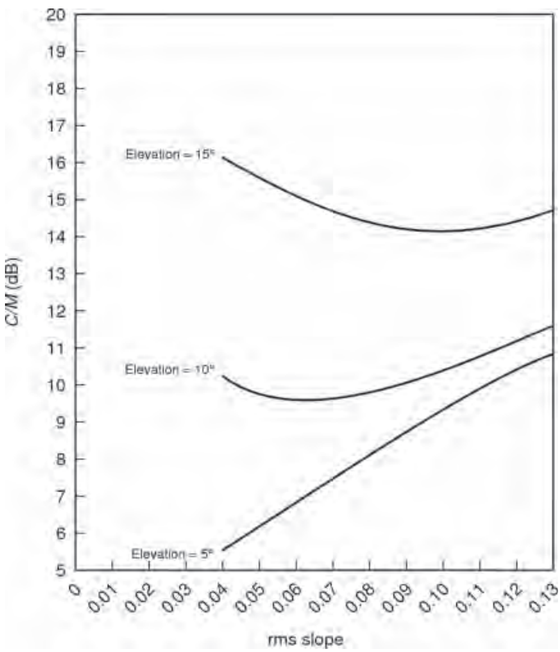


Figure 3.40 Worst case carrier to multipath ratio as a function of rms slope; frequency = 1.5 GHz, circular polarization, antenna gain = 12 dB. (Source: CCIR (ITU-R) report 884-2. Reproduced with permission of ITU.)

multipath noise as the elevation angle increased, and at 10° and 15° elevation as the rms slope is reduced below a threshold. The cumulative fade distribution can be determined for a given carrier to multipath ratio, or when statistics of coherent and diffused components are known (Beckmann and Spizzichino, 1963).

For practical purposes, the combined fading effects for a given frequency, antenna type and probability percentage may be approximated as

$$F_t = F_s(\max) + F_d(\%) \tag{3.40}$$

where $F_s(\max)$ is the maximum fade depth due to specular reflection and $F_d(\%)$ is the diffused component of fade at the desired probability percentage.

Figure 3.41 shows the calculated 99% fade depth for a 14 dB antenna at 1.5 GHz, as a function of wave height for both components. Fade caused by specular components Fs reduces with an increase in wave height until the magnitude becomes negligible beyond a wave height of 2 m, while fade caused by diffused components F_d increases linearly until saturation occurs (The dispersion represents the possible of slope variations). Figure 3.42 depicts fade depth as a function of elevation angle for antenna gain of 8, 14 and 24 dB for wave heights ranging from 1.6 to 3.2 m. The shaded portions indicate the range of slope variations at each wave height. Notice the reduction in fade depth at a given elevation with an increase in antenna gain due to the progressive rejection of the multipath by high gain antennas. A similar effect is seen for each antenna as the satellite elevation is increased – in this case multipath rejection increases with antenna elevation (i.e. as the antenna steers away from the sea surface).

Table 3.11 lists some measured results. A good general agreement is observed between theory and measurements where theoretical predictions were attempted.

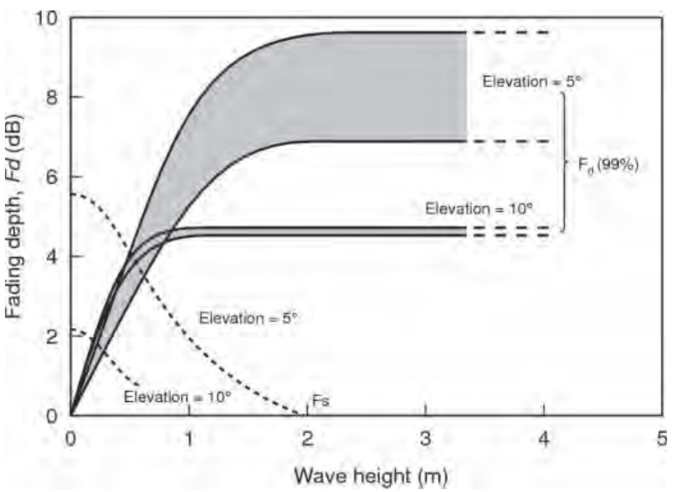


Figure 3.41 Calculated 99% fade depth for a 14 dB antenna at 1.5 GHz, as a function of wave height; F_s = specular component and F_d = diffused component. (Source: CCIR (ITU-R) report 763-3. Reproduced with permission of ITU.)

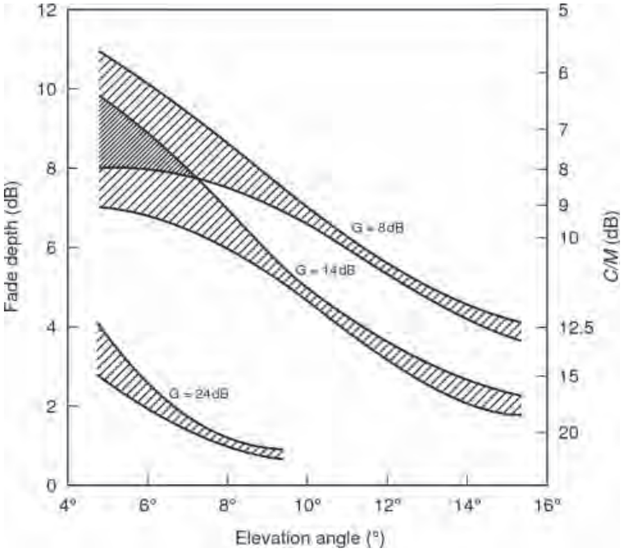


Figure 3.42 Fade depth (99% of time) as a function of elevation angle at various antenna gains (G), for 99% of time. (Source: CCIR (ITU-R) report 884-2. Reproduced with permission of ITU.)

Table 3.11 Measured fade depth at the L band from a number of campaigns, as a function of antenna gain and elevation angle from geostationary satellites

Antenna gain (dB)	Elevation (degree)	Wave height (m) or (sea condition)	Fade depth (dB) not exceeded (99%, except where mentioned)	References
0	6	0.5	6.9–12.8	Higuchi and Shinohara (1988)
3	4	–	12	ITU-R report 763-3, Hagenauer *et al*. (1984)
	19	–	10.5	
5	4	–	14	ITU-R report 763-3, Hagenauer *et al*. (1984)
	19	–	8	
11	4	–	11	ITU-R report 763-3, Hagenauer *et al*. (1984)
	19	–	6.5	
15	5	1–4	7–10	ITU-R report 763-3, Karasawa *et al*. (1986)
20	5	3	4.8	ITU-R Report 763-2
24	5	1.6–3.2	1.5–3	Ohmori *et al*. (1985)
	10	1.6–3.2	1	ITU-R Report 763
25	2	0–4	2 (95% time)	ITU-R Report 884-2; ESA (1977)
	8	(0–4)	1 (95% time)	

Data sources identified in the table.

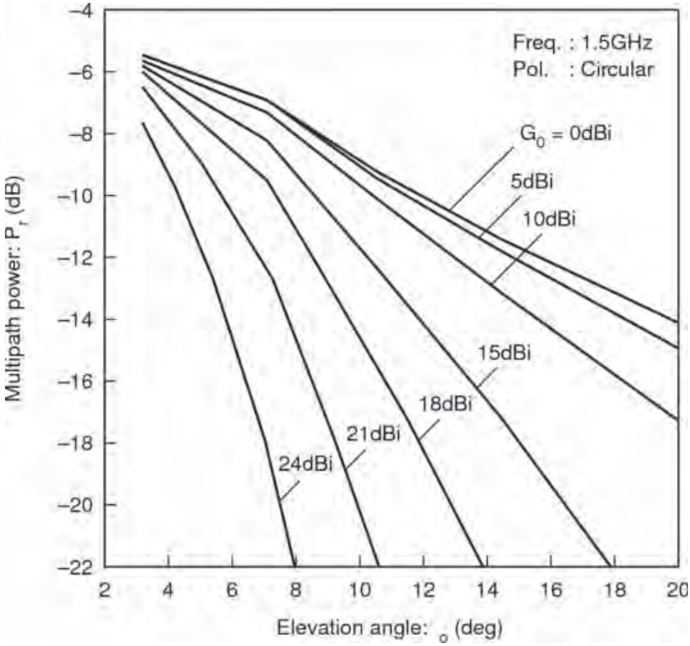


Figure 3.43 Multipath power as a function of elevation angle and antenna gain for circularly polarized wave at 1.5 GHz under slight to moderate sea conditions (wave height 1–3 m). (Source: CCIR (ITU-R) report 884-2. Reproduced with permission of ITU.)

Figure 3.43 shows multipath power as a function of elevation angle for 1–3 m wave height corresponding to slight to moderate sea conditions, applicable in the 1–2 GHz frequency band calculated by using an empirical model proposed by Karasawa and Shiokawa (1988).

Most of the propagation work has focussed on L-band systems, as this is the most used frequency band. Due to the introduction of the S and K_a bands, L band propagation models had to be extended to higher frequency bands. Computed results from one such model indicate that the maximum fade level is unaffected by RF; however, for a specific sea condition, fade level reduces with an increase in frequency in the wave height range 1–3 m. An experiment was conducted to characterize propagation in a maritime environment at K and K_a band using the ACTS (Advanced Communications Technologies and Services) satellite during 1996 (Perrins and Rice, 1997). Similar to the land mobile experiment mentioned earlier the forward link comprised a ground terminal located at the Jet Propulsion Laboratory in Pasadena and ACTS mobile terminal (AMT) (Abbe, Agan and Jedrey, 1996) mounted on a ship in eastern Pacific Ocean where the satellite was observed at an elevation of >40°. The AMT's transmit antenna gain in the 30 GHz band is >20 dBi with a 12° beamwidth in the elevation plane in the 30–60°; and the receive gain is >18.8 dBi with the same beamwidth as the transmit beam in the elevation plane with the receiver G/T > −6 dB/K. At this elevation the multipath was observed to be negligible.

Circular polarization is widely used in mobile satellite systems. When such a wave is reflected at an elevation angle greater than the Brewster angle the polarization sense is reversed. At other angles, the reflected waves become elliptically polarized. Theoretical studies indicate that circular polarization is superior to horizontal polarization in mitigating multipath noise pick-up. But at elevation angles up to $\sim 11°$, vertical polarization can suppress the multipath better.

The fading spectrum of multipath can be determined by theoretical methods. Studies indicate that spectral bandwidth increases when any one of the following parameters is increased – wave height, velocity of a ship, ship roll and pitch movement and elevation angle. Figure 3.44 depicts the theoretical value of a −10 dB power spectral bandwidth of amplitude fluctuations of 1.5 GHz, as a function of elevation angle under typical maritime conditions, that is, wave height 1–5 m, ship speed 0–20 knots and roll 0–30°.

Theoretical calculations based on a model proposed in the literature predict average values of fade duration and fade occurrence interval in elevation angle range of 5–10° as 0.05–0.4 and 5–40 s, respectively (Karasawa and Shiokawa, 1988). The probability distribution of these two parameters approximates an exponential distribution.

A further consideration is the effect of the ship's structure on propagation. Multipath fading occurs through reflections off ship structure, and shadowing occurs when any part of a ship's structure blocks the signal. Both types of fading are affected by a range of parameters – the shape of the ship, antenna location, antenna pattern, satellite azimuth/elevation (relative to the structure) and axial ratios of ship and satellite antennas. The magnitude of fade in is dealt with on a case-by-case basis using theoretical methods such as described or through empirical methods based on measurements. Simulation performed on a large ship model likely to give strong reflections showed a fading loss of the order of 1–2 dB. Loss due to blocking depends on the distance of the blocking structure and its dimensions. As an

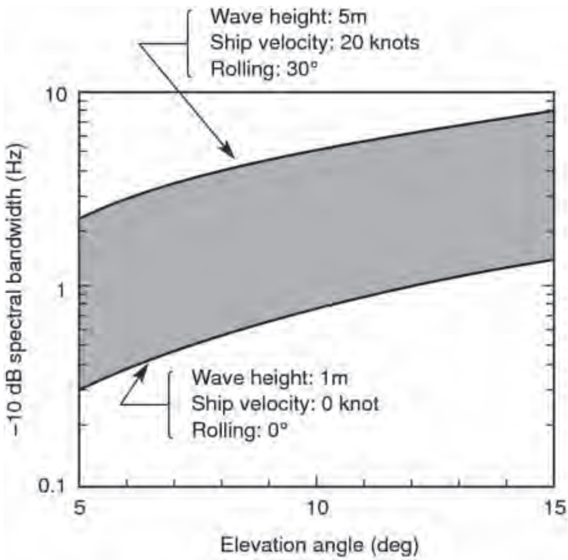


Figure 3.44 Elevation angle versus −10 dB spectral bandwidth of 1.5 GHz multipath fading. (Source: CCIR (ITU-R) report 884-2. Reproduced with permission of ITU.)

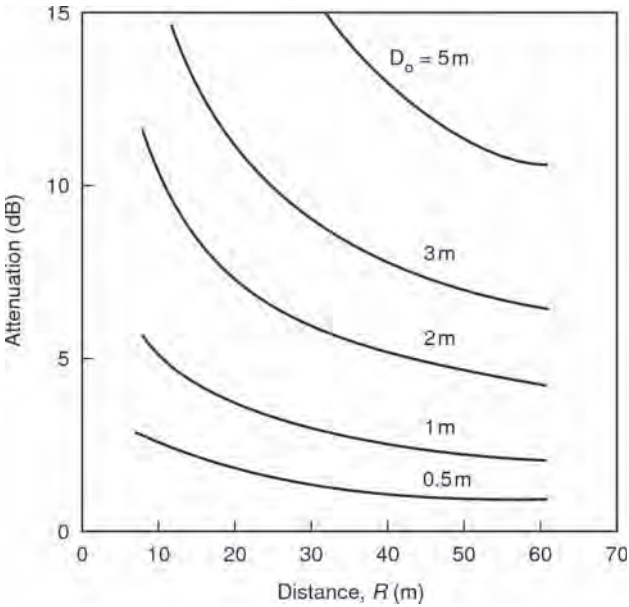


Figure 3.45 L band blocking attenuation due to a column-type structure of various diameters, D_o, from an antenna of 20 dB gain. (Source: CCIR (ITU-R) report 763-3. Reproduced with permission of ITU.)

example, Figure 3.45 shows attenuation due to a column structure of various diameters (D_o) spaced at various distances from an antenna of 20 dB gain (ITU-R report 763-2).

3.3.4 Aeronautical Channel

Propagation effects of aeronautical channels differ from maritime and land mobile propagation because of the high velocity of aeroplanes, their distance from ground and influence of the aircraft body on antenna performance. Aeroplane manoeuvres can affect signal under conditions when aircraft antenna is shadowed by the aircraft structure. When considering helicopters, the rotation of the rotor blades causes a cyclic interruption to the signal path. At present, most of the data exists for geostationary satellite systems, but data for NGSO satellite systems are scarce.

The troposphere constitutes the lowest part of the Earth's atmosphere spanning an altitude of about 8–16 km depending on latitude and season, with an average of about 11 km. About 75% of the mass of atmospheric gases and majority of water content lie within this layer.

Figure 3.46(a) illustrates the attenuation due to oxygen and water vapour at the Earth's surface and altitudes up to 3 km for water vapour and up to 6 km for oxygen at frequencies ranging from 1.5 to 40 GHz at a satellite elevation of 10°. The attenuation caused by water vapour reduces from ~ 1.18 dB at the surface to ~ 0.32 dB at an altitude of 3 km at 20 GHz; whereas it is insignificant at 1.5 GHz (see the arrow on the figure). The attenuation due to oxygen reduces from ~ 0.38 to ~ 0.2 dB at the surface to 0.2 dB and < 0.1 dB at an altitude of 6 km for 20 and 1.5 GHz respectively.

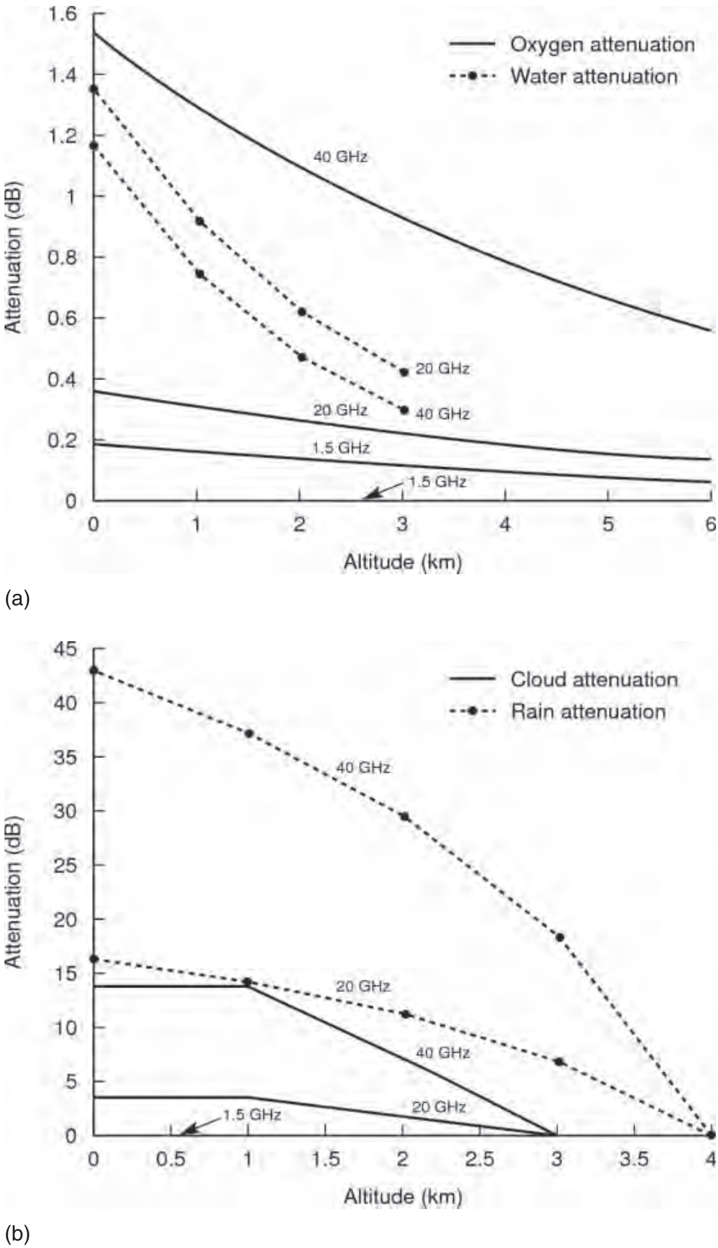


Figure 3.46 (a) Variation of attenuation due to oxygen and water vapour (7.5 g/m^3) as a function of altitude in 1.5–40 GHz range at an elevation angle of 10°.
(b) Variation in attenuation due to cloud (1 g/m^3) and rain (0.1% time, rain climate Zone K, as per ITU-R report 564) as a function of altitude in the frequency range 1.5–40 GHz, at an elevation angle of 10°. (Both parts source: Graphics AR.)

Figure 3.46(b) illustrates variation in cloud (1 g/m^3) and rain (0.1% time – rain climate zone K) attenuation, at 10° elevation (ITU R report 564). Cloud and rain attenuation are insignificant for practical purpose at 1.5 GHz. At 20 GHz, cloud attenuation reduces from 3.48 dB at the Earth's surface to an insignificant level above 3 km, while rain attenuation reduces from ~ 16.5 dB to an insignificant level at an altitude of 4 km. These results indicate that trans-oceanic flights (typical altitude ~ 10 km) are considerably less affected by rain at typical cruising altitude of 10–11 km.

Equation 3.12 also applies to aeronautical channels. The magnitude and characteristics of specular and diffused components depend on type and characteristics of the reflecting or scattering medium below the aircraft. It has been observed that magnitude of multipath caused by the sea surface is considerably more than from the land surface and hence considerable research has been directed towards characterization of aeronautical channels over sea. Measurements conducted to date demonstrate that:

- composite received signals follow a Ricean distribution;
- multipath components exhibit a Rayleigh distribution;
- the magnitude of multipath components are elevation angle-dependent;
- multipath Doppler spectrum possesses a Gaussian distribution;
- fading is frequency selective, due to path delay associated with multipath components;
- the multipath signal is dominated by diffused components – specular components are likely to be negligible for the majority of the time.

When the direct signal received at time t is $s_d(t)$, the reflected component can be represented as $s_r(t - \tau)$ where τ is the time delay of the reflected (or scattered) component whose magnitude depends on the geometry of the path, the reflection coefficient of the Earth's surface and the aircraft's antenna gain. The reflection coefficient depends on characteristics of the Earth below the aircraft – if the surface is sea, then the state of the sea and when land, whether the terrain comprises forest, built-up areas, desert, etc. Signal measurements on aircraft show that reflections and scattering from land is significantly lower than from the sea surface; furthermore, signals have an elevation angle dependence over sea but not over land; as already mentioned, most studies have therefore concentrated on propagation behaviour in flight paths over sea.

Theoretical multipath estimation techniques applicable to the maritime environment can also be applied to the aeronautical channels taking in to consideration the Earth's curvature effects into account. Carrier to multipath ratio is estimated using Fresnel's reflection coefficients for specular reflections and divergence factors (Beckmann and Spizzichino, 1963). Figure 3.47 (ITU-R report 1148) depicts estimated 99% fade levels received through an antenna of gain 7 dBi as a function of its altitude at 1.54 GHz for wave heights 1.5–3 m and elevation angle of 5°; the lower end represents the maritime environment and the upper end extending up to 10 km is an aeronautical environment. Measured carrier to multipath ratio for 0 and 3 dBi antenna for 1.5 GHz MARECS (Maritime European Communications Satellite) transmissions shows a good agreement with theory as demonstrated in Figure 3.48 (see solid circles). Some measured data are without multipath; this type of anomalous condition can arise when multipath signals are shielded by an aircraft's body.

Measured data illustrate that the power spectral density of multipath depends on aircraft speed, elevation angle and ascent/descent angle of the flight. Measurements on a jet aircraft

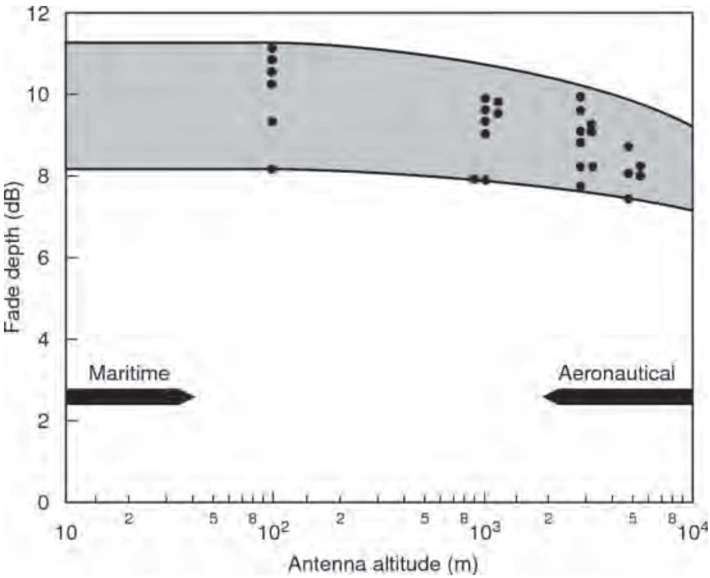


Figure 3.47 99% fade depth at various altitude for 1.5 GHz circularly polarized waves received by a 7 dBi antenna. (Source: CCIR (ITU-R) report 1148. Reproduced with permission of ITU.)

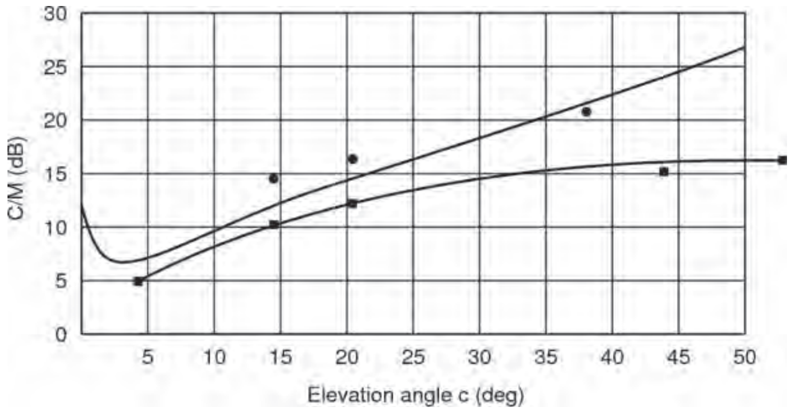


Figure 3.48 Measured and calculated carrier to multipath ratio for 0 dBi antenna (•) and 3 dBi antenna (■). (Source: CCIR (ITU-R) report 1169. Reproduced with permission of ITU.)

travelling at a ground speed of 650 km/h broadside to the satellite and receiving signals at a frequency of 1.6 GHz resulted in a multipath spectrum bandwidth in the range 40–80 Hz, at an elevation angle of 9° and 200 Hz at 31° (Sutton *et al.*, 1973). For level flight, −10 dB spectral bandwidth range was between 25 and 50 Hz at 5°–10° elevation and for ascending/descending flight trajectory the spectrum became skewed [ITU-R report 1148].

The aeronautical channel can be represented as wide-sense stationary uncorrelated scattering (WSSUS) channel (see Section 3.3.2.3 – Wideband) due to randomly changing nature of scatterers over the Earth's surface (Bello, 1973). The model can be used to estimate a number of characteristics of the channel. The Doppler power spectrum of the reflected signal of such a channel is given as,

$$P(f) = \left(\sqrt{2}/B_r\sqrt{\pi}\right) \exp\left(-2f^2/B_r^{\,2}\right) \tag{3.41}$$

The rms Doppler spectrum is defined as twice the standard deviation of the Doppler power spectrum

$$B_r = 4\alpha v \ \sin\ \varepsilon/\lambda$$

Where α = the rms surface slope, v = the aircraft speed at constant elevation angle, ε = elevation angle and λ = wavelength of the carrier.

The Gaussian shape, as well as values predicted by Equation 3.41, agrees well with measurements (Sutton *et al.*, 1973). Figure 3.49 shows the estimated 1/e correlation bandwidth as a function of altitude up to about 10 km, elevation angles of 5 and 10°, frequency of 1.54 GHz and a 10 dBi antenna (ITU-R report 1148). The correlation bandwidth is defined in this case as the frequency separation where the correlation coefficient has reduced to 1/e.

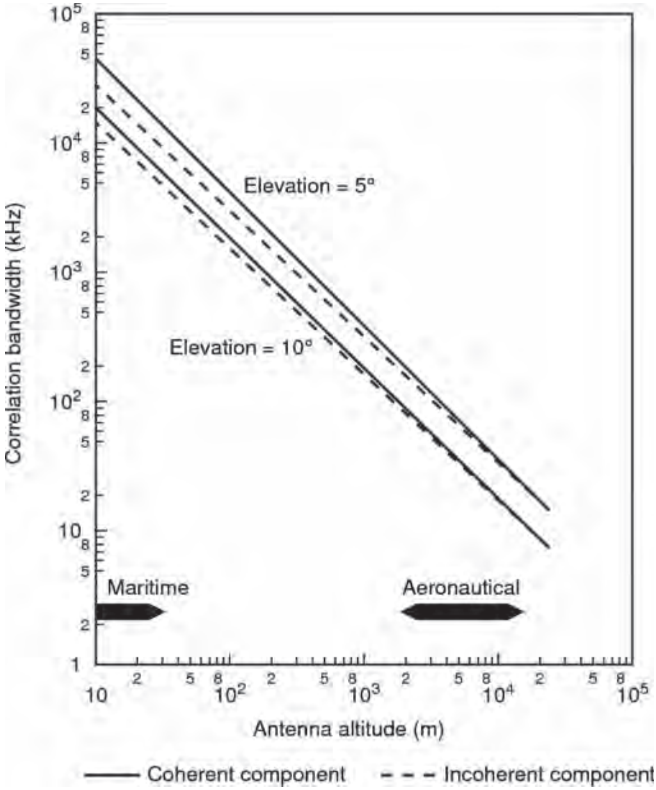


Figure 3.49 Estimated 1/e correlation bandwidth as a function of antenna altitude for an antenna gain of 10 dBi. (Source: CCIR (ITU-R) report 1148. Reproduced with permission of ITU.)

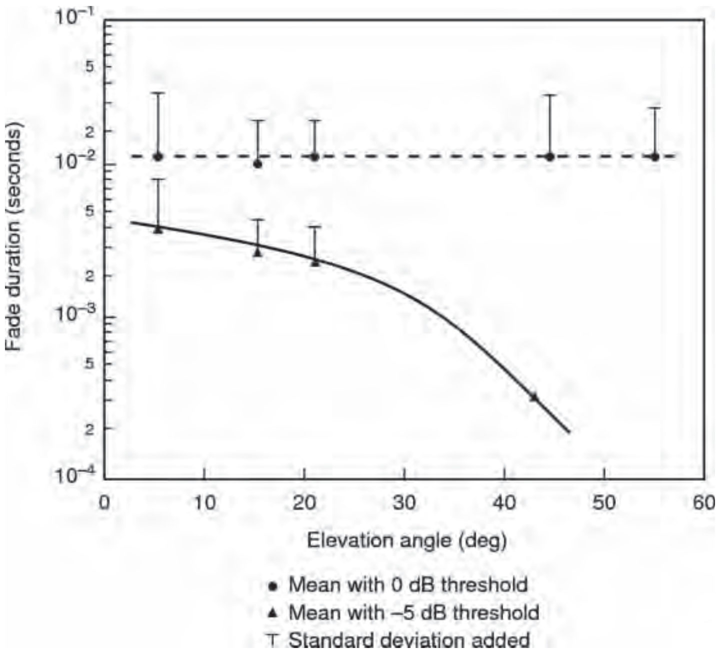


Figure 3.50 Fade duration mean and standard deviation statistics at 1.6 GHz at aircraft altitude of 10 km; antenna gain = 3.5 dB, circular polarisation, data collected over Atlantic Ocean and Western Europe. (Source: Figure 4, CCIR (ITU-R) report 1169, Hagenauer *et al.*, 1987. Reproduced with permission of ITU.)

Note a reduction in correlation bandwidth from several tens of MHz near the sea surface to 20–30 kHz at an altitude of 10 km.

Figure 3.50 shows the measured mean and standard deviation of 1.6 GHz fade duration received through a crossed dipole antenna of 3.5 dB gain over the Atlantic Ocean and in parts of western Europe. The aircraft altitude and ground speed were respectively 10 km and 700 km/s (Hagenauer *et al.*, 1987). The results of the campaign show that the mean fade duration at 5 dB fade level is 4 ms at 5° elevation, decreasing to 0.3 ms at 44°; the mean duration of connection level where the signal is above −5 dB threshold is 30 ms at 5°, increasing to 300 ms at 21°.

Table 3.12 lists a sample of results of a campaign using 1.6 GHz transmissions of the ATS-6 (Applications Test Satellite) satellite. The aircraft at an altitude of 9.1 km travelled at a ground speed of 740 km/h and used a two-element antenna array of 1 dB beamwidth of 20° in azimuth and 50° in elevation (ITU-R report 1148, Schroeder *et al.*, 1976).

3.3.5 System Implications

Signal shadowing and multipath influence system design in a number of ways:

- signal blockage affects link reliability;
- multipath noise causes errors in digital transmission;

Table 3.12 A sample of measured data from ATS-6 transmissions at 1.6 GHz. Land measurements did not show significant elevation angle dependency

Parameter	Measured range (bracketed values for land measurements)	Typical values at given elevation angles		
		8°	15°	30°
Two-sided 3 dB delay spread (μs)	0.25–1.8 (0.1–1.2)	0.6	0.8	0.8
3 dB one-sided correlation bandwidth (kHz)	70–380 (150–3000)	160	200	200
3 dB Doppler spread (Hz)	4–190 (20–140)	5	70	140
3 dB decorrelation time (ms)	1.3–10 (1–10)	7.5	3.2	2.2

(Adapted from Schroeder *et al.*, 1976.)

- frequency-selective fading causes inter-symbol interference when signal bandwidth exceeds the coherence bandwidth;
- Doppler frequency jitter appears as noise causing detection errors; and Doppler frequency shift has to be compensated.

The minimum elevation angle and the number of satellites simultaneously visible from user terminals set a bound on the radio link reliability of the system. The minimum elevation angle is directly related to the available link margin offered by a system; the number of satellites simultaneously visible from a location and their relative separation provide an estimate of the diversity gain.

While constellations are designed to operate at as high an elevation angle as is economically feasible, in order to minimize propagation impairments, it is impossible to eliminate fading, therefore fade countermeasures are invariably applied. They include feedback power level control, static or adaptive channel coding, robust modulation methods and satellite diversity. In a feedback power control scheme, the transmitter power is adjusted to compensate for fade at a receiver through a feedback loop from the receiver to the transmitter. Channel coding is used in all MSS systems and a wide variety of codes is in use. In an adaptive channel coding scheme the code rate is varied according to the state of channel, that is, the code rate is reduced as the magnitude of fade increases, such that BER is maintained at the expense of signal throughput. An adaptive coding scheme was demonstrated in an ACTS experiment, whereby 10 dB of dynamic fade compensation could be introduced to a link. Chapter 4 discusses details of modulation and coding schemes. Such a scheme is used in Inmarsat's BGAN system and DVB S2/RCS + M standards. Figure 3.51 depicts the main features of power control and adaptive code rate control systems.

Note that for portable MSS terminals designed for use in a fixed location, system design considerations are similar to those of VSATs. A trade-off is applied between signal quality, system capacity and terminal and call cost within the confines of the available technology. MSS system capacity is mainly governed by the available bandwidth and forward spacecraft EIRP (i.e. towards mobile). Figure 3.52 demonstrates the sensitivity of space segment

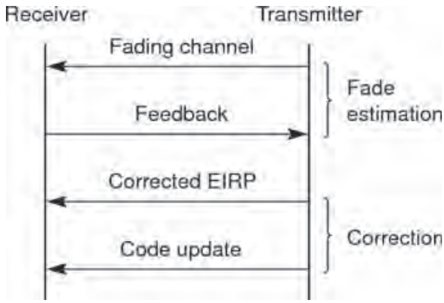


Figure 3.51 Power level fade control and dynamic code control systems

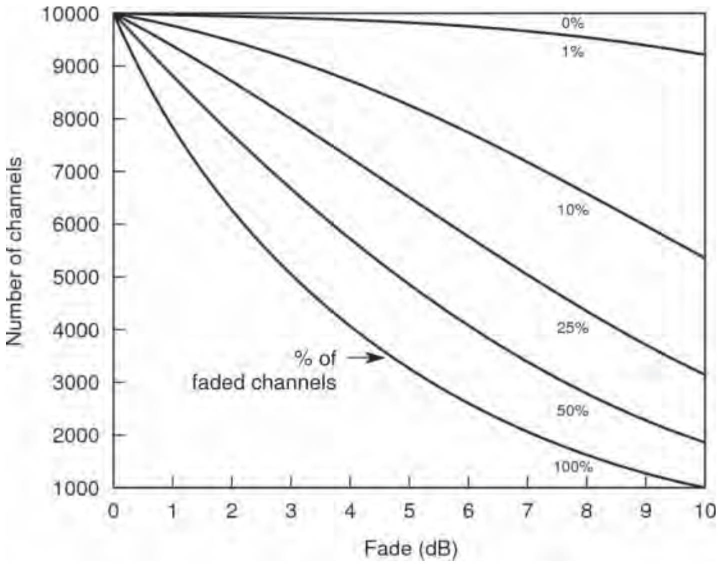


Figure 3.52 Variation in space segment capacity of a 70 dBW EIRP satellite as a function of fade for an MSS using a power-controlled fade countermeasure

power-limited capacity to fade for an MSS system with a total capability of 70 dBW spacecraft EIRP and an EIRP requirement 30 dBW per channel when a power-controlled fade countermeasure is applied. The theoretical forward link capacity of the system without fade is estimated as 10 000 channels. At a fade of 5 dB the capacity reduces respectively to 9750 and 3400 when 1 and 100% of channels fade.

We have already addressed the applicability of path diversity to K_u and K_a band feeder links; deployment of separate antennas in the service link can achieve a limited diversity advantage under multipath dominated and lightly shadowed conditions but little advantage under shadowed condition as shadowing events tend to be correlated. Multiple antennas and complex receivers make the user terminals more expensive and bulkier. In a wide-band system path diversity advantage can potentially be obtained by combining multipath components; however, when the multipath power level is low, ($\sim -20\,\text{dB}$) the advantage is insignificant.

An alternative is satellite diversity. At present, path diversity from separate satellites is rarely used for geostationary mobile satellite systems. Implementation becomes increasingly difficult for applications where receiver antennas are directional, as satellite separation is increased. Note the direct relationship between satellite separation and diversity improvement. Furthermore, deployment of two satellites is expensive; it wastes spacecraft resources as well as geostationary orbital slots. Nevertheless, in safety-critical applications where reliability is paramount these constraints are waived.

Non-geostationary satellite systems have an inherent capability to exploit diversity, because the number of satellites is large and path diversity is a natural by-product. Constellations may be designed such that multiple visibilities of satellites is a system feature, without incurring prohibitive costs. Therefore, a number of non-geostationary systems such as Intermediate Circular Orbit (ICO) and Globalstar have incorporated path diversity to improve link reliability. Figure 3.53 is a pictorial view of path diversity (also see Figure 3.17). Chapter 2 (Section 2.3.8) discusses the path diversity advantages, with examples of measurements, and Chapter 11 describes the application of diversity to specific constellations.

For system analysis and design, it becomes necessary to characterize all elements of a system. Propagation data related to diversity is relatively scarce, mainly limited to investigations of specific systems, and hence diversity advantage would be difficult to characterize. Therefore, theoretical methods, involving simulation such as those mentioned earlier in the section, offer a cost-effective solution.

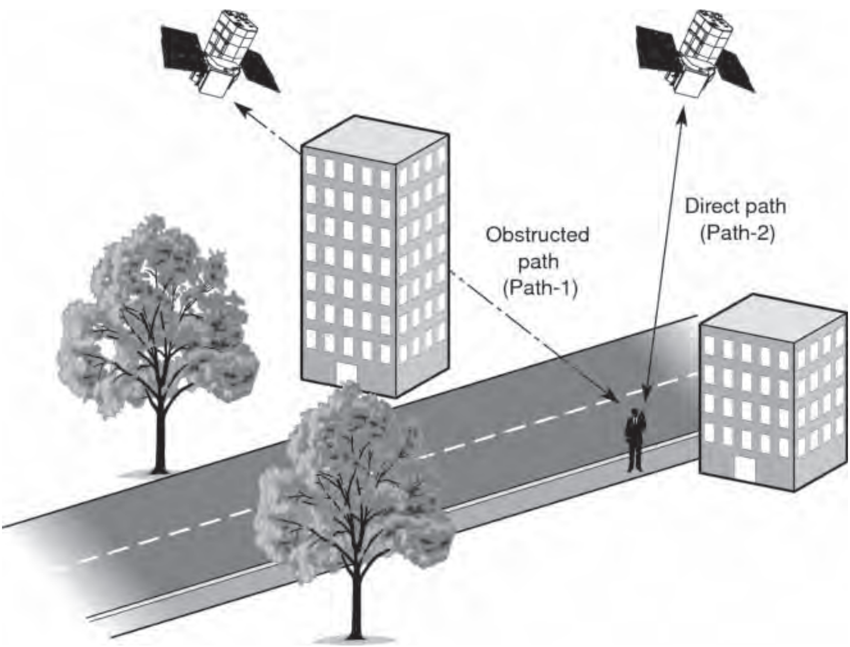


Figure 3.53 A pictorial view of path diversity

Another interesting approach is first to characterize environments of interest and derive a diversity advantage by superimposing constellation geometry. A method called *photogrammetry* proposes a combination of image processing of hemispherical fisheye lens photographs, constellation simulation and a propagation model to predict diversity and other propagation effects at locations of interest, thereby minimizing the necessity of a measurement campaign (Akturan and Vogel, 1997).

3.4 Radio Link Analysis

Radio link design is central in establishing system characteristics and specifications and in developing user terminal and satellite specifications within technological and economic constraints (see Chapter 7). The design is performed effectively with the aid of the transmission equation, which models the satellite radio link behaviour. It encapsulates transmitted power, path loss, propagation loss, interference effects, modulation, coding, multiple access, diversity improvement and signal quality requirements, allowing trade-offs to be applied simply. The subject is well-covered in the literature and therefore we will only introduce essential and interesting aspects relevant for MSS. Table 3.13 lists the key equations, and explains the significance and attributes of each component in a system context. Figures 3.52 and 3.54(a) and (b) demonstrate the utility of these equations applied to an MSS forward service link. Figure 3.52 was explained earlier in this chapter. Figure 3.54(a) represents satellite EIRP required per channel for bit rates up to 70 kbps for a variety of user terminals ranging from hand-held terminals to ship-borne terminals, assuming a E_b/N_o of 5 dBHz to meet the quality objective. Figure 3.54(b) extends the bit rate up to 2 Mbps marking 500 kbps as the cut-off for the existing L band systems.

Consider the model of a mobile satellite communications system. The radio link comprises a forward/return feeder link, a path internal to the satellite and forward/return service links, and intersatellite link(s) when applicable. In addition, there are a number of extraneous paths causing interference. By far the most important part of an MSS link are the service links, as they establish fundamental bounds in terms of terminal size, throughput, traffic carrying capacity, and so on. Depending on the emphasis, various trade-offs can be applied as summarized next:

- Forward service link:

  For a power-limited satellite, a trade-off can be exercised between terminal size, throughput and bandwidth (spectrum).

  For a given user terminal size and throughput, a trade-off is possible between satellite power and bandwidth.

  For a bandwidth limited system, given a terminal size and throughput, a trade-off can be applied between space segment capacity, spot beam size and number, modulation/coding/multiple access scheme (see Chapter 4), etc.

- Return service link:

  A trade-off can be applied between mobile terminal EIRP, throughput and sensitivity of satellite and feeder station.

  For a bandwidth limited system, a trade-off is possible between space segment capacity spot beam size, modulation and multiple access scheme for a given mobile EIRP.

Table 3.13 Link equations and explanation of each component in a system context

Link equations	Parameter	Description	Context
$R_x = P_s + G_s + G_d - 20 \log(4\pi D/\lambda) - L_l - L_i$ (3.42)	R_x	Received power level (dBW)	Fundamental component that determines viability and capability of a radio link Depends on transmitter EIRP, link losses and receive antenna gain Determines signal quality (E_b/N_o) (see Equation 3.46)
	P_s	Transmitter power (dBW)	*User terminal* Radiation effects set an upper bound on handheld user terminal Permissible size, cost, battery power set bound on directional user terminal
	G_s	Transmitter antenna gain (dB)	*Satellite* High gain with multi-spot beams is necessary to sustain MSS service link and enhance frequency reuse *User terminal* Omni-directional antennas are necessary for handheld and small mobile terminals High gain antennas are necessary for portable, broadband systems High gain tracking antennas are necessary for mobile broadband user terminals mounted on ships and aircrafts

(*continued overleaf*)

Table 3.13 *(continued)*

Link equations	Parameter	Description	Context
$EIRP = P_s + G_s$ (3.43)	EIRP	Effective isotropic radiated power (dBW)	A measure of effective power transmitted *Satellite EIRP* Determines the throughput capability of radio link *User terminal EIRP* Maximum limited by safe radiation limit, cost, size and DC power source constraints
	G_d	Receiver antenna gain (dB)	Satellite Antenna gain must be high enough to receive low power transmissions of user terminals User terminal Antenna gain of user terminals govern the receiver sensitivity and hence user data rate
	$20 \log (4\pi D/\lambda)$ (3.44)	Free space loss (dB)	λ = Wavelength; D = distance between transmitter and receiver in same unit as λ
	L_l	Additional link losses, for example fade (dB)	Loss can be caused by a number of factors; see text
	L_i	Loss due to interference (dB)	A measure of extra power required to compensate for interference

$(C/N)_r = Ps + Gs - 20 \log(4\pi D/\lambda) - L_l - L_i + (G_d/T_r) - 10 \log(k) - 10 \log(B)$ (3.45)	$(C/N)_r$	Received carrier to noise ratio	See Equation 3.42 and explanation accompanying R_x above. $(C/N)_r$ is a measure of the carrier power to noise power ratio over a single radio connection Equation 3.46 shows its relationship to E_b/N_o. Note the inclusion of available bit rate R_t
	G_d/T_r	Receiver sensitivity (dB/K)	A measure of receiver sensitivity. This is a fundamental component in determining throughput and signal quality
	$10 \log (k)$	−228.6 dB/K	Boltzmann constant
	$10 \log (B)$	Bandwidth (Hz) in dB	
$E_b/N_o = (C/N) + 10 \log B - 10 \log(R_t)$ (3.46)		E_b/N_o = Energy per bit by noise power spectral density; R_t = bit rate	A measure of quality of service in digital systems Specified on basis of bit error rate (BER), modulation scheme, modem implementation losses, coding type and coding rate Influences the terminal sensitivity specification (G/T)

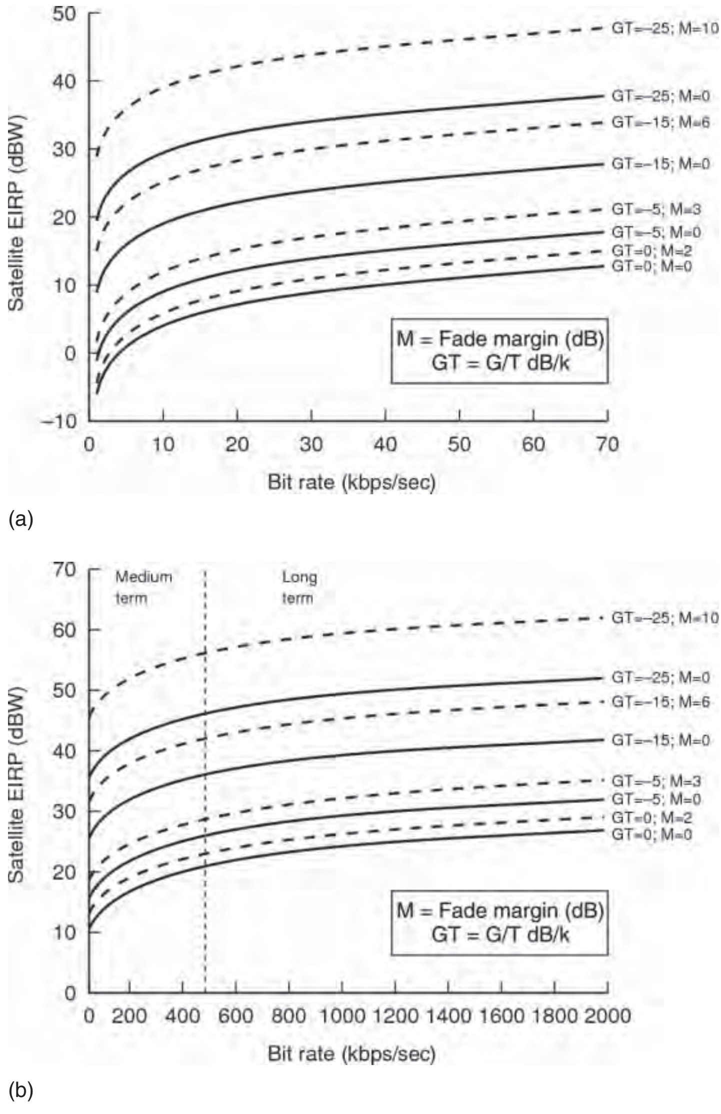


Figure 3.54 User bit rate versus satellite EIRP/channel for a variety of user terminals ranging from hand-held to ship-borne for E_b/N_o of 5 dB-Hz at bit rates up to (a) 70 kbps and (b) 2 Mbps, showing respectively the present and future capabilities of L band MSS (Both parts source: Graphics AR.)

Examples of practical considerations affecting trade-offs include (see also Chapter 7):

- Spacecraft service link EIRP tend to be limited by antenna and high-power amplifier technology;
- Sensitivity (G/T) of mobiles tend to be limited by permissible antenna size (Gain is directly proportional to antenna aperture area);

- Mobile EIRP that is limited by permissible antenna size, mobile battery capacity, HPA size/cost and radiation safety limits, the latter applying, in particular, for hand-held services;
- In the return direction, sensitivity of satellite receiver that influences mobile terminal size, EIRP and service capability.

A regenerative satellite repeater de-couples feeder and service link paths by regenerating the incoming signal before re-transmitting it on a different transmission format (see Chapter 6 for further details). Links can therefore be optimized individually. The overall BER at a receiver for a regenerative transponder can be approximated as:

$$\mathrm{BER_t} \approx \mathrm{BER_u} + \mathrm{BER_d} \tag{3.47}$$

where $\mathrm{BER_t}$ = the total bit error rate, $\mathrm{BER_u}$ = the bit error rate of uplink and $\mathrm{BER_d}$ = the bit error rate of downlink.

In a transparent transponder noise in the up and downlinks are coupled during the translation and therefore overall BER becomes multiplicative. The links must be optimized in their entirety, respecting the limitations of each part of the link. The total E_b/N_o in this case is given as

$$(E_b/N_T) = [(E_b/N_u)^{-1} + (E_b/N_d)^{-1} + (E_b/N_I)^{-1} + (E_b/N_{iu})^{-1} + (E_b/N_{id})^{-1}]^{-1} \tag{3.48}$$

where, E_b/N_T, E_b/N_u, E_b/N_d, E_b/N_I, E_b/N_{iu}, E_b/N_{id} are energy per bit to noise power spectral density of total link, up-link, down-link, inter-modulation, up-link interference and down-link interference respectively.

Note the large number of components that have to be balanced. Recall that the most vital ones are the service links due to their inherent power limitation, severe propagation impairments and susceptibility to inter- and intrasystem interference.

Due to invariant geometric relationships, link parameters such as free space loss, look angles and Doppler Effect of a geostationary satellite system are nearly static for practical purposes. The main variables are propagation loss and interference. Note that mobiles move freely within the service area, and therefore receive and cause variable amounts of interference. The traditional method for interference management involves ensuring adequate interference margin at the worst-affected location, that is, the location where the wanted signal is the weakest and the interfering signal is strongest. When considering intrasystem interference, an operator can tailor frequency reuse, such that dense traffic areas are better protected, or an interference margin is provided to a certain percentage of geographical areas (e.g. 95%). This approach is no worse than provisioning propagation margin based on statistics; thus the overall traded-off can be applied to interference and propagation margin. Table 3.14 shows examples of link budget for FDMA (frequency division multiple accessing), TDMA (time division multiple accessing) and CDMA schemes for hypothetical geostationary satellite systems.

Link analysis of non-geostationary constellations is more involved, due to continuous variations in path length, elevation angle and interference during a call. Figure 3.10 shows the temporal variation in path loss and elevation angle at a given location for LEO, MEO and GEO systems as a comparison.

The wide variation in path loss is further exacerbated by loss due to spacecraft antenna gain variation caused by motion. The net effect is a need to incorporate a large dynamic range for

Table 3.14 Comparative forward link budgets of TDMA and CDMA

(a)

	GSM (TDMA)		MSBN (CDMA)	
	Transportable	Vehicular	Transportable	Vehicular
Ku band up-link				
Fixed station EIRP (dBW)	57.8	57.8	57.8	57.8
Diameter (m)	1.8	1.8	1.8	1.8
Path loss (dB)	207.4	207.4	207.4	207.4
Other losses (dB)	4.5	4.5	4.5	4.5
Uplink C/N_0 (dB-Hz)	73.1	73.1	73.1	73.1
Voice activity factor (fraction of time active)	0.4	0.4	0.4	0.4
Satellite				
Transponder type	Transparent/ Global + 3 spots/ 15 MHz	Transparent/ Global + 3 spots/ 15 MHz	Transparent/ Global + 3 spots/ 15 MHz	Transparent/ Global + 3 spots/ 15 MHz
Total L-band EIRP (dBW)	45	45	45	45
Capacity used	8 MHz; 8 × 1 MHz modules	8 MHz; 8 × 1 MHz modules	8 MHz; 8 × 1 MHz modules	8 MHz; 8 × 1 MHz modules
Satellite G/T (dB/K) in K_u band	−1.4	−1.4	−1.4	−1.4

(*continued overleaf*)

Table 3.14 (*continued*)

(a)

	GSM (TDMA)		MSBN (CDMA)	
	Transportable	Vehicular	Transportable	Vehicular
L band down link				
Satellite EIRP/carrier (dBW)	29.3	38.0	15.8	24.4
Path loss (dB)	188.3	188.3	188.3	188.3
Other losses (dB)	1.1	1.1	1.1	1.1
Shadowing margin (dB)	0	4.7	0	4.7
Mobile G/T (dB/K); Gain (dB)	−9.0, 16	−13.0, 12	−9.0, 16	−13.0, 12
Self-noise loss (dB)	–	–	0.5	0.5
Down-link C/N_o (dB-Hz)	59.5	59.5	45.5	45.4
Inter-system Interference (dB-Hz)	–	–	59.0	63.6
Required E_b/N_0 (dB-Hz)	3.0	3.0	5.0	5.0
Bit rate (Kb/s)	270	270	6.76	6.76
Implementation margin (dB)	2.0	2.0	2.0	2.0
Total C/N_0 (dB-Hz)	59.3	59.3	45.3	45.3
Number of active simultaneous channels	592	80	832	115
Offered service	Speech	Speech	Speech	Speech

(Source: Priscoli and Muratore, 1996. Reproduced with permission of John Wiley & Sons, Ltd.)

the receiver coupled with a dynamic power control of RF channel. Another consideration is the elevation-dependent variability in propagation behaviour, which requires elevation dependent modelling (see Section 3.3.2.4). A further consideration related to propagation is to include diversity effects and methods of combining signals from diverse paths (see Section 3.3.2.3 and Chapter 2, Section 2.3.8). Estimating interference in non-geostationary satellite systems is rather involved, due to the dynamics of the constellation as well as frequency reuse scheme deployed in the network; in fact frequency planning and interference analysis are intertwined. Once propagation effects, diversity improvements and interference loss are established, link analysis can be applied through the appropriate link equation.

Revision

1. Demonstrate with an example, the principle of co-channel frequency reuse by multiple spot beam satellites.
2. Estimate the spectrum required for a 350 spot beam satellite system capable of supporting seven-beam clusters when,
   i. Traffic is distributed uniformly such that each spot beam requires 10 200-kHz channels.
   ii. Three clusters require 20 channels each, with traffic in other clusters uniformly distributed as stated in (i). State assumptions. (Hint: Frequencies cannot be reused within a cluster.)
3. State the difference between forward and return link interference analysis of a mobile satellite system. Demonstrate the difference algorithmically.
4. Describe the differences in interference analysis of geostationary and non-geostationary satellite systems. Outline the salient features of interference model in each case.
5. Suggest a methodology for long term spectrum forecast of a regional mobile satellite system. State its limitations.
6. What are the propagation impairments common to all types of satellite communication systems? Discuss the implications of each on system design.
7. The signal received on a mobile comprises a number of components, the magnitude of each depending on the local environment and mobile category – that is land, maritime and aeronautical. What are the factors that affect the magnitude of the components in each category?
8. Explain the significance of each component of the received signal on system design differentiating between mobile terminal types where necessary.
9. Develop the transmission equation and explain its role in radio link design.
10. Explain the significance of various components of the transmission equation in a system context.

References

Abbe, B., Agan, M. and Jedrey, T. (1996) ACTS mobile terminals. *International Journal of Satellite Communications*, **14** (3), 175–190.

Akturan, R. and Vogel, W.J. (1997) Path diversity for LEO satellite-PCS in the urban environment. *IEEE Transactions on Antennas and Propagation*, **45** (7), 1107–1116.

Arapoglou, P.D., Liolis, K.P. and Panagopoulos, A.D. (2012) Railway satellite channel at Ku band and above: composite dynamic, modeling for the design of fade mitigation techniques. *International Journal of Satellite Communications and Networking*, **30**, 1–17.

Aulin, T. (1979) A modified model for the fading signal at a mobile radio channel. *IEEE Transactions on Vehicular Technology*, **VT-28** (3), 182–203.

Babalis, P.G. and Capsalis, C.N. (1998) A new model for the characterisation of the combined shadowing and multipath effects in satellite systems. *ICT*, **98**, 25–28.

Beckmann, P. and Spizzichino, A. (1963) *The Scattering of Electromagnetic Waves From Rough Surfaces*, Pergamon Press, New York/Oxford.

Bello, P.A. (1973) Aeronautical channel characterisation. *IEEE Transactions on Communication*, **COM-21**, 548–563.

Burzigotti, P., Prieto Cerdeira, R., Bolea Alamañac, A. *et al*. (2008) DVB-SH analysis using a multi-state land mobile satellite channel model. Proceedings of 4th Advanced Satellite Mobile Systems Conference (ASMS 2008), Bologna, Italy.

Butt, G., Evans, B.G. and Richharia, M. (1992) Narrowband channel statistics from multiband measurements applicable to high elevation angle land-mobile satellite systems, multiband propagation experiment for narrowband characterisation of high elevation angle land mobile-satellite channels. *Electronics Letters*, **28** (15), 1449–1450.

Clarke, R.H. (1968) A statistical theory of mobile radio reception. *Bell System Technical Journal*, **47**, 957–1000.

Corazza, G.E. and Vatalaro, F. (1994) A statistical model for land mobile satellite channels and its application to nongeostationary orbit systems, *IEEE Transactions on Vehicular Technology*, **VT-43**, 738–741.

Cox, D.C. and Leek, R.P. (1972) Correlation bandwidth and delay spread multipath propagation statistics for 910 MH z urban mobile radio channels. *IEEE Transactions on Communication*, **23** (11), 1271–1280.

Crane, R.K. (1980) Prediction of attenuation by rain. *IEEE Transactions Communications*, **COM-28**, 1717–1733.

Dooren, G.A.J., Herben, M.H.A.J., Brussaard, G. *et al*. (1993) Electromagnetic field strength prediction in an urban environment: a useful tool for the planning of LMSS. IMSC 93, Third International Mobile Satellite Conference, June 16–18, Pasadena, CA, JPL Publication 93-009, pp. 343–348.

ESA (1977) ESA/ESTEC Contract No 5323/83/NL/JS, DFVLR, Oberpfaffenhofen, D-8031 (in German).

Freeman, R.L. (1996) *Reference Manual for Telecommunications Engineering*, 2nd edn, John Wiley & Sons, Inc., New York.

Gambaruto, E., Richharia, M. and Trachtman, E. (2008) Land vehicular propagation channel characterization of BGAN system. *International Journal of Satellite Communications and Networking*, **26** (2), 85–117.

Goldhirsh, J. and Vogel, W. (1992) Propagation Effects for Land Mobile Systems: Overview of Experimental and Modelling Results, NASA Reference Publication 1274.

Hagenauer, J., Dolainsky, F., Edbauer, F. *et al*. (1984) Multipath Fading Effects and Data Transmission for Small Ship Earth Stations (Standard C), DFVLR Final Report for ESA/ESTEC, DFVLR, Oberpfaffenhofen, Contract No 5323/82/NL/JS, D-8031, p. 223 (in German).

Hagenauer, J., Neul, A., Papke, W. *et al*. (1987) The Aeronautical Satellite Channel. DFVLR Technical Report No NE-NT-T-97-17, DFVLR, Oberpfaffenhofen, May, D-8031.

Higuchi, T. and Shinohara, T. (1988) Experiment of Inmarsat Standard-C system. Fourth International Conference on Satellite System for Mobile Communication and Navigation, IEE, Conference, Publication No. 294.

Hogben, N. and Lumb, F.E. (1967) *Ocean Wave Statistics*, Her Majesty's Stationery Office, London.

Hwang, S.H., Kim, K.J., Ahn, J.Y. and Whang, K.C. (1997) A channel model for nongeostationary orbiting satellite systems. Proceedings of IEEE Vehicular Technology Conference, pp. 41–45.

Ikegami, T., Arakaki, Y., Wakana, H. and Suzuki, R. (1993) Measurement of multipath delay profile in land mobile satellite channels. Proceedings of the Third International Mobile Satellite Conference, IMSC'93, Pasadena, CA, June 16–18, JPL Publication 93-009, pp. 331–336.

Ippolito, L. (1986) *Radiowave Propagation in Satellite Communications*, Van Nostrand Reinhold.

Ippolito, I.J. and Russel, T.A. (1993) Propagation considerations for emerging satellite communications applications. *Proceedings of the IEEE*, **81** (6), 923–929.

ITU (1986) Attenuation by Hydrometers in Particular Precipitation, and Other Atmospheric Particles. ITU-R Report 721-2, Propagation in Non-ionised Media, Recommendations and Reports of the ITU-R, Geneva, p. 5.

ITU (1990) Propagation Data and Prediction Methods Required for Earth–Space Telecommunication Systems. ITU-R Report 564-3, Propagation in Non-ionised Media, Recommendations and Reports of the ITU-R, p. V.

ITU (1992a), ITU-R Rec. 531-2, RPI Series, Geneva.

ITU (1992b) Propagation Data Required for the Design of Earth-Space Land Mobile Telecommunication Systems. P Series Radiowave Propagation, Recommendation ITU-R P.681-2.

ITU (2008) Radio Regulations, Vol. 1, Article 5.

ITU (2009), Propagation Data Required for the Design of Earth-Space Land Mobile Telecommunication Systems. P Series Radiowave Propagation, Recommendation ITU-R P.681-10.

ITU-R (2012) Recommendation ITU-R P.531-11, Ionospheric propagation data and prediction methods required for the design of satellite services and systems.

Jahn, A., Buonomo, S., Sforza, M. and Lutz, E. (1995) Wideband channel model for land mobile satellite system. Proceedings International Conference on Universal Personal Communication ICUPC'94, San Diego, CA, pp. 122–126.

Jahn, A. and Lutz, E. (1994) DLR channel measurement programme for low earth orbit satellite systems. Proceedings International Conference on Universal Personal Communication ICUPC'94, San Diego, CA, pp. 423–429.

Kanatas, A.G., Kanderakis, E.C. and Constantinou, P. (1995) Narrowband land mobile satellite channel modelling. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communication Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 109–114.

Karaliopoulos, M.S. and Pavlidou, F.N. (1999) Modelling the land mobile channel: a review. *Electronics and Communication Engineering Journal*, **11** (5), 235–248.

Karasawa, Y. and Shiokawa, T. (1988) A simple prediction method for L band multipath fading in rough sea conditions. *IEEE Transactions on Communications*, **COM-36** (10), 1098–1104.

Karasawa, Y., Yasunaga, M., Nomoto, S. and Seiokawa, T. (1986) On-board experiments on L-band multipath fading and its reduction by use of the polarisation shaping method. *Transaction IEICE (Japan)*, **E69** (2), 124–131.

King, P.R. (2007) Modeling and measurement of the land mobile satellite MIMO radio propagation channel. PhD thesis. University of Surrey, UK.

Lee, W. (1986) *Mobile Communication Design Fundamentals*, Howard W Sams & Co..

Liolis, K.P., Gómez-Vilardebó, J., Casini, E. and Pérez-Neira, A. (2009) On the statistical modeling of MIMO land mobile satellite channels: a consolidated approach. ICSSC 2009, The 27th AIAA International Communications Satellite Systems Conference, ICSSC 2009, Edinburgh Conference Centre, Heriot-Watt, Edinburgh, Scotland, June 1–4, 2009, 5 pp.

Long, M.W. (1975) *Radar Reflectivity of Land and Sea*, Lexington Books, Lexington, MA.

Loo, C. (1985) A statistical model for the land mobile satellite link. *IEEE Transactions on Vehicular Technology*, **VT-34**, 122–127.

Lutz, E. (1996) A Markov model for correlated land mobile satellite channels. *International Journal of Satellite Communications*, **13**, 333–339.

Lutz, E. (1998) Issues in satellite personal communication systems. *Wireless Networks*, **4**, 109–124.

Lutz, E., Cygan, D., Dippold, M. *et al*. (1991) The land mobile satellite communication channel – recordings, statistics and channel model. *IEEE Transactions on Vehicular Technology*, **VT–40**, 375–386.

Lutz, E., Papke, W. and Plöchinger, E. (1986) Land mobile satellite communications-channel model, modulation and error control. Proceedings of an ESA Workshop on Land Mobile Services ESTEC, ESA, SP-259, June 3–4.

Matricciani, E. (1995) Transformation of rain attenuation statistics from fixed to mobile satellite communication systems. *IEEE Transactions on Vehicular Technology*, **44** (3), 565–569.

Matsudo, T., Minamisono, K., Karasawa, Y. and Shiokawa, T. (1993) A prediction model of signal degradation in LMSS for urban areas. International Mobile Satellite Conference, IMSC 93, The Third International Mobile Satellite Conference, June 16–18, Pasadena, CA, JPL Publication 93-009, pp. 355–60.

Murr, F., Arbesser-Rastburg, B. and Buonomo, S. (1995) Land mobile satellite narrowband propagation campaign at Ka band. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communication Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 134–138.

Ohmori, S., Irimata, A., Morikawa, H. *et al*. (1985) Characteristics of sea reflection fading in maritime satellite communications. *IEEE Transactions on Antennas and Propagation*, **AP-33** (8), 838–845.

Parsons, D. (1992) *The Mobile Radio Propagation Channel*, Pentech Press, London.

Pätzold, M., Killat, U., Laue, F. and Li, Y. (1998) On the statistical properties of deterministic simulation models for mobile fading channels. *IEEE Transactions on Vehicular Technology*, **47** (1), 254–269.

Perrins, E. and Rice, M. (1997) Propagation analysis of the ACTS maritime satellite channel. Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, CA, June 16-18, pp. 201–205.

Pinck, D.S. and Rice, M. (1995) K/Ka-band Channel Characterisation for Mobile Satellite Systems IMSC-95, 8 pp.

Priscoli, F.D. and Muratore, F. (1996) Radio and network comparisons for MSBN and GSM systems in a geostationary satellite environment. *International Journal of Satellite Communications*, **14**, 439–454.

Richharia, M., Kaluvala, N., Axel, J. and Fines, P. (2005) Aeronautical channel model for broadband L-band satellite communications. 23rd Communications Satellite Systems Conference, Organizer American Institute of Aeronautics and Astronauts (AIAA), Rome, Italy.

Schroeder, E.H., Thompson, A.D., Sutton, R.W. *et al*. (1976) Air Traffic Control Experimentation and Evaluation with NASA ATS-6 Satellite. The Boeing Company Report No FAA-RD-75-173, III NTIS, Accession No. D6-44048, Vol. V, NTIS Accession No. D6-44050, National Technical Information Service, Springfield, VA.

Scalise, S., Ernst, H. and Harles, G. (2008) Measurements and modeling of the land mobile satellite channel at Ku-band. *IEEE Transactions on Vehicular Technology*, **57** (2), 693–703.

Sforza, M., Dibernardo, G. and Cioni, R. (1993) Propagation model for the land mobile satellite channel in urban environments. International Mobile Satellite Conference, IMSC 93, The Third International Mobile Satellite Conference, Pasadena, CA, June 16–18, JPL Publication 93-009, pp. 349–354.

Sutton, R.W., Schroeder, E.H., Thompson, A.D. and Wilson, S.G. (1973) Satellite – aircraft multipath and ranging experiment results at L band. *IEEE Transactions on Communications*, **COM-21** (5), 639–647.

Vatalaro, F. (1995) Generalised Rice-lognormal channel for wireless communications, *Electron*. Lett., October, 1899–1900.

Vatalaro, F., Corazza, E.G., Caini, C. and Ferrarelli, C. (1995) Analysis of LEO, MEO and GEO global mobile satellite systems in the presence of interference and fading. *IEEE Journal on Selected Areas in Communications*, **13** (2), 291–300.

Vogel, W.J. and Goldhirsh, J. (1990) Propagation Limits for Land Mobile Satellite Services. AIAA-90-0847-CP, pp. 564–574.

Vogel, W.J. and Hong, U. (1988) Measurement and modelling of land mobile satellite propagation at UHF and L band. *IEEE Transactions on Antennas and Propagation*, **AP-36**, 707–719.

4

Modulation, Coding and Multiple Access

4.1 Introduction

This chapter gives an overview of a range of concepts and techniques related to modulation, coding and multiple access schemes. Topics are covered at a system level from a mobile satellite service (MSS) perspective and treatment kept to a level essential to grasp the sensitivities of components in a system context. Modulation and coding schemes applicable to MSS are widely used in other areas of radio communications and are hence well-documented in the literature, so the emphasis here is rather on their performance in an MSS environment. Some of the recent schemes are addressed in detail. Similarly, the treatment of multiple access schemes focusses on their applicability to MSS.

4.2 Modulation

The first part of this section outlines the requirements of a modulation scheme for its applicability to an MSS system. The next section narrows the choice of modulation schemes to those best suited for MSS and in the final part, the performance of each is reviewed.

4.2.1 MSS Requirements

Radio channel characteristics exert a marked influence on the performance of modulation schemes. In MSS systems, service link characteristics set a bound on system performance (Chapter 3). The main sources of degradation affecting the performance of modulation schemes in mobile communication links are:

- signal fades caused by environment and velocity dependent multipath, characterized by slow ($< \sim 1$ Hz) and rapid ($\sim$ tens of hertz) signal fluctuations;
- signal fades caused by tropospheric effects, applicable at frequencies $> \sim 10$ GHz;
- signal phase fluctuations caused by local oscillators, multipath and Doppler jitter effects;

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

- very low carrier to thermal noise ratio due to small user terminals (UTs);
- non-linearity arising in various system components such as caused in mobile earth stations that use class-C amplifiers;
- large frequency changes attributed to Doppler associated with medium earth orbit (MEO) and low earth orbit (LEO) that have to be adequately compensated.

It is anticipated that many future MSS satellite transponders will be regenerative. On-board applications will require lightweight, small, power-efficient modems and high coding gain.

In coherent demodulation schemes, the carrier is recovered prior to demodulation. Carrier recovery circuits are susceptible to thermal noise and fading and therefore particular care has to be exercised in mobile earth stations. Furthermore, carrier recovery becomes more difficult as radio frequency or data rate increase in thermal noise limited, fading links. In this respect, non-coherent demodulation schemes are more robust.

Demodulation also requires symbol clock recovery for extracting the transmitted bit stream. The requirement becomes more stringent for higher level modulation schemes where it is essential to synchronize time, phase and amplitude of both I and Q channels. Again, the problem worsens as radio frequency and symbol rate increase. From this viewpoint, schemes that have less stringent synchronization requirements are preferred when robust performance is important such as for signalling, although this advantage has to be traded-off against the lower spectral efficiency in comparison.

The choice of a modulation scheme is additionally influenced by the multiple access scheme, which in turn influences spectrum utilization efficiency and is governed by service requirements, propagation considerations, etc. (see Section 4.4). For example, the synchronization need is quite demanding in time division multiple access (TDMA) schemes, due to the need to operate in burst mode, which imposes stringent timing requirements on carrier recovery and symbol/bit synchronization circuits. Additional complexity is introduced in demodulators by the need for dynamic adjustments in frame synchronization due to satellite motion. For identical link and throughput requirements, carrier recovery and synchronization problems are less problematic for circuit-mode communication where the radio carrier remains relatively stable throughout a call. In a code division multiple access (CDMA), spread spectrum modulation remains the choice with a provision to tailor the base modulation as required.

The spectral efficiency of the modulation scheme is vital in modern MSS because of the impending shortage of spectrum, due to rising demands. Higher order modulation schemes, coupled with powerful codes are spectrally efficient but require higher transmission power, leading to the classical power-bandwidth trade-off. Since carrier recovery precedes demodulation these circuits can sometimes set the threshold; rapid synchronization after a prolonged shadowing event is another related problem. This is an area of active research as explained later in the chapter.

Finally, it is essential that modulators/demodulators used in mobile terminals can be mass produced to provide economies of scale.

4.2.2 *Preferences*

Due to the overwhelming advantages offered by digital systems, all the MSS systems use digital transmissions. We will only deal with digital modulation schemes here. Frequency modulation (FM) was used in first generation mobile satellite systems such as the Inmarsat-A service. Single-side band (SSB) systems were considered as a potential candidate due to their spectral efficiency at one stage. FM systems provide a more effective use of satellite power than the SSB systems, but are less spectrally efficient. Frequency shift keying (FSK) continues to be used for low cost and simple MSS applications.

There are three basic digital modulation schemes, depending on the parameter of the carrier frequency altered by the baseband digital stream – amplitude shift keying (ASK), FSK and phase shift keying (PSK). As the names suggest, in ASK the amplitude of the carrier is altered, in FSK the carrier frequency is changed and in PSK the phase of carrier is varied in accordance with the incoming digital stream. Figure 4.1 (Xiong, 1994) illustrates a tree diagram of digital modulation schemes, along with relationships between them.

Modulation schemes are shown divided into two broad categories – constant envelope and non-constant envelope. The relationships between various schemes are also illustrated. Note that some of the schemes have more than one parent, that is, they can be derived from more than one technique. The schemes that can be differentially encoded are marked as D and those detected non-coherently are marked as N. Differential encoding removes the phase ambiguity experienced in coherent detection in which a circuit is unable to distinguish between phase 0 and π radians. In general, constant envelope modulation schemes are preferred, as they offer a more robust performance. Of the constant envelope modulation schemes, FSK schemes are suitable for low bit rate transmissions, due to demodulator hardware simplicity; they have a lower spectral efficiency than other schemes and are therefore not suited for higher bit rate transmissions. PSK modulation schemes have a near constant envelope, but exhibit discontinuity in phase, whereas constant phase modulation (CPM) schemes have a constant envelope with a gradual change in phase, which results in better side-lobe performance. Traditionally non-constant envelope modulation schemes were not used for MSS, but due to the pressing need of better spectral efficiency, multilevel schemes such as quadrature amplitude modulation (QAM) together with powerful convolution codes were introduced in wideband MSS systems.

Common digital modulation schemes used in MSS are:

- binary phase shift keying (BPSK) and its variants such as aviation-BPSK (or symmetric BPSK);
- quadrature phase shift keying (QPSK) and its variants such as offset-quadrature phase shift keying (O-QPSK), aviation-QPSK, minimum shift keying (MSK) schemes, for example Gaussian minimum shift keying (GMSK);
- multi-level FSK and its variants;
- spread spectrum modulation in conjunction with CDMA;

- 16-QAM with turbo-coding; 16-APSK and 32-APSK are new entrants, but yet to be established;
- coded orthogonal frequency division multiplexing (COFDM) modulation scheme (used in radio broadcast systems).

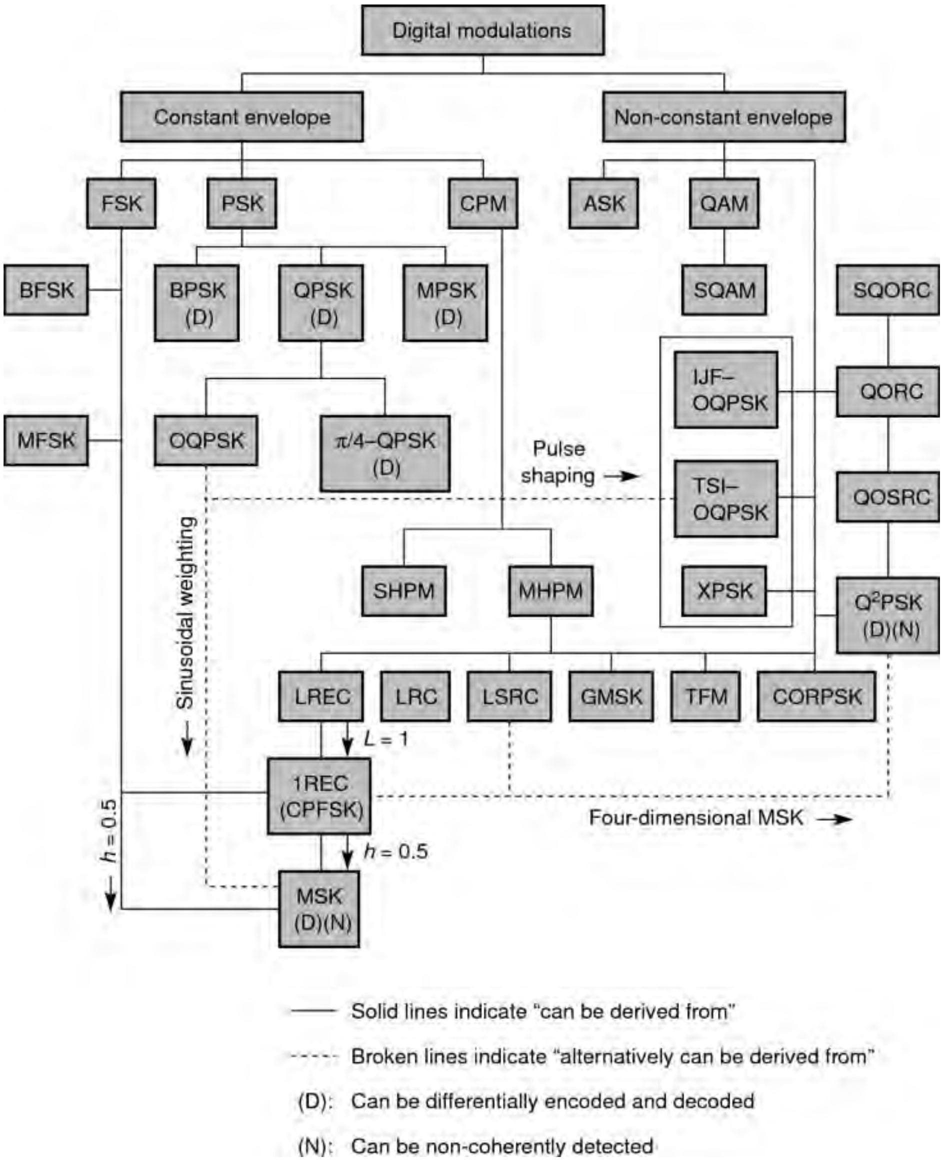


Figure 4.1 A tree diagram of digital modulation schemes. (Source: Xiong, 1994. © 1994 IEEE. Reproduced with permission.)

4.2.3 PSK Schemes

PSK modulation techniques may be divided into three broad categories according to the demodulation technique:

1. reference signals are transmitted with the main signal to aid carrier extraction;
2. the carrier is extracted from the received signals by appropriate signal processing;
3. carrier recovery is not required for demodulation.

Coherent detection of PSK signals gives a better power efficiency than non-coherent detection schemes. For coherent detection, carrier synchronization (or recovery) is essential for extracting phase information of the received signal; this can be problematic in the presence of multipath, low carrier-to-noise ratio and Doppler shifts. Noise introduces phase error in the recovered carrier and timing jitters in the recovered bit stream.

As an example in a BPSK scheme, a phase jitter of $\Delta\phi$ in carrier recovery and Δt in bit synchronization timing, the bit error rate (BER) in Gaussian noise degrades to (see Table 4.1);

$$Pe = \tfrac{1}{2}\,\mathrm{erfc}\sqrt{(E_s/\eta)}\,(\cos^2\phi)(1 - 2|\tau|/T)^2 \tag{4.1}$$

Figure 4.2 shows the effect of phase and time jitter on BER for a number of combinations where phase jitter and time jitter are respectively $\pm 15°$ and $|\tau|/T$ equal to .025.

Differential detection schemes do not require carrier recovery and are therefore robust and capable of rapid synchronization, but at the expense of some degradation in performance. Examples of such schemes are MSK and $\pi/4$ QPSK, differentially coded (D) BPSK, DQPSK (Differentially encoded QPSK), $\pi/4$-DQPSK, DMSK, and others.

Two modes of transmissions are possible in the reference signal transmission method. The reference signal can be either an unmodulated pilot multiplexed with the main transmissions,

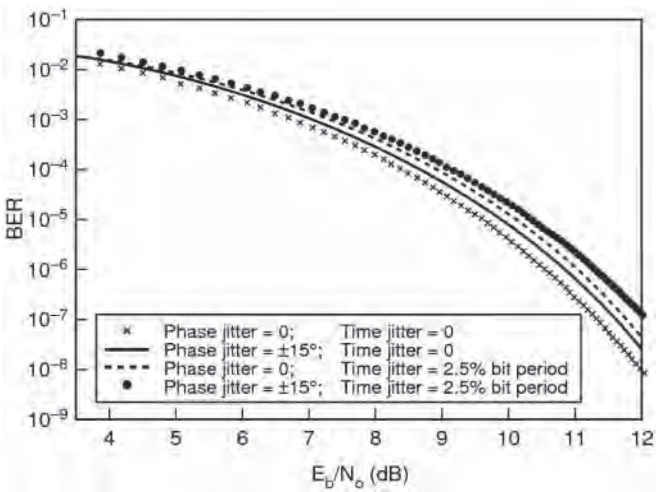


Figure 4.2 E_b/N_o versus BER with combinations of phase and time jitter. (Source: Graphics AR.)

or the signal may be multiplexed with information bits; numerous techniques have been proposed for the latter. A spectral null can be created at the centre frequency of the transmission spectra, where a pilot is inserted. Symbols may be inserted once or periodically in each TDM frame so that fading of data and symbol are correlated. The reference signal transmission scheme is wasteful of power and bandwidth but carrier recovery is simpler. Examples of this technique are transparent tone-in-band (TTIB) and tone-calibration technique (McGeehan and Bateman, 1984; Simon, 1986; Korn, 1989; Davrian, 1987, 1985; Caves, 1991). These schemes do not exhibit irreducible error rate (explained later), and E_b/N_o in the presence of Rayleigh fading is high. Transmissions do not possess a constant envelope and therefore the scheme is sensitive to system non-linearity.

In the carrier recovery method, frequency and phase of carrier is derived by processing the received signal. Well-established carrier recovery techniques include:

- $(C)^M$ method, where C is the carrier signal and M denotes number of symbols,
- Costas loop, and
- Decision feedback loop.

In the $(C)^M$ method, carrier is recovered by passing the received carrier, f_c, through a circuit that raises the carrier to the power of M to give a carrier at M f_c, followed by a divide by M circuit.

The Costas loop, named after its inventor, consists of a two-phase lock loop (PLL) arrangement that uses a common loop filter and voltage control oscillator (VCO). The VCO frequency is kept synchronized to the carrier frequency through a feedback loop comprising a multiplier, whose two inputs are outputs of an in-phase (I) and a quadrature phase (Q) comparators. The multiplier output provides a corrective voltage to maintain the VCO synchronized to the carrier (see Figure 4.3).

A decision feedback loop circuit recovers the carrier through a PLL using feedback from the demodulated signal. The recovered carrier is within the feedback loop, hence the name (see Figure 4.4).

A few schemes that use non-coherent detection and constant envelope signals to minimize sensitivity to the amplitude and phase fluctuations are GMSK, DMSK and DOQPSK. In general, such schemes are affected significantly by fading and suffer irreducible BER, where BER cannot be reduced below a lower bound irrespective of the level of E_b/N_o.

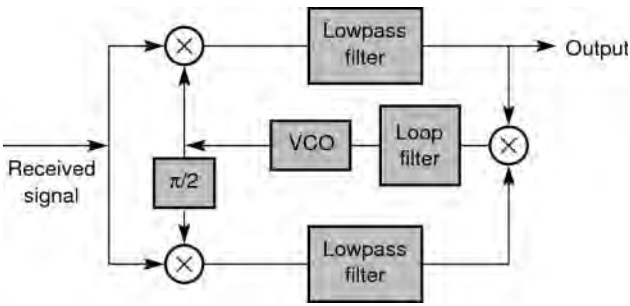


Figure 4.3 VCO using Costas loop

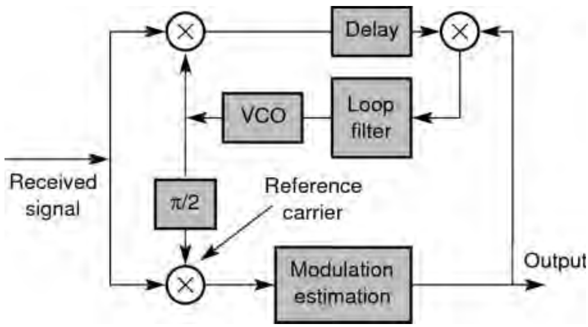
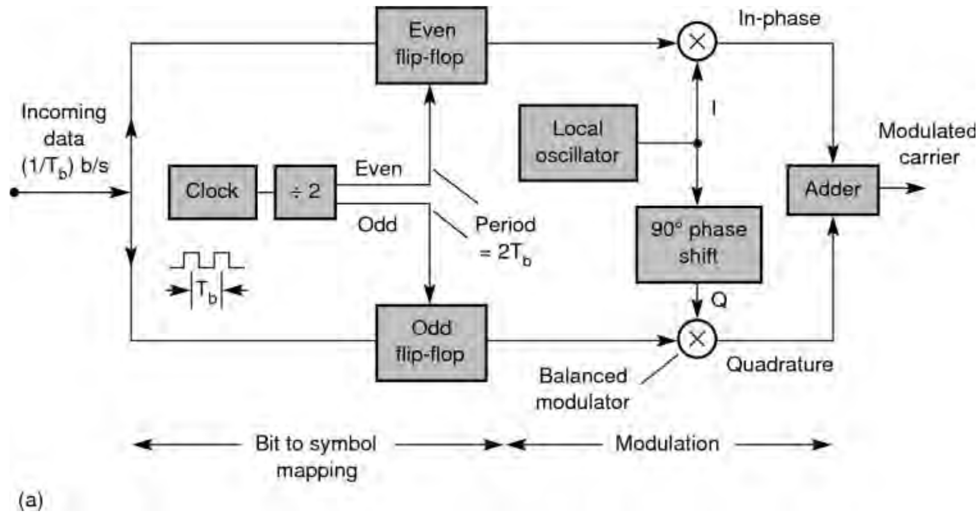


Figure 4.4 A decision feedback loop circuit

Figure 4.5(a) and (b) shows a simplified block schematic of a typical QPSK modulator and demodulator, respectively. The input digital stream is divided into I and Q components. Each stream is fed into a multiplier, the other input of which is an I or Q carrier as necessary. The I and Q components are then summed to accomplish a QPSK signal; at the same time, undesired components produced in multiplication are cancelled out. The signal is divided into two paths at the receiver, each of which is fed into the respective I or Q channel multiplier, the other input of which consists of the recovered carrier. The output of each multiplier is low pass filtered and fed in to an analogue-to-digital converter synchronized to the incoming bit stream through the timing recovery circuit. BPSK modems operate in the same manner as QPSK but only with one arm.

Derivatives of QPSK are generated by altering characteristics of the I and Q channels in various ways to reduce the sensitivity of QPSK signals to non-linearity and reduce the amplitude of side lobes to minimize the occupied bandwidth. For O-QPSK, the Q channel is delayed by half the symbol duration to avoid simultaneous phase transition, which gives a smoother phase transition of the carrier, thereby reducing the spectral side lobes of the transmitted signal. For MSK, I and Q channel timing is the same as O-QPSK, but followed by co-sinusoid spectral shaping for the I stream and sinusoid shaping for the Q stream. The MSK detection scheme is similar to the O-QPSK scheme, with an additional circuit for shaping the digital stream as sinusoid and co-sinusoid compliant with transmissions. GMSK (Murota and Hirade, 1981), another derivative of O-QPSK, uses Gaussian pulse shapes.

Yet another QPSK derivative, known as the $\pi/4$-QPSK scheme (Baker, 1962; Xiong, 1994) has been used in both terrestrial and mobile satellite communication. In this scheme, carrier phase transitions are limited to $\pm\pi/4$ and $\pm3\pi/4$ and therefore envelope variations caused by band limiting and subsequent spectral regrowth due to non-linear amplification are considerably reduced (see next paragraph). This scheme allows differential encoding and differential decoding, which has the advantage of removing the phase ambiguity inherent in coherent demodulation. By shaping the digital pulses as sinusoid and offsetting the transition in the I and Q channels, improved spectral restoration is achieved, which gives an improved performance when operating with saturated non-linear amplifiers. This type of modulation scheme is known as $\pi/4$-controlled transition phase shift keying (CTPSK) (Feher, 1991).



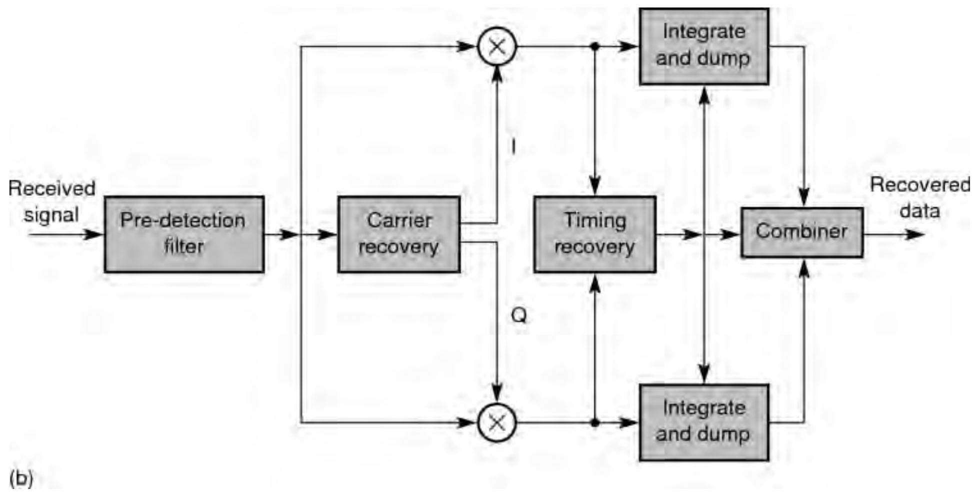


Figure 4.5 (a) A QPSK modulator and (b) a QPSK demodulator

Since modulation translates the baseband spectrum to an intermediate frequency (IF) that is up-converted to RF before transmission, several modulation schemes use a shaped baseband digital signal to minimize side lobes. CPM schemes use smooth changes in phase, which result in lower side lobes; these schemes are also power efficient (Sundberg, 1986) and are therefore a candidate for MSS. Depending on the pulse shape used to vary the phase, the modulation index (h) that determines the amplitude and size of symbol (M), a variety of CPM signals can be obtained. Examples of CPM schemes are continuous phase frequency shift keying (CPFSK), MSK, GMSK and tamed frequency modulation (TFM) (De Jaeger and Dekker, 1978). It has already been mentioned that the MSK scheme is similar to the O-QPSK scheme, with an additional circuit for shaping the digital stream as a sinusoid to smoothen the wave in order to minimize the effects of non-linearities. Similarly GMSK another derivative of O-QPSK, uses Gaussian pulse shapes.

In all well-designed schemes, a filter is placed at the modulator output to suppress side lobes to reduce the occupied bandwidth. Filter roll-off is a trade-off between occupied bandwidth and inter-symbol interference. However, the purpose of the filter may be defeated when spectral regrowth occurs subsequently elsewhere in the system, for example due to band limiting or by system non-linearities in the presence of amplitude fluctuations. Filtered PSK has some amplitude variation and therefore non-linear class-C amplifiers and filtering in mobile earth stations cause spectral regrowth in the transmitted spectrum. O-QPSK has lower amplitude variations and consequently lower spectral regrowth. Figure 4.6 shows the spectrum of an O-QPSK signal before and after transmit non-linearity. The $\pi/4$-QPSK modulation scheme offers good spectral efficiency, and low spectral regrowth when passed through non-linear amplifiers with a capability to use differential detection (Liu and Feher, 1991).

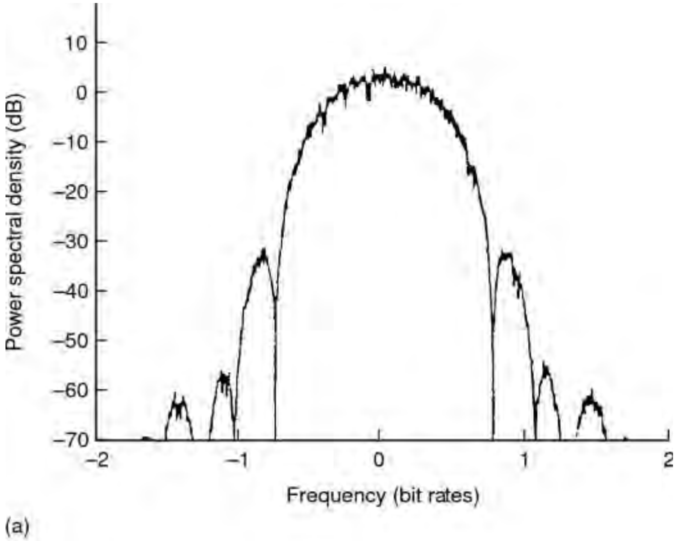
Multilevel frequency shift keying (M-FSK) and its derivatives, such as CPFSK, also offer good spectral and power efficiency (for a review of CPFSK, see Sundberg, 1986). In the M-FSK scheme, sidelobes are generated by discrete changes in frequency at symbol transition; in CPFSK, such changes are smoothed to accomplish reduced side-lobe levels. MSK can be considered as a special case of binary CPFSK.

Due to a shortage of spectrum in the L band, extensive research is in progress to improve spectral efficiency, while retaining robustness in the presence of channel impairments. Inmarsat's Broadband Global Area Network (BGAN) system uses 16-QAM in conjunction with turbo-code (Feldman and Ramana, 1999). The scheme is discussed in some detail later in this section.

The QAM scheme is adversely affected by system non-linearities; in this respect, amplitude phase shift keying (APSK) in combination with pre-distortion at the transmitter provides advantage; the APSK scheme is used in the digital video broadcasting-second generation (DVB-S2, see Chapter 8) standard in a dense arrangement of 16-APSK and 32-APSK to maximize spectral efficiencies. The DVB-S2 standard has been extended to incorporate mobility; since the standard utilizes an adaptable modulation-coding scheme, there is provision for the modulation level to be readjusted to an appropriate lower level if such a high modulation level is not sustainable on any link.

4.2.4 Performance Comparison of Conventional Digital Modulation Schemes

A modulation scheme should provide the desired BER at acceptable energy per bit by noise power spectral density (PSD) ratio (E_b/N_o) and spectrum requirements, leading to the classical power–spectrum trade-off. In general, spectrally efficient modulation schemes are more sensitive to noise and hence require larger E_b/N_o. Figure 4.7 shows constellation diagrams of BPSK, QPSK, 8-PSK and 16-QAM schemes. In an APSK scheme the permissible phases are arranged in concentric circular rings at discrete angular distances to each other. As the distance between the permissible states reduces, the signal becomes more susceptible to noise. Noise sources include modulator/demodulator circuit imperfections, adjacent channel interference, band-limiting effects of filters in transmit/receive sections, thermal noise, intermodulation noise, co-channel intrasystem interference caused by frequency reuse, inter-system interference, phase noise and multipath fading.



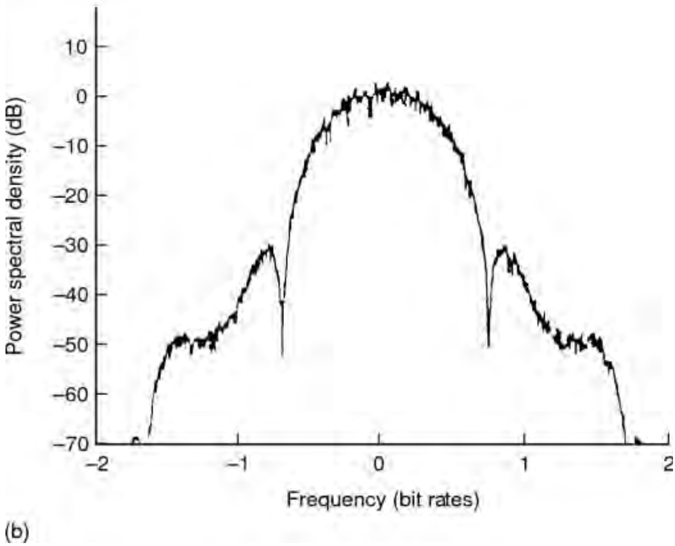


Figure 4.6 Spectrum of an OQPSK (Offset Quadrature Phase Shift Keying) signal before and after a non-linearity such as caused by a power amplifier. (Source: Lodge *et al*. 1987. © 1987 IEEE. Reproduced with permission.)

The PSDs of a few digital modulation schemes used in MSS, namely BPSK, MSK and QPSK, are compared in Figure 4.8.

A number of useful parameters are listed in Table 4.1. Parameters include various types of bandwidths, (E_b/N_o) for BER of 10^{-5}, implementation complexity and immunity to non-linearity. Table 4.2 summarizes the PSD and BER versus E_b/N_o of various modulation schemes, and Table 4.3 lists examples of modulation and coding schemes commonly used in MSS.

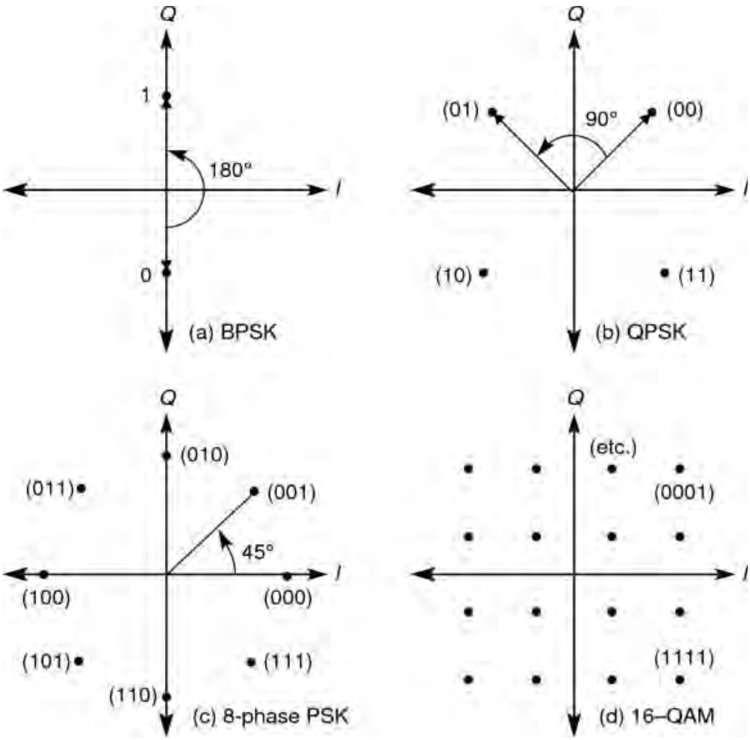


Figure 4.7 Constellation diagram of: (a) binary-phase shift keying, (b) quadrature phase shift keying, (c) eight-phase shift keying and (d) 16-quadrature amplitude keying

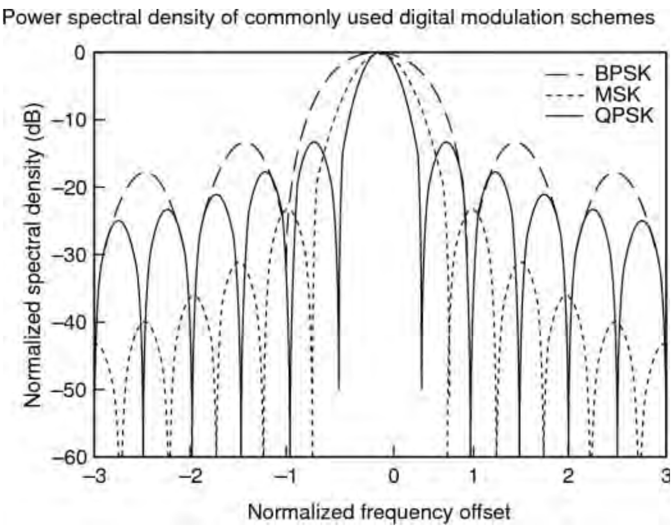


Figure 4.8 A comparison of normalized power spectral density of BPSK, MSK and QPSK. (Source: Graphics AR.)

Table 4.1 A comparison of common modulation schemes

Modulation	99% Power bandwidth (Hz/bits/s)	Null-Null bandwidth (Hz/bits/s)	Half-power bandwidth (Hz/bits/s)	Nyquist bandwidth PSK schemes ($R/\log_2 M$; R = bit rate; M = number of phases)	Eb/No dB-Hz (BER = 10^{-5})	Immunity to non-linearity	Implementation complexity
BPSK	20.56	2.0	0.88	R	9.6	Poor	Very good
QPSK	10.28	1.0	0.44	R/2	9.6	Fair	Good
OQPSK and (π/4) PSK	10.28	1.0	0.44	R/2	9.6	Good	Fair
MSK	1.18	1.5	0.59	–	9.6	Very good	Poor

(Data sources: Xiong, 1994; Amoroso F, 1980.)

Table 4.2 Useful parameters of common modulation schemes

Modulation	Single-sided power spectral density centred at carrier frequency	Null-null bandwidth	Bit error rate in Gaussian noise	Comments
M-ary PSK	$P_sT_s/2\,[\sin \pi fT_s/\pi fT_s]^2$ P_s = carrier power, T_s = symbol duration.	f_b/N	When number of symbols, M, is large, $\mathrm{erfc}[NE_b/\eta \sin^2\pi/M]^{1/2}$ BPSK and QPSK, $\frac{1}{2}\,\mathrm{erfc}(Eb/\eta)^{1/2}$	$M = 2^N$, denotes number of phases; N = number of bits/symbols; M = 2 for BPSK, M = 4 for QPSK; f_b = bit rate
QAM	Same as M-ary PSK	For 16 QAM; $f_b/4$	For 16 QAM; $2\,\mathrm{erfc}\,[0.4E_b/\eta]^{1/2}$	–
M-FSK (M signals orthogonal)	At each discrete frequency; Impulse + BPSK PSD, that is $P_sT_b/2\,[\sin \pi fT_b/\pi fT_b]^2$; where T_b is the bit period.	Mf_b/N	$[(M-1)/2]\,\mathrm{erfc}\,[NE_b/2\eta]^{1/2}$	–
MSK	$8P_sT_b/\pi^2[\cos 2\pi fT_b/1-(4fT_b)^2]^2$	f_b/N	$\frac{1}{2}\,\mathrm{erfc}(Eb/\eta)^{1/2}$	–

Table 4.3 Examples of modulation and coding schemes used in various MSS systems

System	Transmission rate (kbps)	Modulation	Coding	RF bandwidth (kHz)
Inmarsat-A SCPC (historic interest)	Analogue	FM	Not applicable	25 and 50
Inmarsat-Aero SCPC (Pre-BGAN)	21 (original)/10.5 (evolved)	A-QPSK	½ rate convolution FEC; constraint length = 7; 8-level soft decision Viterbi decoder	17.5/10
Inmarsat-B				
Forward signalling (pre-BGAN system)	6	BPSK	½ rate convolution FEC; constraint length = 7; 8-level soft decision Viterbi decoder	10
Circuit mode voice: SCPC	24	O-QPSK	½ or ¾ rate convolution FEC; constraint length = 7; 8-level soft decision Viterbi decoder	20
Inmarsat-C				
Forward link	1.2	BPSK	½ rate convolution coding with interleaving	10
Return link (Message)	1.2	BPSK		5
Inmarsat – BGAN				
Forward link	16 QAM and QPSK			
Return link	16 QAM and $\pi/4$ QPSK			
MSAT	6.75	QPSK, 60% cosine roll-off	Rate ½ and rate ¾ punctured code (K = 7); Viterbi decoding	6.0
Globalstar	–	Spread spectrum/ QPSK	Forward: Rate ½ (voice) Return: Rate 1/3 (voice)	1250
Iridium				
Forward link	50 (burst)	QPSK	Rate 1/3 (voice)	41.67
Return link	50 (burst)	QPSK	Rate 1/3 (voice)	41.67
Orbcom				
Forward link	4.8	DPSK	–	25–50
Return link	2.4	Symmetric DPSK	–	2.5

Because of their flexibility and cost-effectiveness, computer simulations are widely used for performance evaluation. As an illustration, consider a simulation study conducted to compare candidate modulation schemes DMSK, DOQPSK and BPSK-TTIB for a land mobile satellite system (Lodge, Moher and Crozier, 1987). Here, DOQPSK with non-redundant single error correction (SEC) out-performed other schemes in robustness to Ricean fading and amplitude compression by a mobile earth station power amplifier. Figure 4.9 demonstrates the error probability versus energy per bit to noise ratio for the scheme with Ricean fading at various K factors in the presence of thermal noise. The K factor is defined as

$$K = 10\log_{10}(P_m/P_d) \tag{4.2}$$

where P_m = average multipath power and P_d = direct path power.

Figures 4.10 and 4.11 illustrate the simulation model, the spectrum of the Rayleigh process and the transfer characteristics of the mobile terminal.

BGAN system was initially designed for the land portable segment of MSS, and as such, optimized for static radio links; a computer simulation was carried out to assess its

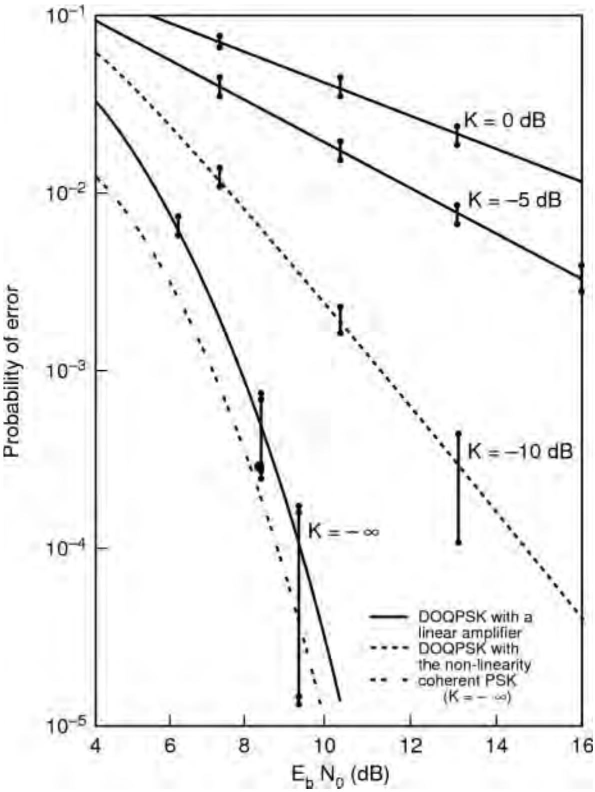


Figure 4.9 Probability of bit error versus energy per bit to noise ratio for DOQPSK with SEC over a Ricean channel in presence of thermal noise. (Source: Lodge *et al.*, 1987. © 1987 IEEE. Reproduced with permission.)

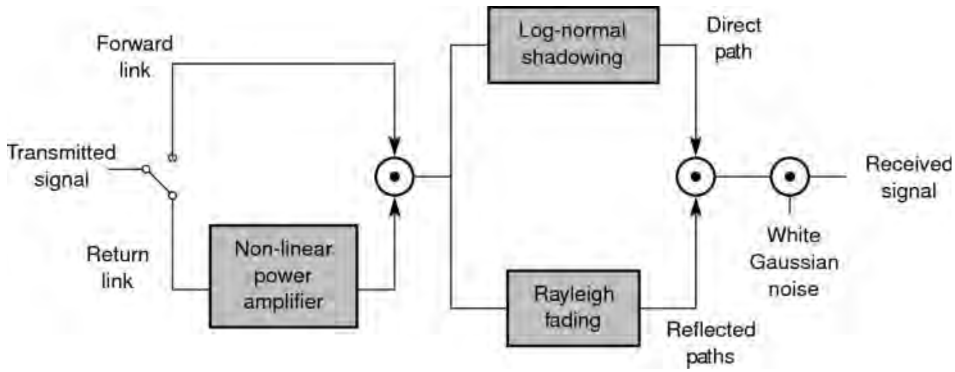


Figure 4.10 A propagation model used for performance comparison of various modulation schemes. (Source: Ball, 1982. © 1982 IEEE. Reproduced with permission.)

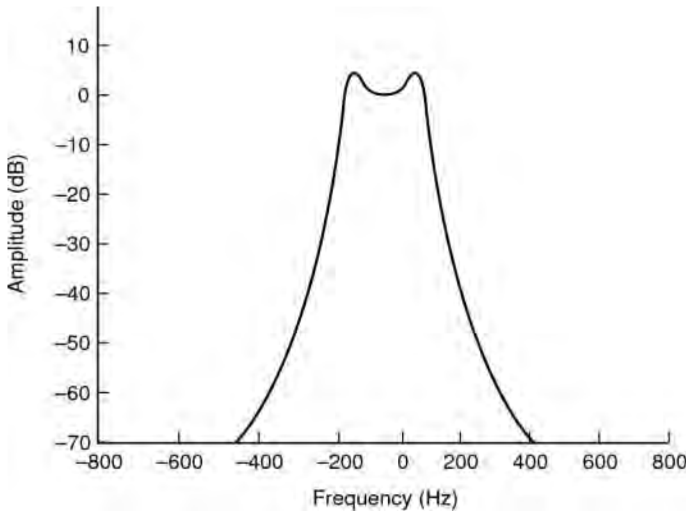


Figure 4.11 Spectrum of the Rayleigh process used in simulation. (Source: Ball, 1982. © 1982 IEEE. Reproduced with permission.)

performance on mobile platforms (Richharia, Trachtman and Fines, 2005). The primary goal was to maximize the reuse of the baseline BGAN technology with backward compatibility and allow reuse of the pre-BGAN antenna/front end at the mobile UTs to facilitate retrofitting of the aeronautical UTs in particular that would avoid extensive refurbishment, licensing and schedule penalty to the airlines. For the maritime UTs the pre-BGAN antenna/front-end specifications were retained but the choice of retrofitting was left to the UT manufacturers. BGAN modulator/demodulator (modem), coder/decoder (codec), channel impairments, propagation channels for each of the three types of operational environment were modelled and end-end physical layer simulation carried out, based on which, the baseline BGAN system design was modified ensuring backward compatibility.

The simulations demonstrated that the performance would deteriorate progressively at low elevation angle ($<\sim 20°$) in all the systems due to an increase in multipath noise necessitating an adjustment to the link margin in proportion to elevation; furthermore the performance of the aeronautical system degraded over calm to moderately active sea surface (but not over land) due to channel dispersion necessitating an equalizer for cancelling the echoes at the UT to ensure high link reliability under all environmental conditions and air corridors; and it was necessary to pre-compensate transmission frequencies from aircraft for self-induced Doppler, as applied in the baseline system. For land mobile UTs, particular care was necessary in carrier synchronization for high speed vehicles and after prolonged shadowing events. In all the cases a nominal direct line of sight path was assumed to be present.

Determination of a suitable fading margin in the presence of propagation impairments is vital for MSS link design. Analytical expressions for BPSK and QPSK were derived to estimate the fade margin at a specified BER and percentage coverage, assuming ideal carrier phase tracking (Davrian, 1985). Figure 4.12 shows the numerically computed results as a function of the K factor, defined here by the authors differently (i.e., reciprocal of K number

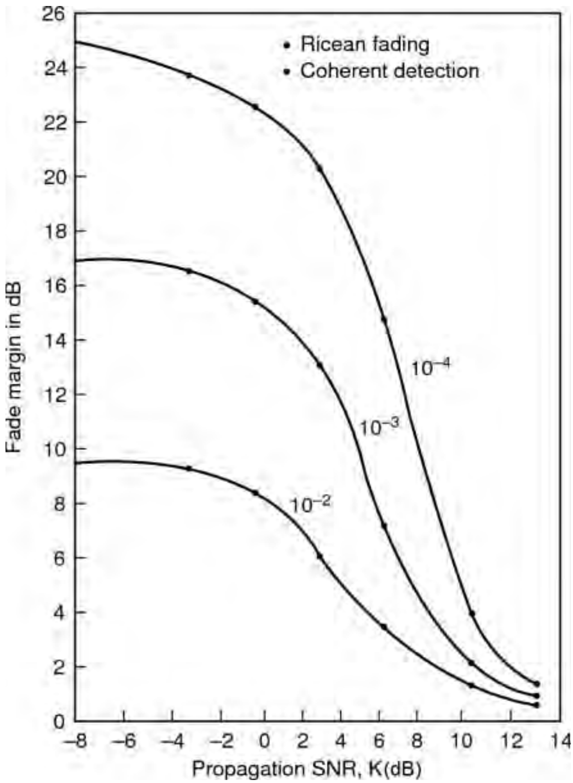


Figure 4.12 Fade margin as a function of K for BPSK and QPSK BER = 10^{-2}, 10^{-3} and 10^{-3}. (Source: Davrian, 1985. © 1985 IEEE. Reproduced with permission.)

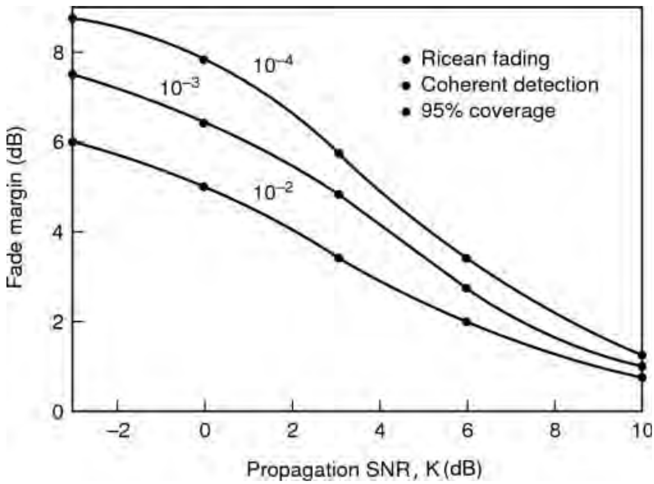


Figure 4.13 Fade margin for 95% coverage in a Ricean channel as a function of K. (Source: Davrian, 1985. © 1985 IEEE. Reproduced with permission.)

the previously defined K number) to Equation 4.2, as

$$K = 10\log_{10}(P_d/P_m) \tag{4.3}$$

where P_m = average multipath power and P_d = direct path power.

Figure 4.13 depicts the fade margin for 95% coverage as a function of K.

Modems are implemented through digital signal processing (DSP) chips that enable reconfiguration of modulation and coding schemes by software. Thus in a multimode terminal modulation-coding can be altered according to the desired radio interface. This implementation also allows investigation and comparison of various combinations of modulation and coding schemes under similar conditions. For example, the performance of a pilot symbol-assisted carrier phase recovery scheme was compared with a differential PSK scheme for application to a helicopter-mounted mobile, which undergoes deep fades synchronized to helicopter blade rotation (Cowley, Lavenant and Zhang, 1997). Flexible implementation of modems and codec is central to the concept of software radio.

4.2.5 Coded Orthogonal Frequency Division Multiplexing (COFDM) Modulation Systems

We have observed that traditional digital modulation schemes suffer in quality in a mobile environment, which worsens as transmission bit rate is increased. A multipath resistant transmission scheme called the COFDM (or simply OFDM) modulation system is therefore used for high quality terrestrial and satellite broadcasts to mobiles (Pommier and Wu, 1986; Alard and Lassalle, 1988; Shelswell, 1995).

As the name implies, COFDM uses coded data, which is frequency division multiplexed with carrier frequencies orthogonal to each other. Incoming data are segmented such that only a fraction of information is contained in each segment, each of which is then transmitted

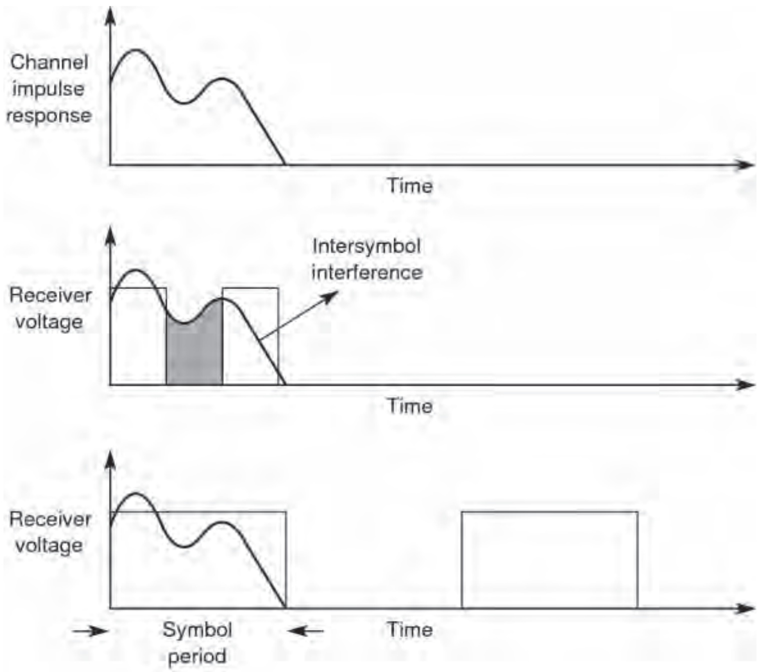


Figure 4.14 Reduction in inter-symbol interference with an increase in the symbol period for a given channel impulse response. A smaller percentage of symbols are affected as the symbol period is increased

on separate RF carriers using a suitable modulation/coding scheme. The reduction in bit rate per carrier increases the symbol period, with a consequent reduction in inter-symbol interference, as shown in Figure 4.14.

In theory, any suitable modulation scheme can be used to transmit the carrier. For example, a $\pi/4$-DQPSK modulation scheme was used in the Eureka 147 project (a European collaborative project – see Chapter 12). Data was differentially encoded at the transmitter with the phase of the reference signal increased by $\pi/4$ for each symbol period, and differentially demodulated at the receiver.

In other words, if the band occupied by the multiplexed carriers is made greater than the coherence bandwidth of the propagation channel (see Section 3.3.3.2), then the probability of all the carriers fading simultaneously, is reduced. To obtain an optimal performance, it is essential that the frequency dispersion characteristics of the propagation channel are well understood. It is necessary to interleave and code the incoming data stream to obtain full benefit of the scheme; where the selection of coding and interleaving depth must also be matched to mitigate the degradation caused by propagation.

To maximize the use of spectrum, the frequencies are made linearly independent, that is orthogonal to each other. Orthogonal carriers permit a reduction in carrier separation and hence the occupied bandwidth. Nevertheless, as COFDM is a wideband system therefore for a radio broadcast service several programmes must be combined to derive the maximum spectrum utilization (see Chapter 12). Mathematically, two signals Ψ_p and Ψ_q are orthogonal when,

$$\int_a^b \psi_p(t)\psi_q^*(t)\,\mathrm{dt} = \mathrm{K} \quad \text{for p = q}$$
$$= 0 \quad \text{for p} \neq \text{q} \tag{4.4}$$

where the symbol * indicates a complex conjugate.

An example of an orthogonal series is sin (mx) for m = 1, 2 … over a period $-\pi$ to $+\pi$. Applied to COFDM, the signals are orthogonal when carrier spacing is a multiple of $1/\tau$, where τ is the symbol period. The receiver consists of a bank of demodulators and a frequency translator, which brings each carrier to the baseband, followed by integration over a symbol period. When carriers are orthogonal integration reduces contributions from unwanted carriers to zero.

Figure 4.15 shows conceptual building units of a COFDM transmitter. Serial input data are converted into a parallel stream each of which modulates a carrier from an orthogonal set, which are finally summed to constitute a composite COFDM signal.

In practice, the boxed part is implemented in software by fast Fourier transform (FFT) DSP chips. Mathematically, the composite signal can be expressed as

$$\mathrm{S_s(t)} = \frac{1}{N}\sum_{n=0}^{N-1} \mathrm{A_n(t)}\, \mathrm{e}^{\mathrm{j}\,[\omega_\mathrm{n}\mathrm{t}+\Phi_\mathrm{n}(\mathrm{t})]} \tag{4.5}$$

where $\mathrm{A_n(t)}$ and $\phi_\mathrm{n}(\mathrm{t})$ are the amplitude and phase of the nth carrier and $\omega_\mathrm{n} = \omega_0 + n\Delta\omega$, ω_0 being the angular frequency of carrier 1.

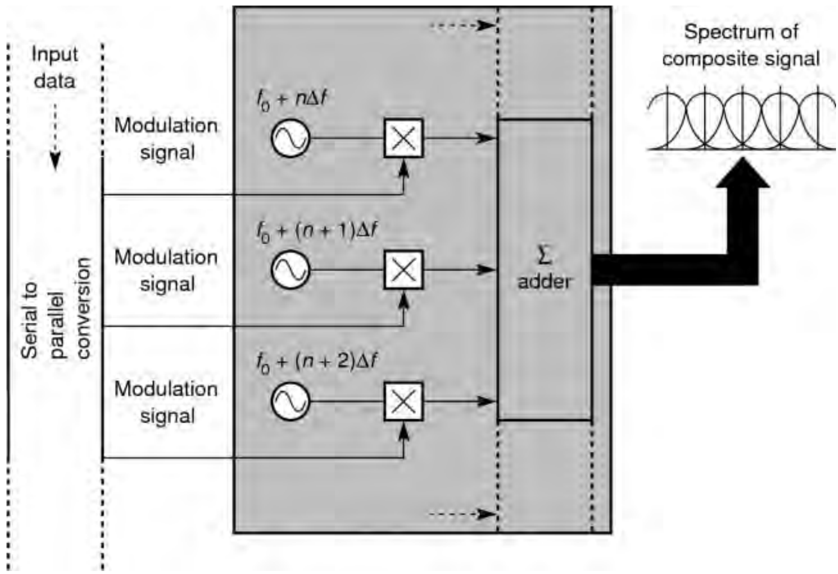


Figure 4.15 Conceptual building units of a coded OFDM transmitter (Shelswell, 1995). (Source: Shelswell, 1995. Reproduced with permission of IET.)

The digital equivalent of the signal can be represented as,

$$S_s(kT) = \frac{1}{N}\sum_{n=0}^{N-1} A_n e^{j}\phi_n e^{j(n\Delta\omega)kT} \tag{4.6}$$

For $\Delta f = 1/NT = \tau$, Equation 4.6 becomes equivalent to an inverse Fourier transform where s(kT) is the time domain representation of the signal. A DSP chip performs an inverse transform on the incoming signal to give sampled time domain signals that are converted to an analogue signal for transmission. To facilitate FFT, the number of carriers N is made equal to 2^n.

A reverse process is applied at the receiver. The received signal is synchronized, digitized and Fourier-transformed to the frequency domain to provide individual carriers with the desired amplitude and phase. One of the practical problems is that of carrier synchronization. One solution is to use a coarse synchronization, followed by a precise synchronization. A coarse synchronization can be achieved by switching off all carriers regularly for a short duration; an amplitude detector can then be used to provide a synchronization pulse when carriers are switched on. Fine synchronization can be achieved by transmitting a reference signal, which can be correlated at the receiver with a replica to achieve an accurate synchronization in time and frequency. Synchronization can be made robust at the expense of power efficiency by introducing guard intervals around each symbol, during which trivial or redundant data is transmitted. Guard intervals offer an additional advantage of reducing the effects of echo or co-channel interference, when the delay of interfering signals is small compared to the symbol period. A guard band of the order of 25% of the symbol period has been observed to be a reasonable compromise for terrestrial environments in the UK. The co-channel interference rejection property enables improvement to the spectrum efficiency of the network.

We have observed that the orthogonal frequency division multiplexing (OFDM) system provides frequency selective multipath resistance by frequency diversity, which, in effect is obtained by increasing the bandwidth of the signal. Thus, it would appear that the system is less efficient than a single carrier system using the same base modulation scheme. For a COFDM supporting only a single programme, this observation may be true; however, when several programmes share the system, spectral advantages become distinct. For example, considering a bandwidth of 1.5 MHz with a base modulation scheme of QPSK and accounting for inefficiencies introduced by imperfect synchronization, coding and guard interval, a transmission rate of ~ 1 bit/Hz or about 1.5 Mb/s is achievable. The band can provide around six stereophonic broadcast channels. With a COFDM network, the same spectrum can be reused several times in a given geographical area to provide the same block of programmes; by contrast, if the programmes were transmitted over the conventional digital schemes to service the same region, the spectrum requirements would be higher due to a lower frequency reuse efficiency.

4.2.6 Spread Spectrum Modulation

Spread spectrum has been used extensively in military communications for a considerable time. Other applications include ranging and radar. A modulated signal is spread in spectrum using a pseudo-random code and retrieved at the receiver with a correlation receiver that uses the same code as the transmitted. Spread spectrum modulation is used with a partner accessing technique called code division multiple access, discussed later in the chapter. Use of distinct uncorrelated codes allows several users to utilize a channel with minimal effect on each other. Users can receive the signal only if the key to the transmitted pseudo-random code sequence is available and hence the system has inherent inbuilt security.

The scheme has received considerable attention for application to mobile communications. In the first and second generation terrestrial systems developed in the late 1980s and early 1990s, schemes such as TDMA were superior in terms of available technology. However, in the 1990s technology advancements resulted in the application of spread spectrum into narrowband and wideband CDMA terrestrial mobile standards. The trend has extended to MSS systems.

Classically, there are two methods for implementing spread spectrum modulation:

- direct sequence scheme;
- frequency-hopped scheme.

In each case, a pseudo-random sequence forms the basis of spread spectrum modulation systems. Figure 4.16 summarizes the main characteristics of pseudo-random sequences – auto-correlation function and PSD. The envelope of the PSD has a magnitude of $(\sin x/x)^2$ comprising a line spectrum spaced at $(1/p)t_0$ where p is the number of bits in the pseudo-random sequence and t_0 is a period of one bit (or a chip). As the chip sequence is not truly random, there is a small component at the carrier frequency. The null of the PSD occurs at $\pm 1/t_0$ (code rate).

Figure 4.17(a) and (b) illustrates the principle of the direct sequence spread spectrum transmit-receive scheme. The message, m(t) is modulated using any standard scheme such as QPSK; the modulated signal is then spread by a spreading function, g(t), up-converted, amplified and transmitted. At the receiver, the down-converted signal is correlated with a replica of the transmitted code, g(t). A correlation peak is obtained when codes match; the resultant signal is band pass filtered and demodulated to obtain the data stream m(t). The transmitted signals are coded to improve performance (see Section 4.3 for channel coding).

In a frequency-hopped spread spectrum system, the transmit frequency is altered in a pseudo-random sequence. A pseudo-random chip generator changes the frequency of a synthesizer. At a receiver the signal is down-converted by a synthesizer synchronized to the transmitted signal frequency, demodulated/decoded to obtain the transmitted bit stream. The technique is illustrated in Figure 4.18(a) and (b).

The property of a code to discriminate interfering signals is determined by the processing gain of the scheme, defined as

$$G_p = B_c/B_m \tag{4.7}$$

Also,

$$G_p = R_c/R_m \tag{4.8}$$

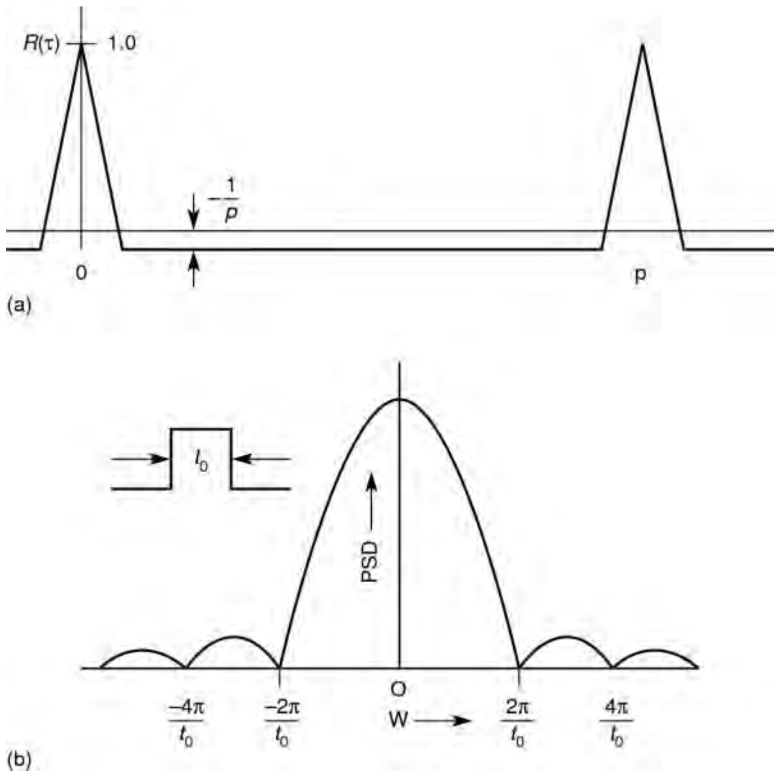


Figure 4.16 (a) Auto-correlation and (b) power spectral density of pseudo-random sequence

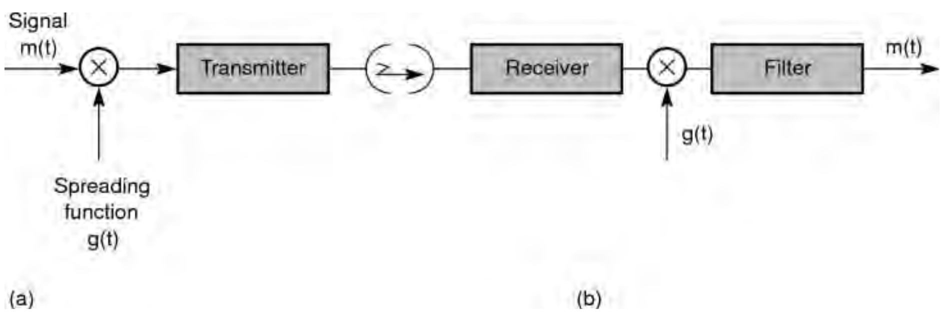


Figure 4.17 Principle of direct sequence spread spectrum scheme (a) transmitter and (b) receiver

where R_c = the chip rate, R_m = the message bit rate, B_c = the occupied channel bandwidth and B_m = the occupied message bandwidth.

As an example, when a BPSK/spread spectrum system is interfered by a single tone of power P_i, the interfering power is reduced by a factor $(P_i/2)G_p$. When the phase of the

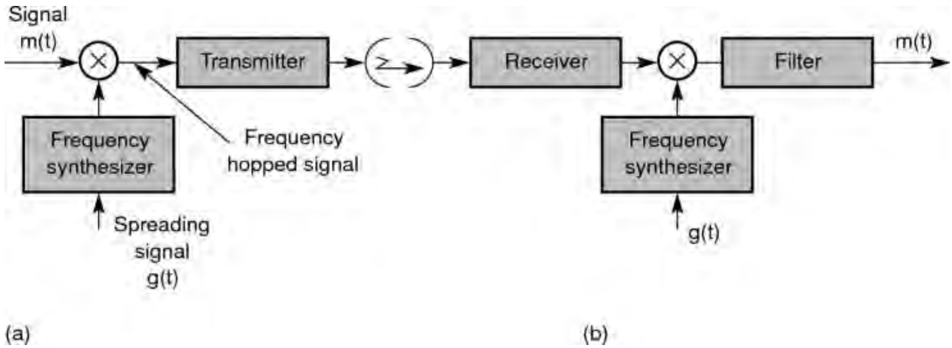


Figure 4.18 Principle of frequency-hopped spread spectrum scheme (a) transmitter and (b) receiver

interferer is taken as a random variable, the error rate is given as,

$$P_e = \tfrac{1}{2}\,\mathrm{erfc}\sqrt{P_s/((P_i/2)G_p)} \tag{4.9}$$

where P_s is the power of the wanted carrier.

In addition to carrier recovery and bit synchronization a spread spectrum demodulator must regenerate the code used for signal spreading. The process involves code acquisition, followed by tracking the code continuously.

4.3 Coding

MSS service links operate with small terminals in a fading environment and a limited satellite effective isotropic radiated power (EIRP) per channel amounting to a severely power-limited link (see a typical link budget in Section 3.4). A robust spectrally efficient modulation scheme alone cannot accomplish the desired link reliability. Invariably, all MSS systems incorporate a channel coding scheme matched to the radio channel and application. In this section, the salient features of coding applicable to MSS are reviewed. The interested reader consult the literature for an in-depth appreciation of this interesting field. In essence, channel coding introduces redundant bits such that BER improves relative to uncoded information for the same power-limited link conditions, the improvement being quantified as a coding gain.

There are two families of code used in MSS systems – *block code* and *convolution code*. Block codes operate on groups of bits organized as blocks, that is, information bits are assembled as blocks before coding (see Figure 4.19(a)). *Code rate* is a measure of the noise rejection strength of the code given as the ratio k/(k + m) where k represents the number of information bits and m the number of redundant bits; a low code rate such as half is more powerful, that is it provides a higher protection to the message bits.

Hamming distance is used to measure the effectiveness of block codes. It is a measure of the minimum number of bits by which two coded words may differ. In practice, linear algebraic codes are preferred due to simpler implementation. Cyclic codes provide algebraic structure amenable to decoding and are therefore common. Examples of codes used in practice include; the Hamming code, the Bose, Chaudhari and Hocquenghem (BCH) code, the Reed–Solomon (RS) code (a type of BCH code) and the Golay code.

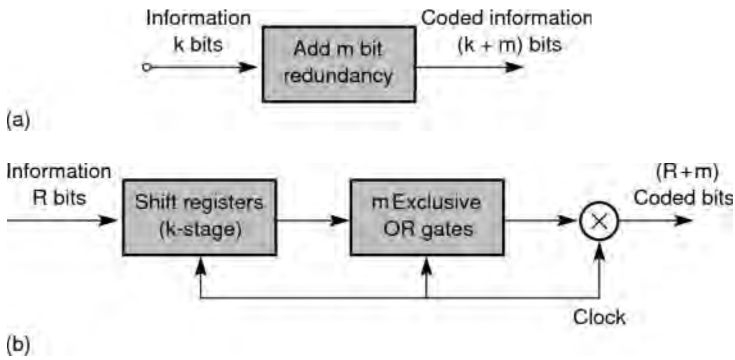


Figure 4.19 (a) The concept of coding and (b) the concept of a convolution coder

Convolution codes are formed by convolving information bits with the impulse response of a shift register encoder. Figure 4.19(b) depicts a conceptual diagram of a non-recursive convolution coder. Convolution codes use information bits stored in a memory and continuously produce coded bits. The constraint length of a convolution code defines the number of information bits – the code memory – that influence the encoder output. The error correcting property of a convolution code improves with an increase in constraint length; however, decoding complexity increases. In a convolution code, the equivalent of Hamming distance is the minimum distance, defined as the minimum number of bits in a code that must be altered to obtain another valid code word. Figure 4.20 (Yasuda, Kashiki and Hirata, 1984) shows the coding gain achievable as a function of bandwidth expansion and code memory at a BER of 10^{-6}, using classical non-systematic convolution codes. Coding gain increases by ~ 0.5 dB for each memory addition, up to about a memory of 6. The coding complexity increases exponentially as $v2^v$, where v is the code memory.

Convolution codes can be decoded by sequential or Viterbi decoding. Both techniques can use either hard or soft decision code. Hard decision decoding uses code words assembled on a bit-by-bit basis, which are each independently derived using a 'hard' decision (e.g. by a comparator). In soft decision decoding, the decoder operates directly on unquantized analogue demodulator output and therefore knowledge of the bit state is not lost prior to decoding and for this reason it outperforms hard decision decoding, though at the expense of complexity. The soft decision decoding rule is to produce a sequence $\{a_r\}$ that is closest, in terms of minimum squared Euclidean distance, to a set of coded sequence $\{C\}$. The sequence set $\{C\}$ is the specific set of code words that can be produced by the coder. The problem is defined more precisely as

$$|r_n - a_r|^2 = \text{Min}\ \Sigma |r_n - a_t|^2 \tag{4.10}$$

where $\{a_r\}$ lies within $\{C\}$ and $\{a_t\}$ is the transmitted sequence; $\{r_n\}$ is the received sequence.

A Viterbi decoder is a near-optimum decoding algorithm when the code generation follows the rule of a finite state machine. From a practical point of view, the Viterbi decoder has a further advantage in that the same decoder chip can be used for decoding various coding rates by a technique known as puncturing. Some coded bits from a coded sequence are deleted according to a predetermined performance pattern. As an example, starting from a

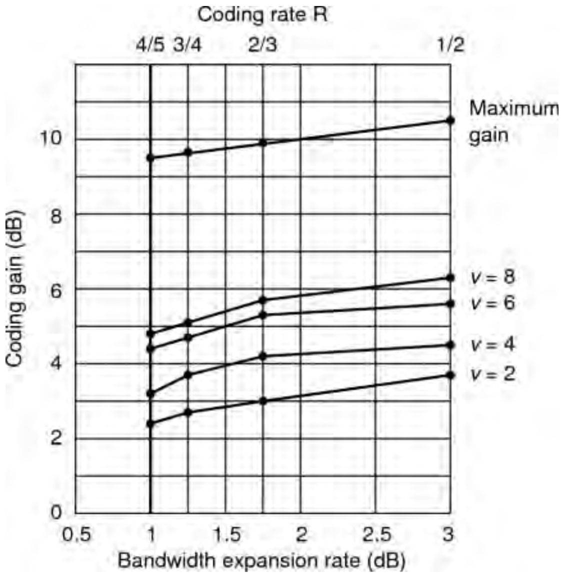


Figure 4.20 Coding gain achievable with three-bit quantization as a function of bandwidth expansion and code memory v at a bit error rate of 10^{-6} using classical non-systematic convolution codes. (Source: Yasuda *et al*, 1984. © 1984 IEEE. Reproduced with permission.)

code rate of half, a two-thirds rate code can be obtained by deleting the third bit; the matrix **P** for achieving the punctured code can be (read, right to left/top to bottom)

$$\mathbf{P} = {}^{\mathbf{11}}_{\mathbf{10}} \tag{4.11}$$

The performance of conventional coding techniques degrades in presence of error bursts lasting longer than the error correcting ability of the code. In MSS links signals fluctuate widely due to motion. Fluctuations also occur due to ionospheric and tropospheric scintillation, fading due to rain and/or cloud, etc. An effective countermeasure for correcting burst errors is *interleaving*. Information bits are dispersed in time such that consecutive bits of a message are far apart. Provided that the separation, called interleaving depth, is greater than the duration of noise bursts, the technique overcomes error bursts.

The RS code is often used for error burst corrections. The RS code operates on groups of s bits called symbols. A block of RS code comprises k information symbols and r parity symbols. Hence, a code word comprises (k + r) symbols. An RS code in conjunction with interleaving can enhance the burst error correction capability considerably. Moreover, decoding of RS codes is relatively simple and therefore RS codes are useful in MSS systems.

In addition to error bursts MSS links suffer from random errors due to impairments such as thermal noise. It is possible to cascade two or more codes – one for correcting random errors and the other for correction of error bursts. This type of coding arrangement, called concatenation, can provide high coding gain with moderate complexity; typically, a block code such as the RS code (outer code) is cascaded with a convolution code (inner code).

A coding family called turbo-code comprising parallel concatenated convolution codes has drawn considerable attention in recent years and therefore we will discuss such schemes in some detail (Berrou, Glavieux and Thitimajshima, 1993). Turbo-codes are decoded by an iterative process, to offer performance close to optimum (i.e. the Shannon limit) out performing convolution code with lower decoding complexity for the same code rate. Furthermore, they offer high coding gain on fading channels. Efficient multilevel modulation schemes in conjunction with turbo-codes offer a solution to effect wideband communication at acceptable satellite EIRP and bandwidth. A 16-QAM scheme with turbo-code is used in the Inmarsat network to provide up to 0.5 Mbps packet-switched service to a variety of terminals on air, land and sea (Trachtman and Hart, 1999; Richharia, Trachtman and Fines, 2005). Turbo codes (TCs) are considered as one of the current state of the art coding technique in addition to a family of code known as Low Density Parity Check (LDPC).

Figure 4.21 shows a basic rate 1/3 turbo-coder. Data stream d_k feeds directly in to a recursive systematic convolution coder C_1 and after interleaving in to another recursive systematic code (RSC) coder C_2, which is not necessarily identical to C_1. Transmitted bit stream comprises of symbol X_k and redundancies Y_{1k} and Y_{2k} and is therefore a rate 1/3 code or it may be punctured to give higher code rate.

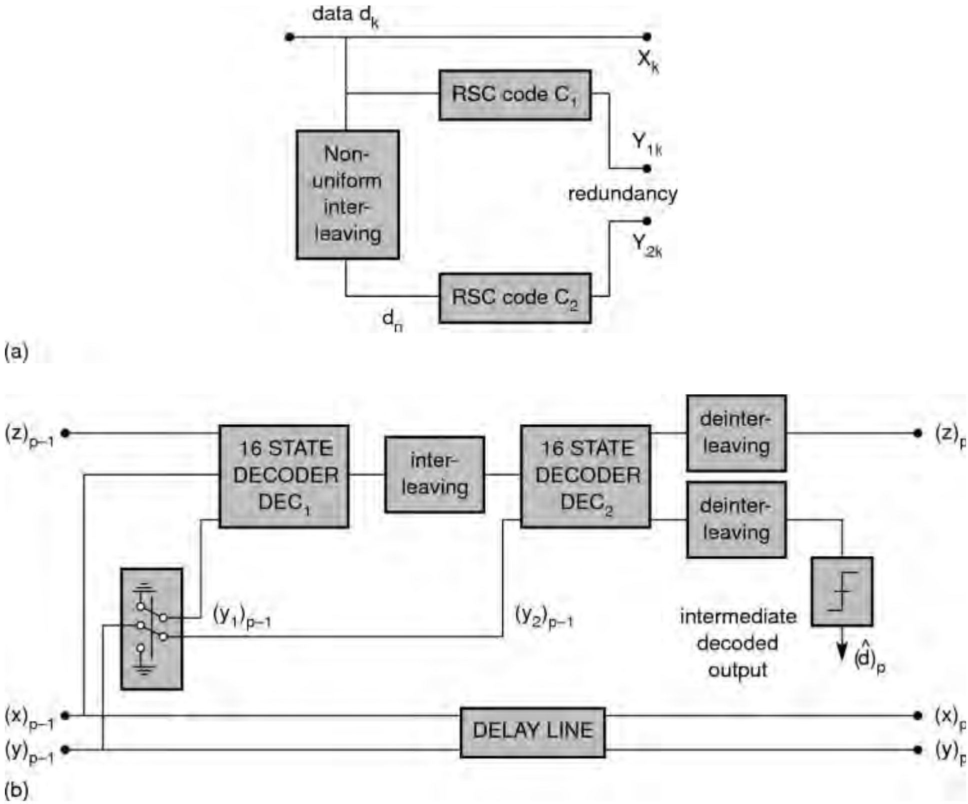


Figure 4.21 (a) A basic one-third rate turbo-coder. (b) A turbo-code decoder. (Both parts source: Berrou *et al*., 1993. © 1993 IEEE. Reproduced with permission.)

The turbo-code decoder comprises two cascaded elementary decoders – DEC_1 and DEC_2 (Berrou and Glavieux, 1996). The decoder DEC_1 is associated with encoder C_1 (Figure 4.21(b)) and DEC_2 with encoder C_2. The redundant bit Y_k is demultiplexed in the DEMUX/INSERTION block (see bottom left-hand side) and sent to DEC_1 when $Y_k = Y_{1k}$ and to DEC_2 when $Y_k = Y_{2k}$. In case redundant bit is not transmitted the DEMUX/INSERTION block sets the corresponding input to analogue zero. The logarithmic of likelihood ratio $\Lambda_1(d_k)$, defined in Equation 4.12, associated with each bit d_k is interleaved as in the transmitter and fed into DEC_2.

$$\Lambda_1(d_k) = \log \Pr\left[\left\{d_k = 1/\text{observation}\right\}/\left\{d_k = 0/\text{observation}\right\}\right]$$
$$\Pr\left[\left\{d_k = 1 \text{ or } 0/\text{observation}\right\} \text{ is a posteriori probability (APP) of bit } d_k\right] \tag{4.12}$$

The decoder DEC_1 uses a modified form of decoding scheme proposed by Bahl *et al.* (1974) and DEC_2 can use Viterbi decoding. The Bahl *et al.* algorithm is preferred to the Viterbi algorithm in DEC_1 because Viterbi algorithm can not provide *a posteriori* probability (APP) of each decoded bit whereas Bahl *et al.*'s algorithm does. However, the algorithm has to be modified for application to a RSC. The feedback loop gives additional redundant information Z_k, thereby improving the performance of the decoder. The name 'turbo' is given due to analogy of this feedback scheme to the principle of a turbo engine. The decoder architecture is modular comprising of a number of decoder units pipelined as illustrated in the Figure 4.21(b) (Le Goff, Glavieux and Berrou, 1994). The input of the pth module at time k consisted of demodulator outputs $(X_k)_{p-1}$ and $(Y_k)_{p-1}$ fed through a delay line and the feedback information $(Z_k)_{p-1}$.

Figure 4.22 shows E_b/N_o Versus BER performance of rate ½ encoder with memory 4, generators G1 = 37, G2 = 21 and parallel concatenation R1 = 2/3, R2 = 2/3 using a Monte Carlo method; the interleaver is a 256 × 256 matrix; a modified Bahl *et al.* algorithm has been used. BER reduces as number of iterations is increased. For 18 iterations, E_b/N_o is 0.7 dB, which is within 0.7 dB of the Shannon limit. The performance of turbo-code improves as interleaver size is increased. Thus the code tends to increase delay as its performance improves. The turbo-code can give several decibels higher coding gain than the standard K = 7 convolution code commonly used in MSS with moderate interleaver size and shows better performance in fading channels. Coding gains of 10 and 11 dB have been reported but this is at the expense of a reduction in data rate or increase in bandwidth and signal delay. BER of 10^{-6} at E_b/N_o as low as −0.6 using a code rate of 1/15 have been reported. Implementation of turbo-code is complex – but use of DSPs has made their implementation feasible. This code is now a part of the third generation terrestrial mobile system standard and is used in one of International Mobile Telecommunication-2000 (IMT-2000) satellite radio interfaces recommended by the ITU (see chapter 8).

Significant research activity is in progress to refine and improve the concept (Divsalar and Pollara, 1997). Trellis-coded modulation (TCM) has been merged with turbo-code to obtain a high coding gain with improved spectral efficiency (see the next section for TCM). The concept of parallel concatenated code has been extended to block codes and interleaving property of turbo-codes has been demonstrated to provide an effective countermeasure for frequency selective fades encountered in wideband transmissions in aeronautical satellite systems (Akhter, Rice and Rice, 1999). Turbo-code, in association with QAM constellations, has demonstrated a high gain in both Gaussian and Rayleigh channels and has outperformed

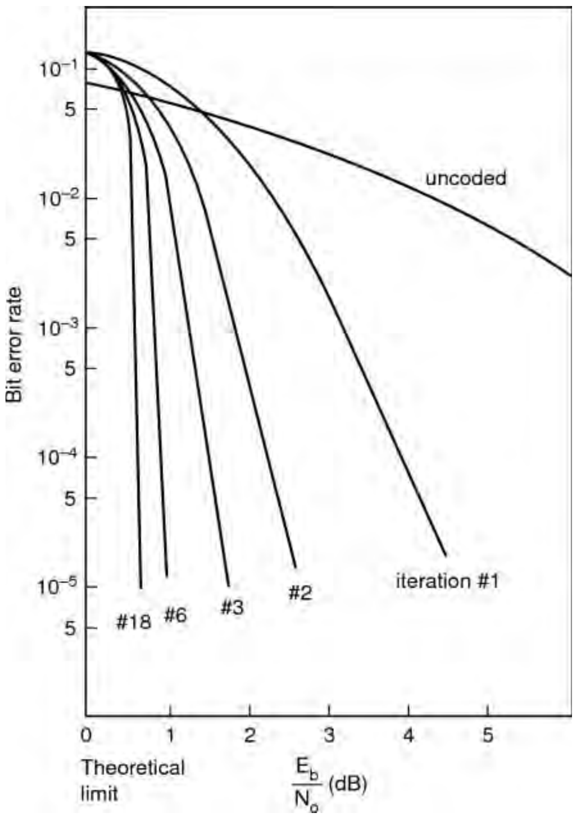
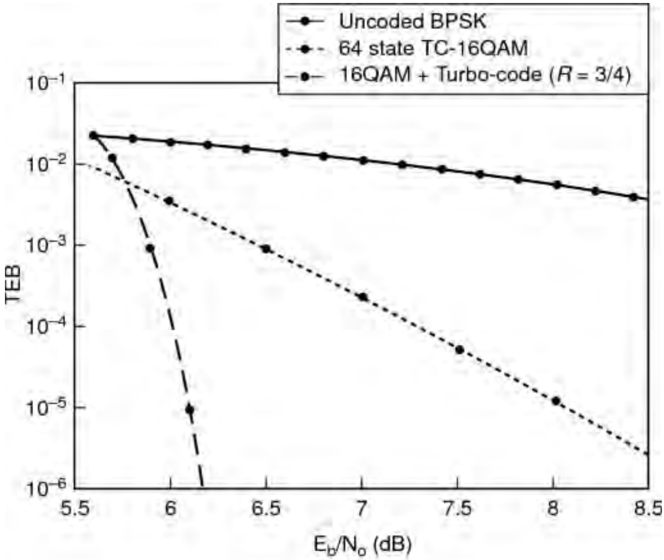


Figure 4.22 E_b/N_o versus BER performance of half rate encoder with memory 4, generators G1 = 37, G2 = 21 and parallel concatenation R1 = 2/3, R2 = 2/3 using a Monte Carlo method. (Source: Berrou *et al*., 1996. © 1996 IEEE. Reproduced with permission.)

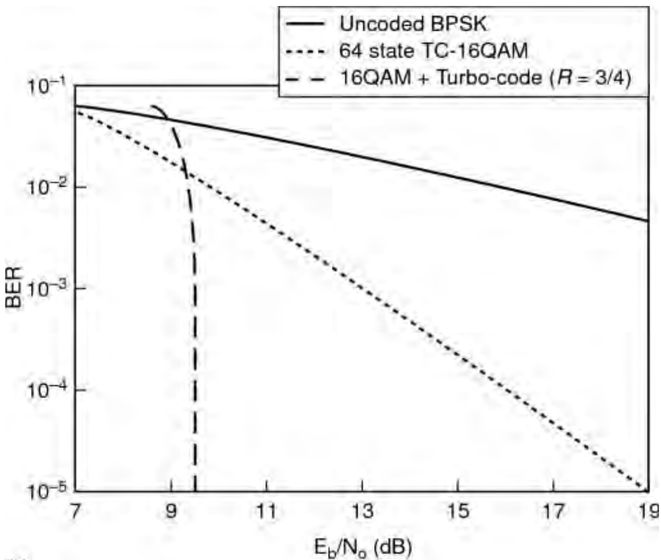
64-state TCM. Figure 4.23(a) and (b) (Le Goff, Glavieux and Berrou, 1994) compares BER of uncoded BPSK, 64-state TC-16-QAM and 16-QAM with three-quarter rate turbo-code in Gaussian and Rayleigh channels, respectively.

Table 4.4 summarizes the main results from this investigation for a Gaussian channel. Note the large coding gain and spectral efficiency possible with turbo-code.

When choosing a coding scheme, the error correcting capability of the code is matched to the type of channel to be encountered and application characteristics. A vast number of codes have been devised over the years; each with a unique feature that offers an advantage under a specific set of operating conditions. For example, codes can be optimized to operate in low or high Gaussian noise environments, in Ricean or Rayleigh fading conditions or combinations of channel environments. In adaptive methods, code rates can be adapted to suit link conditions. Such a scheme was demonstrated under the Advanced Communications Technologies and Services (ACTS) experiment programme sponsored by NASA. Code concatenation is used when different sets of channel conditions co-exist. Inmarsat's BGAN system utilizes an adaptable modulation scheme utilizing a combination of various modulation schemes in combination with an adaptable turbo-code scheme where the code rate is matched to the prevailing channel condition on a connection basis.



(a)



(b)

Figure 4.23 (a) BER of uncoded 8-PSK, 64 state TC-16QAM and 16 QAM with three-quarter rate turbo-code in a Gaussian channel. (b) BER of uncoded 8-PSK, 64 state TC-16QAM and 16 QAM with three-quarter rate turbo-code in a Rayleigh channel. (Both parts source: Le Goff *et al*., 1994. © 1994 IEEE. Reproduced with permission.)

Table 4.4 Coding gain over uncoded and 64-state TCM offered by various combinations of turbo-code rate and modulation schemes in Gaussian noise

Turbo-code rate	Modulation	Spectral efficiency	Coding gain (dB) at BER = $(10)^{-6}$ over uncoded modulation	Coding gain (dB) at 10^{-6} over 64-state TCM
½	16-QAM	2	6.0	2.4
⅔	8-PSK	2	5.5	1.9
¾	16-QAM	3	7.8	2.6
⅔	64-QAM	4	5.8	2.2

(Adapted from Le Goff *et al*., 1994.)

It may be worth recapitulating some interesting features (amongst others) before proceeding:

- Coding increases the bandwidth of a signal. For the same transmitted power, therefore, E_b/N_o of the received signal degrades after coding but coding gain offsets the disadvantage.
- A coding advantage can be achieved only when the carrier-to-noise ratio is adequate for carrier recovery and bit synchronization.
- Low rate codes perform better than high rate codes when BER is high, whereas high rate codes give better performance for medium/low bit rate conditions.
- Under conditions where link continuity breaks or link quality is extremely poor, automatic repeat request (ARQ) schemes (explained later), in conjunction with coding, give an effective solution.

Examples of block codes used in MSS include Hamming, BCH, RS and Golay. Convolution codes in common use have a constraint length of 7 and code rates down vary depending on the applications – typically being of the order of half or three quarters. Table 4.3 lists examples of codes used in mobile satellite systems.

Forward error correction (FEC) codes are used extensively in real-time applications. Time diversity for improving link reliability is often used combining FEC codes with interleaving or ARQ.

Valadon *et al*. (1999) studied a number of modulation/coding schemes for K_a band systems for delivering data rates in the range 32–384 kbps. The investigation included conventional convolution codes, TCM, several concatenated schemes and turbo-codes. Table 4.5 summarizes the candidates selected for further investigation in this particular study.

4.3.1 Trellis-Coded Modulation (TCM)

TCM combines coding and modulation techniques to achieve coding gain without a loss in bandwidth or power efficiency (Ungerboeck, 1987; Divsalar and Simon, 1988). TCM schemes use multilevel modulation, whose states are coded with a finite state encoder to maximize the distance of signal sequences in Euclidean signal space. Signals are decoded using a soft-decision maximum-likelihood sequence decoder. Coding gains of over 6 dB are feasible but these schemes are complex. The name, TCM, stems from the fact that it can be represented by a trellis diagram similar to that used for the binary convolution code.

Table 4.5 A comparison of candidate schemes for broadband personal communications at the K_a band

Coding scheme	E_b/N_o (dB)	Spectrum efficiency
Convolution code (½, 7)	6.8	0.99
⅔-8PSK TCM (16 states) + Reed Solomon (71, 55, 8)	5.7	1.51
⅔-8PSK TCM (16 states) + Reed Solomon (63, 55, 8)	6.1	1.71
¾-16QAM Trellis Coded Modulation (16 States) + Reed Solomon (63, 55, 8)	8.4	2.52
Convolution code (¾, 7) + Reed Solomon (237, 220, 8)	4.4	1.39
Convolution code (¾, 5) + Reed Solomon (237, 220, 8)	4.8	1.39
Turbo-code (rate ½, M = 2, 4 iteration) + Reed Solomon (63, 55, 8)	3.2	0.87

(Adapted from Valadon *et al*., 1999.)

Trellis branches represent coded multi-level modulation signals instead of convolution code symbols. TCM schemes are widely used in voice-band channels for high data rate modems. Note that in conventional coding schemes redundancy is obtained at the expense of bandwidth expansion, but in TCM bandwidth is not expanded; coding advantage is obtained by increasing the number of coded modulation levels without increasing signal bandwidth or power. Take, for example, a rate half convolution code. Due to the addition of one redundant bit per information bit, the transmission rate and hence bandwidth doubles to provide a coding gain of about 5 dB. The TCM scheme can potentially achieve the same gain without compromising on power or bandwidth efficiency. In the case of convolution code, the Hamming distance is maximized to increase resistance to noise, whereas in TCM the Euclidean distance is maximized.

Due to a combination of a convolution code with multiple-level coding, a rather complex demodulator is required. Use of block-coded modulation (BCM) simplifies demodulation (Imai and Hirakawa, 1977; Sayegh, 1986). Some of the well-known BCM decoding algorithms include Euclidean decoding and Berlekamp–Massey decoding. It is also possible to use Viterbi decoding by representing BCM as a trellis (Li, Iwanami and Ikeda, 1993).

The potential advantage offered by TCM has led to a number of experiments for application in mobile satellite communications. A two-thirds rate 16-state trellis-coded modem operating at a rate of 4.8 kbps was investigated for NASA's MSAT-X experimental programme (Divsalar and Simon, 1987). The TCM-coded symbols were further interleaved in 128-symbol blocks to improve performance in the presence of burst errors. A 100% raised cosine pulse shaping was used for matched filtering and Doppler shift estimation. In a Ricean fading channel with a Ricean factor of 10 dB, improvements of 3.1 and 1.6 dB in E_b/N_o ($BER = 10^{-3}$) over a conventional QPSK were observed with and without interleaving.

Similar studies undertaken elsewhere have investigated the performance of concatenated RS code as the outer code and TCM inner code in mobile satellite channels (Francon and Bousquet, 1997).

4.3.2 *Modulation and Coding Trends and Issues*

Deep-space communication systems deal with extremely low carrier-to-noise ratios; and since powerful codes are necessary to contend such conditions they rely on the best available modulation/coding techniques that set a useful benchmark for the industry (Vanelli-Coralli *et al*., 2007). At present such systems prefer TCs and LDPC codes. In addition, international research efforts in this discipline have set guidelines for activities in satellite communications systems. Examples of standardization for the MSS include DVB-S2/RCS + M that applies mainly to the K_u band; GMR-1 3GPP (Geo Mobile Radio Third Generation Partnership Program); and various satellite radio interface such as satellite radio interface-E targeting mobile communications in 1–3 GHz band. We will consider these initiatives as indicative of the current state of technology (see Chapter 8).

The GMR-1 3GPP air interface specification, released in 2009, as a satellite component of the third generation mobile communication system, is an extension of the Enhanced Data rate for GSM Evolution (EDGE). The satellite radio interface E (SRI-E) is an International Telecommunications Union (ITU) recommended interface that supports a multi-spot beam geostationary satellite system to provide world-wide IMT-2000 compatible multimedia services at bit rates up to 512 kbps to fixed, portable and mobile UTs, including the provision of higher data rates on specialized UTs through carrier aggregation. Digital Video Broadcast/Return channel via satellite + mobility (DVB/RCS + M), developed by the digital video broadcast project forum, is the mobile extension of DVB/RCS standard that supports the very small aperture technology (VSAT) segment.

Adaptable coding and modulation schemes that match the prevailing channel conditions of users have emerged as an attractive choice for wide band mobile communications. 16-QAM, 16-APSK and 32-APSK modulation represents the current preference for broadband communication on the basis of the standardization work in progress and evidenced by the techniques used in advanced operational MSS system.

Inmarsat's broadband system, known as Broadband Global Area Network, which provides bit rates up to about 492 kb/s is based on an adaptable modulation-coding scheme (see Chapter 11). Depending on the type of UT and the desired throughput, the modulation scheme utilize QPSK, 16-QAM and $\pi/4$-QPSK schemes at symbol rates of 8.4–151.2 Ks/s with up to 15 turbo-code steps and bandwidths ranging 21–189 kHz. The forward transmit channel units are able to apply different coding rates on an FEC block basis as directed by the system's Bearer Control Layer in order to adapt transmissions to different UT types sharing the forward bearer. The return bearers utilize burst transmissions in slots of either 5 or 20 ms duration, which are described in a return schedule dispatched on a forward bearer. These return schedules also describe the symbol rate and modulation to be used. Variable code rates involve puncturing of the turbo-code generated parity streams using one of a number of pre-defined puncturing matrices, such that the level of redundancy provided by the code is variable in 1 dB steps to enable adaptation to the channel conditions prevailing on the communications link.

A number of refinements are being considered for the application of the OFDM technique to satellite communication, particularly relevant to mobile satellite broadcasting. These include – improvements in channel estimation (e.g. by an optimal distribution of pilot symbols); synchronization issues in a multiple-access (orthogonal frequency division multiple access, OFDMA) receiver, network synchronization issues and fast resynchronization in single-frequency hybrid networks.

Salient features of TC and LDPC coding techniques and issues that are under investigation are summarized as follows (Vanelli-Coralli *et al*., 2007).

TC and LDPC codes provide near optimum performance when the block size is large (e.g. thousands of bits) and can be implemented on DSP chips at a reasonable cost. Both the codes utilize a simple encoder that can produce long code words with good distance properties. While the turbo-coder uses a combination of an interleaver and convolution encoders, LDPC coder generates a parity-check matrix that has a low density of ones in relation to the block size. Both the schemes utilize low-complexity iterative decoding to produce near-optimum performance. The choice therefore depends on the requirements addressing issues such as channel characteristics, target code rate and permissible encoding complexity; some general observations for thermal noise limited links are as follows:

- LDPC can lower the error floor compared to TC when block size is large. However, TC offers a lower complexity at very low coding rate; TC was therefore preferred in the 3GPP2 standard as well as digital video broadcasting-satellite services to hand-held (DVB-SH) and satellite earth stations and systems software-defined-radio (SES SDR) satellite multimedia broadcast standards.
- LDPC perform better than punctured TC at higher coding rates.
- TC can be generally encoded faster than LDPC codes, but there are exceptions.
- LDPC can achieve higher throughputs than TC with parallel decoding architectures.

The DVB-S2/RCS + M utilizes the concept of FEC application at upper layers, where erroneous packets are recovered by the use of erasure RS or LDPC codes, which add redundant packets to achieve very low error rate without the need of retransmissions (see next section).

One of the problems in using such powerful codes is that at such low carrier to noise ratio, the acquisition and synchronization of carrier becomes difficult. An iterative code-aided synchronization can potentially offer a solution – for example, the process of channel estimation can become a part of the recursive decoding. The turbo synchronization process can be aided either by utilizing data bits or by insertion of pilot symbols; the latter can provide an effective solution, particularly for short bursts.

When considering CDMA, code acquisition, tracking and frame synchronization can pose difficulties in low carrier-to-noise ratio and frequency uncertainties. Frame acquisition is central in capturing the pilot symbols embedded within each frame, which in turn are necessary in estimating other channel parameters.

In a mobile environment, random signal blockages, multipath fades and frequency uncertainty due to Doppler induced by UT motion all introduce problems in synchronization. Rapid synchronization following a blockage and a handover is desirable. Techniques that can mitigate this problem include transmission of pilot signals.

4.3.3 Automatic Repeat Request

From the propagation behaviour of MSS channels, we know that there may be a break in the radio connection link due to obstructions in the signal path, sometimes lasting for tens of seconds. Under such anomalous conditions, the FEC schemes are ineffective. ARQ schemes are used when high reliability in message transmission is required and when the associated delay is acceptable. Reliability in message delivery is further improved by storing the message and retrying transmissions at intervals until message delivery is acknowledged. In ARQ schemes, the receiver sends a request for retransmission on detection of an error; if necessary, the process is repeated until success is achieved. Note that in geostationary satellite systems a transmission delay of at least $\sim 1/2$ s is inevitable in implementing an ARQ protocol and therefore delay intolerant services are not suitable for ARQ.

A number of schemes are possible. In the *stop and wait* ARQ scheme, a message is transmitted only when the correct receipt of the previous message is confirmed. In the *Go-back N* ARQ scheme, message transmission is continued until a request for transmission is received, when all the blocks beginning from the block where the error was detected are retransmitted. In the *Selective-request* ARQ scheme, only the corrupted block is retransmitted. The latter scheme would seem appropriate for satellite communications as it is efficient in terms of radio resources.

ARQ schemes are used in store and forward MSS systems such as the Inmarsat-C system and the EutelTRACS system for message delivery (see Chapter 11).

4.4 Multiple Access Schemes

All three basic multiple access schemes – frequency division multiple access (FDMA), TDMA and CDMA – are used in MSS for *circuit-mode* transmissions or in *packet-mode* where the channel resource is shared such that the channel is sufficently loaded. A variety of accessing schemes – often called 'protocols' – are used for *packet-mode* transmissions, which are often characterized by bursty traffic. Such schemes may be classified as *reservation* and *contention* protocols. In a reservation protocol, a channel is reserved for the duration of communication and, in a contention protocol, users are granted access on a first-come-first-served basis. MSS systems combine a number of multiple accessing schemes to match conflicting requirements caused by different characteristics of service and other practical considerations. For example, a request for a channel assignment is sent effectively through a short data burst, whereas voice communication requires a continuous end-to-end connection. Thus, in this case, the requirement can be met by combining a random access protocol such as Aloha for requesting resources with a demand assigned (DA) scheme such as FDMA or CDMA that allocates a dedicated circuit for the call duration.

A multiple access scheme influences critical system parameters such as EIRP, radio spectrum, adaptability to network expansion and mobile terminals' complexity. An access scheme should ideally provide a good radio spectrum efficiency, a capability to manage variable temporal traffic loads, that is, diurnal, seasonal or event-driven, as well as general network growth, a graceful degradation in signal quality in case of an unexpected traffic surge, flexibility in managing a mix of traffic demands, resistance to multipath, shadowing, interference and technology capable of supporting low-cost UTs. Clearly, no single scheme can provide all these features. We will briefly review classical multiple access schemes in the next few paragraphs to refresh our knowledge, followed by a comparison of accessing schemes in an MSS context.

In the FDMA scheme, the available bandwidth is partitioned into a number of segments. Each station is allowed to transmit in one or more segments according to communication needs. The concept is illustrated in Figure 4.24.

When spectrum blocks are pre-assigned to each earth station of a network, the scheme is termed fixed assigned (FA) and if channels are allocated dynamically in response to requests, the scheme is a *demand assigned* (DA) FDMA. In the latter case, when each segment consists only of a single channel, the FDMA scheme is known as a DA single channel per carrier (SCPC)/FDMA scheme. DA schemes are better suited when traffic requirement per user is

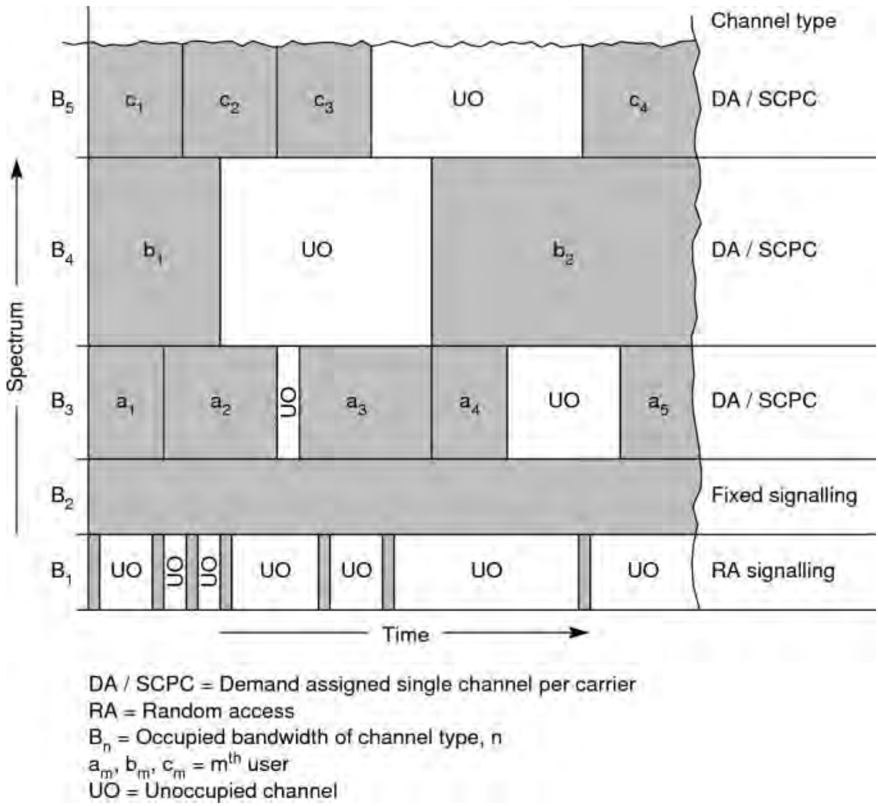


Figure 4.24 A DA/SCPC/FDMA frequency assignment scheme, portraying channels of various types and their occupancy with time

low, a prerequisite of MSS links. Fixed assignment is more suitable when communication demand is high or for continuous broadcasts. In an MSS environment, fixed assignments are used for communication between gateways, network broadcasts and for various types of signalling, such as call initiation.

Channel management in a DA SCPC/FDMA can be either centralized or distributed. In a centralized management scheme, frequencies are maintained and managed centrally in a network control station (NCS); in the distributed management scheme, each participating fixed earth station manages its individual pool of frequencies. The spectrum utilization efficiency of a centrally controlled system is superior, but it is prone to a single point failure. A back-up NCS can provide resilience. Distributed management offers a more resilient architecture, but it is less spectrally efficient; as each station must have individual pools to meet its peak demand, many circuits remain idle after busy hours. Inefficiency increases when traffic requirement per gateway is small and the busy hours of gateways are staggered, a situation that occurs when satellite coverage spans several time zones. In a hybrid scheme, each gateway has a few channels assigned on a permanent basis, while the remaining channels are assigned on demand from a central NCS. This scheme allows a pool of channels to be shared, while retaining the distributed architecture.

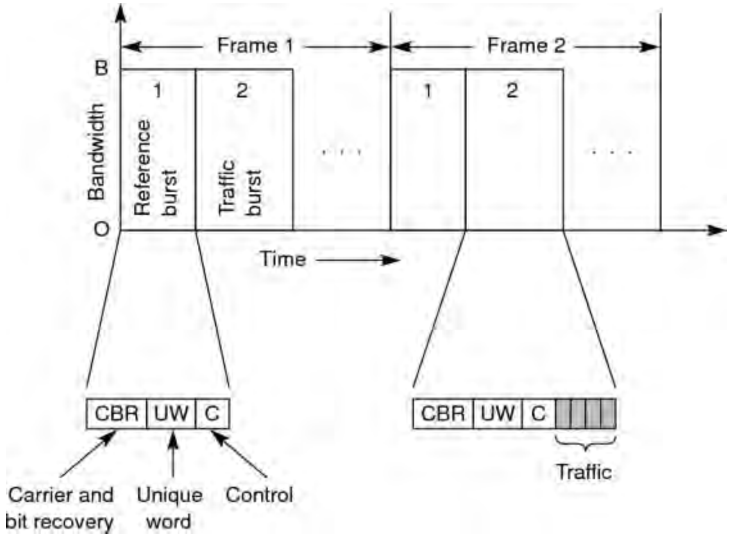


Figure 4.25 A TDMA scheme, comprising a reference burst, traffic burst encapsulated in repetitive frames

In a TDMA scheme, users access the satellite in non-overlapping time bursts, as illustrated in Figure 4.25.

The reference burst consists of a carrier bit recovery (CBR) field for carrier and timing recovery, a unique word (UW) field for burst synchronization and a control (C) field for station identification. The reference burst is followed by non-overlapped time bursts to carry traffic. The sequence is repetitive; where the reference burst synchronizes each frame. Each station transmits within its designated time slot.

The time slots may either be FA or DA. As discussed earlier, DA TDMA is better suited than FA TDMA and a central, distributed or hybrid slot assignment schemes are possible. TDMA network management is involved due to the need to maintain the earth stations in time synchronization. Complexities arise due to time difference in transmission between geographically dispersed earth stations and a satellite's orbital motion, especially in non-geostationary systems. Moreover, synchronization of mobile terminals in fading conditions and Doppler, coupled with a low carrier-to-noise environment requires care in terminal design.

Time division duplex (TDD) as a variant of TDMA used in terrestrial systems for a considerable period a was introduced in the LEO Iridium satellite system. In a TDD scheme, a time slot is used for both transmission and reception, in effect doubling its spectral capacity. The time delay in each direction must be low for the scheme to be effective. This is achievable in LEO systems where time delays are of the order of a few tens of milliseconds.

In CDMA, all users access the satellite without restriction within the full band. All users can co-exist simultaneously when spread spectrum modulation is used, since each user is assigned a unique code, which has low cross-correlation with codes used by others. Each user is able to extract transmissions destined for it by correlating the received signal with a replica of the code used at the transmitter. The correlation function peaks when the codes match, whereas all other transmissions are rejected. With technological developments, the cost of implementing the CDMA system has reduced drastically. It is common to deploy a combination of FDMA and CDMA by segmenting the band appropriately.

The direct sequence CDMA scheme is affected by the so-called *near–far* problem, which occurs when interfering transmissions from near-by UTs are received at a higher level than the transmissions from the wanted UT which may be far away, resulting in excessive BER or even a loss of signal. The effect, however, tends to be lower in frequency-hopped schemes, as the probability of co-channel interference is low. The mechanism of self-interference in these two schemes is slightly different. In the direct sequence scheme, interference from all users is always present and noise-like, whereas in the frequency-hopped system interference PSD will be high (coherent) if the frequency of the interferer falls within the band of the carrier wanted. When the hopping rate is much smaller than the information rate, the interference is coherent, but intermittent; on the other hand, when the hopping rate is much larger, the interference becomes noise-like. One of the problems in the frequency-hopped system is the need to maintain the phase noise within tolerance.

The characteristics of data traffic are generally different to speech and therefore dedicated channel assignments are inefficient. Data is typically characterized by bursts of high activity, followed by long pauses, which would be wasteful if a dedicated circuit were assigned. Note that some types of data traffic are continuous, for example, transfer of large files or a network broadcast or packet streams. Another difference is that data traffic is often less sensitive to message delivery time. Therefore, more efficient use of satellite resource is possible if channels are assigned on per message or per packet basis. A number of data accessing schemes have been developed, each matched to specific channel characteristics and well documented (e.g. Lam, 1979). We have already mentioned previously that schemes range from random access on one end of the scale to reservation TDMA on the other. Random access schemes do not require any network coordination and are therefore quite simple; typical applications are registering a mobile to the network or requesting a channel assignment. The most commonly used random schemes in MSS are Aloha, Slotted Aloha and to a lesser extent Reservation Slotted Aloha. In the Aloha scheme, the user transmits a data packet whenever necessary; in Slotted Aloha the packet transmission is constrained within a time slot; this requires network time synchronization; in Reservation Aloha, a user continues transmission as long as necessary, as negotiated; at the end of transmission, the slot is again open to contention. Reservation TDMA is useful when large amounts of data must be transferred. In this scheme, each user reserves a time before transmission in identical to the conventional TDMA.

Table 4.6 summarizes some commonly used access protocols for various types of traffic used in MSS. The suitability of a protocol depends on a trade-off between permissible message delay, channel throughput, terminal cost and network complexity.

In practice, hybrid schemes offer optimum performance. Typically, frequency division is multiplexed with TDMA or CDMA; packet accessing schemes are combined with circuit-mode DA schemes and fixed assignments are used for network broadcasts. The forward and return links may use different accessing schemes. For example, a FA channel carrying time division multiplexed (TDM) data may be used in the forward link and a TDMA scheme in the return, as synchronization of a continuous TDM stream at the mobile is simple at the same time data aggregation is feasible at the fixed site and synchronization of TDMA bursts is manageable at fixed earth stations. Similarly, a packet access scheme is more efficient

Table 4.6 Traffic characteristics with a suitable protocol

Traffic characteristic	Suitable protocol
Bursty – short messages; e.g. request for a channel, a paging acknowledgement burst	Aloha, Slotted Aloha
Bursty – long messages; e.g. file transfer	Aloha for reservation; a reservation protocol for message transfer
Continuous – e.g. large file transfer, voice	Fixed assigned scheme (circuit mode).

Table 4.7 Examples of accessing schemes

System	Channel function and accessing scheme
Inmarsat (Note: this is a sample and not an exhaustive list; it includes schemes used in Inmarsat's legacy system)	
Inmarsat-B	Forward signalling, network broadcast, telex, inter-station signalling spot beam identifier: fixed assigned TDM
	SCPC voice/data – centrally controlled DA/FDMA
	Return link – request for channel assignment, call acknowledgement, ocean region registration: Aloha
	Return link – low speed data, telex: TDMA
Inmarsat-Aero	Network broadcast, signalling and data in forward direction: time division multiplexed (TDM) fixed assigned transmission
	Circuit mode SCPC – DA/FDMA with distributed channel management
	Request for channel assignment, signalling and data in return direction – Slotted Aloha
	Data/messages in return direction – reservation time division multiple access
Inmarsat C	Forward signalling and messages: fixed assigned TDM
	Return signalling channel: hybrid slotted Aloha with a provision for reserving capacity
	Return message channel: TDMA
Iridium	Multiple access – TDD
Globalstar	Multiple access (forward and return) – CDMA
ICO	Multiple access (forward and return) – TDMA

in the return direction for functions such as channel request and short messages. Table 4.7 gives examples of accessing schemes used in various MSS systems.

4.4.1 Comparison of Multiple Access Schemes

Numerous studies have been undertaken to compare the efficiency of multiple access schemes (Viterbi, 1985; Gilhousen *et al.*, 1990; Giubilei and Miracapillo, 1995; Giovanni, Ferrarelli and Vatalaro, 1995). Table 4.8 summarizes the main characteristics of the most commonly used schemes in MSS.

It is evident that the selection of an accessing scheme involves trade-offs, as each scheme has its strengths and shortcomings. Practical constraints may add further limitations. For

Table 4.8 Main characteristics of common accessing methods

Access scheme	Characteristics
FDMA	Mature technology and low cost Network timing not necessary No restriction on type of modulation or baseband Susceptible to inter-modulation noise introduced by system non-linearity in particular in earth station or satellite high power amplifiers; weak carriers more susceptible Requires tight uplink power control to maintain equitable down-link power distribution between carriers
TDMA	Efficient use of available satellite power and power control is less critical when wideband TDMA carriers used (Note: this is not always the case in MSS); Well suited to digital systems Network timing is essential Fast synchronization of demodulator under fading and power limited conditions difficult
CDMA	Good resistance to interference, allowing closer spacing of satellites in geostationary orbits and tight intrasystem frequency reuse (e.g. between spot beams) Soft handover possible Good resistance to frequency-selective fading and multipath Good signal security Graceful degradation in signal quality Grade of service degradation is gradual and can be perceived by a user without the specific need of a network congestion message Receivers are relatively more complex A-CDMA requires good power control to maximize capacity Wideband transparent transponders (and ground receivers) may get artificially loaded by pick-up of extraneous carriers
Packet access	Well suited for bursty traffic Flexible – can meet demands of sporadic or continuous traffic

example, an asynchronous CDMA may offer an advantage in terms of minimal synchronization requirements coupled with a self-regulating grade of service management. But it may not be possible to integrate the scheme with an operator's existing FDMA system, due to the inherent incompatibility between the systems and the operator's difficulty to modify the existing network.

A scrutiny of requirements reveals a close relationship between accessing and modulation schemes. This will become more evident later in the section when comparing their spectral efficiency. The following factors influence the efficiency of multiple accessing schemes:

- **RF interference:** A certain amount of interference is built into radio link design to allow spectrum reuse. In this context, schemes that offer higher interference rejection have a potentially higher capacity. Note the close relationship between modulation, coding and multiple access schemes.
- **Voice and data activity:** It is established that during a normal conversation, the average occupancy of transmission is about 40%. Similarly, it has been observed that certain

types of 'continuous' data tend to exhibit pauses (Feldman and Ramana, 1999). Therefore average interference power reduces when carrier suppression is used during speech pauses.
- **Variations in traffic mix:** MSS systems provide wide-area coverage that may cover areas with different types of communications, for example data use may be prevalent in developed countries of the service area. Therefore, a multiple access scheme offering traffic adaptability increases net system throughput.
- **Propagation effects:** The MSS propagation environment is characterized by multipath effects and shadowing. Accessing schemes that offer resistance to multipath effects can affect an increase in utilization of radio resources.

In practice, spectrum efficiency has to be considered in conjunction with other practical issues:

- *Maturity of technology* is vital for any commercial systems. Despite several potential advantages, CDMA could not be introduced in early MSSs due to insufficient technical development. For the same reason, TDMA on the forward service link was inconceivable in first generation MSS.
- For established operators, *continuity of service* is essential, which constrains the use of innovative/new technology for existing or legacy service offerings.
- The efficiency of multiple access is influenced by satellite's *radio resource* availability, that is, whether the link is spectrum or power limited.

Earlier, the highest revenue-earning MSS was voice, and consequently, circuit-mode transmissions were dominant. In the past decade, the gap between voice and data usage has bridged very rapidly, influenced largely by the internet and the trend is expected to continue. This has necessitated the introduction of more flexible and efficient upper layer transport schemes, particularly the *Internet protocol* (IP). It is notable that technological limitations have been diminishing rapidly in recent years and hence the schemes that offer an optimal mix of spectrum efficiency, adaptability to traffic growth and capability to manage a different mix of traffic and support efficient transport mechanisms, particularly IP, will offer a distinct advantage. The non-trivial problem of backward compatibility has to be managed by each operator through a carefully chosen evolution path suitable to the operator.

4.4.2 Comparison of Spectral and Power Efficiency

The most common criteria for comparing multiple access schemes are spectral and power efficiency, which essentially translate to the revenue-earning potential of a system. Spectral efficiency is defined here as the number of bits transmitted per RF cycle, that is, bits/s/Hz.

The spectral efficiency η_s of a spread spectrum system is given as (Viterbi, 1985),

$$\eta_s = \left[C/\left(N_0 W_s\right)\right] / \left(E_b/N_0'\right) \left[1 + \left(C/N_0 W_s\right)(M-1)/M\right] \tag{4.13}$$

where C = Total received carrier power from M earth stations, M = number of participating earth stations, N_0 = Total thermal noise power density, W_s = RF bandwidth = $1/T_s$, where T_s = chip duration, E_b = Received energy per bit of information = C/T_b, where T_b = information bit duration, $N_0' = N_0 + I_o$ where I_o = interference PSD.

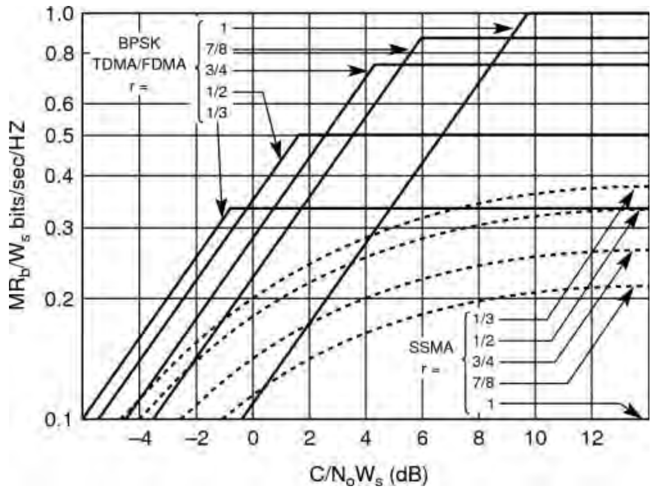


Figure 4.26 Spectral efficiency (bits/s/Hz) versus (C/N_oW_s). (Source: Viterbi, 1985. © 1985 IEEE. Reproduced with permission.)

When interference from other systems exists, the total carrier power C in Equation 4.13 is increased to $(1 + K)C$, where K represents interference from other system.

Figure 4.26 shows spectrum efficiency as a function of C/N_oW_s for BER of 10^{-5} and code rates using BPSK (Viterbi, 1985). The E_b/N_0' for achieving the bit rate ranges from 9.6 dB for the uncoded rate to 4.0 dB for a code rate of one third for practical decoders approaching 3.4 dB asymptotically.

Some conclusions that may be drawn from the figures are:

- CDMA systems are bandwidth limited rather than power limited.
- The limit of spectral efficiency approaches $(E_b/N_o{}')^{-1}$ for high (E_b/N_0) and therefore low code rates improve efficiency.
- The rate of performance improvement reduces as code rate reduces and appears marginal for code rates less than one third.
- The reduction in code rate does not influence the bandwidth because of spreading; the increase in the number of bits due to added redundancy is absorbed by a corresponding reduction of chip rate.

In practice, the following considerations apply (Gilhousen *et al.*, 1990).

- It is well-known that *voice activity* in a typical telephone conversation is about 40% of call duration. By using voice-activated transmission therefore, the interference level is reduced by a factor of 2.5, increasing CDMA capacity by about the same factor.
- In a system deploying *spot beams*, the required isolation between beams where frequencies are reusable is significantly lower in the CDMA than in the FDMA scheme. It is feasible to reuse the same band in adjacent beams. This is possible because contributions to $N_0{}'$ from users in adjacent spot beams can readily be absorbed in a spread spectrum system and interference from spot beams further away becomes progressively lower.

- In FDMA or TDMA systems, *frequency reuse by polarization discrimination* is difficult due to the poor polarization discrimination possible in mobile antennas and polarization reversal caused in a multipath. The CDMA system offers the potential of frequency reuse in an opposite polarization. An increase in capacity of about 60% is possible with a modest antenna isolation of 6 dB.

Introducing the improvements made possible by the factors previously, the total noise received can be modified as follows (Gilhousen *et al.*, 1990):

$$N_0 + I_0 = N_0 + a\rho v\,(N - 1)\left(E_b R_b\right)/W_s \tag{4.14}$$

where, a = 1/number of spot beams (uniformly distributed traffic distribution), ρ = polarization isolation factor, v = voice activity factor, N = number of users, E_b and R_b are information energy per bit and bit rate respectively and W_s = occupied RF bandwidth.

The efficiency η_s is given as,

$$\eta_s = \left(C/N_0 W_s\right)/v\left[E_b/\left(N_0 + I_0\right)\right]\left(1 + a\rho C/N_0 W_s\right) \tag{4.15}$$

Self-interference is one of the main causes of capacity limitation in asynchronous CDMA; transmissions from a user increase the noise to all users. A tight power control is necessary to keep system noise as low as possible, which does not actually eliminate the fundamental self-interference problem, but attempts to reduce it. An improved CDMA technique called band-limited-quasi-synchronous code division multiple access (BLQS-CDMA) proposed by European Space Agency (ESA) researchers tackles the fundamental problem of self-noise itself (De Gaudenzi, Elia and Viola, 1992). BLQS-CDMA offers all the advantages of CDMA with efficiencies comparable to orthogonal accessing schemes, such as TDMA and CDMA. The scheme requires chip clock and carrier frequency synchronization using the information embedded in the CDMA signal structure in the form of a master code. Another feature of BLQS-TDMA is Nyquist pulse shaping of signature chips, giving spectral compactness without compromising on detection performance, leading to simple digital implementation. The key to the improvement is the synchronization of the start epoch of the signature sequence to within $\pm 0.5\,T_c$, where T_c is the chip duration. Each transmitter has its individual signature sequence and a time reference with respect to the network reference. Preferentially phased gold codes (PPGCs) provide optimal performance for quasi-synchronous CDMA (QS-CDMA). By using different generator polynomials, it is possible to generate PPGC families with quasi-orthogonal cross-correlation properties within the same code family and pseudo-random correlation properties among signatures of different families. In a spot beam environment, the same family code can be used in each spot beam and different ones in other beams. Interference from the same beam will thereby be eliminated and interference from adjacent beams that should have some beam isolation will behave as self-interference of asynchronous-code division multiple access (A-CDMA). The probability of error for a perfectly synchronized QS-CDMA $\rightarrow 0$, whereas for A-CDMA the value $\rightarrow$ constant when code length $\rightarrow \infty$. Figure 4.27 (De Gaudenzi, Elia and Viola, 1992) compares the theoretical BER for an ideal QS-CDMA with an A-CDMA for a 15-user system, showing the low sensitivity to self-noise experienced by QS-CDMA. This scheme has been investigated and proposed for a number of mobile systems such as the European mobile system (Jongejans *et al.*, 1993) and a K_a band mobile multimedia system using a regenerative transponder (Valadon *et al.*, 1999b).

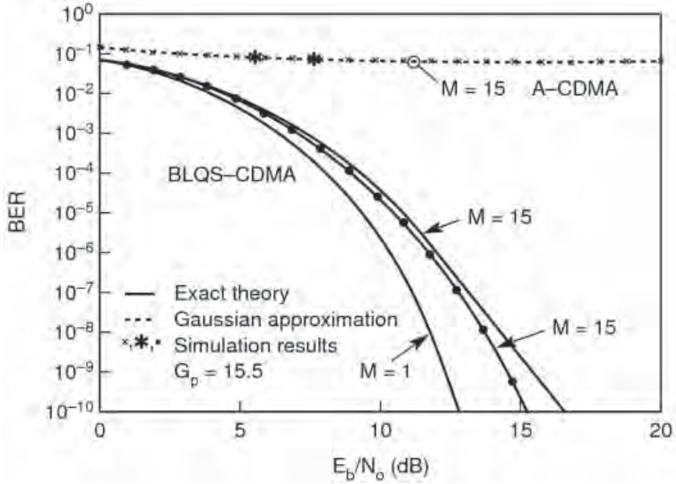


Figure 4.27 A comparison of theoretical BER of an ideal BLQS-CDMA with an A-CDMA for a 15 user system. (Source: Gaudenzi *et al.* © 1992 IEEE. Reproduced with permission.)

Figure 4.28 compares the spectral efficiency, given in bit/s/chip, of these two CDMA techniques for a BPSK base modulation (Giubilei and Miracapillo, 1995). The efficiency of QS-CDMA is r bits/chip, where r = coding rate.

In a thermal noise dominated link, Equation 4.15 also applies to FDMA and TDMA. For these schemes C is the total EIRP of the satellite and B_t the total RF bandwidth (i.e. $W_s = B_t$)

1. Power limited case:

$$\eta_p = \left(C/N_oB_t\right) / \left(E_b/N_0\right), \text{when } MR_b/B_t < G_l r \log_2(m) \tag{4.16}$$

2. Bandwidth limited case:

$$\eta_b = G_l r \log_2(m), \text{when } MR_b/B_t > G_l r \log_2(m) \tag{4.17}$$

where G_l = guard band loss for FDMA and guard time/preamble loss for TDMA, r = code rate, m = PSK constellation dimension, R_b = user information rate, M = number of participating users.

Again, extending the formulation to a spot beam MSS environment and including voice activity advantage, the efficiency for each case is given as

1. Power limited case:

$$\eta_p = \alpha \left(C/N_oB_t\right) / v \left(E_b/N_0\right), \text{when } MR_b/B_t < G_l r \log_2(m) \tag{4.18}$$

2. Bandwidth limited case:

$$\eta_p = \alpha G_l r \log_2(m), \text{when } MR_b/B_t > G_l r \log_2(m) \tag{4.19}$$

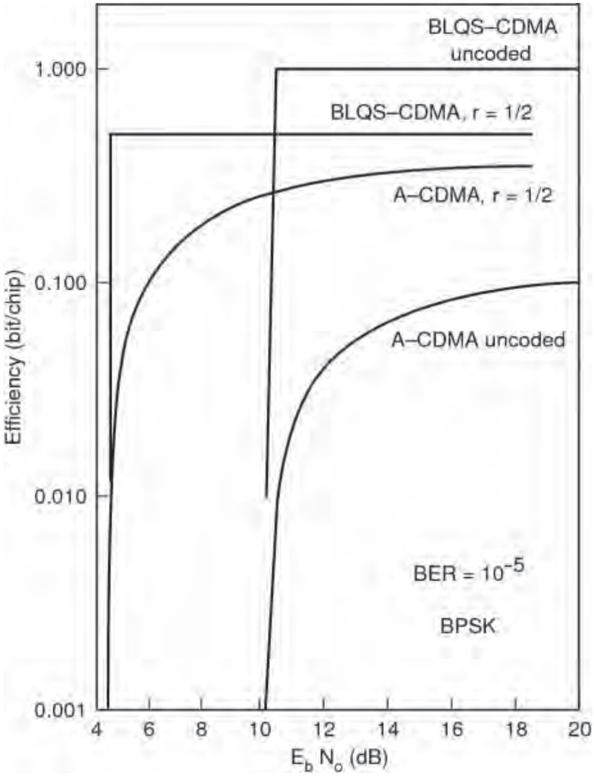


Figure 4.28 A comparison of spectral efficiency (bit/s/chip) of QS-CDMA with an A-CDMA. (Source: Giubilei and Miracapillo, 1997. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

where:
α = frequency reuse factor,
v = voice activity factor.

Figure 4.29 (Gilhousen *et al*., 1990) compares the spectral efficiencies of a CDMA and FDMA for a nine-beam system using the following link parameters.

1. CDMA: $E_b/(N_0 + I_0) = 2.5$ dB; v = 0.35; code rate 1/3, constraint length = 9.
2. FDMA: $E_b/N_0 = 8.4$ dB; 8-DPSK (differential phase shift keying), m = 8; convolution code at r = 2/3; constraint length = 5, v = 0.35; $\alpha = 2$.

Note the increase in capacity of CDMA with increase in $C/(N_0W_s)$, while the FDMA efficiency is limited at 2 bits/s/Hz. This is in contrast to Figure 4.26 where CDMA efficiency is lower than TDMA or CDMA.

For a multiple satellite/multiple spot beam system, interference component is built into the total noise budget of the system. When considering non-geostationary satellite systems,

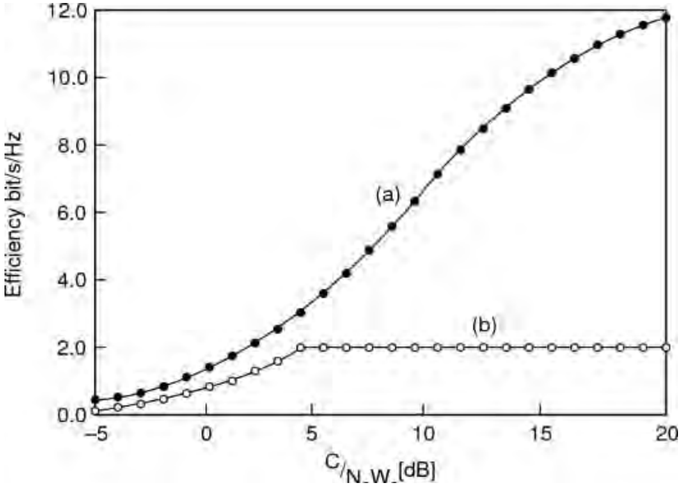


Figure 4.29 A comparison of spectral efficiency of a CDMA (plot a) and FDMA (plot b) for a nine-beam system considering voice activity, antenna discrimination and polarization reuse. (Source: Gilhousen *et al*., 1990. © 1990 IEEE. Reproduced with permission.)

intrasystem interference becomes time-dependent due to constellation dynamics, causing interference levels to become time variant and probabilistic due to random signal fluctuations caused by fading, power control, etc. The multiple access scheme efficiency of such systems is also influenced by frequency planning and code design. Chapter 7 discusses various trade-offs when considering the efficiency of system in totality.

Selection of a multiple access scheme for MSS requires a number of practical considerations. The criteria differ according to the emphasis placed by the system designers. For example, the TDMA scheme was chosen in preference to the CDMA scheme for the proposed intermediate circular orbit (ICO), whereas Globalstar designers preferred the CDMA system (see Chapter 11).

Consider, for example the rationale for selection of TDMA for ICO (Hart, Goerke and Jahn, 1995). Note that some of the conclusions may be contradictory to the material presented previously; such contradictions have been the subject of several interesting debates:

- Wideband measurements showed that frequency selective fades affects bandwidth beyond 5 MHz and hence the fade resistance advantage in favour of CDMA could not be realized within the narrowband available for ICO operation.
- On the basis of an earlier investigation, TDMA was considered more efficient (Meidan, 1994).
- TDMA can benefit from satellite diversity and soft handover, as much as CDMA.
- Satellite links are power limited, and self-interference in CDMA links reduces link margins as the system approaches full capacity; the use of orthogonal CDMA increases capacity, but achieving signal orthogonality is difficult when satellite diversity is used, due to differential path delay.
- In TDMA, use of diplexer can be avoided if transmit and receive bursts are arranged to occur in different time slots, whereas diplexers cannot be avoided in CDMA due to the

need for continuous transmission and reception. Diplexers tend to add losses in the front end, thereby reducing a receiver's sensitivity.
- The inaccuracy introduced in power control loop due to propagation delay is likely to affect significantly the capacity of an asynchronous CDMA.
- Large interference into a CDMA signal can cause outage to all users when such interference occurs.
- Due to wideband characteristics of CDMA, the probability of interference outage in CDMA is higher than in narrowband TDMA, where interference can be counteracted by reassigning the affected call to an interference-free channel.
- TDMA can better manage non-uniform traffic distribution in a spot beam environment, as it allows the peak capacity in a beam to be increased and switched between spot beams.
- TDMA disadvantage vis-à-vis CDMA in terms of return link power can be reduced by using non-linear amplifiers at the mobile whereas CDMA requires linear amplifiers.
- Spectrum sharing between CDMA systems can be problematic, as wideband transmissions from one system will affect the capacity of the other, whereas in TDMA the band segmentation traditionally used has minimal impact on the capacity of other operators.

In contrast, a study conducted in Italy comparing CDMA and TDMA concluded that neither was distinctly superior to the other (Priscoli and Muratore, 1996). The authors took two specific systems, the mobile satellite business network (MSBN) and a satellite extension of the global system for mobile communications (GSM), and compared radio and network aspects of the network. MSBN is a CDMA system developed by ESA for general applications whereas GSM uses the FDM/TDMA (frequency division multiplex) technique. ESA's L-band land mobile (LLM) satellite transponder, a multi-spot, geostationary satellite, was used as the payload. Comparison criteria included flexibility and number of available channels, EIRP, propagation environment and intersystem interference. It was observed that MSBN is more flexible than GSM; GSM generally has an advantage with regard to capacity when considering transportable mobile, but the reverse is true for vehicular mobiles. When considering the transportable terminal, the return link was the limiting factor with regard to the capacity for both technologies. For the GSM system satellite EIRP was the limiting factor, whereas for the MSBN, self-noise and downlink thermal noise were the limiting factors. The CDMA system proved to be more resistant to shadowing loss in terms of system capacity for an average mobile power of 5 W; increasing the shadowing margin from 3 to 7.2 dB reduced the TDMA capacity by $\sim 67\%$, from 192 to 64, whereas the reduction in capacity was $\sim 40\%$, from 221 to 135, for the CDMA scheme. The study assumed that the power control and frame synchronization were perfect; and hardware complexity considered was not included as an evaluation parameter. The forward and return link budgets and the maximum number of simultaneous channels are shown in Table 4.9.

The views and conclusions of system designers who have chosen CDMA will differ. Some of the advantages claimed by CDMA proponents include potential use of low power emissions from MES, the possibility of reducing call drop-outs by use of soft handover, the capability to use multipath signals to improve path diversity, and others (see Table 4.8). Examples of operational CDMA systems are the EUTELTRACS/OmniTracs and Globalstar systems. For example, contrary to the opinions above, the Globalstar CDMA system uses techniques such as interleaving, open- and close-loop power control design and diversity management to retain the advantages offered by the terrestrial CDMA system. The system

Table 4.9 Modulation and coding used in trade-off analysis

	BPSK	8-DPSK	BPSK	CDMA
Code rate	½; constraint length (K) = 7	⅔, TCM 16 state	¾; K = 5	¼; K = 7
Signal constellation size	2	8	2	2
Channel bandwidth (KHz)	13.5	3.4	9	1000
Satellite EIRP/channel (dBW)	16.7	21.7	18.2	16.4
Bit error rate	10^{-5}	10^{-5}	10^{-5}	10^{-5}
E_b/N_0 (dBHz)	3	8	4.5	–
$E_b/(N_0$ + self-interference)	–	–	–	2.7
Ricean factor (dB)	10	10	10	10
Margin (dB)	4	4	4	4

(Source: Giubilei and Miracapillo, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

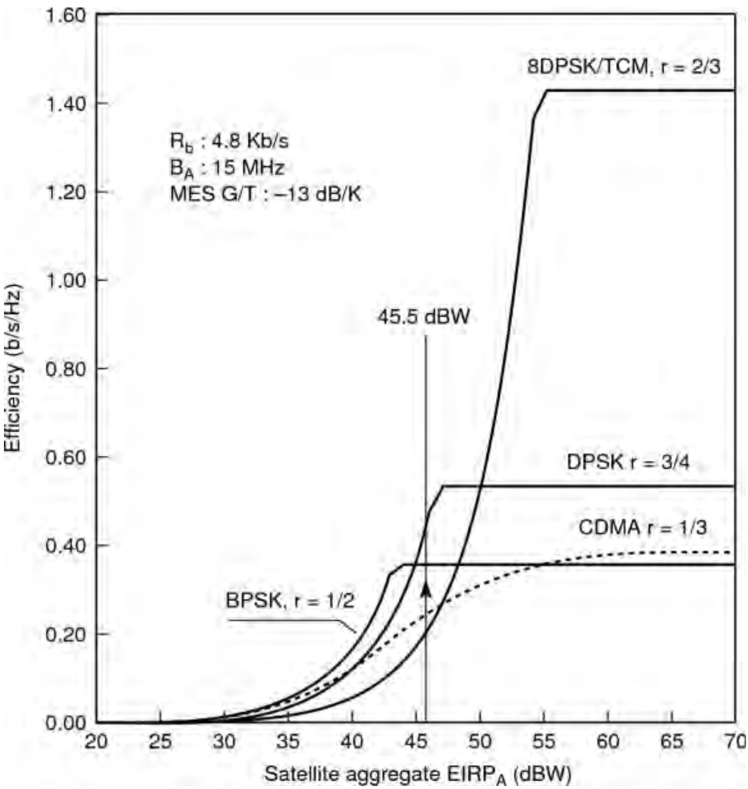


Figure 4.30 Access scheme trade-off analysis for ESA's L band multi-beam payload aboard ARTEMIS. (Source: Giubilei and Miracapillo, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

Table 4.10 System parameters used for the trade-off analysis

Parameter	Value (kbps)
Information rate per user	4.8
Voice activity factor	0.4
Filter roll-off factor	0.4
Available EIRP (dBW)	45.5
MES G/T (dB/K)	−13
Satellite transponder type	Transparent

(Source: Giubilei and Miracapillo, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

uses a combination of FDM/CDM (code division multiplex) and orthogonal signal multiple access techniques. Chapter 11 provides details of Globalstar system design features including elements of the CDMA waveform.

The multiple access choice depends on the emphasis placed by the system designers on applicable constraints. The accessing scheme is only one element of a satellite system; other considerations include service requirements, spectrum availability, interference management technique/specifications, spacecraft characteristics, such as transponder type and antenna complexity, UT requirements, constellation geometry, etc. (see Chapter 7). For example, Figure 4.30 shows the results of access scheme trade-off analysis done for ESA's L band multi-beam payload aboard ARTEMIS (Giubilei and Miracapillo, 1995), showing the efficiency for CDMA and FDMA with a number of modulation schemes and code rates listed in Table 4.9.

TCM parameters were taken from proposals in NASA's MSAT programme for land mobile communications. The trade-off was conducted within the specifications of the payload, that is transponder type, satellite EIRP, G/T, cross-polarization and adjacent beam isolation and other system parameters listed in Table 4.10.

For the case under consideration, FDMA/BPSK offers the best choice. Eight PSK FDMA becomes effective at higher levels. The performance of the CDMA scheme is pessimistic, as the voice activity advantage and polarization reuse advantages are not included. The triangular point shows the operating point of the payload that supports 445 simultaneous users using a BPSK/FDMA scheme.

Revision

1. i. What are the channel degradations in a mobile satellite link that affect the performance of the modulation sub-system?
   ii. Outline the problems in a coherent demodulator associated to such degradations.

2. i. Which modulation schemes are generally preferred in mobile satellite systems?
   ii. Briefly outline the principle of each modulation scheme with the help of constellation diagrams.
   iii. State the reasons to prefer these modulation schemes over others.

3. i. Differentiate between coherent and non-coherent demodulation.
ii. Suggest at least two techniques of carrier recovery in a coherent demodulator.
iii. With the help of block schematics describe the functioning of QPSK modulator and demodulator.

4. i. What are the strengths and limitations of OFDM applied to a broadband mobile satellite system?
ii. What are the limitations in the use of the scheme in MSS systems, noting that the scheme has been widely proposed for mobile satellite broadcast systems?
iii. Explain the process of OFDM generation and reception.
iv. Outline the transmission and reception methods of direct-sequence and frequency-hopped spread spectrum systems.

5. i. Differentiate between convolutional and block codes. Which of the two methods should be preferred for a narrowband mobile satellite system for applications such as messaging and machine-machine communication?
ii. Outline the principle of a typical convolution coder.
iii. Outline the principle of interleaving and state at least two scenarios where such a scheme can be applied advantageously.
iv. Suggest at least two scenarios where code concatenation can be used advantageously.

6. i. Explain the functioning of turbo-coder and decoder.
ii. Compare the main attributes of turbo- and LDPC codes.

7. i. Which accessing schemes are best suited for transmission of: (a) continuous streams such as voice and (b) bursty traffic, such as a bank transaction? State reasons for your choice.
ii. Explain, with reasons, the factors that influence efficiency of multiple accessing schemes for MSS.

8. i. Outline the principles of accessing schemes used in MSS.
ii. Compare the efficiencies of the classical multiple accessing schemes as applied to MSS for power and bandwidth limited links, taking advantage of voice activity and spatial reusability. State assumptions in the quantitative analysis.
iii. Compare the applicability of CDMA and TDMA for a medium bit rate mobile satellite system and suggest your own preference. State the assumptions in making the comparison and in your assessment.

References

Akhter, M.S., Rice, M. and Rice, F. (1999) Turbo code performance over aeronautical channel for high rate mobile satellite communications. Sixth International Mobile Satellite Conference, June 1999, pp. 287–291.

Alard, M. and Lassalle, R. (1988) Principles of modulation and channel coding for digital broadcasting for mobile receivers. EBU Collected Papers on Concepts for Sound Broadcasting into the 21st Century, August, pp. 47–49.

Amoroso, F. (1980) The bandwidth of digital data signals, *IEEE Communications Magazine*, **18** (6), 13–24.

Bahl, L.R., Cocke, J., Jelinek, F. and Raviv, J. (1974) Optimal decoding of linear codes for minimising symbol error rate, *IEEE Transactions on Information Theory*, **IT-20**, 248–287.

Baker, P.A. (1962) Phase modulation data sets for serial transmission at 2000 and 2400 bits per second, *Transactions of the American Institute of Electrical Engineers Part 1: Communication and Electronics*, **81** (3), 166–171.

Ball, J.R. (1982) A real time simulator for mobile radios, *The Radio Science and Electronics Engineer*, **52**, 475–478.

Berrou, C. and Glavieux, A. (1996) Near optimum error correcting coding and decoding: turbo-codes, *IEEE Transactions on Communications*, **44** (10), 1261–1271.

Berrou, C., Glavieux, A. and Thitimajshima, P. (1993) Near Shannon limit error-correcting code: turbo codes. Proceedings IEEE International Conference on Communications, Geneva, Switzerland, pp. 1046–1070.

Caves, J.K. (1991) An analysis of pilot symbol assisted modulation for Rayleigh fading channels, *IEEE Transactions on Vehicular Technology*, **VT-40**, 686–693.

Cowley, W.G., Lavenant, M.P. and Zhang, W. (1997) A mobile satellite modem for helicopter applications. International Mobile Satellite Conference, pp. 479–484.

Davrian, F. (1985) High performance digital communications in mobile channels. Proceedings of the IEEE Vehicular Technology Conference, pp. 114–118.

Davrian, F. (1987) Mobile digital communications via tone calibration, *IEEE Transactions on Vehicular Technology*, **VT-36** (2), 55–62.

De Gaudenzi, R., Elia, C. and Viola, R. (1992) Bandlimited quasi synchronous CDMA: a novel satellite access technique for mobile and personal communication systems, *IEEE Journal on Selected Areas in Communications*, **10** (2), 328–343.

De Jaeger, F. and Dekker, C.B. (1978) Tamed FM, a novel method to achieve spectrum economy in digital transmission, *IEEE Transactions on Communications*, **COM.26** (5), 534–542.

Divsalar, D. and Pollara, F. (1997) Turbo Trellis coded modulation with iterative decoding for mobile satellite communications. Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, CA, June 16–18, Co-sponsored by NASA/JPL and DOC/CRC, Jet Propulsion Laboratory, Pasadena, CA, JPL Publication 97-11, pp. 333–339.

Divsalar, D. and Simon, M.K. (1987) Trellis-coded modulation for 4800–9600 bits/s over a fading mobile satellite channel, *IEEE Journal on Selected Areas in Communications*, **SAC-5** (2), 162–175.

Divsalar, D. and Simon, M.K. (1988) Multiple trellis coded modulation (MTCM), *IEEE Transactions on Communications*, **36** (4), 410–419.

Feher, K. (1991) Baseband pulse shaping techniques for non-linearly amplified π/4-QPSK and QAM systems. Proceedings of Onboard Processing and Switching Conference, NASA, Lewis Research Center, Cleveland, OH, November.

Feldman, H. and Ramana, D.V. (1999) An introduction to Inmarsat's new mobile multimedia service. IMSC'99, Sixth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre and the Jet Propulsion Laboratory, pp. 226–229.

Francon, M.G. and Bousquet, M. (1997) Performance analysis of concatenated Reed- Solomon Trellis coded modulation over mobile satellite channel. Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, CA, June 16–18, Co-sponsored by NASA/JPL and DOC/CRC, Jet Propulsion Laboratory, Pasadena, CA, June 16, JPL Publication 97-11, pp. 303–308.

Gilhousen, K.S., Jacobs, I.M., Padovani, R. and Weaver, L.A. (1990) Increased capacity using CDMA for mobile satellite communication, *IEEE Journal on Selected Areas in Communications*, **8** (4), 503–514.

Giovanni, E.C., Ferrarelli, C. and Vatalaro, F. (1995) Analysis of multiple access techniques in multi-satellite and multi-spot mobile satellite systems. Fourth International Mobile Satellite Conference, Ottawa, pp. 482–487.

Giubilei, R. and Miracapillo, L. (1995) The L-band land mobile (LLM) system: access schemes trade-off. The Fourth International Mobile Satellite Conference, Ottawa, June, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.

Hart, N., Goerke, T. and Jahn, A. (1995) A discussion on mobile satellite systems and myths of CDMA and diversity revealed. Fifth International Mobile Satellite Conference, Ottawa, Canada, June, pp. 469–475.

Imai, H. and Hirakawa, S. (1977) A new multilevel coding method using error- correcting codes, *IEEE Transactions on Information Theory*, **IT-23** (3), 371–377.

Jongejans, A., Rogard, R., Mistretta, I. and Ananasso, F. (1993) The European mobile system (EMS). Proceedings Third International Mobile Satellite Conference, Pasadena, CA, June, pp. 305–310.

Korn, I. (1989) Coherent detection of M-ary phase-shift keying in the satellite mobile channel with tone calibration, *IEEE Transactions on Communications*, **COM-37** (10), 997–1002.

Lam, S. (1979) Satellite packet communication – multiple access protocols and performance, *IEEE Transactions on Communications*, **COM-27** (10), 1456–1466.

Le Goff, S., Glavieux, A. and Berrou, C. (1994) Turbo-codes and high spectral efficiency modulation. Proceedings of IEEE ICC'94, New Orleans, LA, May 1–5, pp. 645–649.

Li, H.B., Iwanami, Y. and Ikeda, T. (1993) Performance of a multidimensional BCM scheme with fading estimation based on time correlation. IEEE ICC'93, Geneva, Switzerland, May, pp. 443–447.

Liu, C.L. and Feher, K. (1991) $\pi/4$ QPSK modems for satellite sound/data broadcast systems, *IEEE Transactions on Broadcasting*, **37** (1), 1–8.

Lodge, J.H., Moher, M.L. and Crozier, S.N. (1987) A comparison of data modulation techniques for land mobile satellite channels, *IEEE Transactions on Vehicular Technology*, **VT-36** (1), 28–35.

McGeehan, J.P. and Bateman, A.J. (1984) Phase-locked transparent tone-in-band (TTIB): a new spectrum configuration particularly suited to the transmission of data over SSB mobile radio networks, *IEEE Transactions on Communications*, **COM-32**, 81–87.

Meidan, R. (1994) To spread or not to spread, this is the question. Proceedings International Vehicular Technology Conference, VTC'94, Stockholm, Sweden, pp. 56–59.

Murota, K. and Hirade, K. (1981) GMSK modulation for digital mobile radio telephony, *IEEE Transactions on Communications*, **COM-29** (7), 1044–1050.

Pommier, D. and Wu, Y. (1986) Interleaving or spectrum spreading in digital radio intended for vehicles, *EBU Review*, **217**, 128–142.

Priscoli, F.D. and Muratore, F. (1996) Radio and network comparisons for MSBN and GSM systems in a geostationary satellite environment, *International Journal of Satellite Communications*, **14**, 439–454.

Richharia, M., Trachtman, E. and Fines, P. (2005) Broadband global area network air interface evolution. 11th K_a and Broadband Communications Conference and 23rd AIAA International Communications Satellite Systems Conference (ICSSC), September 25–28, 2005, Rome, Italy, p. 12.

Sayegh, S.I. (1986) A class of optimum block codes in signal space, *IEEE Transactions on Communications*, **COM-34** (10), 1043–1045.

Shelswell, P. (1995) The COFDM modulation system: the heart of digital audio broadcasting, *Electronics and Communication Engineering Journal*, **7** (3), 127–136.

Simon, M.K. (1986) Dual-pilot tone calibration technique, *IEEE Transactions on Vehicular Technology*, **VT-35** (2), 63–70.

Sundberg, C.E. (1986) Continuous phase modulation: a class of jointly power and band-width efficient digital modulation schemes with constant amplitude, *IEEE Communications Magazine*, **24** (4), 25–38.

Trachtman, E. and Hart, T. (1999) Research elements leading to the development of Inmarsat's new mobile multimedia services. IMSC '99, Sixth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/ Industry Canada and Jet Propulsion Laboratory, pp. 209–212.

Ungerboeck, G. (1987) Trellis-coded modulation with redundant signal sets part I: introduction, *IEEE Communications Magazine*, **25** (2), 5–21.

Valadon, C., Rosmansyah, Y., Tafazolli, R. and Evans, B.G. (1999a) The SWAID project: deriving powerful modulation and coding schemes for future satellite multimedia systems. IMSC '99, Sixth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory, pp. 262–267.

Valadon, C.G.F., Verelst, G.A., Taaghol, P., Tafazolli, R. and Evans, B.G. (1999b) Code-division multiple access for provision of mobile multimedia services with a geostationary regenerative payload, *IEEE Journal on Selected Areas in Communications*, **17** (2), 223–237.

Vanelli-Coralli, A., Corazza, G.E., Karagiannidis, G.K. *et al*. (2007) Satellite communications: research trends and open issues. Proceedings of International Workshop on Satellite and Space Communications, 2007, IWSSC'07, September 13-14, 2007, pp. 71–75.

Viterbi, A.J. (1985) When not to spread spectrum – a sequel, *IEEE Communications Magazine*, **23** (4), 12–17.

Xiong, F. (1994) Modem techniques in satellite communications, *IEEE Communications Magazine*, **32** (8), 84–98.

Yasuda, Y., Kashiki, K. and Hirata, Y. (1984) High-rate punctured convolution codes for soft- decision Viterbi decoding, *IEEE Transactions on Communications*, **COM-32** (3), 315–319.

5

Fixed Earth Stations and User Terminals

5.1 Introduction

In the context of this chapter, fixed earth stations and user terminals (UTs) are systems or devices that provide radio connectivity between users. Terrestrial users access an MSS (mobile satellite service) network through *gateways*, which are large fixed earth stations interfacing terrestrial system(s) with the satellite network; mobile users communicate through *mobile or portable terminals*, which vary in size and appearance from pocket telephones to large ship mounted terminals using 0.5–1 m antennas (see Figures 1.6a–c). Depending on an individual's perception and background, mobile earth stations are called mobile terminals, UTs, satellite phones, communicators, ground terminals, and so on. We will refer to them generically as UTs unless a clear distinction is necessary. Gateways are large fixed earth stations similar to those used in FSS (fixed satellite service) stations at the physical layer, but incorporate a number of higher layer functional features specific to MSS. At the physical layer gateways adapt terrestrial signal format and protocols to enable efficient transport over the satellite channels in both the forward and the return directions taking account of the characteristics of the satellite network. We will use the terms gateway and earth station interchangeably, although this interchange is not valid when a fixed station is configured as a space segment relay – for example when configured simply to forward receive messages to other earth stations in a store and forward configuration (see Figure 1.4b). Only a brief description of gateways, limited to the radio part, is given here, as they are described in a number of textbooks (e.g. Richharia, 1999; Maral and Bousquet, 2009). Similarly, tracking, telemetry and command (TT&C) stations are well-covered in the literature and hence, are excluded.

In accordance to the theme of the book the focus here is on mobile terminals as they are specific to MSS; there has been a rapid evolution and phenomenal interest in this technology recently. Mobile VSATs (very small aperture terminals) have received enhanced attention in recent years. Features of this mobile type are mainly addressed in Chapter 13. There is a keen public interest regarding the biological effects of RF (radio frequency) radiation on people

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

due to the proliferation of mobile and satellite phones and therefore a section is devoted to this topic summarizing the current state of knowledge.

5.2 Gateways

Gateways provide a radio connection between the terrestrial network and the mobile elements of an MSS system. They are designed to receive and process a large number of simultaneous calls, which necessitates a high degree of linearity in amplifiers and mixers to minimize adverse effects of intermodulation. Gateway receivers operate over a wide dynamic range of the received signal, as signals transmitted from mobiles undergo fading, and are affected by variations in satellite antenna gain over the coverage area. Adaptive mobile power control and gain shaping of the satellite antenna footprint alleviate demanding dynamic range requirements to a certain extent (Goldberg, 1996). Gateways support numerous network functions such as call set-up, radio resource management, user database management, mobility management, switching and support to network and business management centres.

Figure 5.1 shows the main hardware support entities of a large earth station. The antenna system is usually a large parabolic dish several metres in diameter with low noise temperature.

The terrestrial interface processes and reformates the outgoing signal for the satellite network and performs the reverse operation on the incoming signals. The arrangement and functionalities of the 'baseband processing and control unit' depend on the requirements of the supported terrestrial network, that is PSTN (Public Switched Telephone Network), PDN (Public Data Network), Internet or PMN (Public Mobile Network) and the satellite

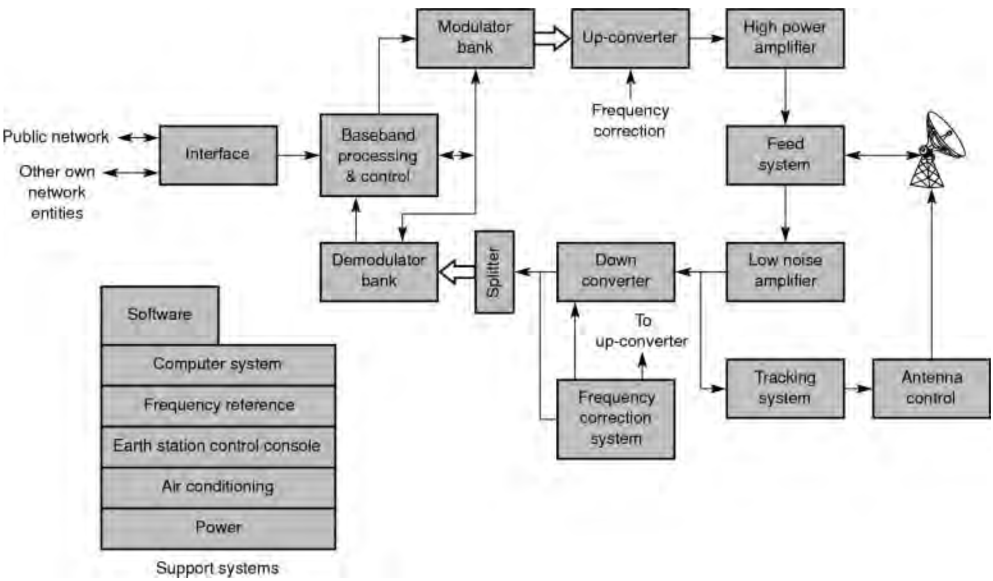


Figure 5.1 Main hardware entities of a large earth station

air interface. It consists of a variety of sub-systems to perform functions that include data segmentation, multiplexing, coding and introduces satellite air-interface protocols and messages; the control unit also manages calls, radio connectivity and network broadcasts. Typically satellite air-interface incorporates a number of layered functionalities to manage each call and the network as a unit. These network messages are transmitted on separate logical channels. The revamped signal is coded, modulated, up-converted, amplified in high-power amplifier and fed via a feed system into a parabolic antenna typically 8–10 m in diameter for a C band feeder link. Signals are received by the same antenna/feed system, amplified in a low noise amplifier (LNA), down-converted and demodulated/decoded to reproduce the message. In systems where a reference signal is transmitted for frequency synchronization, receivers incorporate an automatic frequency correction loop. Earth stations require a number of subsystems for support – a variety of software and computer systems, frequency reference sub-system, a control console, air conditioning and a reliable power source with a back-up. Earth station frequency generator is synchronized to a reference such as derived through the GPS (global positioning system) system. A gateway additionally provides call records and traffic related data to the business centre and shares information with customer management system for customer authentication, etc. Such communication is typically done over independent communication links.

In fixed stations of LEO (low earth orbit) or MEO (medium earth orbit) systems, several tracking antennas are used – one for communication and the others to acquire a rising satellite and/or to receive space diversity signals.

The tracking system is an essential feature of a gateway. Common tracking techniques include step-track, intelligent track and monopulse. Step-track systems are simple and low cost and therefore widely used in MSS gateways. They are susceptible to amplitude fluctuations, which may be caused for a variety of reasons, such as scintillation, rain and multipath (Richharia, 1986a and 1986b). Thus, step-track systems become unreliable in regions where occurrence of scintillation and rain fading are dominant, for example equatorial regions. Step-track systems incorporating intelligence offer a compromise between cost and accuracy. Such a system filters out tracking errors by using a predictive model of the satellite motion. Monopulse systems are highly accurate and agile and therefore best suited for earth stations supporting applications where rapid response with high accuracy is essential, such as spacecraft operations and network control.

Gateways have to comply with various RF and network-related functional specifications to ensure that the performance is compatible with that of others in the network. Typical RF parameters where compliance is mandatory include antenna side-lobe patterns, antenna gain/system noise temperature (G/T), EIRP (effective isotropic radiated power), frequency and EIRP stability, phase noise, frequency response and group delay. Examples of network related functions are call set-up and tear down and mobility management.

The demand assigned SCPC (single channel per carrier) frequency division multiplex accessing (FDMA) scheme is a common MSS access method. When using such an accessing scheme, the earth station installs a bank of single channel units commensurate with traffic requirements. A channel unit typically comprises a voice codec/transcoder and data multiplexer, scrambler, FEC (forward error correction) encoder, frame synchronizer and modulator. For a TDMA (time division multiple accessing) system, a timing generator is essential in accordance with the network's TDMA synchronization technique.

A fixed earth station that acts as a network control station (NCS) supports forward and return link signalling. Forward signalling to mobiles is usually over a time division multiplexed (TDM) broadcast channel; since the RF transmission is continuous, this scheme is easier to demodulate at mobiles. Signalling supported in the return direction includes random access burst transmissions for call set-up and in-band signalling for call management. Gateways exchange signalling with the NCS for allocation of radio resources and mobility management. The method of communication differs; some operators prefer to communicate on radio channels, others on terrestrial links, and yet others combine the two.

Depending on the architecture, an MSS network may incorporate either a central or a distributed management. In a centrally managed system, an NCS manages a number of functions, such as transmission of bulletin board, spot beam identifiers and frequency reference pilots; radio resource and mobility management; performance monitoring and communication with systems' business management entities. Earth stations rely heavily on application software for control and system management.

MSS systems generally use narrow-band SCPC carriers, typically 5 kHz for voice, and hence the system is susceptible to adjacent channel interference in case of relatively small errors in the transmission frequency. The most significant sources of frequency error in the forward direction are short- and long-term drift in earth station and satellite oscillators, and Doppler frequency variations due to satellite motion. Stable unmodulated pilot signals are transmitted in both the directions by one or more participating gateways that are used as a reference by the other earth stations to compensate frequency errors. It is not possible to cancel frequency errors introduced by such a large UT population – the task of compensating motion-induced Doppler is, therefore, performed autonomously in the mobile terminal itself.

5.3 User Terminals

UTs are low-cost, low-capacity and lightweight devices with a capability to operate in a multipath and interference environment. Terminals are generally aesthetically pleasing, easy to operate and incorporate power-saving features to prolong battery usage per charge. They are frequency agile, comply with electromagnetic interference (EMI) standards, support signalling, network protocols and power control, and manage security. Terminals with speech capability use low bit rate voice coders in the range 2.4–4.8 kbps, compliant with the specified speech quality. To support interworking with terrestrial systems, terminals have to operate in multimode. Terminal size must comply with physical limitations.

UTs may be categorized in a number of ways. They may be categorized according to their operational environment as:

- personal;
- portable and
- mobile.

Personal sets are carried on person; these are typically pocket-sized/handheld units, similar to mobile telephones (< 200–300 g). *Portable sets* use a directive antenna in order to obtain a higher throughput and as such are suited for application at fixed positions. *Mobile UTs* may be mounted on ships, aircrafts or land vehicles including railway with a capability to operate while the mobile is in motion.

Categorization of UTs may also be made by transmission rate as follows:

- low bit rate (0.02–4.8 kbps);
- medium bit rate (4.8–64 kbps);
- high bit rate (64 kbps to 2 Mbps);
- broadband (2–10 Mbps) and
- ultra broadband (> 10 Mbps).

Terminals can also operate in an asymmetric mode whereby, depending on the application, one of the two service links operates at a higher bit rate.

UT antenna gain, G/T and radiation pattern are fundamental electrical properties that set limits on throughput. The product of transmit antenna gain and transmitter power defines the EIRP transmitted from the mobile and so sets throughput in the return direction; whereas the ratio of receive antenna gain and system noise temperature defines the system G/T and so sets throughput bound in the forward direction for a given satellite EIRP. Radiation pattern of the antenna affects the magnitude of transmissions to and reception from the unwanted directions and hence interference caused or received by the mobile, as well as magnitude of multipath noise. Thus, we note that the antenna system is a critical element.

The size of antenna is constrained by the mounting space and the permissible weight of the terminals. Ship-borne antennas offer the least constraint in respect of size and weight for large ships, although the antenna mounting platform requires stabilization. The space constraint on terminals for small ships and yachts is severe. Aircraft antennas must exhibit low air drag, rapid tracking and comply with strict air safety regulations. Land mobile vehicle-mounted antennas have to comply with strict size, profile and cost constraints. Land portable antennas should be lightweight and low cost, and finally, hand-held terminal antennas must be a few centimetres in size, lightweight and low cost, with minimum RF leakage for compliance with health safety regulations.

5.3.1 Antennas

Before discussing various categories of terminal we will briefly review UT antennas and tracking systems.

Scattered and reflected radio signals around a vehicle are picked up by mobile antennas, causing random fluctuations, known as multipath noise, on the main signal. Multipath noise increases with antenna beam width as more extraneous signals get captured.

Tracking is required for medium gain antennas when the satellite is non-stationary or in applications where a call is supported on moving vehicles; however, portable and nomadic UTs used in geostationary MSS systems do not require antenna tracking as they are installed at fixed locations and antenna beam width of such UTs is so large that satellites appear stationary.

The ionosphere and multipath cause random polarization rotation to linearly polarized electromagnetic waves. One method for minimizing polarization mismatch is to introduce polarization tracking; however, this poses demanding requirements on antenna design. A more effective solution is to use circular polarization, as random depolarization has insignificant impact on it. In practice, circularly polarized antennas exhibit some ellipticity, which is measured as axial ratio. Figure 5.2 shows maximum mismatch loss as a function

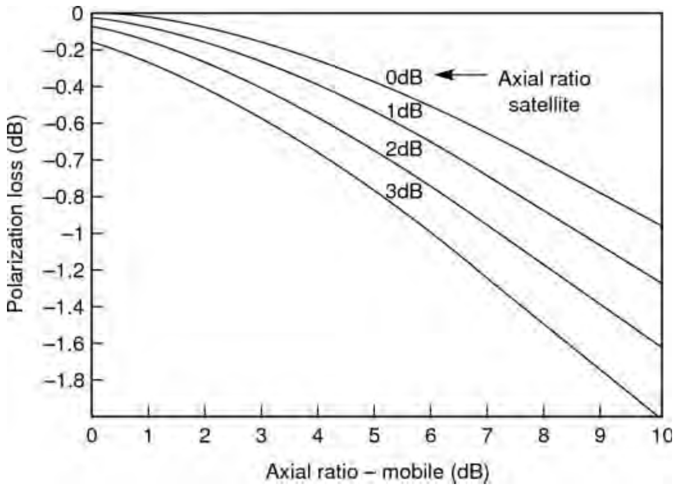


Figure 5.2 Maximum mismatch loss for various axial ratios of satellite and mobile antennas (Source: Graphics AR.)

of the axial ratios of the satellite and mobile antenna. Typical magnitudes of ellipticity for satellite antennas are of the order of 2 dB and so we note from the figure that for a mismatch loss of 0.5 dB mobile antenna can exhibit an axial ratio of 4 dB.

Typical antennas used in L band mobile terminals, together with their characteristics, are summarized in Table 5.1. In an operational MSS, the antenna gain and side-lobe performance are specified indirectly allowing manufacturers to optimize their terminal in the manner best suited to them. Thus antenna gain is specified in terms of EIRP and G/T while side-lobe performance is defined as a generic mask.

Table 5.1 Front-end characteristics of L band user terminals

Category	Antenna tracking	Gain (dBi)	G/T (dB/K)	Antenna type	Application examples
High gain	Step-track, intelligent track, open loop for acquisition	15–21	−10 to −4	Parabolic dish on ships or micro-strip array for land portables/aircrafts with mechanical or electronic steering	High data rate service for ships, aircrafts and land portables
Medium gain	Fixed or tracking	4–15	−23 to −10	Short backfire, phased array, helical, micro-strip	Voice and medium speed data for land portables, aircraft, etc.
Low gain	Fixed	0–4	−30 to −23	Dipole (drooping), quadrifilar, microstrip	Hand-held voice service, low data rate applications on ships, aircraft, etc.

The use of K_u and K_a bands enables smaller antennas in the 'high-gain' category. Note that the G/T of terminals in these bands to some extent is weather dependent because of the dependence of antenna noise temperature on water content along the signal path.

Common mobile antennas include crossed dipole antennas, helical antennas of various types, microstrip patch antennas, phased arrays and parabolic dishes. Parabolic dishes are generally used where space is not a major consideration, such as on large ships. Phased array antennas are used where aerodynamic drag should be minimized, and reliability along with high tracking speed is essential. Therefore, they are used in aeronautical and high-speed land vehicle and railway installations. Other types of antennas, such as crossed dipole, helical and patch antennas are used for land mobile and maritime communications.

The crossed dipole is made of two half-wavelength antennas placed at right angles to each other and fed with equal amplitude signals, which are $\pi/2$ apart in phase. Therefore dipoles require a power divider and a 90° phase shifter for feeding the two dipoles. The antenna transmits in circular polarization with a near omni-directional pattern in the azimuth plane and an elevation angle pattern that has a maximum in the Z-axis or 90° elevation. By bending the dipoles and adjusting their distance from the ground plane, it is possible to tilt the boresight angle in any desired direction for optimizing the dipole to operate at any specific elevation. Figure 5.3 is a diagram of such a dipole. The bandwidth of this type of antenna is relatively narrow.

The helical antenna consists of a number of turns of wire wrapped around a dielectric material or in air, mounted over a ground plane as shown in Figure 5.4.

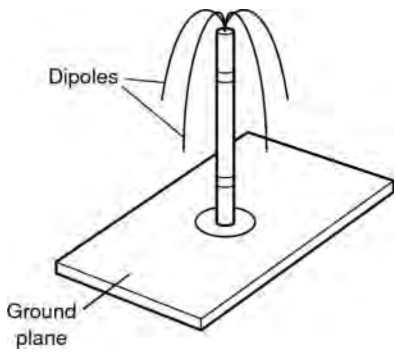


Figure 5.3 A crossed-dipole antenna

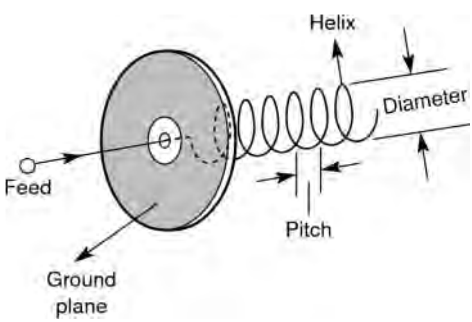


Figure 5.4 A helical antenna

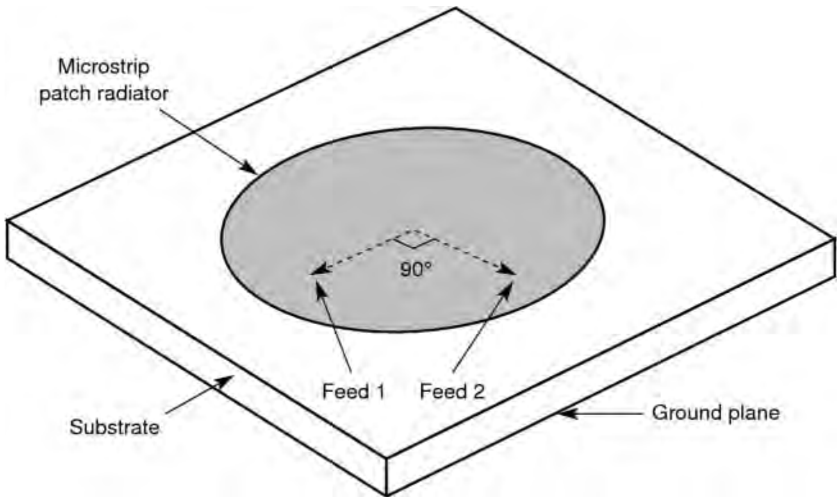


Figure 5.5 A circular patch antenna

The antenna operates with circularly polarized waves in the Z-axis direction (at right angles to the ground pole) and exhibits a wide bandwidth of around 200% with a medium gain. The gain and beam width of the antenna are proportional to the number of turns of the helix. A quadrifilar antenna comprises four helical antennas spaced equally around a cylinder. Typically, the ground plane has three times the diameter of the cylinder. The antenna exhibits a much wider bandwidth than a single helical antenna. The four helixes are fed with equal amplitude signal with phase shifts of 0, 90, 180 and 270°, which make the arrangement relatively complex. Antennas etched on microstrip, known as patch antennas, are useful when a low profile is essential, for example for vehicle mounting. Figure 5.5 depicts a single circular patch antenna for producing circular polarization.

The patch can be excited in basic or higher order mode from two feed points. The resonance frequency of the patch varies inversely with the radius of the circular patch and the relative dielectric constant of the substrate. For a substrate of dielectric constant 3, the patch diameter is 5 and 10 cm at 1 and 2 GHz, respectively. Higher order excitation can be used to maximize gain at any given elevation angle. Microstrip technology is suitable for mass production and offers a low-cost solution to personal communications.

Phased arrays comprise an antenna array in which the amplitude and phases of exciting signals are varied electronically to steer the main lobe. The boresight of a far field pattern depends on the amplitude and phases of excitation, which when changed in response to a tracking error signal affect an enormously agile and reliable tracking system.

The design of vehicular land mobile antennas is challenging as they must be compact, low profile and low cost. Various types of antennas have been studied for land mobile applications (Milne 1995; Shafai *et al.*, 1995).

In general, antenna gain, multipath, blockage, polarization characteristics and antenna noise temperature of low profile antennas degrade as the elevation angle is reduced.

Therefore, antenna gains are over designed at higher elevation angles, or alternatively, their gain is maximized in the elevation angle range where the operational satellite is likely to appear within the service area, a condition that suits a regional geostationary system well.

The gain of a low-profile antenna in the direction of the satellite depends on the effective projected area in the direction of the satellite and therefore varies as $\sin(\eta)$, where η is the elevation angle. In array antennas, commonly used in medium−gain terminals, electrical boundary conditions do not support transmission of circularly polarized waves close to the horizon. Furthermore, the antenna axial ratio and gain degrade at low elevation angles. For example, a low-profile antenna of two wavelengths in size has a maximum gain limit of 7 dBi and an axial ratio of 11 dB at an elevation of 15° (Milne, 1995). We have already mentioned the influence of antenna radiation on multipath.

Antenna performance is sensitive to vehicle structure and the height of the antenna above the ground plane. Reflection and diffraction, which depend on the curvature and effective area of the ground plane on the vehicle roof, cause ripples in the elevation antenna pattern. The magnitude of the ripple decreases as the elevation angle is increased and is more pronounced for large beam width antennas.

The antenna noise temperature of low-profile antennas depends on the elevation angle and environment around the antenna. Measured antenna temperatures at elevation angles of 30−60° range between 30 and 50 K when there is a clear line of sight, increasing to 50−85 K in the presence of the wooded skyline (Milne, 1995). The system noise temperature is ≈ 200 K when diplexer (DIP) and LNA units are mounted close to the antenna or made an integral part of the antenna assembly.

Portable antennas require maximum gain in the broadside direction without necessarily a low profile, in contrast to vehicle-mounted antennas, dealt in the preceding paragraphs, Microstrip radiators offer a viable solution. It is possible to achieve near omni-directional patterns and gains of 2−6 dBi from a single microstrip radiator, while medium gain can be achieved by an array.

Hand-held terminals require an omni-directional pattern in the azimuth within the elevation angle range of interest, for example 5−90° for geostationary satellites and 30−90° for a non-geostationary satellite system and typical gains of 0−1 dB with an axial ratio < 5 dB. Quadrifilar helix and patch antennas can meet such requirements. An L band quadrifilar helical antenna proposed for operation with a geostationary satellite is illustrated in Figure 5.6 (Caballero *et al*., 1995).

The half-wavelength helix has an axial length of 0.4−0.5 wavelength in air, a total height of 15.5 cm and an external diameter of 2.4 cm without a radome. The antenna is fed from the bottom through a wideband balun to balance the antenna and a matching circuit, comprising a quarter-wavelength transformer and a circuit with two parallel stubs. The head−antenna interaction is an important consideration in antennas used in hand-held terminals (Chuang, 1994; Toftgard *et al*., 1993). Studies show a 10 dB or more attenuation due to obstruction from the head and so the antennas were designed to operate above the head as shown in Figure 5.7.

In applications where a tracking antenna is necessary, either mechanical or electronic steering with phased array is used. Phased arrays, commonly used in aircraft, offer a low-profile solution giving high steering speed, adaptability to different environments, reliability and

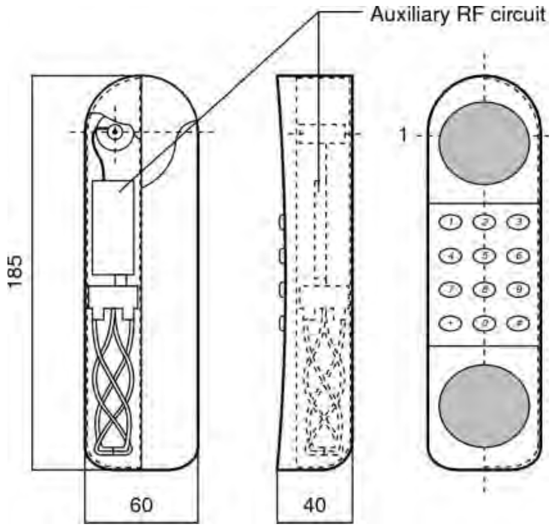


Figure 5.6 An L band quadrifilar helical antenna. (Source: Caballero *et al*., 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

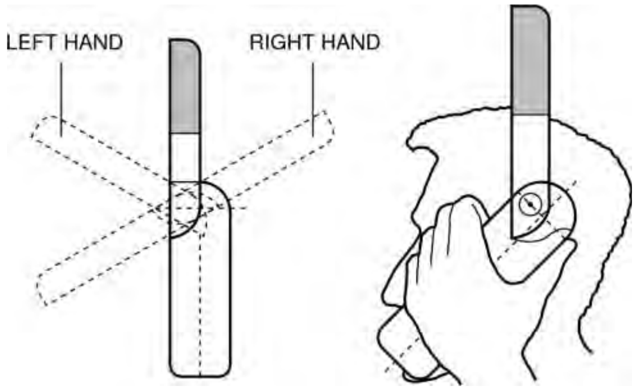


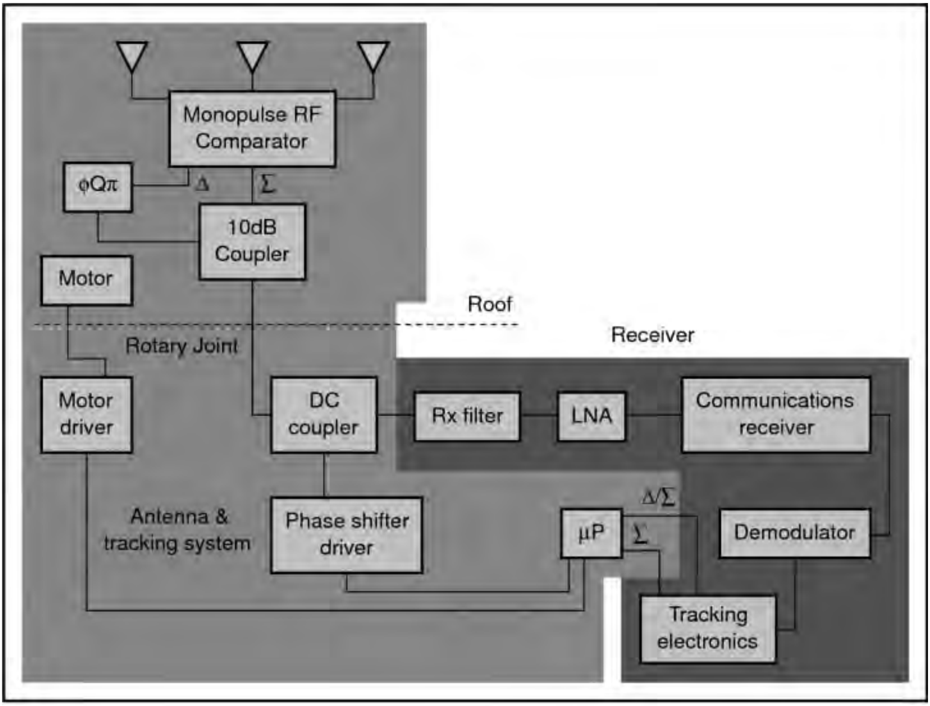
Figure 5.7 Antenna position during a call. (Source: Caballero *et al*., 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

lower power consumption than their mechanical counterparts, but at present the technology is not amenable to low-cost applications. Phase shifters used in phased arrays are lossy and therefore increase the antenna noise temperature that has to be compensated by increasing their size. An increase in the number of phase shifters such as PIN diodes tends to increase the cost. However, the antenna design lends itself to high volume, repeatable mass production with potentially low cost. At present the use of such antennas is limited to applications where cost is not the primary consideration. Low-loss mechanically steered antennas have an advantage with respect to cost. Typically such antennas comprise a planar microstrip array, which is mechanically steered in azimuth by low-profile stepper motors. For regional geostationary satellite systems, the elevation pattern can be shaped to the desired elevation angle range.

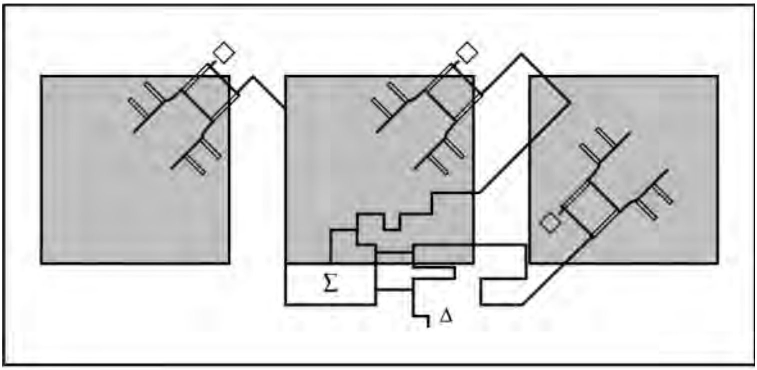
Tracking and acquisition of satellites in a mobile environment are demanding due to the vehicle motion and, additionally, satellite movement in non-geostationary satellite systems. Angular speeds of up to 60°/s and acceleration of more than $50°/s^2$ may be required necessitating updates at rates in excess of 6 Hz. A simple tracking system is necessary to maintain low cost. Step-track is a simple method where the tracking receiver maximizes the received signal level by stepping towards it using small exploratory steps but is unsuitable when signal level changes. The technique is unreliable in the presence of signal fluctuations and is characterized by a low response time and hence unsuitable for tracking rapidly moving satellites. Medium-gain (10–12 dBi) L band monopulse antennas for vehicle use have been developed by the ESA (European Space Agency) for a European mobile satellite system and MSAT programme (Garcia *et al*., 1993). The ESA design used microstrip technology for its potential of low cost and mass production while the MSAT design is based on a modulated monopulse scheme developed by JPL for the MSAT programme (Jet Propulsion Laboratory, 1988). In a conventional monopulse technique, two balanced receiver chains and a dual-channel rotary joint are required – one for the sum signal (Σ) and the other for a difference signal (Δ), which makes the system bulky and expensive. In the *modulated monopulse scheme*, the sum and difference signals are multiplexed, thereby removing the need for two receiver chains. The monopulse beam former produces the sum and difference signals that are fed into a monopulse modulator comprising a phase shifter and a directional coupler. The difference signal amplitude modulates a square wave and gets added to the sum signal in the directional coupler; the multiplexed signal comprises signals at fc ± fm, where fc is the centre frequency and fm is the modulation frequency. The signals are down-converted (DC), divided and detected separately to provide tracking correction in a microprocessor-based tracking processor. The error signals are used to drive stepper motors, which move the antennas to the desired direction in steps of 7.5° at speeds up to 40°/s. There is a provision in the tracking processor to obtain data from an external sensor to accommodate tracking when the signal is lost due to shadowing. Figure 5.8(a) shows a block diagram of the full system, including the receiver. The antenna comprises three patches providing a gain of 10–12 dBi etched on a low permittivity substrate to reduce antenna mass to ensure low inertia and reduce the number of elements. The circularly polarized patches are fed into the RF comparator to provide sum and difference signals as shown in Figure 5.8(b). The tracking system was used to track signals from Inmarsat (International Maritime Satellite) providing accuracies of 0.72–3.90° during test runs.

Open-loop techniques are not susceptible to signal fluctuations as they derive error signals from external sensors such as a flux-gate compass and gyrocompass or a vehicle's inertial navigation system. The error signals for each update are derived by taking the difference between the current and previous estimated position of the satellite and updating the antenna position based on the knowledge of the current vehicle position.

A magnetic compass needle is the simplest sensor, but it is inaccurate, sluggish, and more importantly without an electrical interface. A flux-gate sensor can estimate a north direction to an accuracy of a few degrees through a toroidal coil and a sensing coil. DC current is passed through a coil wound over a highly permeable core, causing the toroid to be



(a)



(b)

Figure 5.8 (a) A block diagram of the tracking system and (b) sum and difference signals. (Both parts source: Garcia, *et al*., 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

magnetized and produce magnetic flux. Two sensing coils are wound orthogonally to each other on the toroid – one coil to sense the north–south component and the other to sense the east–west component. Each coil comprises two oppositely wound segments at opposite ends. A current flows through the sensing coil whenever there is a difference in flux density between the two segments. The sensor has to be calibrated for a difference between true and magnetic north as well as for a vehicle's magnetic environment.

The fibre-optic gyro outputs angular velocity or a mobile's orientation using an optical phenomenon known as the *Sagnac effects*, which states that the phase difference produced between light waves travelling in opposite directions of an optical loop is proportional to the angular velocity of the loop. The sensitivity of the sensor depends on the area of the loop and therefore multi-wound coils are used in practice. The fibre-optic sensor is insensitive to acceleration, has a rapid response and a large dynamic range. A phase modulation-type optic sensor is a commonly used sensor. The sensor comprises a loop, a solid-state laser source, a phase modulator and a photodetector for phase detection (Ohmori *et al.*, 1998).

Open-loop tracking systems generally require some form of closed-loop tracking for initial acquisition. Step-track systems are often a cost-effective method of tracking for such a purpose. Combining both tracking methods provides a higher level of accuracy and reliability. The open-loop technique is established in aeronautical systems, as inertial sensors on-board aircraft are available, except in some smaller planes and helicopters where they are difficult to interface. In such cases, low-cost three-dimensional sensors are necessary. A low-cost technique for such applications developed by the Communications Research Centre proposes the use of a geomagnetic field together with satellite antenna pointing vectors (Sydor and Dufour, 1993). The magnetic field sensor provides the instantaneous geomagnetic field in a three-dimensional space by use of an orthogonal adaptive magnetic sensor. The raw output of the sensor is processed in a neural network that adapts the outputs to magnetic perturbations on the aircraft. The output provides changes to aircraft pitch, roll and yaw, which are mathematically transformed to satellite look angles. The correction is derived by estimating the difference between the wanted and actual look angles. Figure 5.9 illustrates the experimental set-up used for proof of this concept.

5.3.2 Hand-Held UT

A number of operators have targeted the hand-held telecommunication market in recent years with the expectation that satellite phones would be as appealing as terrestrial mobile phones in certain geographic areas and to a certain types of user. Typical services on offer include voice, circuit and packet-switched data, facsimile, paging and messaging while supplementary services include position location, conditional call diversion, caller identification, voice messaging, and others. Users are generally not interested in technology or systems, and expect a similar quality of service (QoS) on satellite systems as that offered by terrestrial mobile phones. However, as satellite systems are unable to offer coverage in heavily shadowed environments, identical service quality is not feasible under all conditions. It is anticipated that user cooperation and awareness is essential, so that users can derive maximum benefits from satellite technology. System designers on their part try to minimize the adverse impact of shadowing as much as is practical by using system features such as short messaging, alerting when the UT is heavily shadowed, adaptive power control, signal interleaving, coding and sending QoS alerts.

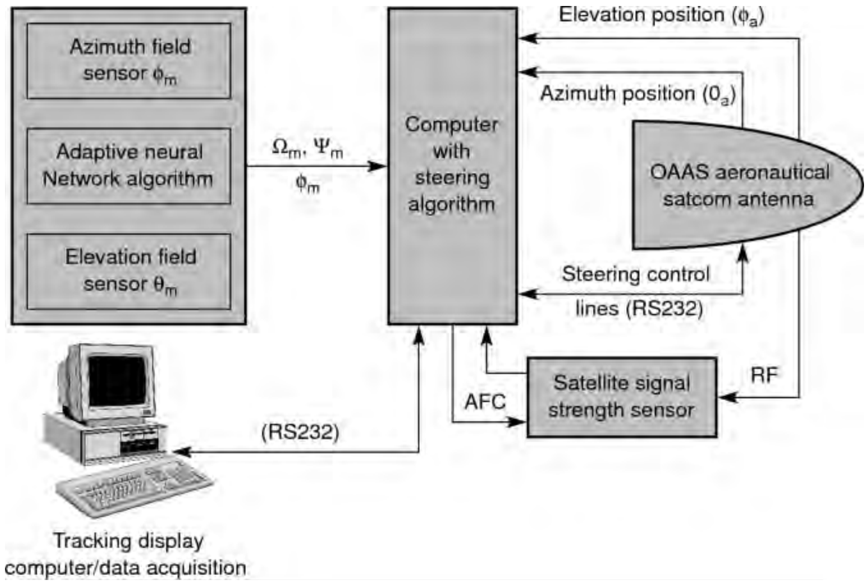


Figure 5.9 Experimental set-up of a low cost tracking system for aeronautical applications. (Source: Sydor and Dufour, 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/ Industry Canada and Jet Propulsion Laboratory/NASA.)

A number of unique requirements arise for the design and production of hand-held satellite phones. Some of the critical drivers are cost of the terminal, service, multimode support, user interface, aesthetics, prevailing fashion, size, weight, battery lifetime and antenna design.

The EIRP and G/T requirements are fundamental RF parameters that influence the size and cost of a UT. Hand-held UTs use small non-tracking antennas and transmitted power is limited due to the need to maintain transmitter power within radiation safety limits (see Section 5.4). Table 5.2 lists a set of RF requirements of hand-held and notebook sized units

Table 5.2 RF parameters of a TDMA terminal for operation with a HEO system

	Terminal type	
	Hand-held	Notebook
Antenna gain (dBi) (edge of coverage)	1–2	8.5
Average transmit power (mW)	<500	250
Peak EIRP (dB W)	7	8.5
G/T (dB/K) (edge of coverage)	−22.5	−15

(Source: Stojkovic and Alonso, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, co-sponsored by Communications Research Centre/ Industry Canada and Jet Propulsion Laboratory/NASA.)

Table 5.3 Typical service requirements of a hand-held terminal

Service	PSTN voice Satellite and terrestrial cellular support *Optional service* Duplex data (including Internet): up to 20 kbps/2.4 kbps voice codec Facsimile High penetration paging Navigation and location-based service
Band	L (1.5/1.6 GHz) and S (2.5 GHz)
Average transmit power	0.25 W (averaged over 6 min)
Antenna size and polarization	Diameter = 10 mm, height = 100 mm; circular polarization
Digital signal processing capability (MIPS: Million Instructions Per Second)	100
Telephone size	Less than 300 cc
Talk time between charges	8–10 h
Idle time between charges (standby)	100 h

(Source: Stojkovic and Alonso, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

for a proposed HEO (highly elliptical orbit) satellite system that provides high elevation angle visibility in Europe (Stojkovic and Alonso, 1995). Table 5.3 lists service and terminal requirements of a typical hand-held service.

The talk time for each battery recharge is an important practical consideration, especially for users who stay away from the base for a long time. Current available stand-by operation of about up to 100 h is available from satellite phones. To improve talk and stand-by time, battery drainage should be reduced in addition to improving battery technology. A power amplifier consumes a significant amount of battery power and improving its efficiency is vital.

The rate of improvement of battery technology is much slower than that of electronic components. Nickel cadmium (NiCd) used earlier are banned in many countries due to toxicity in cadmium and it has been superseded by Nickel metal hydride (NiMH) batteries wherein Cadmium has been replaced by a non-toxic alloy. A number of new technologies are under development. Lithium-ion is an emerging battery technology; although expensive at present, it has the potential to reduce battery sizes and weight considerably, offers a longer life time and is more environmentally friendly in terms of disposal. Further, improvements in energy density are possible when Lithium polymer batteries have been developed. Leakage of harmful electrolytes is not possible with this technology, as it uses a polymer instead of liquid electrolyte and enables cells to be shaped. Table 5.4 compares batteries of four technologies.

There is also interest in introducing replaceable alkaline batteries as the power requirements of personal communication products fall. This is already evident in pagers, many of which use a single replaceable battery. However, considerable power reduction will be necessary before it becomes a viable option in mobile phones.

Table 5.4 Comparison of three rechargeable battery technologies

	Nominal voltage (V)	Weight energy density (W h/kg)	Volumetric energy density (W h l^{-1})	Power (W/kg)	Efficiency (%)	Cycle life
						Number of cycles
Nickel cadmium	1.2	40–60	50–150	150	70–90	1500
Nickel metal hydride	1.2	30–80	140–300	250–1000	66	500–1000
Lithium-ion	3.6	150–250	250–360	1800	~99 (estimated)	1200–10000
Lithium-ion polymer	3.7	130–200	300	>3000	~99.8 (estimated)	500–1000

(Source: Data from http://en.wikipedia.org/wiki/Rechargeable_battery.)

In addition to improving battery life, manufacturers are improving battery power management by including features such as 'doze', 'sleep' and 'suspend', which reduce battery drainage, and introducing intelligent batteries with features such as visual communication of battery status to users or passing battery data on to host systems for intelligent management. Such features can extend battery life (e.g. the user can budget battery usage and reduce the probability of data loss in case of a low-battery shut down).

The complexity of a terminal is increased by a need to compensate for the Doppler effect and support active power control. The size, mass and cost of terminals are minimized by the use of VLSI (very large-scale integration) and mass production techniques. As mentioned, one vital consideration in an overall context is the degree of user cooperation built into the system. If users are aware of the limitations and willing to move to an unobstructed area during a call where the links are more reliable the system can operate with a modest link margin. In locations where satellites communication is the only source, the operator can expect a greater degree of cooperation. A plug-in outdoor antenna module can ensure indoor operation. More recently the use of a terrestrial retransmission scheme – known as Auxiliary Terrestrial Component (ATC) in the US and Complementary Ground Component (CGC) in Europe – have been proposed for MSS. In this scheme, multi-mode terminals switch to the terrestrial component in areas of dense population such as cities where satellite transmissions are heavily shadowed (see Chapter 7).

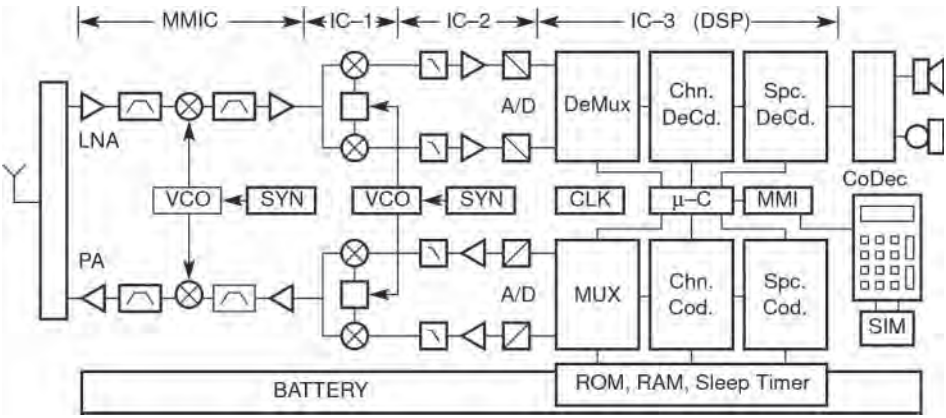
Due to the large number of supplementary features available in a mobile phone, it is essential that the human-machine interface be as simple as possible. Typically a modern satellite phone looks and feels similar to its terrestrial counterpart, which incorporates a small keypad with a graphical user interface; other features include voice-activated commands, a large memory, muting, auto-answer, hands-free operation, subscriber identity module (SIM) card support, integration with a personal digital assistant (PDA) or a personal computer for e-mail, web browsing, a built in camera, GPS receiver, radio, and so on.

5.3.2.1 Technology

Satellite phones manufacturers have tended to maximize reuse of terrestrial mobile technology to benefit from economies of scale. The discussion in this section addresses those technologies where commonalities are possible and evident. A multimode receiver comprises a common front end, followed by a programmable modulator/demodulator (modem) and baseband units. Several personal satellite communication systems operating in L and S bands have been developed to operate with terrestrial systems such as GSM/GPRS (global system for mobile/general packet radio service and/or CDMA (code division multiple access) systems in various parts of the world.

A possible architecture of a hand-held terminal is illustrated in Figure 5.10(a). Typical antenna sizes of the first generation satellite phones are about 100 mm in height and 10 mm in diameter, and typical EIRP and G/T requirements are about 500 mW and −25 dB/K, respectively (see Section 3.4). Figure 5.10(b) illustrates Inmarsat's handheld telephone for voice and data service. A DIP is a critical unit in the front end of a receiver as it isolates transmit and receive chain but it also introduces a loss at a critical point of the chain where loss must be minimized. The power amplifier and LNA use monolithic microwave integrated circuit (MMIC) technology and it is possible to integrate the entire RF section in a single MMIC.

Modem, voice processing including voice recognition, user and peripheral interface control and housekeeping functions are best performed by digital signal processing (DSP) chips



(a)

(b)

Figure 5.10 (a) An architecture of a hand-held terminal (Source: Stojkovic, I. and Alonso, J.E. (1995). IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.) and (b) a handheld unit for operation with a geostationary MSS. (Source: Inmarsat. Reproduced with permission of Inmarsat.)

and microcontrollers. The DSP chips require highly linear analogue-to-digital converters with a large dynamic range. The attempt is to convert analogue signals to digital as close as possible to the RF, which brings digital processing advantage at an early stage.

Battery technology is important, not only for capacity but also because of weight considerations and other less obvious needs such as environment friendliness. The sleep mode is necessary to conserve battery power where only the essential subsystems are active. With the current state of technology, battery weight is expected to dominate the handset weight.

In terrestrial systems, handsets have reached the lower limit of size. The chips are amongst the smallest part of the set in comparison to the size of the display or battery. However, further reduction in battery drainage can be achieved by reducing the power requirements of chips. Reducing the number of chips improves power consumption, as more is lost in the PCB (Printed Circuit Board) interconnections than within the chip itself. Furthermore, by integrating more functions in chips, chip packaging and fabrication costs reduce. Another approach is to reduce supply voltage, which increases talk and standby times. Alternatively, retaining the same battery supply, more circuits can be packed into the same volume to improve performance and functionalities by way of more flexible software.

Let us consider some implementation scenarios noting that the technology is evolving rapidly and manufacturers tend to prefer propriety solutions. Typically, two to three chips may be used in both RF and baseband in terrestrial cell phones. The intention would be to introduce single-chip design using 0.35 μm CMOS (Complementary-metal-oxide-semiconductor), which would allow easy customization as well as reduction in power. Making changes to the architecture of a customized design is easier if only a single chip is used. Increasing the number of chips requires more changes and hence increases the cost and turn-around time. Customization is often used by smaller phone manufacturers in order to distinguish their product for marketing. On the RF side, the effort is to reduce the number of fixed components such as resistors and capacitors down to say, 100 from 300, making the RF part cheaper and easier to assemble. Various approaches are used for signal processing of analogue and digital signals. Some companies use separate chips for each; others mix functions. Separate digital and analogue signals avoid cross-talk and allow an independent process for each. CMOS technology, which has a leading edge, is easier to implement by separating analogue and digital technologies (Agrawal, 1998).

Technologies for integration of different RF bands such as GSM (900 MHz) and DCS 1800 (1800 MHz) as well as combination of different modes (GSM and DECT) is established. This occurred in two stages. In the first stage – similar to the integration of PDA with a mobile phone – telephones of two standards were assembled together on a single package. The goal was to integrate them at a chip level. Integration of DCS (Digital Cellular System) with GSM was less problematic as baseband technologies are similar, the main difference was in RF; whereas, integration of GSM with DECT (Digital Enhanced Cordless Telecommunications) was more challenging due to their entirely different technologies. Problems in integrating PDA with GSM included incompatibility of PDA operating systems with GSM, which operates in real time, and need for higher voltage to drive the PDA LCD (Liquid-Crystal Display) that negated the battery saving features of a GSM set. Manufacturers have also integrated on a single chip baseband processing with PDA functions and GPS receiver to provide new functionality. Evolution of such hybrid terrestrial technologies has led to the introduction of dual mode satellite/terrestrial terminals.

Chip level integration of satellite/terrestrial standards has been demonstrated. US companies Infineon Technologies, SkyTerra and TerreStar Networks are reported to have developed a multi-standard mobile platform based on Infineon's software defined radio (SDR) technology that use mass-market devices costing about the same as terrestrial cellular-only devices to operate with multiple cellular and satellite-based communications technologies including GSM, GPRS, EDGE (Enhanced Data Rate for GSM Evolution), WCDMA (wideband code division multiple access), HSDPA (High-Speed Downlink Packet Access), HSUPA (High-Speed Uplink Packet Access) and GMR1-2G/3G (Newsrelease, 1 April 2009, www.infineon.com)

Rapid advancements in cellular technology have led to very strong competition between manufacturers and rapidly changing handset designs for various reasons including higher performance and quality and because of changes in customer taste. A similar trend has been observed in satellite phone markets. These changes place a heavy demand on manufacturers and in particular those in the VLSI area where typical turnaround times for chip manufacture are about 9–12 months. Each development poses stringent requirements on cost, size and power consumption and the need to deal with multiple standards. The typical obsolescence time of a chip reduced from two to three years in the mid-1990s to less than a year by 2000. One technique to manage such rapid changes is the use of development platforms, which can provide a low-risk, rapid route for chip development.

A microprocessor card known as subscriber identity module is the personalized part of a UT. The SIM provides user authentication, radio transmission security and stores user data. The entire subscriber's information, which the network authenticates for granting network access, is security protected. The SIM card can be ported across terminals, giving the user the flexibility to migrate between terminals, as the network only authenticates the data stored in the SIM as the user's subscription account.

Figure 5.11 shows a breakdown cost of material and manufacture for a 16 GByte memory mobile smartphone available in 2012 timeframe using 'teardown' data from a mobile phone obtained by IHS iSupply Research (www.isuppli.com). The display and touch screen costs dominate and with memory and wireless sections constitute over 50% of the total costs. The reader should take this example as indicative only, as such data depend on the manufacturers' preferences, models' capacities and are sensitive to the products' lifecycles.

Examples of Message Signals

Mobile terminals are used for a variety of communications, which include voice, facsimile, e-mail, file transfer, compressed video, web content, and others. MSS systems are primarily digital with a provision to interface with legacy analogue terrestrial systems for backward compatibility.

In a telephone network data are sent through a modem. The incoming data stream is modulated to a format compatible for the telephone line; at the receiver signals are demodulated to retrieve data. To harmonize data transport over the telephone networks the International Telecommunications Union-Telecommunication (ITU-T – formerly, International Telegraph and Telephone Consultative Committee or CCITT). specifies a range of protocols that cover analogue and digital modems and interface in V-Series Recommendations (Data communication over the telephone network) (ITU-T, 2013). Recommendations V.10–V.34 specify *voice band modems* and *their interfaces*. For example, the V.34 standard supports up to 33.8 kbitps bidirectional data transfer. Recommendations 35–V.39 apply to *wideband modems. V*.35 is a layer 1 standard to support speeds up to is 2 Mbps. Recommendations V.40–V.49 specify *error control* and *data compression* methods. Recommendations

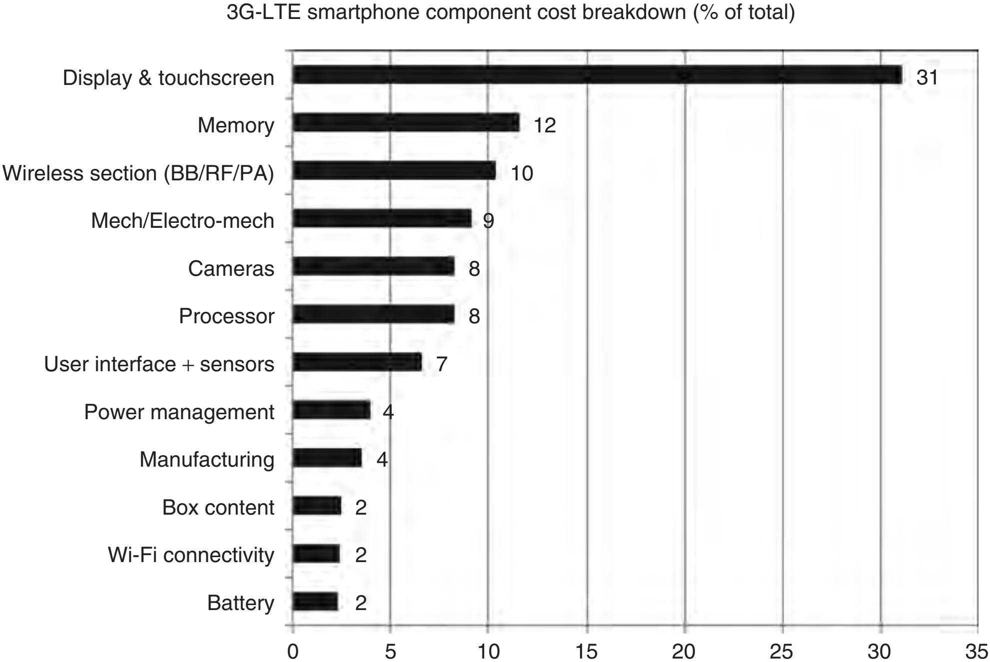


Figure 5.11 Cost of material and manufacture of a typical smartphone supported on the third generation (3G) and Long Term Evolution (LTE) networks obtained through a teardown analysis. (Data source: IHS iSupply Research – www.isuppli.com.) The LTE network is recognized by the ITU as one of the fourth generation cellular technologies

V.60–V.99 cover s*imultaneous transmission of data* and *other signals. Interworking with other networks* is covered in recommendations V.100–V.199. The recommendation V.110 covers terminal adaptor functions for the connection of terminals having interfaces conforming to V-series to the ISDN (Integrated Services Digital Network). Recommendations V.400–V.499 cover *digital circuit modems*.

Similarly, facsimile (fax) standards are available. The ITU-T adopted the Group 3 standard (in 1980) for the digital transmission of images through modems, which provided a threefold improvement in speed compared to Group 2 standard machines. Originally, the Group 3 fax was developed for transmission of 2.4–9.6 kbps data. In 1984, the ITU-T adopted a Group 4 fax standard to supersede the Group 3 fax by permitting error-free transmission of documents over digital networks, such as ISDN, at speeds up to 64 kbps. At such rates, transmission time for a single page is less than 10 s. Despite competition from Internet, facsimile is still widely use.

Electronic-mail or e-mail comprises text, graphics and sometimes sounds and animated images exchanged between users through a computer network, Internet being a prime example. E-mail is used most extensively by remote mobile users who may also access their company intranet via the Internet. E-mail systems are now an indispensable part of business and personal communications. The X.400 suite of ITU-T Recommendations define standards for Data Communication Networks for Message Handling Systems.

Internet e-mails follow Internet standards specified by the Internet Engineering Task Force (IETF). Examples of commonly used e-mail protocols include Simple Mail Transfer Protocol (SMTP) or RFC 5321, a protocol for transmission of e-mail; Post Office Protocol 3 (POP3) and Internet Message Access Protocol (IMAP4) are e-mail retrieval protocols used by e-mail clients. A basic format of e-mail is available in RFC 5322.[Note: Request for Comments (RFC) is a memorandum published by the IETF, describing methods applicable to the working of the Internet].

Voice Coding

Bandwidth efficiency is paramount in an MSS system. Traditional voice coding techniques such as pulse code modulation (PCM) or adaptive differential pulse code modulation (ADPCM) are too inefficient for the purpose, offering bit rates in the range 16–64 kbps. *Waveform coding* techniques attempt to reproduce the transmitted waveform by sampling, and coding the sampled signals; hence there is a lower limit on the sampling speed, in accordance with the Carson rule. By contrast, *source coding or parametric coding* techniques extract specific characteristics of speech for digital coding. Speech consists of a sound-carrier that is modulated by the intelligence component. The bandwidths of speech parameters are narrower, which allows the coding rate to reduce in the range 0.04–4 kbps. A hybrid method combines the best elements of waveform and source coding.

In general, the complexity of the coding technique increases as the coded bit rate is reduced. DSP chips reduce the cost to acceptable limits. MSS voice telephony systems operate at a coding rate in the range 2.4–8 kbps allowing RF bandwidth of 5 kHz. In linear predictive coding (LPC), commonly used in MSS, speech is modelled as a time-varying linear predictive filter. Examples of speech LPC coding techniques that have been used in mobile communications are adaptive predictive coding with maximum likelihood quantization (APC-MLQ) at 16 kbps, which is used in the Inmarsat-B (legacy) system; multi-pulse excited LPC (MPE-LPC) at a bit rate of 9.6 kbps, used in the Inmarsat aeronautical system; improved multi-band excitation (IMBE); code-excited linear prediction (CELP) and its family, such as VSELP (Vector Sum Excited Linear Prediction), used in the IS-54 North American system at 7.95 and 6.7 kbps, and PSI-CELP (Pitch Synchronous Innovation-Code Excited Linear Prediction) at 3.45 kbps in the Japanese PDC (Personal Digital Cellular) system; regular pulse excitation and long-term prediction (RPE-LTP) used in the GSM system at 13 kbps, etc. Recent enhancements to algorithms such as CELP have brought down the bit rate to 2.4 kbps.

A measure of distortion in the digitized version of speech is quantization distortion caused by speech digitization, measured as the quantization distortion unit (QDU). One unit is the equivalent of distortion resulting from converting speech to a 64 kbps PCM and then reconverting it back. QDU performance standards have been recommended for system designers, for example ITU specifies a limit of 14 QDU for any international connection. The distortion is distributed across the network, for example in a 5-4-5 distribution; the originating country is allowed 5 QDU, the international transit 4 and the terminating country 5. In an international MSS, multiple countries and operators are often involved. Furthermore, the interaction of disparate low bit rate codecs in tandem is not yet fully understood and hence existing standards do not necessarily address all aspects of mobile systems. New concepts, such as the introduction of an impairment factor in ITU-TG.113, are being introduced to address such concerns.

Effort is under way to develop a generic codec by programming a number of codes on a single DSP, enabling access to multiple networks. It is feasible to adjust the coding rate to suit channel conditions – thereby improving the channel capacity by adjusting the code rate according to channel conditions.

As the performance of voice coders degrades due to increases in bit error, one consideration in the selection of a codec is its susceptibility to errors. Noise rejection of the handset is another important consideration. At the receiver, local noise can enter through the air gaps around the earpiece (and the other ear!). At the transmit end, noise entering along with speech can cause problems for a number of handset components – voice codec, voice activity sensors, speech interpolation systems and echo suppressors, amongst others. A voice activity detector can be falsely triggered, causing unnecessary RF transmission. Hands-free terminals have a compounded problem due to significantly higher noise pick-up. A noise reduction algorithm with echo-cancellation reduces the problem at the expense of processing delay. It is important that during type approval, an opportune time to capture such problems, mobiles are tested in environments close to reality.

5.3.3 Mobile Terminals

5.3.3.1 Ship-Borne Terminals

Ship-borne terminals were the first to be introduced for civilian use, as maritime safety was the dominant theme when the first civilian MSS was introduced in 1980s. Depending on the size of the ships and communication requirements a variety of ship-borne terminals can be used. Large ships operating in the L band can deploy parabolic dishes typically 0.8–1.2 m, whereas yachts and small boats accommodate small low-gain antennas.

Two examples of ship earth stations are presented here. Inmarsat-B system was introduced around 1993 to provide voice and digital communications at bit rates up to 64 kbps including safety services to ships via Inmarsat's global network of satellites and land earth stations. Its front-end specification is reusable in the next generation products and hence its RF part remains representative. Inmarsat-C system, introduced in 1991 to support data communication including safety services at bit rates up to 600 bps to small ships and land mobiles, continues to be in use. Inmarsat-B terminals use large tracking antennas around 1 m in diameter, whereas Inmarsat-C terminals are lightweight with small non-tracking omni-directional antennas. Some of the main characteristics of these terminals are summarized in Table 5.5 (Inmarsat, 1991(a) and (b)).

Large Terminals

There are various versions of the Inmarsat-B system – land transportable, vehicle mounted, ship-borne, and so on. In general large ship-borne terminals comprise an outdoor deck unit consisting of a stabilized, tracking antenna system to which essential RF electronics are connected; and a below-deck indoor unit where the main signal processing is performed and to which peripherals such as a telephone, personal computer, and so on are attached either as its integral part or as a plug-in. A block diagram of the main units of such a ship earth station is shown in Figure 5.12.

The outdoor unit comprises a parabolic tracking antenna mounted on a stabilized antenna platform. The antenna is housed in a radome for protection against the extremely harsh weather conditions experienced on ships. The electronics comprise subsystems that must

Table 5.5 Main characteristics of Inmarsat-B and Inmarsat-C terminals (Inmarsat, 1991a,b)

Parameter	Specifications	
	Inmarsat-B	Inmarsat-C
Transmit frequency range (MHz)	1626.5–1646.5	1631.5–1646.5
Receive frequency range (MHz)	1525–1545	1530–1545
Transmit EIRP (dB W)	25–33	14
Receive G/T (dB/K)	−4	−23
Typical antenna size (m)	0.9	0.1
Typical high power amplifier type	Class-C	Class-C
SCPC channel pairing	No	No
Synthesizer step (kHz)	10	2.5
Data service (kbps)	9.6/16/64	0.6
Voice-band data (kbps)	Up to 2.4	Not applicable (NA)
Voice coding	16 kbps; APC	NA
Voice channel modulation/coding	O-QPSK; ¾ FEC – constraint length = 7	NA
Voice channel transmission rate (kbps)	24	NA
Voice channel bandwidth (kHz)	20	NA
Forward voice activation	Yes	NA
Forward power control	Yes	NA
Satellite forward link EIRP (SCPC) (dB W)	16	21 (messaging/signalling)
Telex channel (forward link) access/modulation; coding; transmission rate (kbps)	TDM/BPSK; ½ FEC; 6	NA
Telex channel (return link) access/modulation; coding:transmission rate (kbps)	TDMA/O-QPSK; ½ FEC;24	TDM/BPSK; ½ FEC; 0.6
Forward signalling channel characteristics	As Telex	TDM/BPSK; ½ FEC; 0.6 (messaging/signalling)
Return request channel access/modulation; coding:transmission rate (kbps)	Aloha; O-QPSK; ½ FEC; 24	Aloha; BPSK; ½ FEC; 0.6

O-QPSK, offset-quadrature phase shift keying; BPSK, binary phase shift keying.
(Data source: Inmarsat, 1991a and Inmarsat, 1991b.)

be mounted close to the antenna to minimize insertion loss, that includes DIP, low-noise amplifier, down-converter, high-power amplifier and electric motors for tracking. Usually they are housed on the antenna mount. Antenna tracking and stabilization are maintained through feedback from the ship's motion sensors and a tracking receiver. A step-track system is commonly used for satellite tracking. A four-axis stabilization system comprises an azimuth/elevation axis mounted on the X−Y axis. Motion sensors comprising accelerometers, rate and level sensors provide roll and pitch information. In another type of stabilization scheme, flywheels are used to provide the inertia for maintaining antenna stabilization, but flywheel systems are complex and heavy. A three-axis stabilization system is also used in some traditional designs (Hoshikawa, 1988).

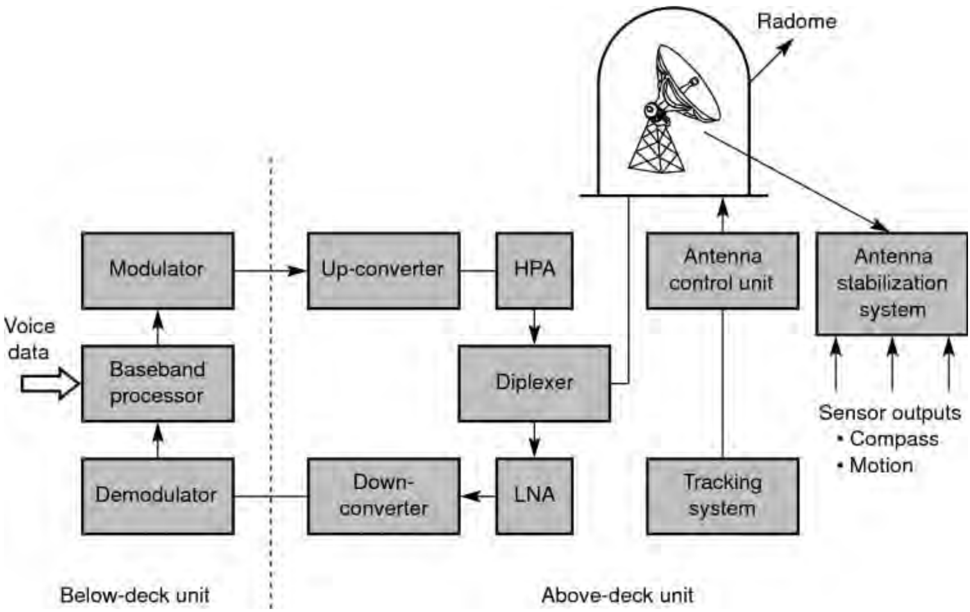


Figure 5.12 Main functional entities of a ship earth station

The main components of a channel unit housed in the below-deck equipment are shown in Figure 5.13. The scrambler removes discrete spectral lines from transmissions by using a pseudo-random signal. The initial state of the scrambler shift register is changed randomly during each call set-up to improve security and privacy.

The Inmarsat-B terminal can support frequency division demand assigned voice and data services. The voice service uses APC (adaptive predictive coder) at 16 kbps; transmissions are voice-activated, and accessed using demand assigned frequency division multiplexed 20 kHz spaced SCPC channels. The frequency synthesizer can tune in steps of 10 kHz for SCPC channel assignments within the ranges shown in Table 5.5. Telex communication is received from a TDM broadcast channel at 6 kbps and transmitted via a 24 kbps TDMA return channel. Signalling and bulletin board operate at 6 kbps and return signalling is transmitted at 24 kbps. Signalling and terminal control are managed by a microprocessor-controlled unit. Class-C non-linear power amplifiers are used to achieve high efficiency.

Small Terminals

The Inmarsat-C system was developed for low cost, low bit rate (600 bps) communication services to small ships, yachts and land vehicles; the system was later adapted to operate in aircrafts. The main characteristics of the terminal are summarized in Table 5.5.

The Inmarsat-C UT, weighing 1–4 kg, uses a store and forward transmission scheme as well as virtual circuit-mode connections. In the store and forward mode, messages are formatted and transmitted in a simplex mode at a time when a channel becomes available. In the virtual-circuit mode, a permanent or a semi-permanent connection is established for a

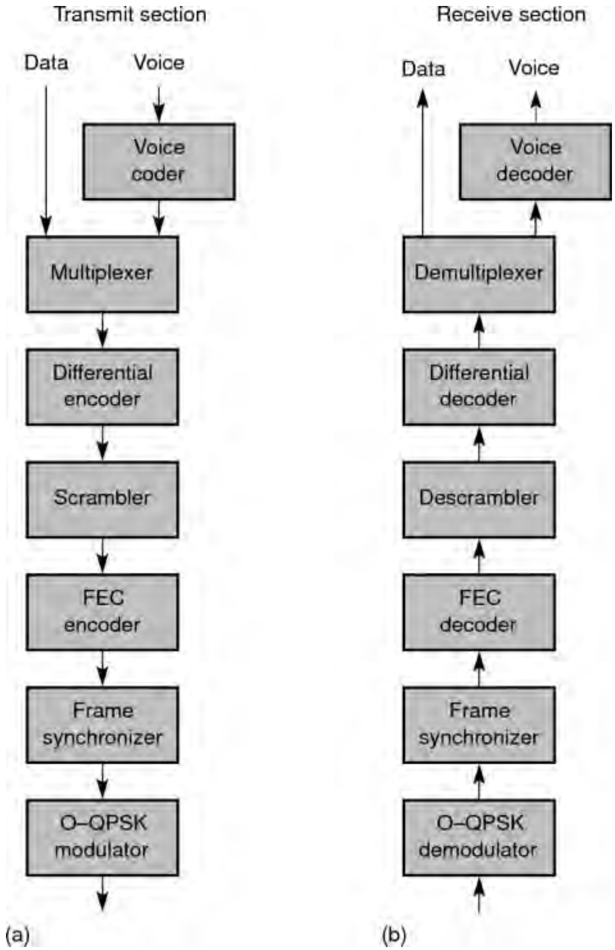


Figure 5.13 Main radio components of a below deck equipment

call. Other features of the system allow duplex circuits, polling, data reporting and group call reception. The transmission rate is 1200 symbols/s with half rate FEC convolution coding. As terminals deploy omni-directional antennas, multipath noise is severe. Interleaving with ½ rate coding is used as a countermeasure to multipath, while an ARQ (Automatic Repeat Request) scheme is used to combat deep fades that break continuity.

Due to their light weight, terminals can be easily mounted on vehicles and vessels; hand-carried versions are also available; there is usually a standard interface for connecting to a personal computer. Some models incorporate a built-in console for creating messages. Figure 5.14 (Inmarsat, 1988) represents a block diagram of an Inmarsat-C mobile terminal. The terminal comprises two main functional parts: data terminal equipment (DTE) and data communications equipment (DCE). The DTE is used to format messages and transfer them to the DCE for transmission to the network. The DCE stores the message until a channel is available for transmission. A microprocessor is used for various control purpose; a scrambler

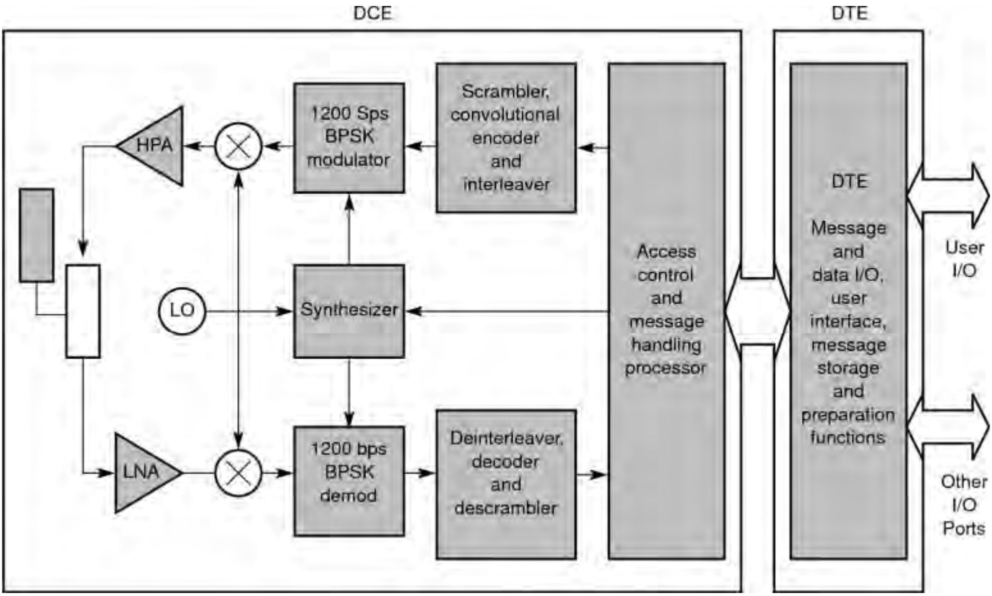


Figure 5.14 Inmarsat-C terminal block diagram. (Source: Inmarsat, 1988. Reproduced with permission of Inmarsat.)

is used for dispersing the signal; and a frequency synthesizer tunes the RF in steps of 2.5 kHz. A reverse operation is performed in a receiver.

Various types of omni-directional antennas are used, such as quadrifilar, helix, crossed drooping dipole and microstrip patch. Quadrifilar antennas, protected by a radome, are commonly used for ship terminals and microstrip patches are used where a low profile is desirable.

The terminal receives network information from a TDM broadcast transmitted by an NCS. A number of gateways are deployed in the network. These gateway stations broadcast signalling and messages signals to mobiles, while the NCS manages the network in entirety. Mobile terminals use the Slotted Aloha scheme to access the network.

5.3.3.2 Land-Mobile and Portable Sets

Due to price sensitivity of a land mobile service, the UT and service costs are of paramount importance for the success of a land mobile service. Vehicle-mounted, transportable, railway and portable sets belong to this category. Figure 5.15(a–d) illustrates a variety of applications of BGAN (Broadband Global Area Network) land portable UTs.

At present, land mobile and portable sets can provide throughput ranging from a few hundred bits per second to 500 kbps, depending on their antenna size, the space segment capability of the MSS and the market segment. Portable and vehicle-mounted land-mobile sets were introduced in the early 1990s as a part of the MSS technological evolution, and while their size continues to shrink, throughput has increased up to 500 kbps.

Figure 5.15 A user portable user terminal used in various applications: (a) enterprise, (b) aid, (c) government and (d) construction. (Source: Inmarsat. Reproduced with permission of Inmarsat.)

The design of such sets is technically challenging due to stringent size limitations and the severe propagation conditions experienced on land. Omni-directional or near omni-directional antennas with antenna gain $< 3-4$ dBi are lightweight and low cost with limited throughput, whereas medium-gain antennas with gains of the order of $4-15$ dBi are complex when tracking is incorporated, heavier and more expensive but can provide higher throughput. Demands on the complexity and functionality of electronic units are similar to those of other types of mobile terminals. In the remaining part of the section, a terminal of this category, capable of being used in a number of environments is described (Fuji *et al*., 1995), followed by design considerations of a portable multimedia UT used in Inmarsat's BGAN is described. This is followed by description of K_a band terminals used in the ACTS (Advanced Communications Technology Satellite) programme to illustrate an architecture of the next generation mobile systems (Abbe *et al*., 1995).

The mobile terminal developed to operate in the AMSC/TMI network (American Mobile Satellite Consortium/Telesat Mobile Incorporation) (a first generation system) has the flexibility to operate on land or marine mobiles, as a transportable or fixed site installation, and deploys a medium-gain or a high-gain antenna system, as necessary. The specification of the mobile terminal is summarized in Table 5.6.

The terminal can support circuit-switched voice and data services at 2.4 and 4.8 kbps respectively, mobile roaming, net radio and Group 3 facsimile services. A cellular portable phone can be plugged in to provide MSAT voice services. The terminal has a minimum EIRP of 15.5 dB W and a minimum G/T of −16 and −12 dB/K with medium- and high-gain antennas respectively (see Table 5.7).

The terminal comprises an antenna, a trans-receiver unit and a hand-held phone. Various types of antennas are used, each suited to a specific category, as listed in Table 5.7. Disc and dome antennas are used for roof-mounted applications on land; transportable antenna systems for trucking applications; flat antennas for briefcase sized sets; and fixed antennas for remote fixed location applications.

Table 5.6 Characteristics of mobile terminal

Parameter	Specifications
Transmit band (MHz)	1 626.5–1 660.5
Receive band (MHz)	1 525–1 559
Channel spacing (kHz)	6
Channel tuning increment (kHz)	0.5
Channel rate (kbps)	6.75
Modulation/filtering	QPSK/60% cosine roll-off
Scrambling	15 stage PN sequence (ITU-R report 384-5)
FEC coding; decoding	½ and ¾ punctured FEC, K = 7; Viterbi
Signalling	
forward; access/modulation/rate (kbps)	TDM/DQPSK/6.75
return; access/modulation/rate (kbps)	Random access/DBPSK/3.375 TDMA/DBPSK/3.375
Data rate (kbps)	2.4 or 4.8
Interface	
handset – voice	4-W with serial async keypad and LCD signals
data interface	RS232 C (CCITT V24 or V28)
facsimile interface	CCITT group 3, 2.4 kbps voice band signal
IB interface	CCITT X.25
Power supply (V) volt/amps	12 (Range 11–16)/6
Size/weight trans-receiver unit	12″(D) × 8″(W) × 2″(H) 5 lbs

DBPSK, differential binary phase shift keying; DQPSK, differential quadrature phase shift keying.
(Source: Fuji *et al*., 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

Table 5.7 Characteristics of various types of front end

Antenna type	Disk	Dome	Transportation	Fixed site	Transportable
Approximate size and weight: cm (inches) kg (lbs)	Diameter = 34.8 (13.7)	Diameter = 17.3 (6.8)	Diameter = 29.2 (11.5)	Width = 50.0 (19.7)	Width = 35.6 (14)
	Height = 4.8 (1.9)	Height = 16.8 (6.6)	Height = 18.3 (7.2)	Depth 50.0 = (19.7)	Depth = 35.6 (14)
	Weight = 2.4 (5.3)	Weight = 1.4 (3)	Weight = 2.5 (5.5)	Height = 5.1 (2)	Height = 10.9 (4.3)
				Weight = 6.4 (14)	Weight = 9.1 (20)
Maximum EIRP (dB W)	16.5	16.5	15.5	16.5	16.5
Receive G/T (dB/K)	−16	−16	−13	−10	−12
Gain category	Medium	Medium	High	High	High
Elevation angle range (°)	25−60	15−60	25−60	5−90	15−90

(Source: Fuji *et al.*, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

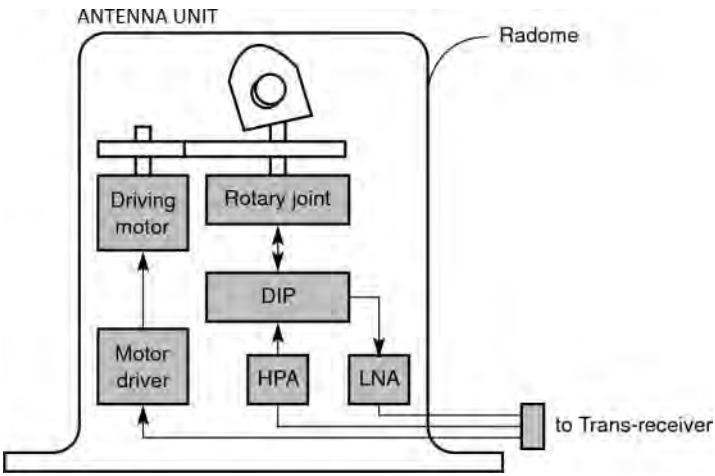


Figure 5.16 Main blocks of a dome antenna unit. (Source: Fuji *et al.*, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

The main components of a dome-type antenna unit are shown in Figure 5.16. The antenna comprises an axial mode helical antenna mounted on an inclined plane to serve the desired elevation angle range. Other components of antenna units are a rotary joint, a LNA, HPA (high power amplifier), DIP and motor for tracking.

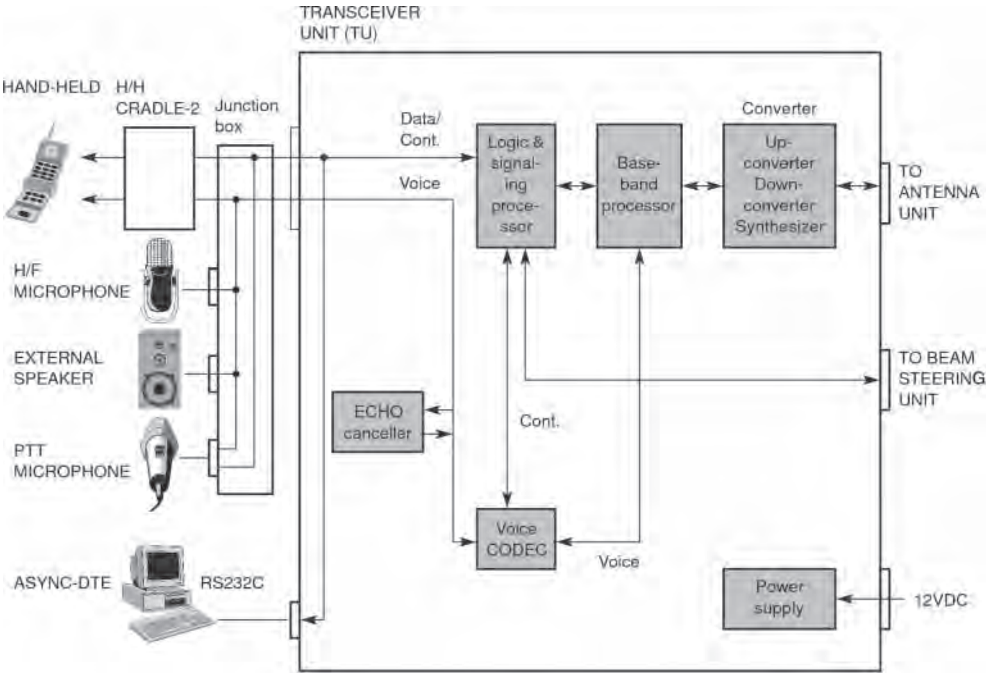


Figure 5.17 The main units of trans-receivers. (Source: Fuji *et al.*, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

The main subsystem of a trans-receiver unit is shown in Figure 5.17.

The unit, common to all types of antenna, comprises up and down frequency converters, a baseband processor, logic and signal processor, a power supply and an echo canceller for hands-free operation. The frequency synthesizer is a dual tuning type phase lock loop synthesizer driven by a direct digital synthesizer with low spurious and phase noise characteristics. The baseband processor, comprising a digital signal processor, performs modulation/demodulation plus framing/deframing. The demodulator uses differential detection with maximum likelihood sequence estimation with a Viterbi algorithm. The voice codec uses the IMBE technique that provides 6.4 kbps data. The logic and signalling processor controls the main functions of the mobile terminals, including network access and control. The junction box is an interface to a number of peripherals, including a cradle for hand-held telephone.

In Chapter 11 we will introduce Inmarsat's L band BGAN system, which provides data rate up to about ½ Mbps to portable and mobile UTs from an L band fourth generation geostationary satellite system representing the current state-of-the-art technology. A number of technical considerations were used in developing UTs for this network. An attempt was made to maximize the reuse of the front end (antenna and LNA) of previous generations for the maritime and aeronautical UTs. A radically new approach was necessary for the remaining parts of the UT to support a dynamically adaptable air interface based on variable modulation and coding schemes with modulation scheme extending from QPSK (quadrature phase shift keying) up to 16QAM (quadrature amplitude modulation) together with powerful and adaptable turbo-codes; and the need to support circuit and packet mode transmission in presence of multi-path fading and other impairments. Portable UTs had to be battery operated, low weight and low cost. Here we will highlight the major technical issues to be considered in the design of such a portable multimedia UT (Skinnemoen, 2003).

The UT antenna size and amplifiers are chosen such that the EIRP, G/T and linearity requirements of the system are satisfied within the cost target constrained by the limited volume of production. The UT includes capabilities to: decode dynamically changing modulation-coding schemes; manage the associated radio resource management functions; support power saving features such as sleep mode. Furthermore, the UT should incorporate: interfaces such as Bluetooth, USB (Universal Serial Bus), LAN (Local Area Network) or WLAN (Wireless LAN); a position location function using GPS; efficient voice processing algorithm as specified in the system; and robust synchronization and acquisition techniques to counteract channel impairments.

The antenna size is limited to 10–12 cm due to the limited size of the UT with a need to support both transmit and receive bands. A single patch antenna provides about 7.7 dBi gain and results in a G/T of approximately −18.5 dB/K. This should be adequate to support an adaptable air interface that has a dynamic power variation of the order of 15 dB (Satellite EIRP ranges ~ 25–40 dB W towards the mobile). The EIRP of the terminal is estimated as 10 dBW to support the target throughput of a few hundred kbps so that the HPA can be sized around 2 W.

Since the production volumes are low in comparison to terrestrial mobile phones, development of new chip-set are not economically viable and hence the design has to be based on commercially available chip with a provision for system-on-chip or ASIC (application-specific integrated circuit) design. Thus the hardware would comprise an (analogue) RF part with a digital baseband. To enable indoor operation there should be a provision to attach an external antenna. The software would take into consideration the propagation delay of the order of 250 ms and include features to minimize log-on with minimum exchanges. Since spot beam sizes are hundreds of kilometres wide a GPS receiver is integrated in the receiver; additionally this feature is necessary to support the legal requirement imposed in many countries and the system includes provision for open-loop timing and location information. Accurate location information also facilitates location-based billing arrangement. However, it is recognized that due to the limited link budget of GPS in comparison to BGAN and possible obstructions around a user, the location accuracy may be compromised under some conditions and this may impact initial acquisition of the receiver i.e. when a mobile is switched on.

To comply the linearity requirements, efficiency of power amplifiers has to be compromised resulting in higher heat dissipation. Since cooling surface is limited, efficient heat dissipation is, therefore, essential. Features to minimize power drainage in the UT include avoiding unnecessary transmission and reception when there is nothing to receive and to use sleep mode wherein all but the bare minimum subsystems are switched off. With typical battery capacity of the order of a few ampere-hours the average drainage should be expected to be of the order of a few tens of milliamps and hence the stand-by time would be 24–48 h.

To support IP suite including the Internet and applications such as VoIP (Voice over Internet protocol) and multicast, performance enhancement proxies (PEPs) and QoS settings the UT must include an efficient software platform. Since the system is compatible with 3G core network, the terminal supports USIM (Universal Subscriber Identity module) card so as to support roaming between terrestrial and satellite networks.

TCP (Transmission Control Protocol) performance is known to be adversely affected on mobile satellite channels and hence its performance should be optimized with a PEP. Further QoS optimization may be applied after a better understanding of QoS issues with regards to inter-layer interaction from application layer down to the physical layer. An adjoining issue includes support of ARQ protocol on top of TCP, which can result in a double loop and unless taken into considerations, may negate the efficiency of the ARQ protocol.

A mobile terminal was developed under the ACTS programme to demonstrate K_a band MSS technology. Known as the ACTS Mobile Terminal (AMT), the terminal was meant to be a proof-of-concept K_a band terminal developed by NASA (National Association of Space Administration) at the Jet Propulsion Laboratory. The AMT was intended to demonstrate speech and data transmission and act as a precursor to the development of terminals for aeronautical, maritime, land and personal communications. The AMT included features to support the characterization of K_a band propagation in a mobile environment. The terminal operated at 30 GHz in the up-link and 20 GHz in the down-link and at data rates of 2.4, 4.8, 9.6 and 64 kbps. Some vital system-level considerations are as follows:

- geostationary system – location 100°W;
- high gain spot beams;
- regenerative transponders;
- terminal operational elevation angle – 30°;
- FDMA scheme;
- estimated 98% link margin for an average year – 1.2 dB at 30 GHz and 1 dB at 20 GHz; noting that far higher attenuation can occur during heavy rain;
- high shadowing loss;
- high Doppler and Doppler change rate due to vehicle motion – of the order of 3 kHz and 370 Hz/s, respectively;
- high phase noise in the communication channel due to high noise closer to carrier centre frequency that requires robust modulation scheme against phase noise.

The key technical challenges were:

- development of high gain tracking antennas;
- compensation for rain attenuation and;
- compensation for high Doppler, frequency uncertainties and phase noise.

To counter the rain attenuation, a rain compensation algorithm (RCA) is used. The dynamic algorithm involves pilot power measurement and beacon measurement at the fixed station followed by an exchange of information between the terminal and a decision to lower the data rate from 9.6 to 4.8 or 2.4 kbps in the presence of rain, thereby increasing the link margin from 3 to 6 dB.

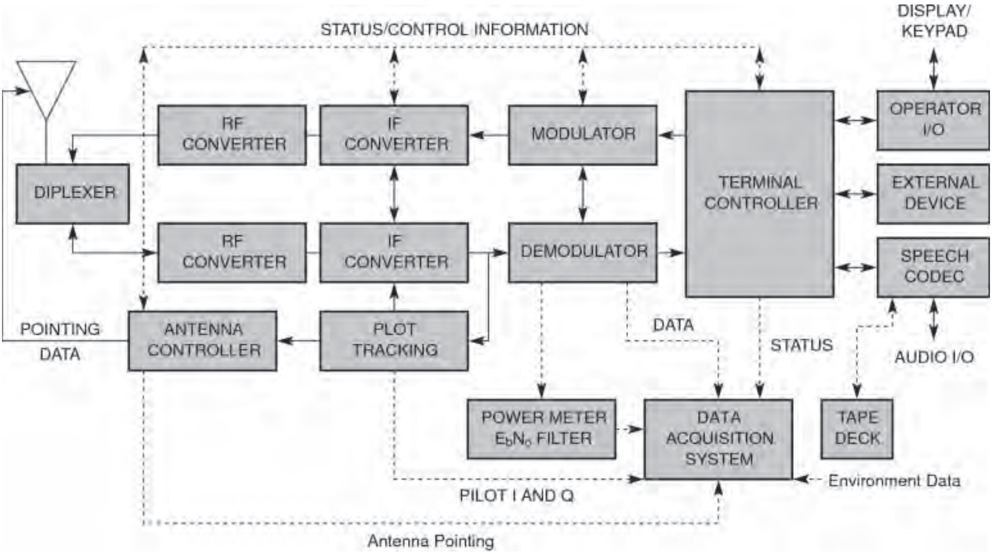


Figure 5.18 Block diagram of ACTS mobile terminal. (Source: Abbe *et al*., 1995. The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

Figure 5.18 shows a block diagram of the AMT.

The terminal controller is responsible for all the coordination mentioned below:

- translation of communication protocols to executable instructions, for example timing and handshake between speech coder, modem, user interface and external devices during call set-up, end of call and data rate change;
- operation and execution of the RCA algorithm;
- control of intermediate frequency (IF), RF and antenna electronics;
- system monitoring and interface to data acquisition system and
- support of test functions during experimentation, for example correlation, bit error measurement.

The AMT modem has been developed with the aim of minimizing the effects of the phase noise from the satellite. A DPSK (Differential Phase Shift Keying) modulation scheme is used because of its simplicity and robustness to phase noise and deep short-term fading. A half rate convolution coding with interleaving is used. An offset of $\pm$ 10 kHz is tolerated by the modem. The required E_b/N_0 for a BER (bit error rate) of 10^{-3} at 9.6 kbps is 6.6 dB in Gaussian noise, degrading by 1 dB in the presence of phase noise.

In addition to up and down conversion, the IF converters track the pilot and pre-compensate for Doppler. The pilot provides a frequency reference for the terminal and is useful for

Doppler pre-compensation of the transmitted signal. The pilot signal is also used in the RCA unit, for antenna tracking, and by the terminal propagation data acquisition system. An IF frequency of 70 MHz is used in the receiver and 3.373 GHz in the transmitter in compliance with fixed earth station hardware.

Two types of vehicle antenna were developed – a reflector-type elliptical antenna and an active array comprising integrated HPA and LNA units as MMIC. The reflector antenna is enclosed within an ellipsoidal radome of 23 cm/9″ base diameter and a height of 9 cm/3.5″. The active array minimizes RF loss. The terminal has a minimum EIRP of 22 dB W, a G/T of −8 dB/K and a bandwidth of 300 MHz. The reflector permits TWTA (Travelling Wave Tube Amplifier) to operate at powers lower than 1.5 W and operates with a mechanical tracking system. The antenna is dithered around the bore-sight by about a degree at a rate of 2 Hz while the pilot signal strength is measured. The signal level, together with the vehicle's inertial information extracted from a turn rate sensor, is used to estimate the tracking error. The system is able to maintain tracking in the presence of fade lasting up to 10 s. The small antenna diameter and narrow beamwidth enable the use of such a tracking scheme. Finally, the data acquisition system acquires and displays measured data related to propagation, communication link, pilot and data signal state, noise, antenna pointing, vehicle heading, etc. for performance evaluation.

5.3.3.3 Aeronautical Terminals

Aeronautical terminals have to comply with strict regulatory guidelines to ensure that the installation does not cause detrimental effects to on-board electronics. The Aeronautical Radio Incorporation (ARINC) specifies first, second and third generation L band satellite unit terminal characteristics for civil aviation respectively in the ARINC 741, ARINC 761 (Inmarsat Swift64) and ARINC 781 (Inmarsat SwiftBroadband) standards (see ARINC web site: www.arinc.com/). For example, Aeronautical terminals, often called Aeronautical Earth Stations (AES) for historic reasons, are categorized into four classes. Class 1 AES support low speed service with low gain antenna; class 2 AES support passenger voice service with high gain antennas; class 3 AES support voice and high data speed; and class 4 AES offer voice, high speed and low speed data services in ARINC 741. The standards also specify the characteristics of AES subsystems – satellite data unit, RF unit, DIP, LNA, HPA, antenna and beam steering electronics.

Aeronautical terminals used in the Inmarsat legacy network are described here as they remain representative of this category of mobiles (Inmarsat, 1990). Note that the RF part of these UTs has been reused in the second and third generation Inmarsat systems. There are two basic types of services supported – within each there are various service options compliant with the ARINC standard, that is; low data rate, high data rate, voice or a combination of these. Low data rate terminals deploy omni-directional antennas and high data rate terminals deploy medium gain antennas of about 12 dB. Table 5.8 summarizes the main characteristics of these two types of terminals.

Figure 5.19 depicts the main units of a high gain AES; for a low gain antenna the functional units are similar but with different characteristics. A phased array antenna provides electronic steering through a beam steering control unit for which the antenna correction inputs are provided by a programme track sub-system. The programme track system, embedded within the data processing unit, derives tracking corrections by calculating the difference

Table 5.8 The main characteristics of two classes of Inmarsat aeronautical service terminal

Parameter	Terminal characteristics	
	High gain	Low gain
Transmit frequency range (MHz)	1626.5–1660.5	1626.5–1660.5
Receive frequency range (MHz)	1525–1559	1525–1559
Transmit EIRP (dBW) (nominal)	25.5	15.5
Receive G/T (dB/K)	−13	−26
Antenna characteristics		
typical antenna type	Phased array	Helical
antenna Gain (dBi)	∼ 12	∼ 0
tracking	Programme track	None
hemispherical coverage (%)	>75	>85
Typical high power amplifier class; power (W)	Class-A; 60	Class-C; 40
SCPC channel pairing	Yes	Not applicable (NA)
Synthesizer step (kHz)	2.5	2.5
Low rate data service (kbps)	0.3	0.3
Voice channel modulation; coding	A-OQPSK; ½ FEC	NA
Voice channel transmission rate	4.8/9.6	NA
Voice channel bandwidth (kHz)	17.5	NA
Forward voice activation	Yes	NA
Satellite forward link EIRP at start of call (SCPC) (dBW)	21.5	NA
SCPC channel		
access/modulation; coding; transmission rate (kbps)	DA-FDMA/A-QPSK; ½ FEC; 21	NA
Data channel (return link)		
access/modulation; coding; transmission rate (kbps)	R-TDMA/A-BPSK; ½ FEC; 0.6	R-TDMA/A-BPSK; ½ FEC; 0.6
Forward signalling channel characteristics		
access/multiplexing; coding; transmission rate (kbps)	Packet mode/TDM; ½ FEC; 0.6	Packet mode/TDM; ½ FEC;0.6
Forward data channel characteristics		
access/multiplexing; coding; transmission rate (kbps)	Packet mode/TDM; ½ FEC;0.6 and 10.5	Packet mode/TDM; ½ FEC;0.6
Satellite forward link EIRP (data) (dBW)	12.5	21.5
Return data channel characteristics		
access/modulation; coding; transmission rate (kbps)	Random access or R-TDMA; A-BPSK/A-QPSK; ½ FEC; 0.6/10.5	Random access or R-TDMA; A-BPSK; 0.6
Return request channel		
access/modulation; coding; transmission rate (kbps)	Aloha/A-BPSK; ½ FEC; 0.3	Aloha/A-BPSK; ½ FEC; 0.3

(*continued overleaf*)

Table 5.8 (*continued*)

Parameter	Terminal characteristics	
	High gain	Low gain
Typical application	Aeronautical administration communication (AAC) Aeronautical operational control (AOC) Air traffic control (ATC) Aeronautical passenger communication (APC):voice	Aeronautical administration communication (AAC) Aeronautical operational control (AOC) Air traffic control (ATC)

(Data source: Inmarsat, 1990.)

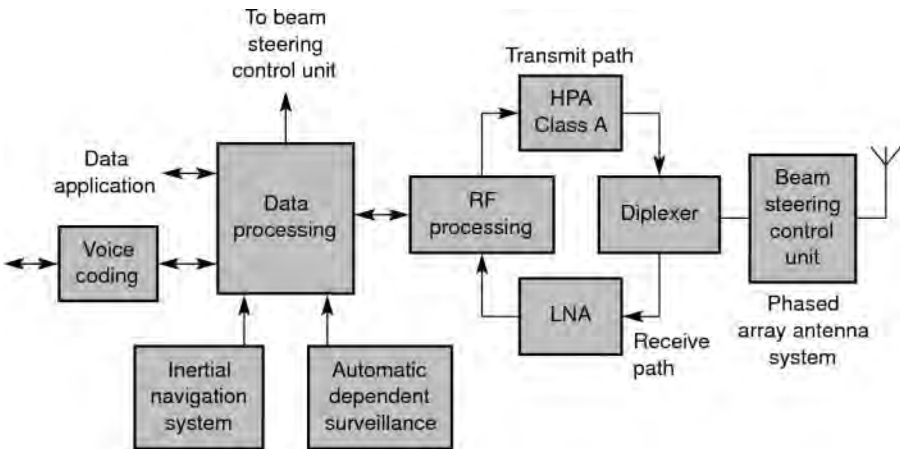


Figure 5.19 Main units of a high gain AES. (Source: Inmarsat, 1990. Reproduced with permission of Inmarsat.)

between the estimated satellite position and the aircraft's position obtained from the on-board inertial navigation system. To minimise intermodulation noise in the case of multi-carrier transmissions the HPA comprises a linear amplifier of about 60 W, which can support a maximum of two to three voice channels. The RF processing unit interfaces with base-band signals and includes up-down conversion, modulation-demodulation and coding-decoding. Data processing involves reformatting, protocol conversion and functions to suit the service and application. The automatic dependent surveillance system inputs the required aircraft data for remote ground surveillance.

An AES receives signalling messages on a broadcast channel transmitted from a ground earth station (GES) of the network and, and in turn uses a return (or R) channel for signalling

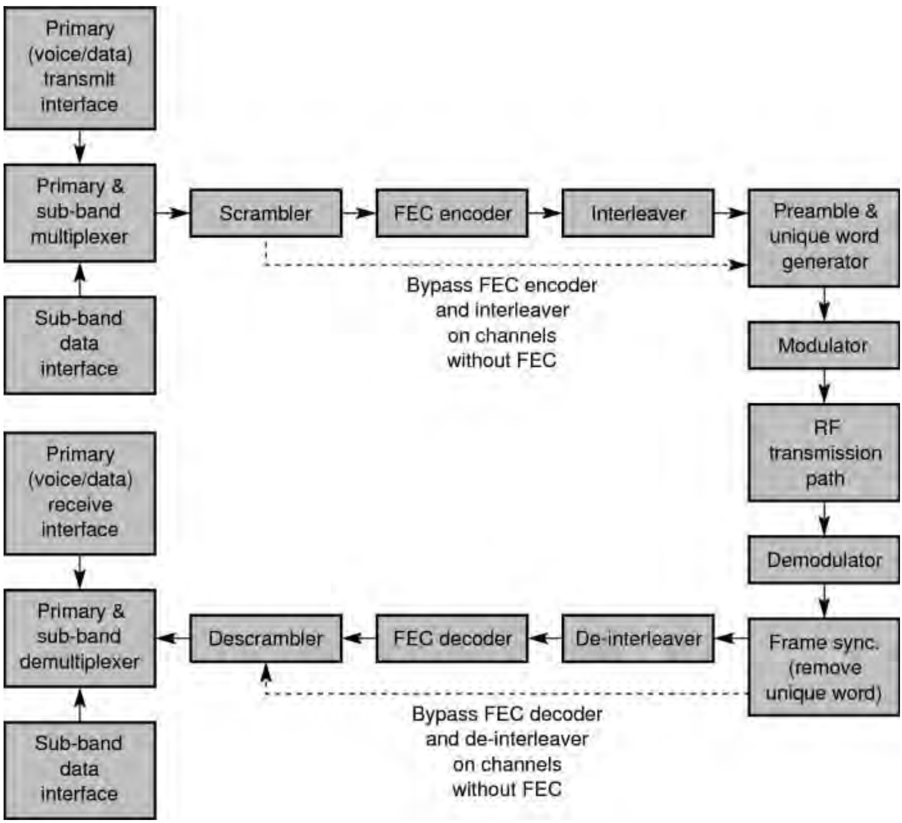


Figure 5.20 Voice channel unit of AES. (Source: Inmarsat, 1990. Reproduced with permission of Inmarsat.)

(and data transmissions) rates of 600 bps, 1.2 kbps or 10.5 kbps to 600 bps is used at log-on. The GES responds at the same rate, informing the AES of the appropriate bit rate for further communication, derived from the signal quality of the request burst and the satellite in use. TDMA channels, operating at the same rate as R channels, support data applications.

A voice channel unit, shown in Figure 5.20, comprises a voice or data unit and interface, a scrambler, an encoder, an interleaver to minimize the impact of error bursts, a preamble and a unique word generator for timing and identification. A reverse set of operations is performed at the receiving unit. There is a provision to bypass FEC encoder/decoder at the two ends.

High gain antennas are usually phased arrays installed either at the top of the fuselage or on both of its sides. Top-mounted antennas increase air drag to a certain extent but can view the satellite without the so-called blind zone (or 'keyhole') associated with phased arrays, whereas side-mounted antennas have lower air drag but have keyholes. Keyhole areas can be reduced by positioning phased array units around an elevated mount.

5.4 Environmental Issues

With heightened awareness of mobile communications, and in particular terrestrial cellular systems, a number of environment-related issues and concerns have been raised:

- **Aesthetics**: Aesthetics relates to the look and feel of the UTs and network infrastructure particularly the base station antennas. Base station antenna sites have undergone severe scrutiny due to their unsightly appearance affecting city skylines and potential radiation risks. A number of innovative ideas are under study and development. In one such experiment, an antenna system was camouflaged as a tree, blending unobtrusively with its surroundings. Satellite gateways are far fewer and generally sited well away from population centres and therefore do not pose concerns from this viewpoint. UT designs continue to change rapidly and they are now considered as a fashion accessory – innovative ideas include pendant and wristwatch phones.
- **EMI from mobile telephones to other equipment**: Concerns have been raised regarding the EMI effects on health and safety equipment such as hearing aids, cardiac pacemakers, equipment in hospitals, on aeroplanes, ships, and etc., and in a more general sense, of everyday equipment such as fixed telephones, radio, television, etc.

  In situations where the EMI effects of mobile terminals on sensitive medical equipment are unknown, the safe approach has been to forbid their use in areas such as hospital theatres, etc. This approach is not feasible in public places, where a mobile user may answer a telephone next to an individual with an implanted device. The best course of action is to ensure that such medical devices have high resistance to EMI. The effects of EMI on equipment such as hearing aids is less critical; controlled tests on more vital equipment such as implanted pacemakers do not show adverse effects of RF radiation; clearly, however, making such devices resistant to EMI is in the public interest and it is essential to introduce regulations for EMI immunity for critical health and safety equipment.
- Concern regarding the effects of RF radiation on people continues to receive significant public attention. For this reason, the next section reviews the subject.

5.4.1 Biological Effects

In 2011 over 4.6 billion people used mobile telephones world-wide and the user population is expected to increase dramatically in the next decade. Therefore, a better understanding of mobile phones' biological effects on humans is imperative, and if adverse effects are observed, adequate safeguards in the design of mobile systems and phones are crucial.

At present, interaction mechanisms for low-level exposure to emissions from hand-held mobile phones and their biological effects are not well understood. In the absence of conclusive evidence, concerns regarding the adverse effects on humans of RF radiation from mobile phones continue to be raised in public by individual or anecdotal claims or through inconclusive or inadequately substantiated reports, and so on. Claims pertain to the possibility of contracting cancer, DNA damage, heating of the brain tissue, headaches and tiredness – amongst others. However, none of these claims has been proven or reproduced under controlled test environments but public concern remains entrenched. International Agency for Research on Cancer dealing with cancer research has called electromagnetic fields from mobile phone as 'possibly carcinogenic'. The World Health Organization intends to 'conduct a formal risk assessment of all studied health outcomes from radiofrequency field exposure by 2012' (WHO, 2013).

A number of research programmes are on-going throughout the world at various institutions, such as the World Health Organization, to understand more quantitatively the effects on humans of RF emissions in the frequency bands (~ 800–2200 MHz) and power (10–1000 mW) used by various cellular standards. Lin (1997) and Foster and Moulder (2000) give good summaries of the work applicable to mobile communications. The results applicable to terrestrial mobile systems are equally applicable to mobile satellite communications, as the frequency ranges are similar and the same international safety standards apply.

A commonly used health and safety RF exposure standard, ANSI/IEEE C95.1–2005 (IEEE, 2005), promulgated by American National Standards Institute (ANSI), provides guidelines for exposure limits. Another international standard specifying exposure limits is due to International Commission on Non-Ionizing Radiation Protection (ICNIRP, 2009). For example, the maximum permissible exposure (MPE) in the 300–3000 MHz range for controlled environments specified in the ANSI/IEEE standard, is f/30 W/m^2; and for an uncontrolled environment MPE is f/200 W/m^2 in the frequency range 400–2000 MHz.

Specific absorption rate (SAR) is a dosimetric measure defined by the National Council on Radiation Protection and Measurements (NCRP) of the USA as 'the time derivative of the incremental energy absorbed by (or dissipated in) an incremental mass contained in a volume of a given density' (NCRP, 1981). The permissible value for SAR is ~ 0.4 W/kg averaged over the whole body or ~ 8.0 W/kg for any 1 g of tissue for a controlled environment and around one-fifth less for an uncontrolled environment.

SAR, along with other quantities including incident field and induced field, is used as a measure of RF exposure. Due to the significance of such measurements for research and radiation compliance of equipment and the difficulty in making accurate measurements, particularly when the radiation source is close to the head, considerable effort is under way to improve the measurement technique (called, dosimetry). The measurements can be done either experimentally, theoretically or a combination of the two. In absence of a universally acceptable procedure for such measurements there tends to be variability in the reported results. Figure 5.21 (Balzano, Garay and Manning, 1995) illustrates SAR measurements made by isotropic electric probes inside a fibreglass human head filled with liquid dielectric brain-equivalent material, for radiation caused by a mobile phone transmitting 0.6 W through an extended antenna, as shown. Notice that the maximum SAR of 1.1 W/kg is at the ear lobe nearer to the antenna, close to the feed point of the antenna; the level is much lower elsewhere; and the maximum SAR is within acceptable limits.

A number of SAR measurements within the brain tissue of the phantom human head have been reported in the 800–900 MHz band (cellular band) using 0.6 W transmissions from communication devices placed 1–3 cm from the surface of a model head. There is considerable variability in SAR value reported (0.44–6 W/kg); however, most of the results complied with the ANSI/IEEE C-95.1 guidelines (revised since the experiments were conducted) except for a few reports that show SAR in excess of the limits. The most pessimistic results reported are as high as 6 W/kg. Possible reasons for the wide variability are the difficulty in making accurate SAR measurements, the dependency of measurements on the position of the antenna with respect to the head, and the influence of a device's hardware on

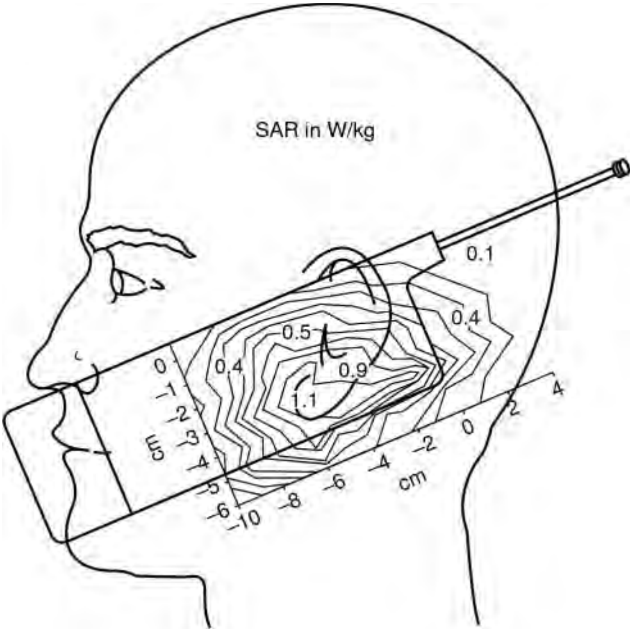


Figure 5.21 SAR measurements inside a phantom human head from a mobile phone transmitting at a frequency of 835 MHz and a power of 0.6 W. (Source: Balzano *et al.*, 1995. © 1995 IEEE. Reproduced with permission.)

its radiation characteristics. It is noted that the SAR values are lower for longer antennas (in terms of wavelength) than shorter ones – this is due to higher current regions being higher up on the antenna and therefore further away from the head for longer antennas.

Interestingly, measurements on human models standing next to trunk-mounted mobile telephone antenna transmissions at 835 MHz transmitting at 1 W show exposure levels well within the ANSI standard at any point. Field strength measurements made by researchers on standard off-the-shelf mobile phones in the UK show compliance with radiation standards.

Existing measurement methods can be tedious for large-scale measurements such as when required for compliance and type approval of mobile phones. Hence, computer-assisted numerical methods are under investigation. The difficulty here lies in modelling the human body and transceivers accurately. For example, to simplify computing requirements, mobile phones have been modelled as a metal box with an antenna, without considering the influence of internal circuitry, sub-assemblies, packaging or the external geometry of the mobile phone. Magnetic resonance imaging (MRI) based models have been used for simulating the human head; and finite-difference, time-domain numerical methods have been used for numerical computation. Simulated SAR results in the brain region for mobile phones operating in 800–1800 MHz using antennas of one-quarter to half wavelength vary considerably due to the differences in modelling assumptions, but reported results for 0.6 W transmissions (0.65–1.97 W/kg) generally show compliance with ANSI standards (1995 version) for radiation received within the brain region. However, power deposition in superficial tissues can exceed 1.6 W/kg. Results pertaining to the reduction in SAR with increased distance and increase in antenna size confirm experimental observations. The user's physical environment has been observed to influence SAR measurements. The presence of a reflective ceiling can redistribute SAR, reducing the peak SAR inside a human head. A reflective vertical wall, by contrast, can increase the peak SAR within the head by up to 100%. A number of problems need to be solved before a standard technique is developed; problems include the influence of tissue inhomogeneity and size of volume element on SAR distribution in a model.

Adverse thermal effects on the functioning of the human body at high intensity are well known. But at the levels at which mobile phones operate, it is unlikely that radiation causes thermal effects. Heating of the brain by microwave radiation at the very worst is estimated to be no more than that caused by mild exercise. Hence, research efforts have converged on the effects that may be caused by long-term use, including lifelong exposure to low-level radio waves.

Tumour induction from mobile phone radiation has been a topic of considerable interest. The results conducted on animals in the frequency range reported to date, with one exception, indicate that tumour growth is not caused by RF radiation; however, RF radiation in the range 915–2450 MHz does accelerate tumour growth if the tumour is initiated first by a cancer promoter. In another study, life-long exposure of rats to pulsed microwaves showed no difference on their general health or life expectancy; however, a statistically significant increase was observed in primary malignancies at the end of life, which required further research. A study conducted specifically for mobile telephones operating at 915 MHz reported no increase in tumour growth.

Several studies related to the nervous system have concentrated on the neurovascular construct known as the blood-brain barrier (BBB), which maintains the physiochemical environment of the brain within the narrow limits essential for life, following a 1990 report of radiation effects on the BBB. The conclusions of such studies remain controversial, in particular, for low-level radiation.

Microwave-induced auditory sensation, which pertains to auditory sensation in humans and animals from microwave pulses, is a well-known microwave effect. It has been observed that if a microwave pulse or pulse train is delivered above a threshold power density (e.g. 400 mJ/m^2 at 2.45 GHz) to a human subject, it is heard as a click for a single pulse or as a series of clicks at the pulse repetition rate. At present, it is believed that the audio sensation is caused by acoustic pressure induced by thermo-elastic waves that are produced by soft tissues in the head on absorption of microwave radiation. The acoustic pressure travels to the inner ear by bone conduction where it is treated in the same manner as normal sound. It is believed that the threshold microwave auditory effect has insignificant impact on hearing apparatus. However, at present, there is little data regarding the long-term effects of the microwave auditory effect on either the central nervous system or the middle and inner ear hearing apparatus. Research is in progress to gain insight into the possible effects from mobile telephones.

The formation of cataracts through exposure to microwave radiation is yet another concern. It is generally accepted that prolonged high-level microwave radiation causes cataracts in the eyes of laboratory animals, but below a threshold, there is no such effect, no matter the duration of exposure. The conditions that trigger the cataract are not well understood.

A valuable investigative tool for studies related to humans is epidemiology in which study of incidence of a disease is studied across communities. Studies conducted over specific groups such as military microwave workers and air force veterans, exposed to radiation in their occupation show are inconclusive – two studies showed statistically non-significant results and one indicated statistical significance. Other epidemiological studies that have investigated effects of microwave on the growth and development of children from parents exposed to microwaves, the incidence of cancer, mortality rate generally show no adverse effects. A study conducted in 1994 to investigate the near-term mortality rate of mobile phone user in 1994 showed no difference in relation to non-users. Studies conducted in 1999 on 300 000 mobile users show no effect on mortality. All other epidemiology studies to date have yielded similar results. One of the problems in all studies to date is the lack of measurements of microwave exposure levels, which lead to large uncertainties. In those studies in which a statistically significant rise of adverse effects was observed, the results were diluted by cross-impact from other sources. Furthermore such studies must be conducted over a decade or more, making them difficult to control and interpret. Thus, it is difficult to conclude of any positive or negative association from the studies so far.

Researchers conducted in 1997 at the Royal Adelaide Hospital in Australia observed that mice exposed for 18 months to microwave emissions at the frequency and intensity used in digital phone were twice more susceptible to cancer of the lymph system than others that were not exposed. However, three other teams were unable to reproduce the results, including a team at Brooks Air Force Base in Texas, which found no evidence of increase in cancer to exposed mice, genetically made susceptible to breast tumours, for 20 h each day over 18 months.

In 1995, a research team of the University of Washington, Seattle reported DNA break, similar to the type caused by carcinogenic chemicals or X-rays, in rats exposed to microwave radiation. Two independent studies conducted by research teams in St Louis, Missouri and Belgium subsequently failed to reproduce the result.

In 1998, a Swedish study demonstrated that people making more frequent telephone calls were more likely to complain of headaches and tiredness. However, it has also been observed that people's awareness biases such observations; for example it has been reported that people in Norway were twice as likely to complain of such symptoms when compared to Sweden where radiation effects of mobile phone attracted a lower publicity.

While such investigations progress, mobile telephone manufacturers continue to improve techniques for minimizing exposure to radiation. Techniques include shielding of radiation in the direction of the brain and mounting antennas away from the head. Trials conducted in National Physical Laboratory (NPL), UK for the science magazine *New Scientist* used a simulated human head – a human skull covered by simulated flesh – to study shielding provided by two commercially available shielding devices. Field strengths were measured inside the head using two makes of mobile telephones, one each from Motorola and Nokia. A reduction of 16–48% in field strength was observed when the antenna of the mobile phones was kept pointing downwards, the amount of reduction depended on the make of the mobile phone. The shielding was less effective in each case when the antenna was pointed upwards. In all cases the quality of reception degraded when a shielding device was used. The NPL study concluded that the most effective method of reducing radiation to brain was to use a mobile phone, which has the radiator mounted on a belt that reduces the field strength within the head by 94%, though the radiation to other parts of the body showed an increase.

While there is general agreement on certain adverse effects of high levels of microwave radiation, it is evident from data available to date that the results on low level RF transmissions as encountered from existing mobile phones have yet to conclusively demonstrate adverse effects. Several investigations are in progress to improve our understanding of the subject.

Revision

1. Describe the hardware entities and functions of an MSS gateway.
2. Outline the characteristics of various types of antennas used in MSS user terminals and assess their suitability for hand-held, land vehicular, maritime and aeronautical mobile platforms.
3. Discuss the tracking options for mobile platforms installed on ships and passenger aircraft.
4. With the help of block schematic describe the main features of: (i) handheld UT, (ii) a large ship earth station and (iii) an aeronautical earth station.
5. Discuss the trade-offs in designing a land portable multimedia UT to support 500 kbps transmissions in an L-band MSS system.

References

Abbe, B.S., Agan M.J., Girardey, C.C. and Jedrey, T.C. (1995) Channel and terminal description of the ACTS mobile terminal. IMSC 93. The Third International Mobile Satellite Conference, Pasadena, CA, June 16–18, JPL Publication 93-009, pp. 387–392.

Agrawal, A. (1998) Mobile Chips: Towards a Common Platform. *Telecommunications* (Mar.), 86–88.

Balzano, Q., Garay, O. and Manning, T.J. (1995) Electromagnetic energy exposure of simulated users of portable cellular telephones. *IEEE Transactions on Vehicular Technology*, **44**, 390–403.

Caballero, J.E., Badenes, J., Fernandez, J., *et al*. (1995) Hand-held terminal antennas for personal satellite communications. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 351–356.

Chuang, H.R. (1994) Human operator coupling effects on radiation characteristics of a portable communication dipole antenna. *IEEE Transactions on Advanced Packaging*, **42**(4), 556–560.

Foster, K.R. and Moulder, E.J. (2000) Are mobiles safe? *IEEE Spectrum*, **37**, 23–28.

Fuji, T., Tsuchiya, M., Isota, Y. and Aoki, K. (1995) Design and performance of mobile terminal for North American MSAT network. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 365–369.

Garcia, Q., Martin, C., Del Valle, J.C., *et al*. (1993) Microstrip monopulse antenna for land mobile communications. IMSC'93, The Third International Mobile Satellite Conference, Pasadena, CA, June 16–18, JPL Publication 93-009, pp. 575–580.

Goldberg, B.G. (1996) Technical Challenges for a Crowded Wireless Environment. *Satellite Communications* (Dec.), 32–35.

Hoshikawa, T. (1988) Inmarsat ship earth station type RSS401A. *Anritsu Technical Journal*, **56**, 41.

ICNIRP (2009) Statement on the Guidelines for Limiting Exposure to Time-Varying Electric, Magnetic and Electromagnetic Fields (up to 300 GHz). International Commission on Non-Ionizing Radiation Protection.

IEEE (2005) IEEE Std C95.1, 2005. *IEEE Standard for Safety Levels with Respect to Human Exposure to Radio Frequency Electromagnetic Fields, 3 kHz to 300 GHz*, IEEE.

Inmarsat (1988) Inmarsat-C System Definition Manual.

Inmarsat (1990) Inmarsat-Aeronautical System Definition Manual.

Inmarsat (1991a) Inmarsat-B System Definition Manual.

Inmarsat (1991b) Inmarsat-M System Definition Manual.

ITU-T (2013) Web site available at: www.itu.int/rec/T-REC-V/en (accessed August 22, 2013)

Jet Propulsion Laboratory (1988) MSAT-X Quarterly No 13, January.

Lin, J.C. (1980) The microwave auditory phenomenon. *Proceedings of the IEEE*, **6**, 67–73.

Lin, J.C. (1997) Biological aspects of mobile communications fields. *Wireless Networks*, **3**, 439–453.

Maral, G. and Bousquet, G. (2009) *Satellite Communication Systems – Systems, Techniques and Technology*, 5th edn, John Wiley & Sons, Ltd, Chichester.

Milne, R. (1995) Performance and operational considerations in the design of vehicle antennas for mobile satellite communications. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 329–333.

Mobile Europe (1997) *When Less is More*. Mobile Europe (Feb.), 7 (2), pp. 27–30.

NCRP (1981) Radio Frequency Electromagnetic Fields: Properties, Quantities and Units, Biological Interaction and Measurements. Report 67, National Council on Radiation Protection and Measurements, Bethesda, MD.

Ohmori, S., Wakana, H. and Kawase, S. (1998) Vehicle antennas, *Mobile Satellite Communications*, Chapter 4, Artech House, Boston, MA, pp. 87–123.

Richharia, M. (1999) *Satellite Communication Systems*, Macmillan Press Ltd, Basingstoke and London (jointly published with McGraw-Hill, New York).

Richharia, M. (1986a) An improved step-track algorithm for tracking geostationary satellites. *International Journal of Satellite Communications and Networking*, **4**, 147–156.

Richharia, M. (1986b) Design considerations for an earth station step track system. *Space Communication and Broadcasting*, **4**, 215–228.

Shafai, L., Moheb, H., Chamma, W. and Barakat, M. (1995) Low profile antennas MSAT applications. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 334–339.

Skinnemoen, H., Johansen, T.V. and Eriksen, E. (2003) Ultra-portable multimedia satellite terminals for the BGAN satellite system. 2003 IEEE 58th Vehicular Technology Conference, 2003, VTC 2003-Fall, October 6–9, 2003, Vol. 4, pp. 2678–2682.

Stojkovic, I. and Alonso, J.E. (1995) Key issues for user terminals operating with HEO satellites. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 370–375.

Sydor, J. and Dufour, M. (1993) Aeronautical satellite antenna steering using magnetic field sensors. IMSC 93, The Third International Mobile Satellite Conference, Pasadena, CA, June 16–18, JPL Publication 93-009, pp. 593–598.

Toftgard, J., Horusleth, S.N. and Andersen, J.B. (1993) Effect on portable antennas of the presence of a person. *IEEE Transactions on Advanced Packaging*, **41**(6), 739–746.

WHO (2013) Electromagnetic Fields and Public Health: Mobile Phones, Fact sheet No. 193 June 2011, www.who.int/mediacentre/factsheets/fs193/en/ (accessed August 22, 2013).

6

Spacecraft

6.1 Introduction

High-power satellites with sensitive on-board receivers are central to the concept of mobile satellite services (MSSs) as these requirements are a prerequisite for supporting services to small mobile terminals. Therefore the advent of mobile satellite communications had to wait until satellite and earth station technology had developed to an extent that satellite transmitters could provide the requisite power and earth stations had shrunk to a size supportable on mobiles. Both components of technology have since matured such that satellite phone and broadband services are widely available. The next generation systems are enhancing the capabilities to provide higher throughput to the individual, land, maritime and aeronautical communities. While a majority of the existing services operate at the L and S bands, the emerging systems also utilize the K_a band in order to alleviate the spectrum shortage of the lower bands. In this chapter, we will outline the satellite technologies that enable the MSS.

Figure 6.1 shows the main sub-systems of a spacecraft.

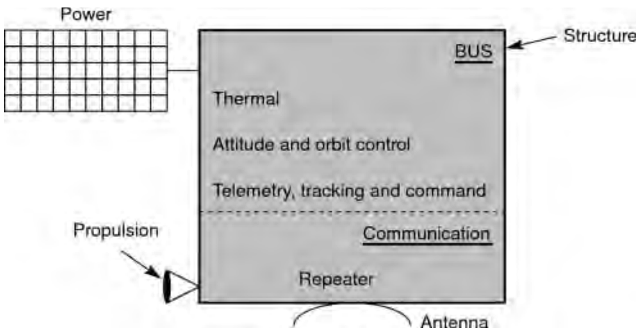


Figure 6.1 Main sub-systems of a spacecraft

The main communication unit of a satellite (or payload) comprises an antenna system and a repeater (or transponder). When an intersatellite link (ISL) is used, the payload also supports an ISL payload. A payload requires support from a number of sub-systems collectively

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

known as *bus*, which comprise a structural sub-system for housing components, a thermal sub-system for temperature maintenance, a propulsion system for orbit-raising and orbital maintenance and an electrical power sub-system. Spacecraft status monitoring, tracking and commanding are performed by the telemetry, tracking and command system – the spacecraft's link to the ground control centre.

Keeping to the theme of the book, the majority of the chapter will address the payload, as it has a direct bearing on the capabilities and architecture of an MSS system. At present, transparent (or 'bent pipe') satellite transponders are the norm, but regenerative transponders and on-board processing sub-systems are gradually finding favour as the technology matures because of the ensuing advantages. The chapter includes: a brief overview of transparent transponders but with the emphasis on regenerative transponders, as the technology of the former is well documented in several textbooks; a look at the influence of orbital altitude on spacecraft design; system aspects of intersatellite links; and a summary of the current state of technology. To begin with, some enabling spacecraft technologies are explored, followed by issues related to MSS.

6.2 Satellites for MSS

There are a number of MSS system requirements, which influence the architecture an MSS payload:

Operational frequency: At present, the L and S bands are prevalent; K_u band secondary allocations are used in at least one MSS system and a majority of fixed satellite service (FSS) mobile very small aperture terminal (MVSAT) systems; K_a band MSS-MVSAT system is a recent introduction;

Deployment of enhanced spot beam technology is mandatory to facilitate efficient use of radio resources, provide desired effective isotropic radiated power (EIRP) and G/T with capabilities to alter beam shape, size and position in response to needs;

Spacecraft EIRP and G/T in the service link should be adequate to serve mobile terminals;

Efficient spectrum utilization is mandatory to mitigate congestion;

Dynamic distribution of spacecraft EIRP and spectrum across the service area is essential to respond to changes in traffic pattern;

Full eclipse operation is desirable for low earth orbit (LEO), medium earth orbit (MEO) and international geostationary orbit (GEO) satellite systems; regional MSS operators using a GEO system may not operate at full power if the orbital location is chosen such that the eclipse occurs late at night.

Propagation delay: For those MSS systems that intend to provide full global coverage, low-delay terrestrial-like services or a fault-tolerant/survivable space segment architecture, LEO or MEO systems are preferred.

These requirements cascade down to the following payload requirements:

1. *Transponder*
   a. Available spacecraft EIRP must be high enough to be able to serve present and future traffic far enough into the future – ideally to the end of spacecraft life; this necessitates efficient high power amplifiers and narrow reconfigurable/steerable spot beams.

b. Flexible real-time power and bandwidth allocation between spot beams to permit allocation of space segment resources flexibly (i.e. traffic driven).
c. Benefits of on-board processing technology such as improved radio link performance and switching should be assessed against limitations of the technology and cost trade-offs.
d. Suitability of ISLs to enhance network connectivity should be assessed to enhance connectivity.

2. *Antenna system*
   a. Spot beam technology with a provision of steerability is essential in the service links to enhance frequency reuse, reduce mobile EIRP, make effective use of spacecraft transmitter power and provide flexibility in resource allocation.
   b. Only a few large fixed earth stations are necessary in the feeder links and therefore a single or a few spot beams are enough to meet capacity and frequency coordination requirements in the feeder radio link.
3. *General*
   a. Technology amenable to mass production is essential for supporting large non-geostationary earth orbit (NGEO) constellations.
   b. Higher integration of payload components such as output combiner with antennas, band pass filters with low noise amplifier (LNA), solid state power amplifier stacks, etc. is desirable.
   c. High DC power is mandatory to meet the large power requirements of the payload, necessitating large and efficient solar arrays and power system.

Eclipse operation is a bus requirement but is included here as it is critical to LEO and MEO MSS systems. The number of eclipses is large for LEO and MEO satellites due to their orbital geometry, unless the orbital geometry minimizes or eliminates eclipses (see Chapter 2). Spacecraft batteries undergo a considerable number of charge/discharge cycles due to eclipse, which reduces the lifetime of the batteries and hence useful life of the satellite.

6.2.1 Transponders

Due to the simplicity, reliability, technological maturity and flexibility in changing signal format and accessing schemes, a majority of existing MSS use transparent transponders. Inmarsat, Globalstar and ORBCOMM systems use transparent transponders, whereas the Iridium system incorporates a regenerative payload.

There is a wealth of information on conventional transponders and hence the topic is treated here only at a refresher level. Regenerative transponders are discussed in some detail because of their future potential as well as their relatively scant coverage in the literature.

Figure 6.2(a) and (b) depicts the main functional blocks of transparent and regenerative transponder, respectively.

On-board processing is used for a variety of functions – regeneration, spot beam formation, on-board switching, active filters, etc. and can be performed at baseband, IF or radio frequency (RF); note that satellites with on-board processing need not employ a regenerative transponder. Many operational satellites use signal processing only for signal-routing and beam-forming network.

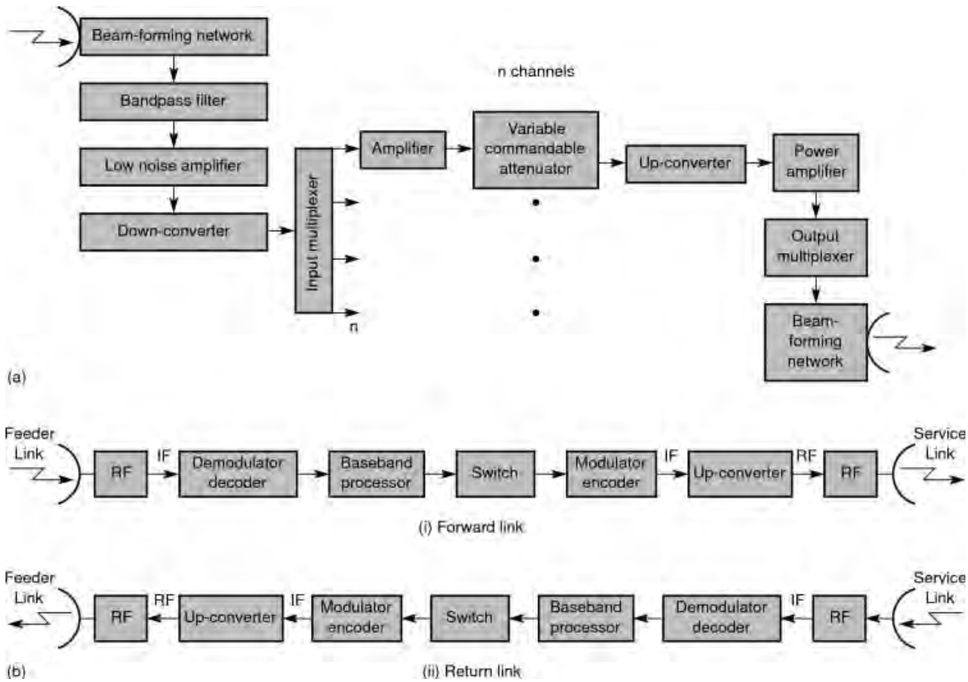


Figure 6.2 (a) A block diagram of the forward link of a transparent transponder. (b) Block schematic of an MSS payload with on-board processor, showing (i) forward link (ii) return link; beam forming network is assumed to be a part of RF block

6.2.1.1 Transparent Transponder

Figure 6.2(a) shows a conceptual block diagram of a transparent transponder. The received signals are band pass filtered to eliminate out-of-band signals, amplified, down-converted to IF, subdivided into channels, each routed to a specific spot beam. The signals are then amplified, upconverted and amplified to the specified transmit level, recombined in the output multiplexer and fed into a beam-forming network, which processes each channel so that they form the appropriate beam on transmission from the antenna system. When regenerative transponders are used, IF signals are demodulated, decoded and processed at baseband where they may also be routed. The processed baseband signals are coded and modulated, then up-converted, amplified and fed into the output multiplexer and a beam-forming network for onward transmission.

A number of beam-forming methods are discussed later. Consider some recent developments applicable to MSS payloads.

To serve a large number of spot beams flexibly, it becomes necessary to segment the spectrum into narrow bands for flexible and spectrally efficient routing; a finer granularity is better suited in this respect. Narrowband Surface Acoustic Wave (SAW) filter banks for demultiplexing, followed by switches and an analogue beam-forming network, provide flexible allocation of power and bandwidth. But due to the limited granularity achievable from SAW filters (~400 kHz), the channel bandwidths are coarse, which limits its traffic-matching capability.

The granularity can be improved by an analogue time domain Fourier transform technique called the Chirp Fourier Transform (CFT). The technique can be further augmented by digital methods to obtain channels narrower than 100 kHz, permitting flexible allocation of small numbers of frequency division multiple accessing (FDMA) channels.

In an advanced adaptation, the hybrid CFT and digital de-multiplexing technique use channel-by-channel switching, thus achieving the ultimate granularity in channel-to-beam routing. Consider a system where the outputs of such a de-multiplexer are fed into a digital narrowband beam-forming network capable of generating a very large number of repositionable beams peaking on individual users. An accurate user location is necessary, an algorithm for which can be included as a part of the digital processing required for digital beam forming. The technique can offer channel-level processing including power level control, active interference suppression, enormous adaptability to traffic variations and frequency reuse, high use of satellite and mobile RF as the user is at gain peak and transparency for introduction of new services. This type of system is at a conceptual stage awaiting technological development.

6.2.1.2 Regenerative Transponder

The advantages and theory of regenerative transponders are established for over two decades, but their introduction in practice has been gradual. A regenerative transponder performs essential signal processing such as demodulation, modulation, transponder or beam routing and switching, as well as advanced processing such as error correction/coding, reformatting of data, interference reduction, rain fade compensation, packet switching, and so on (see Figure 6.2b).

A regenerative transponder provides considerable performance advantage over its transparent counterpart under similar link conditions. In a conventional transponder, all types of link noise are amplified or attenuated by the same amount as the wanted signal, and therefore noise components add up in power. In a regenerative transponder, the received signal is regenerated to baseband and thereby the noise components in the up- and down-links get de-coupled; the total bit error rate (BER) in such a case is given as

$$e_t = e_u + e_d \tag{6.1}$$

where e_t = total bit error, e_u = up-link BER and e_d = down-link BER.

Figure 6.3(a) compares behaviour of up-link and down-link carrier-to-noise ratios for these two classes of transponder for the same link quality assuming a linear channel without coding (Maral and Bousquet, 1984).

When the up-link and down-link carrier to noise ratios are of the same order of magnitude, a regenerative transponder shows a clear advantage. This is generally the case for MSS links, and in particular, the return link when mobile terminals can transmit very low EIRP levels. Note that the advantage is reduced as up-link carrier to noise is increased, which would imply that the advantage is not notably significant in the forward link. The advantage of regenerative transponder is significantly increased when inherent non-linearity of the satellite channel is included; non-linearities are mainly introduced at the outputs of earth station and satellite transmitter output. Between 2 and 5 dB gains are possible even in the presence of high up link carrier to noise ratio density (see Figure 6.3b). The advantage in the link margin can be traded off against interference in an interference limited situation and hence

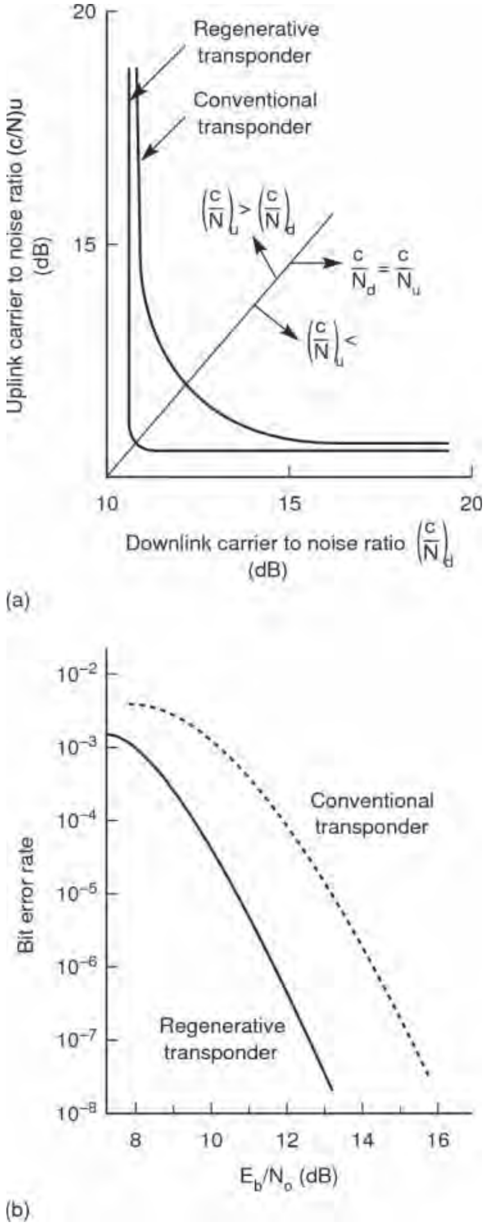


Figure 6.3 (a) Relationship between up-link and down-link carrier-to-noise ratio for conventional and regenerative transponders to provide the same radio link quality. (b) Relationship between BER and E_b/N_0 for conventional and regenerative transponders. (Source: Brugel, 1994. Reproduced by permission of IOS Press.)

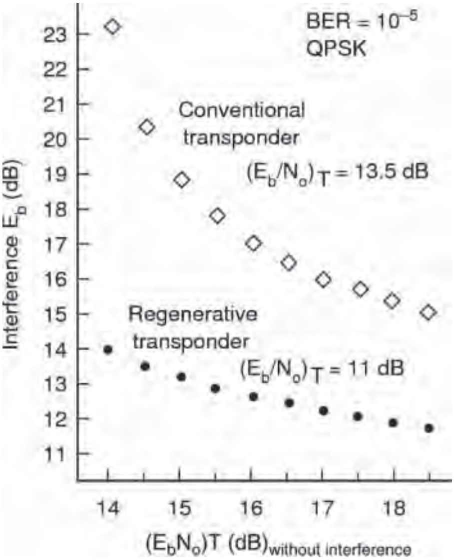


Figure 6.4 A comparison of interference tolerance of conventional and regenerative transponder. (Source: Brugel, 1994. Reproduced by permission of IOS Press.)

regenerative transponders can be used to increase system capacity in such an environment (see Figure 6.4). The decoupling property of regenerative transponders gives resistance to uplink and downlink signal fading. Furthermore, up- and down-links can be matched separately giving an advantage of several decibels in power resulting in an increase in overall system capacity.

Consider an MSS system using a K_a band feeder link and L or S band service link. The feeder link is stable for a majority time (>95%) when the predominant noise is Gaussian. For the remaining periods, the link fades to varying degrees due to rain or scintillation in periods of seconds, minutes; whereas the service link consists of a large number of short duration (2–3 min) randomly varying channels. As the characteristics of each segment are different, different types of coding schemes are desirable to maximize the coding gain. A regenerative transponder allows such a scheme.

The main advantages and disadvantages of regenerative transponders are summarized in Table 6.1.

It has been estimated that, under average operating conditions, improvements of the order of 8 dB are possible by regenerative transponders over conventional transponders in the presence of severe up-link rain fading. The calculations have assumed equal up-link and down-link carrier-to-noise ratios under nominal conditions; carrier-to-noise ratio of the down-link 10–15 dB above the detector threshold; and a fading of around 12 dB in the up-link. Under more realistic link conditions, improvements in the order of 10–12 dB have been demonstrated.

A regenerative transponder with baseband processing permits reformatting of data and therefore the up-link and down-link multiple access schemes can be different and optimized to suit each. For example, in a time division multiple accessing (TDMA) system operating through a conventional transponder, the forward link burst rate is limited by the mobile earth

Table 6.1 Main advantages and disadvantages of regenerative transponders

Advantages	Disadvantages	Comments
Improved channel quality	Increase in payload complexity	Extent of disadvantages will reduce as technology matures
Interference resistance	Heavier payload	Recent breakthroughs include development of high volume MMIC capability and improvements in radiation hardening of digital components
Higher capacity	Larger power drainage	Most K_a band systems plan use of on-board processors
Flexible and dynamic routing of messages	Difficult to change signal format, accessing schemes, modem/codec, etc. with current technology	Reconfigurable payload offer flexibility in changing signal format
Better network interconnectivity		
Allows use of smaller terminals		
Permits optimization of up and downlink independently		
Access security		

station's G/T and demodulator performance. A regenerative transponder eliminates such a limitation by allowing the downlink accessing scheme to be better matched to mobile earth station capability, for example by grouping channels into a manageable set of time division multiplex (TDM) streams at a rate commensurate with the mobile Earth station's capability, while leaving the feeder station the flexibility to transmit in TDMA. Figure 6.5 represents a configuration of a data adaptation regenerative transponder that combines a number of high rate TDMA service links and distributes data at low rates to a large number of mobile users, thereby improving feeder link spectrum efficiency.

A regenerative transponder offers the possibility of traffic routing, switching and on-board processing and thus the name 'intelligent' satellite or a 'switch in the sky'. Switching is done at RF, IF or baseband. RF and IF switching have been in use for well over a decade dating back to the Intelsat VI and VII satellites (1990s) for supporting TDMA traffic. Incoming bursts of data from one or more spot beams are routed to the destined beam using a pre-programmed switch matrix; the switch control unit is programmable through ground commands, providing flexibility to respond to traffic demands and increasing the spacecraft capacity by ~30% compared to a TDMA system without satellite switching.

RF or IF switched TDMA systems provide coarse interconnectivity – when a TDMA slot is not filled, the capacity remains unutilized; if switching is performed at message and packet level, messages/packets from all beams destined for a given beam, can be combined and routed to the appropriate beam, thereby improving resource utilization and network connectivity. A conventional time-space-time switch consisting of input and output buffers interconnected with a switching matrix can be useful.

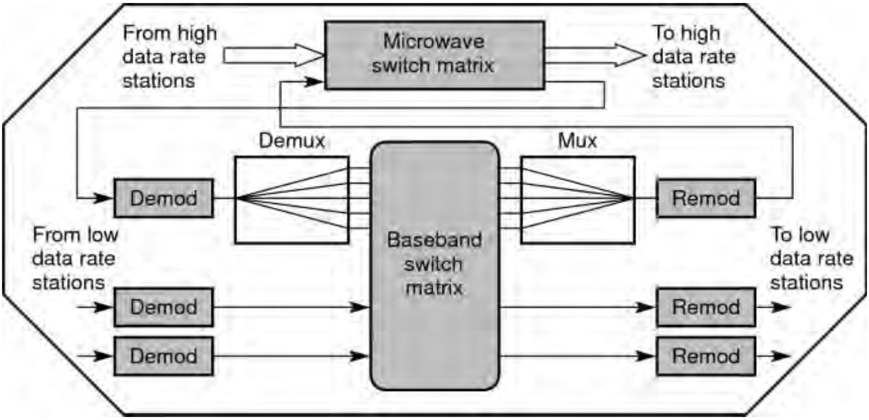


Figure 6.5 Data rate adaptation using a regenerative transponder. (Source: Brugel, 1994. Reproduced by permission of IOS Press.)

Call set-up and routing can be autonomous or managed by the ground network control centre (see Chapter 7). On-board call set-up can minimize latency and reduce ground network complexity at the expense of a more complicated on-board architecture. In addition to switching, considerable processing is necessary for routing and call establishment. The on-board processor must be able to interpret signalling protocols, extract and decode destination addresses from packets, store messages on a temporary basis during call set-up and establish a routing path, which may be transferred over an ISL. In an MSS system, it is also necessary to invoke the network mobility management function, which can be achieved in a manner similar to that on ground (e.g. use of home location register/visitor location register (HLR/VLR) databases). An additional on-board task is to convey to the network business management centre call record data for billing, diagnostics and fraud monitoring and traffic data collection.

Radio resource management can also be transferred on-board, at least partially. Considerable capacity enhancement is possible in a multi-spot beam carrying non-uniformly distributed variable traffic across beams when resources are allocated dynamically (Richharia *et al.*, 2006). Consider a system where the service spreads across several time zones; the business traffic peak tends to ripple from eastern spot beams to the western beams, according to business hours. Traffic patterns may get severely distorted in the event of an unusual event such as a sports event and military action. In such cases, allocation of fixed spectrum/power in spot beams is wasteful – a dynamic and flexible sharing of resources can improve the resource productivity (see Chapter 9).

The architecture of a baseband processing satellite to support a public network must be compatible with the network's transmission format, for example digital video broadcasting (DVB), Internet protocol (IP), air traffic management (ATM), etc. The signalling and protocol within the satellite network should be transparent to the core terrestrial protocols. At present, there are no standards established for protocols within a satellite's processor; standardization of protocols should enable interconnection of satellite systems.

Consider a few representative (but not exhaustive) baseband switching techniques in a multiple carrier TDMA system. Three methods are exemplified – carrier switching,

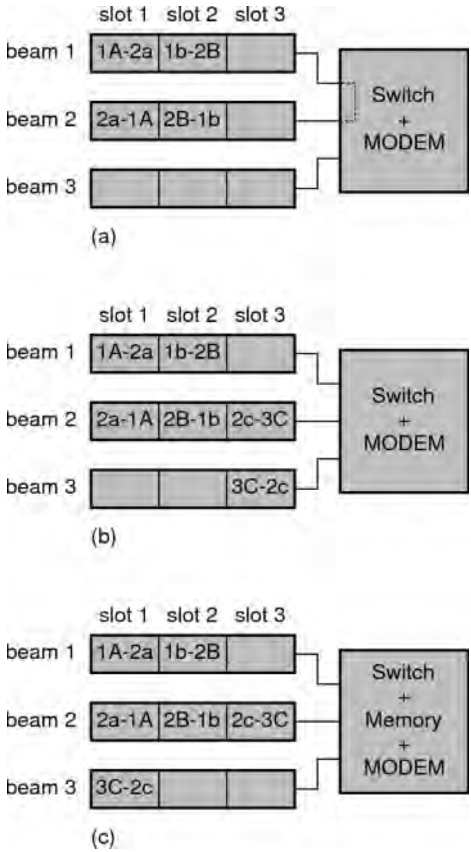


Figure 6.6 Three proposed baseband switching methods (a) carrier switching, (b) simultaneous-slot switching and (c) slot switching. (All parts source: Taira, *et al*., 1998. Reproduced with permission of John Wiley & Sons.)

simultaneous-slot switching and slot switching as illustrated in Figure 6.6 (Taira *et al*., 1998). The TDMA scheme consists of three spot beams with one carrier per beam and three slots per carrier.

In a *carrier switched* system, the source and destination TDMA slots are interconnected via carriers switched to the source and destination beams, excluding connectivity with spot beam 3. A request originating in beam 3, destined for beam 2, is dropped because there are no carriers available in beam 2. The carrier switch can be controlled either by the ground or on-board network control centre.

In a *simultaneous-slot switched* system using the same call pattern, a call request from beam 3, destined for beam 2, is assigned to slot 3 even though slots 1 and 2 of beam 3 are empty. The assignment corresponds to the slot available in beam 2. In this method, the message is switched to the appropriate beam within the baseband processor (BBP).

In a *slot-switched* system, a beam 3 call request is assigned the first available slots in each beam, that is, slot 1 in beam 3 and slot 3 in beam 2. Thus, messages have to be stored

Table 6.2 Simulation conditions for comparing the capacity performance of various MC-TDMA baseband switching schemes

Parameter	Characteristics
Air interface	
RF	2.6/2.5 GHz
Modem	$\pi/4$ shift QPSK/coherent demodulator
Multiple access	Multi-carrier TDMA; five slots/carrier
Transmission rate	70 kbps
Information rate	5.6 kbps (voice) 32 kbps (data)
Error correction	Convolution coding (constraint length 7, coding rate ½); Viterbi decoding (3 bit, soft decision)
Capacity simulation parameters	
Frequency reuse	Reuse every third beam
Capacity/beam	120 voice, 20% mobile-mobile (i.e. 40 channels)
Call duration	Exponential with a mean of 120 s
Call arrival	Poisson distribution
Traffic density in each beam	Uniform
Congestion handling	Call blocked and cleared during congestion
Call blocking rate	3%

(Source: Taira, *et al*., 1998. Reproduced with permission of John Wiley & Sons, Ltd.)

on-board in order to direct them to the correct slot in each beam and hence on-board memory is necessary, as illustrated in the figure.

The authors compared the on-board traffic handling capacity of these schemes using a computer simulation assuming a single dimension 12-spot beam over Japan (13 m diameter dish), S band mobile system and other characteristics summarized in Table 6.2; and they compared the hardware size and power consumption of the on-board processor for each. The results revealed that the carrier switching scheme is the least efficient, while the performance of the other two schemes is very similar; the processor hardware size and power consumption are minimum for the simultaneous-slot method.

6.2.1.3 Regenerative Transponder Technology

Technology of the regenerative transponder is relatively new compared to the transparent transponder. Many next generation MSS and FSS satellites deploy such transponders. A general set of requirements for the equipment is summarized as follows.

- Lightweight to minimize weight penalty;
- Radiation hardened, reliable and compact;
- Compliant with performance specification of parameters such as speed, accuracy, reliability, noise immunity, BER performance, level stability, etc.;
- Provision of autonomous checks to ensure correct functioning;
- Fault tolerant hardware and software to minimise occurrence of faults/errors;
- Low power drainage.

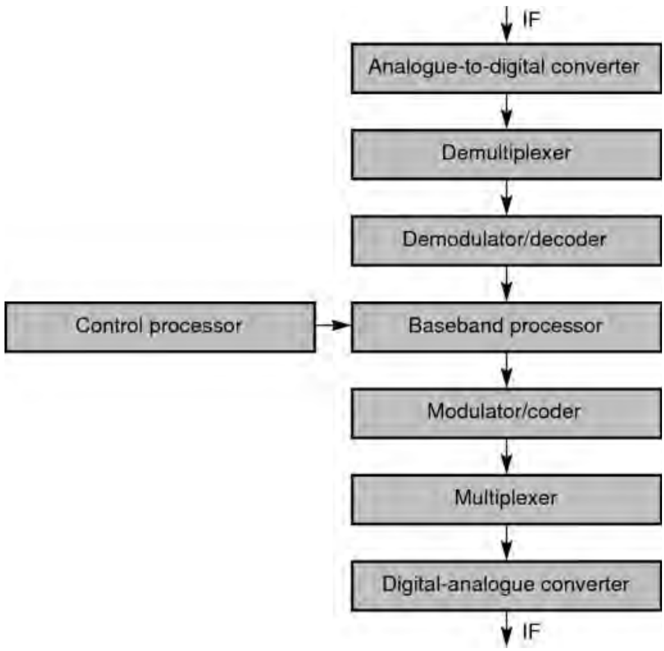


Figure 6.7 Main communications block of baseband processor architecture

BBP architecture can be implemented in various ways, but as illustrated in Figure 6.7, the main communications related functional blocks comprise:

- Modulator-demodulator
- Coder-decoder
- Analogue to digital converters
- Call processing and control sub-system
- Switch

Modulators can be implemented either directly at the down-link frequency or at IF; the techniques are relatively straightforward for the more conventional modulation schemes such as phase shift keying (PSK), except that the hardware must be space-qualified. The most complex implementation is that of the demodulator and baseband signal processor.

Demodulators use conventional demodulation principles; coherent demodulation, being the most efficient, can therefore be expected as a preferred choice. In addition to managing the well-known problems of carrier and clock recovery, the implementation problem on-board a satellite is complicated by the necessity to demodulate a number of incoming carriers. If a separate demodulator was deployed for each carrier, the size and power requirements of the demodulator bank would become prohibitive. Instead, multiplexed carriers are demodulated in a group, thus the hardware size is reduced – this is achieved by a technique called multi-carrier demodulation. Frequency division multiplexed carriers are digitized and separated into individual carriers, which are then filtered and demodulated individually by standard digital signal processing techniques. Multi-carrier demodulators may be implemented in an analogue form or digitally by the *Fast Fourier Transform* (FFT) technique in a device known as a trans-multiplexer.

Analogue domain FFT implementation can be achieved by a bank of contiguous tunable SAW filters in conjunction with a chirp oscillator; thereby, transformed signals are channelized simultaneously. The signals are then digitized for further processing.

Due to differences in transmission sources, synchronization requires care. The differences in the received carriers can impose a limit on the data rate and frame duration of individual carriers.

Figure 6.8(a) shows an example of an on-board processor proposed for an MSS payload to support a geostationary multiple carrier TDMA system for hand-held telephone and medium rate data service, as specified in Table 6.2 (Takeda *et al.*, 1997). Figure 6.8(b) shows the architecture of the forward and return processor for the proposed system.

The system incorporates baseband switching and mobile-to-mobile communication with signal regeneration and on-board call set-up. In calls to/from a fixed direction, on-board call setup (in cooperation with gateway) is used but without signal regeneration. Call set-up is managed by the control processor. De-multiplexers can demodulate up to 100 FDM (Frequency Division Multiplex) carriers per beam, while the switch can handle more than 100 carriers per beam. All four processors are implemented in application-specific

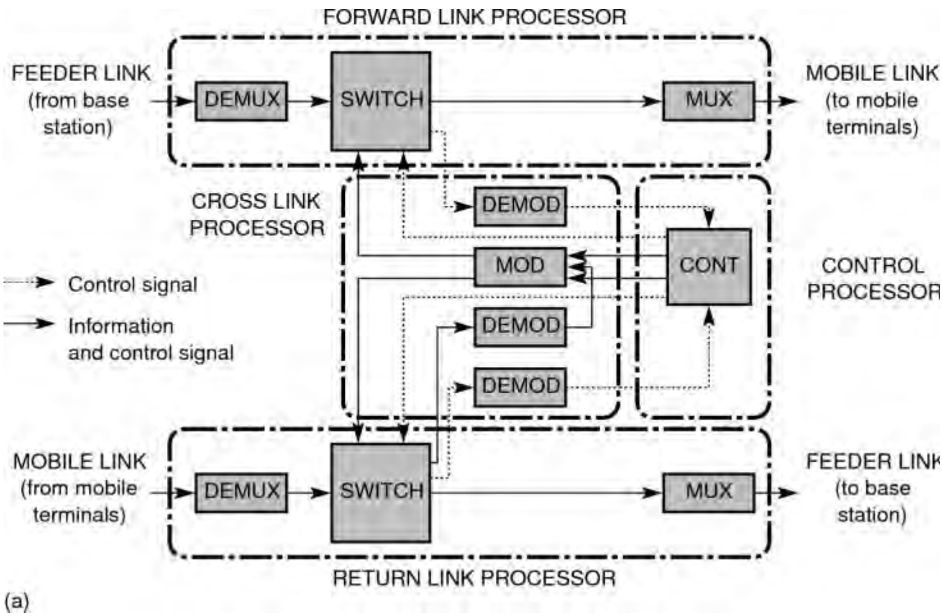
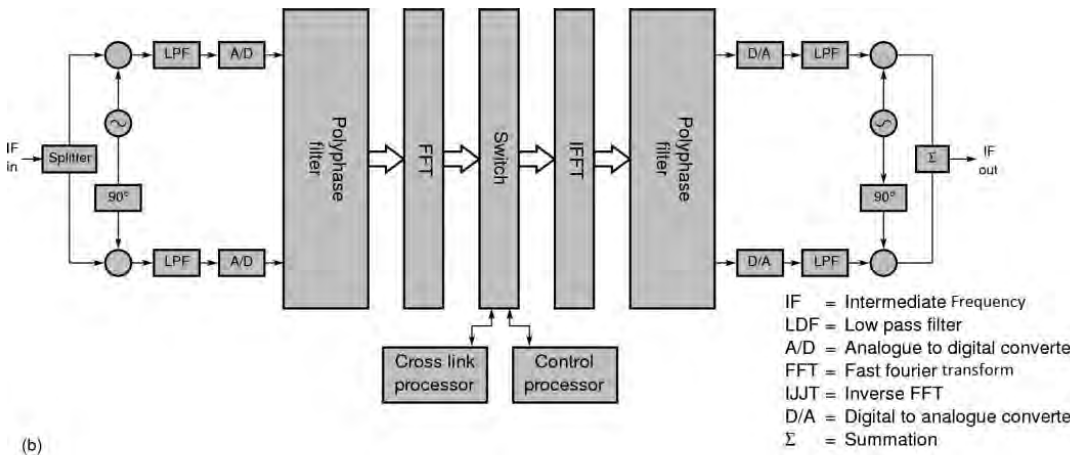


Figure 6.8 (a) A block diagram of an on-board processor proposed for an MSS payload to support a geostationary multiple carrier TDMA system for hand-held telephone and medium rate data service, as specified in Table 6.2



re 6.8 (b) Architecture of the forward and return processor for the proposed system. (Both parts source: Takeda *et al*., 1997. Fifth I Mobile Satellite Conference, Pasadena, California, June 16–18 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97- lsion Laboratory, Pasadena, California, June 16, 1997.)

integrated circuit (ASIC) and utilize digital signal processing. The forward and return processors illustrated in Figure 6.8(b) utilize polyphase FFT in preference to other techniques such as the bulk-FFT method and multistage method because it minimizes the number of multiplications per sample, thereby reducing the hardware size. The polyphase filters comprise finite impulse response (FIR) filters attenuating to 35 dB in the adjacent channel. The FFT used $N = 128$, which represents the number of carriers that can be de-multiplexed simultaneously and governs the size of the circuit. The pipeline junction method is used to minimize complexity. The polyphase inverse FFT method is used in the multiplexer to simplify the overall design. Modulator and demodulator units are implemented digitally (Figure 6.8a). The control unit performs all demand assigned channel allocation. Its functions include analysis and processing of channel control packets, call set-up and mobility management, that is, registration of the mobile terminal position, timing adjustment for each mobile, call control including channel search, assignment, carrier monitoring, channel release and RF transmission management. The controller's software can be uploaded from the ground through the feeder link. The controller can also generate standard timing signals; provide statistics of channel connection, and perform other functions. A redundant processor is used due to the criticality of the sub-system.

Baseband processing is achieved by integrated circuit chips at various levels of integration, that is large scale integrated (LSI), very large-scale integration (VLSI), etc. Most of the processing is implemented in the digital domain, but demodulation can use analogue processing. These devices can be fabricated by a number of solid state technologies. One of the main challenges in a space environment is radiation hardening of devices to minimize gradual degradation in performance due to radiation dosage. In a geostationary orbit, the main sources of radiation are electrons trapped in the Van Allen belt (see Figure 2.15). Other sources of radiation are solar flares and cosmic rays. Even with good shielding (e.g. 4–5 mm aluminium shields), on-board devices must be able to withstand a total radiation dose of 3×10^5 rad (Si) over a 10-year period. Additionally, digital devices are susceptible to single event errors such as latch up and soft errors due to ionization and high energy particles contained in cosmic rays. An adjunct advantage of miniaturization is a reduction in radiation dosage per component and hence an improvement in radiation hardness. The radiation hardness for a regenerative transponder can be specified as follows (Kato, Arita and Morita, 1987):

Latch up: free,
Total dosage hardness: 1×105 (Si),
Soft error rate: 1×10^{-5} error/bit day.

The main fabrication technologies expected to provide components for regenerative transponders are compared, vis-à-vis component density, power consumption, speed, susceptibility to single-event, soft errors and latch-up error due to radiation and availability in the year 2000 time-frame, in Table 6.3 (Kato, Arita and Morita, 1987). As regenerative transponder technology is evolving rapidly because of a number of regenerative satellites planned in the near future, particularly in the USA, the interested reader is encouraged to consult the most recent literature. Here we summarize the general trends as envisaged in the 1990s and should therefore be taken as indicative.

Table 6.3 Si and GaAs technologies

Material	Device	Total radiation dose (Rad (Si))	Soft error performance	Latch up performance
Si	Bipolar	$10^5 - 10^7$	Poor	Very good
Si	CMOS/bulk	$10^5 - 10^6$	Good	Poor
Si	CMOS/SOS	$\sim 5 \times 10^4$	Very good	Very good
GaAs family	Metal semiconductors (MES) Hetro-junction bipolar transistors (HBT) High electron mobility transistors (HEMT)	$>10^7$	Poor	Very good

(Adapted from Kato *et al*., 1987.)

The most commonly used technologies are Silicon (Si) or Gallium Arsenide (GaAs). Complementary-metal-oxide-semiconductor/Silicon-On-Sapphire (CMOS/SOS) technology provides the best performance, except for immunity to total dosage (a measure of the radiation absorbed by a device in space over the period of mission – see, Poivey, 2013 for a space hardening assurance methodology). However, the component density of devices using this technology is lower and development cost is high; therefore, they are better suited to special applications such as military. CMOS bulk devices have relatively poor latch-up performance, but several promising development approaches for hardening CMOS bulk devices are underway. GaAs and bipolar technologies are most susceptible to soft errors. It is expected that all the technologies will be able to provide the desired radiation hardness, but CMOS/SOS technology will continue to remain expensive.

CMOS technology achieves a higher number of gates per chip, larger storage density in terms of bits per chip compared to bipolar and GaAs technologies and is expected to remain as such in the near future. CMOS devices have the lowest power consumption, but in terms of propagation delay–power product GaAs show better performance than other technologies. Gate delays are the least for GaAS, followed by bipolar and CMOS.

It can be expected that for low and medium speeds, CMOS devices will be preferred and for high speed, bipolar or GaAs may be preferred, leading eventually to GaAs as the preferred choice.

For analogue on-board modems, the choice of two fabrication techniques is available – monolithic microwave integrated circuit (MMIC) that is suited for lower frequency bands up to ~4 GHz, and microwave integrated circuits (MIC) suitable for higher frequency bands. The MMIC device choice is either a Si bipolar device processed by a technology called super self-aligned process technology (SST) or a GaAs device. Below ~2 GHz, SST processed Si bipolar devices were preferred in the 1990s due to their proven radiation hardness in terms of total dosage. Between ~1 and 6 GHz, GaAs MMIC can be used. For modems above ~4 GHz, MIC technology can reduce the hardware significantly.

A survey of device technology used in RF devices conducted in 2005 shows GaAs as the dominant technology for low noise amplifier, modulator and RF switch; and silicon-germanium (SiGe) dominant for phase modulator. The reader interested in RF device technology for space applications can refer to Leon (2010), which provides details on RF technology, device structures, device physics, fabrication processes and materials used in RF devices.

A number of experimental missions have flown in recent years for technology validation and communication concepts. Examples of such missions are the USA's Advanced Communications Technology Satellite (ACTS), Japan's ETS-VI (engineering test satellites) and COMET experiment, Italy's ITALSAT (ITALian SATellite) programme and ESA's (European Space Agency) ARTEMIS (Advanced Relay and TEchnology Mission Satellite) and ALPHASAT satellites. On-board processing has been used for quite considerable time dating back to military satellites such as FLTSAT-7 and FLTSAT-8 (FLTSAT: FLeeT SATellite)(launched in 1989 – still operational in 2013). Amongst commercial systems, the Iridium system is a forerunner in extensive use of regenerative transponders including the ISL.

A number of K_a band proposals of 1990s intended to use regenerative transponders. An example of a recent implementation is the regenerative transponder of the satellite SPACEWAY® 3 (launched over North America in 2008), based on European Telecommunications Standards Institute (ETSI) and Telecommunications Industry Association (TIA) Regenerative Satellite Mesh A (RSM-A) standard for the K_a band (ETSI, 2004; TIA, 2006). The SPACEWAY® 3 satellite includes features such as spot beam forming, dynamic beam mapping and adaptive resource control. The regenerative payload using on-board packet processing provides full-mesh, single-hop connectivity between two or more terminals without a terrestrial support. The on-board processor routes IP packets to their destination and provides advanced broadband IP networking services including support for quality of service (QoS) needed for voice and video applications and on-board bandwidth-on-demand for IP traffic (Gopal and Arnold, 2009). Figure 6.9 shows an implementation of RSM-A architecture – showing the distribution of IP network, packet control protocol and packet lower layer functions across satellite, Network Operations and Control Center (NOCC) and terminal segments with respect to the core RSM-A functions. The core RSM-A functions relate to physical and link packet networking; the ground based NOCC provides

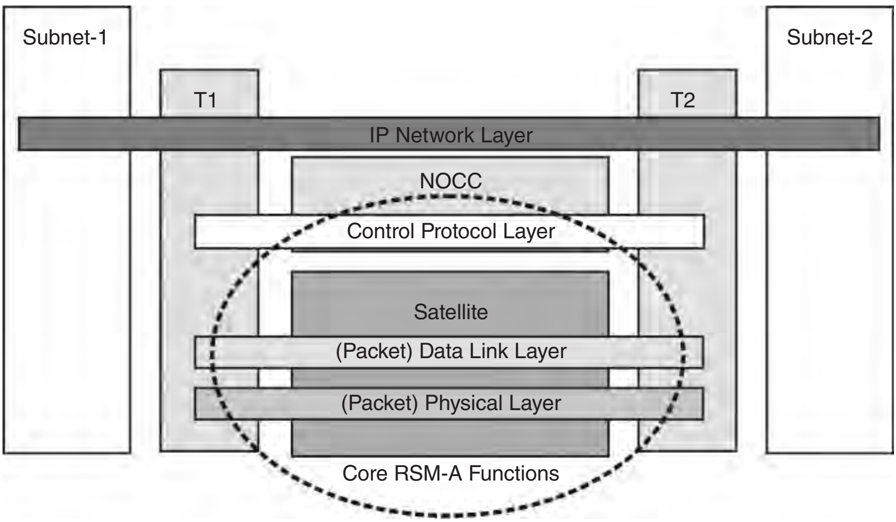


Figure 6.9 RSM-A core functions and a system instance. (Source: Gopal and Arnold, 2009. Reproduced with permission of IET.)

addressing, QoS signalling and multicast support for making network level decisions, which are communicated over IP protocols to the end user networks designated as subnet 1 and subnet 2.

NASA's ACTS programme was intended to demonstrate advanced technologies for next generation satellites with four main technology thrusts – K_a band system design and component development, on-board multi-beam communication processing system, multiple dynamically-hopping high gain antenna system and advanced network control architecture combining Earth and Space segments. Figure 6.10(a) and (b) (Campanella, Pontano and Chitre, 1990) illustrates ACTS' communications payload and BBP unit respectively.

The ACTS multi-beam communication package operates at 30 GHz in the uplink and 20 GHz in the downlink and includes a regenerative transponder comprising a BBP connected to a multi-beam antenna system. There are two modes of operation – in the baseband processor mode (BPM), communication is set up using the BBP, and in the microwave switch mode, the traditional satellite switched TDMA scheme is utilized. The BBP mode supports a lower bit rate and is called a low bit rate system (LBR); the satellite switched mode supports a high bit rate (HBR) and is known as a high bit rate system.

The ACTS communication package uses a TDMA scheme, which, for the LBR, operates at burst rates of 27.5 Mbps (available in uplink only) or 110 Mbps comprising respectively 1728 and 432, 64 kbps, channels assigned within 1 ms frames. There are three fixed spot beams directed towards Cleveland, Miami and Tampa; and two hopping beams covering two large north-eastern US regions that also provide spot coverage to 13 other states. The total capacity for each beam is 2000 channels, determined by the on-board memory. An adaptive coding scheme is used to make the link robust in the presence of rain fades or other degradation such as interference. Transmissions are normally uncoded; however, during adverse conditions the coded mode can be activated by ground commands. The LBR mode uses convolution coding with consequent data rate reduction. When both parts of the links are coded, the half-rate Forward error correction (FEC) code provides a margin of 10 dB at the expense of reduction in throughput by a quarter. In the HBR mode, link compensation is obtained by transmitter power control. The BBP demodulates the serial minimum shift keyed (SMSK) signal and, if commanded, decodes the incoming TDMA bursts. Each channel is then routed through a 3×3 time-space-time switch. The LBR system is controlled by a master control station (MCS) that communicates with users via an order wire (OW) for call set-up and with the ACTS multi-beam communication package for controlling BBP and antenna scanning. To establish a call, the user sends a request via the OW contained in the reference burst to the MCS giving its identity, the destination identity and the number of 64 kbps circuits required. The MCS determines TDMA bursts compliant with the request, routing, and beam dwell control message, and transmits the message to the BBP and hopping beam antenna control of ACTS in an uplink burst. The message is stored in an on-board memory, while the assignment information is sent by the MCS to the source and destination stations over the OW. The signalling messages are protected by FEC and repetition to minimize the probability of error. A synchronized burst time plan change is then executed by the BBP, hopping beam system and affected Earth stations. At the end of a call, the 'on-hook' message is sent by the calling or called party over the OW, which is used by the MCS to return the capacity to the space segment (Naderi and Campanella, 1988).

The satellite switched transmission mode, that is the HBR system, operates at 220 Mb/s to interconnect the three fixed spot beams mentioned. The satellite switch can also be

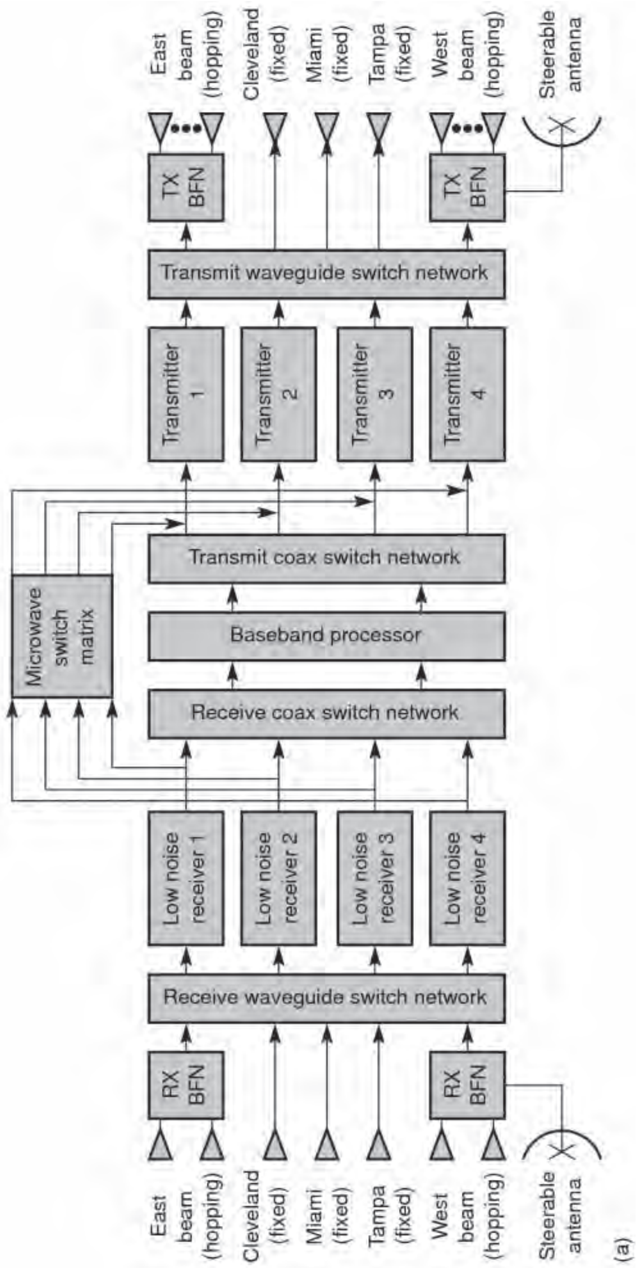


Figure 6.10 (a) ACTS multi-beam communication system

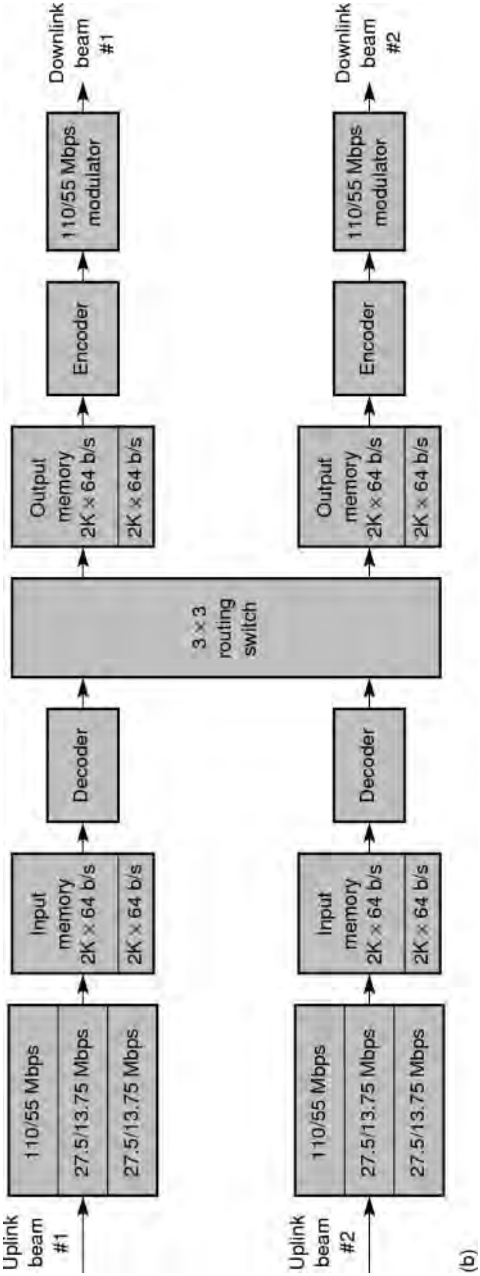


Figure 6.10 (b) ACTS baseband processor. (Both parts source: Campanella, Pontano and Chitre, 1990. Reproduced with permission of American Institute of Aeronautics and Astronautics, Inc.)

commanded to remain fixed for conducting experiments. ACTS baseband hardware makes use of custom LSI to minimize the size, weight and power of the BBP, combining a number of technologies depending on the complexity, size and speed needed for the unit.

Table 6.4 shows the function, application, technology, number of LSI chips, total radiation dose and comments on chip dimension (Gargione, 1990).

The hopping beam-forming network uses, as a basic element, fast ferrite switches capable of changing state in less than a microsecond. The 30 GHz LNA use high-electron mobility transistors (HEMT) to achieve a 3.4 dB noise figure. The payload technology was later used in the development of the Iridium system by Motorola, one of the technology participants.

NASDA (National Space Development Agency, Japan) and MPT (Ministry of Posts and Telecommunications, Japan) have a long-term programme for development of advanced satellite communications technology and systems through ETS and COMmunications and broadcasting Engineering Test Satellites (COMETS) programmes. A number of on-board technologies, such as multiple narrow spot beams (0.3°) and their interconnectivity using on-board switching, an S-band mobile communications payload, S, K_a and optical wavelength ISLs, were flown on ETS-VI but unfortunately, due to a problem in the apogee-kick motor, the satellite did not reach geostationary orbit. ETS-VI did not include a regenerative transponder; call by call routing was designed to be routed at RF or IF through a satellite-switch controller, which obtained routing information through the tracking, telemetry and command (TT&C) channel (Kawai, Tanaka and Ohtomo, 1991). The follow-on satellites in the COMETS series were expected to extend the multi-beam mobile technology to the K_a and millimetre band (50/40 GHz) with regeneration, and demonstrate S and K_a band LEO–GEO ISLs (Oshima *et al.*, 1993). For example, the mission of ETS-VIII launched in 2006 is to facilitate uptake of digital communications, including communications with mobile phones and other mobile devices. The 3 tonnes (gross weight) satellite 40 m in diameter deploys two large deployable antenna reflectors and two solar array paddles each about the size of a tennis court, is one of the world's largest geostationary satellites making mobile communications more reliable. Work is in hand to develop deployable antenna reflectors with metal-mesh, high-power transponders and on-board processors for application to other large space structures.

Other technologies being developed are satellite-based multimedia broadcasting system for mobile devices to provide CD-quality audio and video, more reliable voice and data communications, global positioning of and broadcasting to moving objects such as cars; faster disaster relief, and others. The mission also intends to experiment satellite-positioning technology, using a high-precision clock system, involving ETS-VIII and global positioning system (GPS), through the reception of signals transmitted from the clock (JAXA, 2013).

ESA's Olympus satellite, launched in 1989, demonstrated a number of technologies – a 18/12 GHz (up/down) high power direct broadcast system, a five-beam K_u band communication payload using a 4×4 on-board switch matrix, and the primary 30/20 GHz package for advanced communication experiments. A 12/20/30 GHz multiband propagation payload provided beacons for propagation data collection. The advanced communication experiment payload, using two narrow spot beams of 1°, was used in a double-hop on-board simulator experiment, where the on-board simulator, using a SAW-based multi-carrier demodulator for demodulating a group of 24 QPSK/FDMA (QPSK, quadrature phase shift keying) channels was placed on the ground. The demodulated signals were reformatted and transmitted back to the spacecraft as a TDM stream (Loo and Hayes, 1993). The experience

Table 6.4 Technology used in BBP custom LSI

Function	Application	Technology	Quantity	Radiation hardening level (total dose in RAD)	Comments
Maximum-likelihood convolution decoder	Input channel decoding	CMOS	4	10^4	3-micron gate length; ~30 000 transistor
Memory update controller (MUC)	Control memories	CMOS	24	10^6	5-micron gate length
Serial-parallel/parallel-serial	Input channel, output channel	CMOS	28	10^6	5-micron gate length
Encoder	Output channel encoding	MOSAIC-ECL	2	10^6	3-micron emitter width; 10 × 10 cell array; ~2 000 transistor array
Serial to parallel	Input channel	MOSAIC-CML	6	10^6	3-micron emitter width; 10 × 10 cell array; ~2 000 transistor array
Parallel to serial	Input memory, routing switch, output channel	MOSAIC-CML	32	10^6	3-micron emitter width; 10 × 10 cell array; ~2000 transistor array
Correlator	Input channel synchronization	MOSAIC-CML	6	10^6	3-micron emitter width; 10 × 10 cell array; ~2 000 transistor array
Timing and control: A	Input channel synchronization	MOSAIC-CML	4	10^6	3-micron emitter width; 10 × 10 cell array; ~2 000 transistor array
Timing and control: B	Clock distribution	MOSAIC-CML	4	10^6	3-micron emitter width; 10 × 10 cell array; ~2000 transistor array

Note: MOSAIC (Motorola oxide isolated self-aligned implanted circuits), a novel silicon process technology; ECL = Emitter Coupled Logic.
(Adapted from Gargione, 1990.)

gained by this simulated on board processing experiment was utilized for the ITALSAT project, a programme sponsored by Italy for advancing payload technology (Bellaccini and Rozera, 1991) ITALSAT has three main payloads. Two of the payloads use regenerative transponders; the third and the most advanced payload consists of a K_a band regenerative transponder using an on-board BBP for routing calls to any of the six spot beams directed towards Italy's main metropolitan areas. Signals from the six spot beams are multiplexed, grouped in threes, demodulated in coherent QPSK demodulators and fed into a 6×6 baseband switch, which operates in channel increments of 32 kbps for being routed to the desired beam. The switch is controlled by an on-board demand assigned multiple access controller that obtains call set-up information from signals uplinked on the TT&C channel. The baseband switch, in addition, synchronizes the demodulated bursts and generates a reference burst for network synchronization. K_a band uplink operates at 147.456 Mbps. The satellite was launched in January 1991.

ARTEMIS (Advanced Relay and Technology Mission) system, sponsored by ESA, built on technologies developed for the ITALSAT programme. The main thrust was to test an optical ISL and an L band mobile communication system using regenerative transponders. The spacecraft used a platform integrated control system to increase reliability, reduce weight and reduce power consumption using autonomous on-board processing to provide functions such as processing of on-board telemetry, management of power distribution, attitude and orbit control management, thus simplifying ground control requirements. The ARTEMIS system (in operation since 2003) has achieved the following objectives (ESA, 2011):

- Transfer of data from LEO satellites (SPOT-4: *Satellite Pour l'Observation de la Terre*: French for Earth Observation Satellite) in optical and RF bands.
- Satellite mobile communications over Europe, Northern Africa and part of the Middle East.
- Navigation services for the EGNOS (European Geostationary Navigation Overlay Service) system.

ESA's ALPHASAT satellite programme in addition to a powerful L-band capability that supports Inmarsat's operational system includes four hosted ('piggy-backed') payloads (ESA, 2012), that support laser communication for low earth orbit to geostationary orbit communication and a K_a band transmitter to transfer data to ground; Q/V band communication and propagation to research the use of these frequencies for future applications and propagation characterization; an advanced sensor known as Startracker to gain experience on its application for accurate and autonomous attitude acquisition; and an environmental testing and radiation sensor.

Other notable developments have been in the defence sector but much of the technology tends to remain classified. As an example, the US Department of Defense's Fleet satellite system flight-tested on-board processing technologies on FLTSAT-7 and FLTSAT-8 as long ago as 1986 and 1989 respectively. The on-board processing system consisted of a demodulator at the very high frequency (VHF) using a SAW FFT technique. The on-board processor consisted of an intelligent controller capable of assigning and routing calls autonomously, that is, without assistance from a ground station.

Iridium satellites with extensive use of on-board processing and ISLs demonstrate the maturity of regenerative transponder technology.

Table 6.5 summarizes the main regenerative technologies and antenna systems demonstrated for MSS during the 1990s (Brugel, 1994); note that all of these supported the Ka band

Table 6.5 The main regenerative technologies and antenna systems demonstrated during 1990s

	OLYMPUS	ITALSAT	ACTS	ETS VI	Artemis	IRIDIUM
Country/ Organization	European Space Agency	Italy	USA	Japan	ESA and Italy	USA/ Motorola
Launch	1989	1991	1992	1994	1996	1998/1999
Frequency band (GHz) (communications)	K_a band (30/20), Ku-band (18/12 and 14/12)	K_a band	K_a band (30/20)	K_a band, Ku band (14/12); S band (2.6/2.5)	K_a band (30/20); L band, K_u band	L band (K_a band ISLs)
Antenna system	DBS K_u band: two spots; Specialized services payload: five spots; Advanced communications payload: two spots (1°)	2 m, six spots, 5 and 3.5 m antenna	K_a band – three fixed spots, two pairs hopping spots, one fixed spot: various beam steering modes; beam size 0.27° and 1°	K_a band – 13 spots (0.3°); C-band – one spot; S-band five spots	3 m, multi-spot	3.2, 2.3, 0.8 m multi-spot
Regeneration capability	4 × 4 IF switch; ground based on-board simulation tests	6 × 6 baseband switch	4 × 4 IF switch; 3 × 3 baseband switch	16 × 12 IF switch – with on-board switch controller	Baseband switch	Baseband processor

Demodulator	SAW-CFT ground based	Coherent QPSK at 147 Mbps	Digital demodulation (SMSK)	Not implemented	Not available	N/A
Traffic routing technique	SS/TDMA controlled through ground commands	On-board DCP	BBP controlled from ground	Satellite switch controlled from ground	On-board autonomous	On-board
Solid state technology	CMOS for SAW-CFT	CMOS for demodu-lator and DCP, GaAS for soar cells	CMOS, MOSAIC-ECL and MOSAIC-CML for BBP (see Table 6.4)	Satellite switch: CMOS; SSPA: GaAs, IF switch: GaAS	SAW-CFT	GaAs VLSI
ISL	GEO-LEO Ka band	None	None	S Band (2.3/2.1 GHz); MM wave (43/38 GHz); Optical	GEO-LEO data relay	Ka band LEO-LEO

(Adapted from Brugel, 1994.)

and the experience gained from these missions has contributed towards the next generation MSS being deployed during this decade.

6.2.2 Antenna Systems

Invariably, all the MSS systems being planned or in operation deploy multiple spot beam antennas to maximize spacecraft EIRP and spectrum reuse. Technology for generating over 300 spot beams in the L band using large reflectors of up to 20 m diameter and multiple feed combinations is available. The peak gain of each spot beam is of the order of 30–40 dB with an isolation of 20 dB between reusable spot beams permitting terrestrial cellular-like frequency reuse grid. A number of techniques have been proposed to generate spot beams with such stringent requirements.

Direct radiating arrays use active radiating elements allowing considerable flexibility and failure resistance. The transmit and receive arrays can be mounted on the same aperture or separately. Separately mounted antennas are simpler to design but their deployment is complex, whereas composite antennas are complex in design with simpler deployment. The latter may consist of dual-frequency transmit/receive patches or separate interleaved radiating elements.

A number of reflector antenna designs have been proposed. In a multi-feed reflector antenna, the transmit and receive feed arrangement may be optimized separately. For example, in the receive section, a focussing reflector antenna using beam synthesis requires the smallest feed arrangement. As amplitude control at feed does not have any efficiency impact, the technique uses optimal weighting of pre-amplified signals from a few feeds only (Normand *et al.*, 1988). On the transmit side, the antenna may be fed with clusters of feed each element being fed with a single power amplifier. This arrangement requires feed switching to cope with changes in traffic load. Alternatively, a semi-active multi-matrix arrangement such as used in Inmarsat-3 and 4 satellites may be used. The technique has been used in the ACes (Asia Cellular Satellite) programme for generating 140 spot beams using a planar feed assembly comprising 88 cup-dipole radiators. This type of arrangement has good power efficiency with small reflector and feed sizes. In such an arrangement (see Figure 6.11), feeds are shared between several beams and fed from the same power amplifiers via a Butler-like matrix. The matrix adjusts the amplitude and phase to create spot beams at the desired Earth locations. Another approach for shaping antenna beams has been the use of shaped reflectors.

Lightweight passive large deployable antennas have been introduced in the past by Hughes in missions such as American Mobile Satellite Consortium (AMSC) satellites. This lightweight single-piece flexible 7×5 m antenna weighs around 20 kg and can be deployed passively. The antenna is curled and fitted to the nose of a satellite during launch and when released on station, the ‘springback’ antenna springs into its original shape.

The Astro Aerospace Corporation have developed low-cost, lightweight, deployable reflector technology for all offset-fed antenna sizes used on modern MSS geostationary satellites, meeting the stringent passive intermodulation (PIM) and electrostatic discharge (ESD) specifications. The reflector can be stored in a very compact form, giving an aperture size to stowed length ratio from 15–25 when the aperture varies from 6–25 m. Figure 6.12 shows the stowed and deployed form of the reflector (Thompson, 1997).

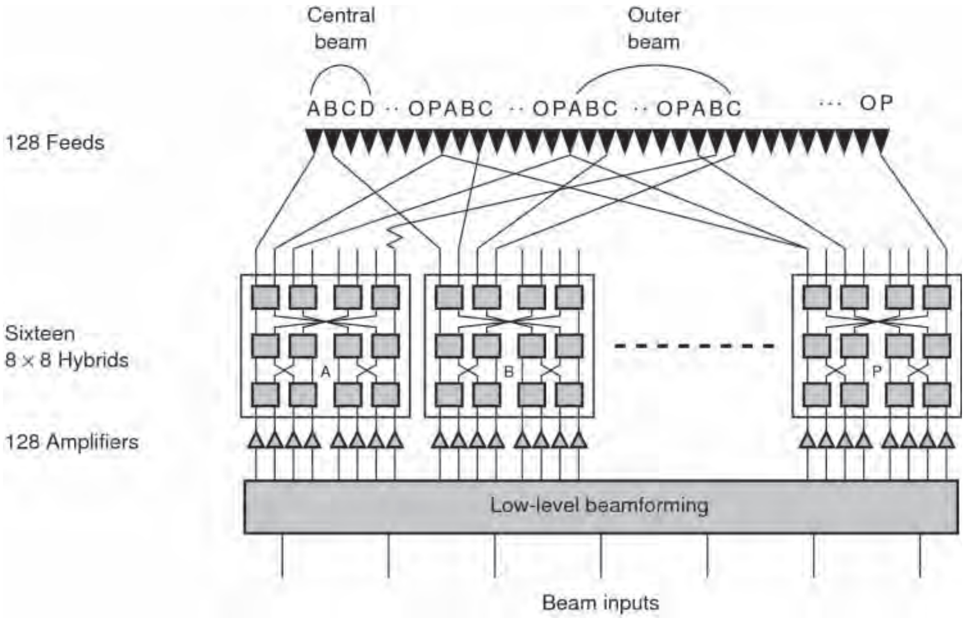


Figure 6.11 Multi-matrix spot beam generation. (Source: Benedicto *et al*., 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

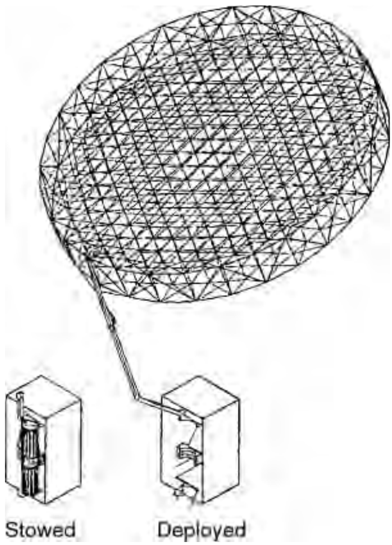


Figure 6.12 Stowed and deployed form of a reflector. (Source: Thompson, 1997. The Fifth International Mobile Satellite Conference, Pasadena, California, June 16–18 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97-11, Jet Propulsion Laboratory, Pasadena, California, June 16, 1997.)

Considerable effort is under way to introduce phased arrays on satellites. Static phased arrays have already been introduced in satellites such as used in the Globalstar network. Reconfigurable arrays that enable operators to change the shape of a beam according to demands or after spacecraft relocation have recently been introduced in MSS satellites such as Inmarsat-4. The Teledesic system planned to use phased array antennas to provide an Earth-fixed frequency assignment scheme, which requires satellites to track Earth-fixed regions.

The development of a potentially useful class of space structure known as an 'inflatable deployable structure', capable of deploying antennas of up to 30 m, is being sponsored by NASA. The structures could operate in the frequency range 0.3–88 GHz (Freeland *et al.*, 1998; Lou, 2000). A successful space flight of an inflatable antenna has already been demonstrated, in May 1996. NASA is reported to have contracted the development of technologies for such a structure to create inflatable habitat for inter-planetary missions to a US company (news item, www.parabolicarc.com, April 2013). One of the main difficulties is to construct 10–20 m antennas at low weight and cost; the use of inflatable antennas could reduce by 'orders of magnitude the cost, mass and launch volume of large space structures'.

6.2.3 Effect of Orbital Characteristics on Spacecraft Design

Up to the early 1990s, geostationary orbits were an obvious choice for MSS networks (see also Section 7.3.5). Since then, the introduction of non-geostationary orbits in MSS network architecture has resulted in a radically different network design approach. Both orbital characteristic and MSS network architecture influence non-geostationary spacecraft design, by imposing specific constraints. Orbital characteristics of interest are the number of eclipses, radiation environment, altitude and inclination. A network requirement of further interest is the use or not of ISLs and the extent of a satellite's on-board participation in network features such as mobility management and routing.

The extent of service support during a sun eclipse at the satellite affects the design of a satellite's power system, which must be able to operate during an eclipse when solar cells cannot generate electricity. Rechargeable batteries are used to provide power during blackouts; however, the problem for low/medium orbits can become severe in some types of orbits due to a higher number of charge-discharge cycles and relatively shorter periods for recharging the storage batteries. The number of eclipses in circular orbits increases as orbital altitude is reduced, with the exception of a Sun-synchronous orbit when the orbital plane is chosen to be normal to the direction of the Sun where satellites do not experience an eclipse (see Section 2.2.3). A geostationary satellite undergoes around 90 eclipses a year, with a maximum eclipse duration of 72 min, whereas a LEO satellite at an orbit of 1000 km, which has an orbital period of about 100 min can get eclipsed for about one-third of the orbit (~34 min) when the orbital plane is parallel to the Earth–Sun vector. Up to 5000–5500 eclipses can occur in a year in a LEO system in some cases. Note that the recharge time available to LEO satellites is considerably lower than that of a geostationary orbit whereas the number of charge–discharge cycles is significantly higher. Since the life-time of a battery is governed by the number of charge–discharge cycles and the depth of discharge, a battery that can operate for 12 years at a discharge depth of 0.8 in a geostationary orbit could have a lifetime of only about 5 years at 50% discharge depth. The geostationary orbit would experience 1080 eclipses in the period up to the end-of-life whereas the LEO could undergo up to 25 000 eclipses.

Since electrical power system can occupy up to 30% of a communication satellite's mass the goal is to utilize light weight efficient battery technology. A number of battery technologies are available – nickel-cadmium (NiCd), nickel-hydrogen (NiH_2), silver-zinc (AgZn), lithium ion (LiION). nickel-metal-hydride (NMH_2) has limited application and a sodium-sulfur (NaS) battery is under development.

The choice of battery technology depends on the mission life time, power requirement, satellite size, mass and charge cycle efficiency and cost (Hill, 2011). Nickel-hydrogen is known to be more reliable, smaller mass and has a longer lifecycle than nickel-cadmium batteries, which are their main competitor. At present lithium ion and sodium sulfur batteries cannot meet lifecycle and recycling requirements, and their operation is difficult. Spacecraft powers are foreseen to increase to 20–30 kW range with 15–20 year lifetime posing stringent demands on battery longevity and reliability. Nickel-hydrogen appears to be the preferred choice for such missions (Nelson, 1999).

The space radiation environment depends on the orbital altitude. In this respect, the highly elliptical orbit (HEO) satellites are most susceptible as they pass through a high radiation environment during parts of their orbit. LEO and MEO altitudes are selected so that these high radiation belts are largely avoided. Satellites in low earth orbits are affected by drag, in particularly if their array size is large, and hence there is a limit on the DC power generation capacity of LEO satellites. For a given terminal type, service and geographical traffic spread, a satellite's EIRP per channel reduces with a reduction in altitude due to a lower path loss, and at the same time the satellite's traffic capture area diminishes, reducing the satellite's peak load (see Chapters 2 and 7). Electronic devices are susceptible to sudden short failures in a radiation environment (see Section 2.2.3; see also, Leon, 2010), causing a software glitch and therefore fault-tolerant software is necessary, especially in critical aspects of spacecraft operation.

The number of gateways can be reduced considerably by using ISLs. This feature is useful for LEO satellite systems that may require hundreds of gateways for world-wide connectivity without ISLs. Associated with ISLs is a need for signal regeneration and packet-switching for establishing routes, and this requires considerable signal processing and computing power. Compared to inclined orbit, incorporating ISLs in polar orbit is less demanding due to a relatively simple ISL geometry. ISLs have also been proposed for geostationary satellite systems to improve connectivity. Mobility management is quite involved for non-geostationary orbits due to satellite movement with respect to users. If the network architecture requires on-board processing or radio resource management this imposes considerable processing requirements on satellites.

A significant effect of altitude on spacecraft requirement arises due to the strong influence of altitude on coverage area. Figure 6.13 depicts the applicable geometry. It can be readily shown that the edge-of-coverage carrier to noise ratio (C/N) at the spacecraft from a user transmission is given as (Egami, 1995):

$$(C/N) = (\eta/32)\ (E_h/kT)\ (D/H^2) \tag{6.2}$$

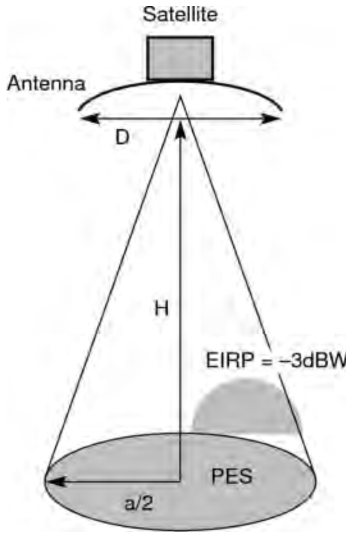


Figure 6.13 Geometry for estimating spacecraft antenna. (Adapted from Egami, 1995.)

where η = antenna aperture efficiency, E_h = EIRP of mobile terminal, k = Boltzmann constant and T = satellite's system noise temperature, D = antenna diameter and H = satellite altitude.

From Equation 6.2, we note that for a specified EIRP and satellite system noise temperature, the uplink C/N depends on the $(D/H)^2$. As an example, Figure 6.14(a) and (b) demonstrates the size of antenna for a hand-held system that typically has an EIRP of −3 dBW, assuming satellite system noise of 300 K and $\eta = 60\%$. Note the large antenna size required for geostationary satellite orbit (GSO). This was one reason why GSO systems were not favoured for hand-held service in the early 1990s when large antenna technology was not developed to acceptable standards.

The cell (or spot beam) diameter can be derived by geometry as

$$a = 2H \tan \theta/2 \tag{6.3}$$

where θ = satellite antenna 3 dB beam width of receive antenna ~70 λ/D.

LEO and MEO satellites are launched in clusters for economy and to complete the constellation as soon as possible for the service (see Section 6.5). Satellites must comply with the volume and shroud specifications of the launch vehicle.

Due to the typical size of constellations (e.g. number of satellites in Iridium constellation = 66), traditional methods of satellite manufacturing are inefficient. Only a few geostationary satellites are required to attain the desired coverage – typically one to two satellites for regional systems and four to five for international systems, and moreover, satellites can be deployed independently to provide service to large areas. Satellites can be assembled one at a time over two to three years and with rigorous tests at each step; replacement and diagnosis of defective parts could be tedious when spacecraft require disassembly. On the contrary, LEO and MEO constellations require significant numbers of satellites, if not all,

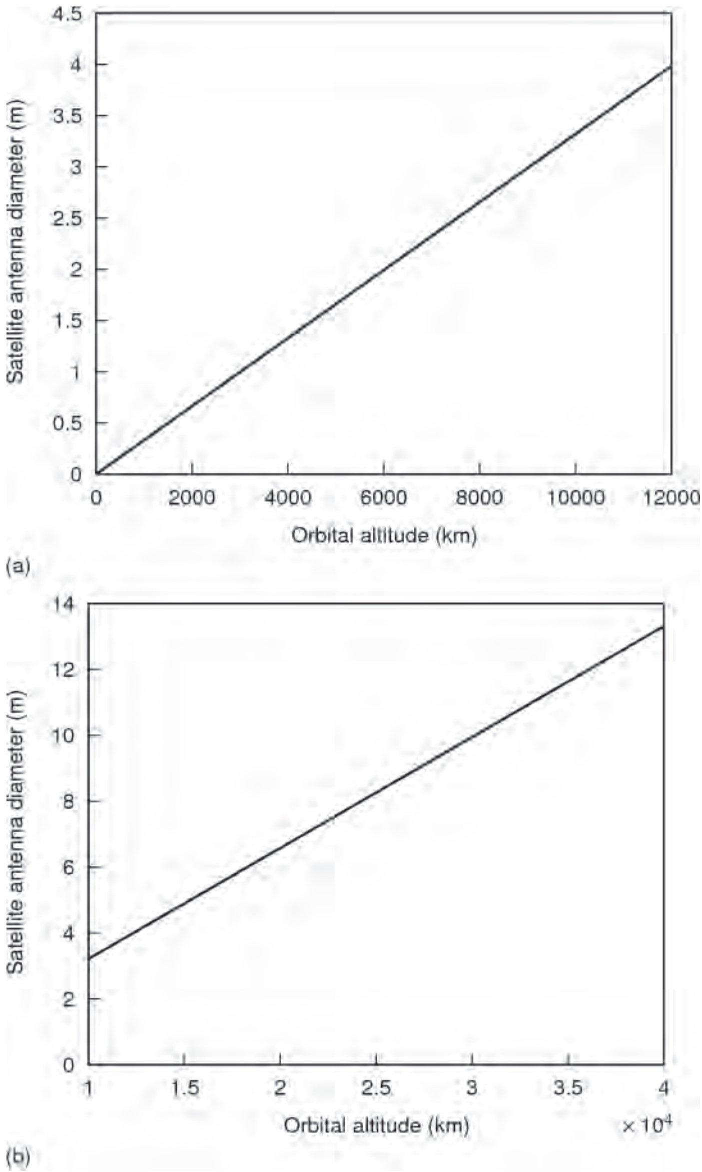


Figure 6.14 Satellite altitude versus spacecraft antenna diameter to support handheld communications at altitude (a) up to 12 000 km and (b) ranging between 10 000 and 40 000 km (Both parts source: Graphics AR.)

deployed quickly for reliable coverage, requiring a radically different assembly approach. We will consider the production approach of a number of recent systems as illustrations (Kiernan, 1996).

Iridium satellites were assembled with techniques used in automobile assembly lines, resulting in a satellite being produced every five days at the peak of activity in 1998, with up to 10 satellites being assembled in a day. The production time of each satellite was around 35 days. Motorola engineers undertook production aspect into consideration from the design phase with the view that 80–90% of cost, quality and production time is influenced by satellite design. The production was viewed in its entirety from component construction through to spacecraft delivery and on to the launch site. Components were produced to allow just the quantities needed for the current assembly – this ‘lean production’ is used by Japanese car manufacturers. The technique allows continuous refinement of the assembly process. Other techniques involved – use of off-the-shelf components, use of easily pluggable and interchangeable circuit boards, and rigorous low-level tests to minimize system-level problems. Before the start of production, extensive electrical, software, acoustic and vibration tests were performed on two full-scale satellite models to minimize the tests on production runs. Experience gained through assembly of the first batch of satellites was used for refining further production runs. The production line operated with about 50 engineers working in three shifts continuously at 15 different stations. Each satellite required about 160 printed boards that were assembled in one factory, sent to the next one where they were assembled and tested as sub-systems, and then sent to the final assembly line where they were integrated as satellite units.

The ORBCOMM satellites manufactured by Orbital Sciences Corporation and Teleglobe follow a different strategy. The ORBCOMM satellites are small 100 lb satellites, 34 in all, launched eight at a time. The process begins by using a rigorous programme for accepting components to minimize failures. All satellites are identical, even though it was possible to make some satellites less sturdy than others due to lower mechanical stress on satellites kept on top during the stacked launch – this approach is believed to have saved considerable manufacturing effort and time. Like the Iridium approach, only the first two satellites of the series were tested rigorously. Some of the tests were eliminated from later models, as the spacecraft were already characterized. Removing such tests reduced the manufacturing time by an estimated 50%. The goal was to produce a satellite every 10 days.

The Globalstar constellation of 48 satellites was manufactured by Loral Space and Communication. The company designated design engineers to remain assigned to a product through to its full production run. Mechanical and electrical interfaces were standardized to allow quick changes of failed sub-systems. Each module of the modular design was tested thoroughly before being assembled as a spacecraft. In the production line, satellites were moved from one assembly point to another and tested at each ‘functional island’. The satellites were always moved forward. In the case of a defective satellite, it was removed from the production line, and therefore did not affect the production flow.

6.3 Intersatellite links

The ISL technology, evaluated since the 1980s, was introduced in a commercial system in the 1990s (IJSC, 1988). Several future K_a and V band geostationary satellite systems plan to use such links in the future.

The topic of ISLs has been revisited a number of times since the 1980s but it was only in the late 1990s that the technology was deployed in commercial satellite systems. The Iridium system is a prime example in the use of ISLs.

The idea of ISLs has been around for a considerable time, having been studied for the Intelsat-IV A application. At that time, the interest was in links between geostationary satellites or between LEO and geostationary satellites, such as used in the TDRS (tracking and data relay satellite) programme. There were a number of disadvantages, which precluded their application to commercial systems. The ISL package would cause an unacceptable increase in the weight of the satellite and require on-board processing, which was not developed sufficiently in the 1980s. Moreover, ISLs would add additional noise to the severely power-limited downlinks, and the additional ISL propagation delay was not desirable for many applications. However, ISLs have been in use for non-commercial applications and experimentation since 1976. Table 6.6 lists the satellites that are known to have used ISL along with the year and ISL application (Muri and McNair, 2012). Notice that only the Iridium system has been used for mobile communication. In addition, following the success of the Silex experiment, the ESA within the Advanced Research in Telecommunications Systems Programme (ARTES-7) is implementing the European Data Relay Satellite System (EDRS) using laser and Ka band ISLs to relay data from LEO satellites. The system with commercial goals is expected to be operational by 2015 with plans to expand the operational system by a service called Globenet (Witting *et al*., 2012).

Table 6.6 Examples of the usage of intersatellite links

Launch Year(s)	Satellite(s)	ISL Frequency	Application
1972–1978	OSCARs 6–8	146 MHz	Amateur radio
1976	LES-8 and 9	36, 38 GHz	Demonstrate long range digital communication
1983–2013	TDRSS	C, Ku, Ka	Relay LEO satellite data to ground
1985–1995	Luch	UHF (ultra high frequency), Ka	Relay LEO satellite communication
1994	ETS-6	2, 23, 32 GHz, Optical	Experimentation/demonstration
1997	Navstar Block IIR	UHF	Navigation – cross link data transfer for autonomous operation
1997	Iridium	23 GHz	Mobile communication
1998	Comets (ETS-7)	2 GHz	Demonstrate communication between LEO earth observation satellite and a GEO satellite
1994–2003	MilSTaR I/II	60 GHz	Military communication
1998	SPOT-4	Optical	Data transfer via Artemis
2001	Artemis	S, Ka Optical	GEO system supporting data relay to several LEO satellites
2002	Envissat	S-band	Data transfer via Artemis
2002	Adeos-II	2, 26 GHz	Data transfer via Artemis
2005	OICETS	Optical	Data transfer via Artemis
2010	AEHF SV-1	60 GHz	Military communication
2015	Iridium Next	23 GHz	Mobile communication

The technology has now matured with successful commercial operation of over a decade. Advantages of deploying ISLs are as follows:

- ISLs can provide efficient connectivity when source-destinations are wide apart.
- The cost of routing can be optimized.
- They are a means of bypassing other carriers thus minimizing dependence.
- ISLs allow re-routing of packets if a node is congested; this node may be either a gateway or a satellite.
- Spare capacity can be redistributed between co-located or separated satellites. This feature also improves redundancy.
- The use of ISL on geostationary satellites can eliminate the need for multiple-hop satellite transmissions, or alternatively terrestrial connection, when connecting widely separated sites. This approach also reduces propagation delay, and saves valuable spectrum by avoiding satellite – ground-satellite hops.
- When used in a geostationary system, ISLs allow an increase in the elevation angle of the earth station by allowing it to use a higher elevation satellite for connection to a low elevation satellite. Such an arrangement could, for example connect London and Tokyo with 15° elevation visibility using an ISL between two satellites separated by 30°. Alternatively, such an arrangement can provide high elevation coverage for MSS within a region by creating a single virtual satellite by connecting two well-positioned satellites. Such a scheme allows two widely separated MSS regions to be coupled in to a single network and thus opens an interesting new MSS architecture.
- ISLs can be used to create a satellite cluster where, for example an MSS satellite could connect directly to an FSS satellite, bypassing a terrestrial connectivity.
- ISLs can increase the coverage of LEO satellite systems by connecting a LEO satellite to a satellite at a higher orbital altitude. An example of such a scheme is NASA's TDRS system. Here, LEO satellites gather weather and Earth resource data and transmit them to the ground via geostationary satellites facilitating an instant analysis. The ISLs are also used for LEO satellite command and control.
- ISL can also be used for interconnecting different satellite systems for lease or for mergers; if an ISL facility is built in, it reduces the risk to investors if the full capacity of a satellite is not utilized, as it is easier for the satellite to become a part of another consortium. One could envisage the use of a low-cost buffer satellite to interconnect different operators' satellites, if the ISLs are not compatible, as there is no agreed standard on ISL architecture (Morgan, 1998).

6.3.1 Frequency Bands

Most of the ISL frequency bands have been allocated in the region of the oxygen and water absorption bands, which are unsuitable for Earth–satellite paths because of the high absorption loss. Figure 6.15 shows the attenuation as a function of frequency. The first absorption band is due to water vapours centred around 22.2 GHz, where theoretical vertical one-way attenuation is of the order of 0.5 dB for moderate humidity and the second absorption peak of ~150 dB at moderate humidity at around 60 GHz is caused by oxygen. The attenuation is proportional to the Cosec of the elevation angle. Even if these bands were used in terrestrial systems, the signals would suffer considerable attenuation at ISL altitudes, thus minimizing the risk of receiving or causing interference to ISLs. Table 6.7 summarizes the main ISL

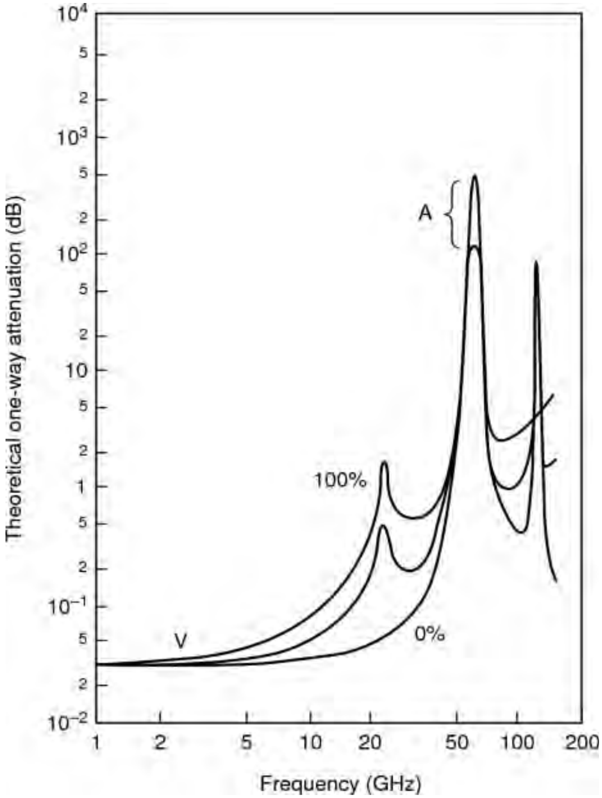


Figure 6.15 Theoretical one-way attenuation for vertical paths through the atmosphere (assumptions: US standard atmosphere for July, 45° latitude). The centre curve represents moderate atmosphere, the top and bottom curves the limits for 0 and 100% relative humidity. V – Vertical polarization; A – limit of uncertainty. (Source: ITU-R Rep 390-4. Reproduced with permission of ITU.)

Table 6.7 Frequencies allocated for ISLs

Frequency (GHz)	Bandwidth (MHz)	Allocation type	Note
22.55–23.55	1000	Co-primary	Radio astronomy use
24.45–24.75	300	Co-primary	–
25.25–27.5	2250	Co-primary	Geostationary uses limited
32–33	1000	Co-primary	Used for radio navigation and deep space communication
54.25–58.2	3950	Co-primary	–
59–64	5000	Co-primary	–
65–71	6000	Co-primary	–
116–134	18 000	Co-primary	–
170–182	12 000	Co-primary	–
185–190	5000	Co-primary	–

Note 1: All bands are shared; Note 2: Consult latest RR tables, as allocations can be revised.

frequency allocations. The reader should refer to the latest issue of Article 8 of the Radio Regulations for the most recent allocation.

Optical frequencies are also used for ISLs. Optical links provide a significantly larger bandwidth but require a precise intersatellite tracking. Optical links can support data rates of the order of 10 Gbps as opposed to microwave links that can support data rates of the order of 250 Mbps with existing technology. The International Telecommunications Union (ITU) does not regulate optical frequency allocation for ISLs. Optical wavelengths depend on the chosen laser source. Typical examples of lasers that can be used in ISLs are diode pumped Nd:YAGs (wavelength 1.064 to 0.503 μm), indium gallium arsenide phosphide lasers (wavelength 1.3–1.5 μm) and aluminium gallium arsenide lasers (wavelength 0.8 μm).

6.3.2 Implementation Issues

ISLs may use tracking or fixed antennas (telescopes for optical ISLs) depending on the ISL geometry, antenna beamwidth and target orbit control accuracy. The TDRS system uses tracking antennas and so do the Lincoln satellites, which operate in an inclined orbit at about 14° separation. Fixed antennas simplify the ISL system. Such links are used in the US government's geostationary Milstar system and the Iridium LEO system. An ISL tracking system reduces the station-keeping/attitude control requirement on satellites in contrast to fixed antennas that require satellites to remain within a tight bound.

Optical AlGaAs lasers were used in the Silex experiment (semiconductor laser ISL experiment) for connecting SPOT-4 remote sensing satellites with the ARTEMIS satellite.

The Japanese Optical Inter-orbit Communications Engineering Satellite included the capability to link with ARTEMIS. ETS-6 had planned S and K band intersatellite experiments but the satellite failed to reach a geostationary orbit for which the mission was intended. Experiments included acquisition, tracking and control of laser links with plans to establish links with ESA's ARTEMIS satellite. The proposed Wideband European Satellite Telecommunications (WEST) system was intended to use optical ISLs. The Teledesic system intended to use 60 GHz ISLs connecting up to eight adjacent satellites. The proposed EDRS will use a combination of optical and K_a band ISLs.

Although laser links provide high EIRP due to the signal cohesiveness, the receiver must track the source, imposing a stringent station-keeping and stability requirement on the spacecraft. One solution is to deploy a two-stage tracking system wherein a wide beam is used for initial acquisition; after acquisition, a fine tracking system takes over. A further consideration in optical links is that of solar interference, as solar rays can destroy the receivers. The receiver must, therefore, be isolated when such events are expected – as for example during satellites' Sun acquisition.

6.4 Emerging Technologies

A number of spacecraft technologies emerged for commercial applications in recent years. These technologies offer a number of improvements applicable to all types of communication satellites; but since each satellite is seen as an asset, buyers are reluctant to accept a new technology to minimize delay and risk to a commercial venture. Declassification of military technology in 1990s has enabled developments conducted in military programmes such as Milstar to reach commercial markets. These emerging satellite technologies are expected to make satellites smaller for a given performance requirement, smarter, more complex and yet cost effective (Foley, 1995).

A number of electrical propulsion technologies have been introduced recently, enabling an increase in satellite life by reducing dependence on chemical fuel for orbital maintenance. As an example, the Hughes Galaxy 3R satellite required 800 lbs less station-keeping fuel by using a Xenon engine – an electrical technique that has 10 times the specific impulse of existing chemical fuels. A xenon ion propulsion engine ionizes xenon gas by passing it through an electrified metal grid resulting in an efficient generation of thrust. Other recent techniques are the arc-jet system and stationary plasma thrusters. A plasma thruster allows a reduction of in-orbit mass by 15–20%.

Weight reduction is being introduced by deploying lightweight antennas and lighter material on the spacecraft body and payload. For example, aluminium honeycomb and graphite for the spacecraft body have been introduced by the Hughes aircraft company. In the electronics section, alloys that combine aluminium with lighter metals such as lithium and beryllium are under consideration. Use of newly developed stronger graphite fibre in composites is expected to reduce the overall weight by providing desired structural strength with less material.

Space Systems/Loral has introduced GPS receivers on LEO satellites such as Globalstar for autonomous attitude control and position determination. Thereby, the need for on-board sensors for attitude determination is eliminated and ground operations cost is reduced as the satellites are able to estimate their ephemeris. Research is under way to extend the concept to geostationary satellites. Other innovations introduced include better packaging with improved on-board processing that permits faster processing in a smaller space and gallium arsenide solar cells for higher power.

The time for introducing space-hardened computer technology has been reduced from 15 years to about 2 years, which gives the satellite industry a considerable advantage in benefitting most from research efforts in the computing field. This rapid turnaround is expected to be of paramount importance for future projects, which envisage the use of powerful on-board computers for functions such as fast packet-switching to support IP and ATM technologies.

Considerable progress has been made in MMIC and ASIC technologies, enabling a reduction in size, mass and costs of these circuits by a factor of 2–5 and, additionally, making satellites more amenable to mass production techniques, an essential ingredient for rapid constellation deployment (discussed later). MMICs are integrated circuits on low-loss dielectric and used for functions such as amplification, mixing, etc., and ASIC are integrated circuits developed to perform specific tasks such as computation or electronic functions of attitude control sub-systems. Weight reduction of up to 100 kg has been reported for spacecrafts. ASIC and MMIC enable the design of sub-systems in modules that may be reused for different missions.

Other areas where significant development has occurred are the Gallium Arsenide solar cell, Travelling Wave Tube Amplifier (TWTA) and battery technologies, which together enable introduction of very high power satellites (~10-20 kW). Gallium-arsenide cells are more efficient than silicon cells, enabling generation of higher power in the same surface area. Due to a large dissipation in TWTA, of the order of 50%, the devices become very hot, therefore cooling becomes essential. At higher power, the traditional conductive cooling is not effective; therefore, radiation cooling is used instead. In this method, heat is dissipated directly to space by using long cylindrical TWT (Travelling Wave Tube) and fins that protrude outside the spacecraft into cold space. Satellites of 15–20 kW, capable of 4.5–6 kW EIRP, are already operational and industrialists are forecasting power levels of 30 kW capable of transmitting EIRP of >9 kW in the near future.

We have already discussed a number of developments progressing in payload technologies. Examples of developments in antennas and related technologies of interest are large lightweight antennas, phased arrays, multi-port tracking amplifiers and advanced beamforming techniques. Developments related to transponder technology include on-board processing hardware and software and, to repeat, in the areas of MMIC and ASIC, packaging and material technologies.

Spacecraft K_a band technology is being deployed aggressively for next generation fixed and mobile personal communications. We have already discussed a number of enabling K_a band MSS-related initiatives, which have laid the foundations for such a development.

6.5 Launching Satellite Constellations

While the launching techniques for geostationary satellites are better known and documented, techniques used in launching non-geostationary satellite constellations are less so. Here we briefly review the launch sequence of geostationary satellites for completeness, followed by deployment description of two specific LEO satellite constellations – Iridium and Globalstar.

A satellite is launched into an orbit by injecting it into space at the required altitude in a direction and velocity appropriate for the altitude. Figure 2.5 (Chapter 2) shows the theoretical relationship between altitude and velocity for circular orbits, for altitudes up to the geostationary orbit.

As mentioned before, orbits used for MSSs lie within specific altitudes above ~700 km. Thus, a launch vehicle must impart a velocity of 7.4 kmps for an altitude of ~900 km (LEO), 4.6 km/s at an altitude of 12 460 km (MEO) and 3.0747 km/s at a geostationary altitude. The velocity increment of a launch vehicle can be optimized by maximizing the ratio m_f/m_0 where m_f is the mass of expanded rocket fuel and m_0 is the total mass of the launch vehicle. Therefore, most conventional launchers use multiple stages; each stage is jettisoned after spending its fuel reducing m_o to maximize thrust from the next stage.

The inclination of an orbit is governed by the latitude of the launch station, and is given as

$$\cos(\mathrm{i}) = \sin(\varepsilon)\cos(\theta) \tag{6.4}$$

where i = inclination, ε = azimuth of launch and θ = latitude of launch site. The minimum inclination equal to the latitude of the launch site is obtained by launching a satellite in an easterly direction ($\varepsilon = 90°$). If the required inclination of the final orbit is less than the latitude of the launch station, orbit corrections have to be applied, expending valuable fuel. Therefore, launch sites for equatorial orbits are chosen to lie as close to the equator as possible. A novel solution to this problem is to launch a satellite from a transportable vessel at sea, thereby allowing a launch from any chosen latitude, including a location on the equator.

Geostationary satellites may be launched directly into its orbit but it is more common to do so in the sequence outlined below. Velocity increment is least when a satellite is launched from a *parking orbit* to an intermediate elliptical orbit known as *transfer orbit*, which has an apogee at geosynchronous altitude and then injected into a geosynchronous orbit. This principle is called *Hohmann Transfer*, after Hohmann, who proposed the concept in 1925. The launch is conducted in the following sequence:

1. Launch into a parking orbit;
2. Transfer into the transfer orbit;
3. Inject into an inclined geosynchronous orbit; The inclination depends on the latitude of the launch site;
4. Orbital adjustments for the inclination to approach zero;
5. Drift satellite to the desired orbital location;
6. Conduct orbital manoeuvres to position the satellite at the desired location.

When the satellite has achieved a (near) geostationary orbit, the solar panels are deployed (for a three-axis stabilized satellite), the Sun and the Earth are acquired, and the communication antennas are deployed. The satellite sub-systems are checked out and the satellite commissioned.

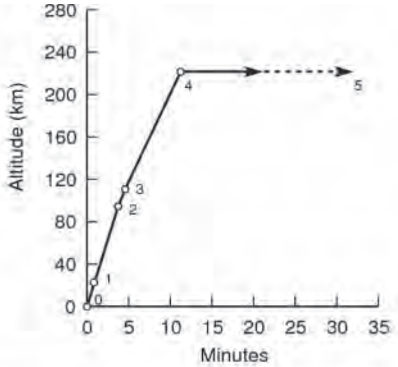
Figure 6.16 shows the launch sequence of a geostationary satellite and the typical time required to achieve major mission milestones up to injection into the transfer orbit.

A satellite is launched from the ground in a vertical direction to minimize air friction; its guidance system tilts the rocket to achieve an easterly direction for obtaining maximum advantage of the Earth's motion. The rotational velocity of the Earth at the equator is 450 m/s. To launch a satellite in a retrograde orbit would require an additional 600 m/s velocity. The launch flight plan depends on the launcher. To ensure that the position of a satellite is favourable in terms of its position relative to the Sun and that the satellite is visible from satellite ground stations, satellites must be launched within a time window called the *launch window*.

Deployment of a non-geostationary constellation quickly and economically is vital for commercial success. Therefore launching multiple satellites per launch are common – current vehicles can launch 3–12 satellites per launch in low Earth orbit. Satellites may be placed in low or medium Earth orbit directly or via a parking orbit. Typically, a launcher deposits clusters of satellites at regular intervals in a parking orbit. When a precise orbital position of the cluster has been determined, each satellite is manoeuvred to the desired altitude and inclination, by firing on-board thrusters at the appropriate time for a specific duration at the appropriate time, through ground commands. Support of a network of tracking stations dispersed throughout the world is essential for the purpose.

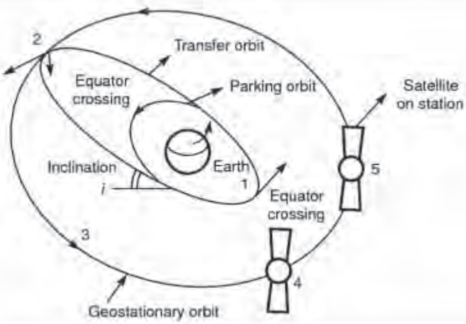
The Iridium constellation comprises 66 satellites in six 780 km altitude circular polar orbital planes (see Section 11.4.1) and the Globalstar constellation in its final deployment was planned to comprise 56 satellites (including eight in-orbit spares) in eight 1414 km altitude inclined orbital planes (see Section 11.4.2). Both consortia targeted the deployment of the full constellation in about 15–18 months.

Consider the Delta II launch of Iridium satellites from Vandenberg Air force Base in California (Kiernan, 1997a). Satellites are launched in groups of three. The first satellite is jettisoned at an altitude of 638 km in 3130 s (about 52 min) after lift-off, followed by the remaining two at intervals of 200 s. The satellites gradually drift apart over several days. About an hour and 40 min after the launch, the first radio contact with each satellite of



Event number	Event
0	Vertical lift-off
1	Guidance system begins tilting rocket towards east
2	First-stage drop-off
3	Second-stage ignition
4	Horizontal insertion into parking orbit 185 to 250 km
5	Second and third stages fired at equator to acquire transfer orbit

(a)



Event number	Event
1	Velocity increment to acquire transfer orbit; satellite spun for stabilization: attitude manoeuvres done before apogee-kick motor firing
2	Apogee-kick motor fired to give necessary velocity increment; orbit circularized and inclination reduced to near zero
3	Satellite despun
4	Three-axis stabilization acquired; antennas deployed/unfurled
5	Minor orbit corrections performed to correct residual orbital errors, in-orbit tests and position satellite on station

(b)

Figure 6.16 Launch sequence of a geostationary satellite. (Both parts source: Richharia, 1999. Reproduced by permission of Palgrave-MacMillan.)

the cluster is made by controllers stationed at Motorola's satellite control centre in Chandler, Arizona, when satellites are in view of Iridium's tracking station in Oahu, Hawaii. The contact is meant to check whether satellites are in the designated orbit and functioning normally. There are, in all, four tracking stations located in Hawaii, Yellowknife and Iqualit in the Northwest territory of Canada and Snjoholt in Iceland. To avoid the possibility of exhausting on-board batteries, solar arrays are deployed at an early stage, followed by deployment of communication antennas for ground and intersatellite communications. About three hours into the mission, satellites test their secondary antennas to be used in the initial checkout. The satellites are switched from the secondary antennas to the main antennas in the fifth orbit, giving an increased communications throughput. In the first two days, the satellites' primary antennas and modems are checked out, batteries recharged, software upgraded if required and feeder link and ISL performance checked out. About 48 hours into the mission, satellites begin ascending to the final location to an altitude of 780 km, firing low-thrust electrothermal hydrazine thrusters. The final orbit is achieved gradually in about two weeks. Thrusters are fired first over poles to raise the orbit and then over the equator to circularize the orbit. Once on station, L band communication links are activated and tested. Functions such as subscriber call handling, call handover between beams and satellites are checked out. Other functions such as paging, call forwarding and billing are tested, followed by beta testing, where average users test the system. Software upgrades are anticipated every six months or so, as practiced by terrestrial cellular system operators. A system simulation facility set up at the Chandler plant is utilized for solving technical problems. The facility uses an Iridium satellite model and simulated radio links allowing tests of satellite functionality under different conditions including different commands.

Globalstar planned deployment through a number of different launchers but some readjustments had to be done due to a launch failure (Kiernan, 1997b). The first two launches in batches of four were to be made from Cape Canaveral in Florida, each on a single Delta 2 booster. The next three launches were planned in batches of 12 on Zenit boosters. The final three launches were to be made on Soyuz boosters, supplied by a Russian–French consortium, ferrying four satellites per launch. As the intention here is to present the launch process rather than the sequence of deployment or the company's strategy in this respect, we will not attempt to follow the chronology. Note that the launch of a large number of satellites in a single launch can impact the schedule and cost significantly in the case of a launch failure. One Zenit launch failure resulted in a loss of 25% of the Globalstar constellation but the setback was recovered through astute planning.

For Delta 2 launch, four satellites held in a canister in the launch vehicle are jettisoned almost simultaneously at an altitude of 1250 km and in the same orbital plane. The initial few manoeuvres are initiated by each satellite autonomously using on-board computers. The manoeuvres include extension of the magnetometer boom, acquisition of the Sun and the Earth, stabilization to avoid tumbling and deployment of solar arrays to avoid battery depletion. When in orbit, the satellites are controlled by Globalstar's Operations Control Centre in San Jose, California, using tracking stations in Texas, France, South Korea and Australia. A preliminary health checkout of each spacecraft is made to ensure that the vital satellite functions, such as attitude control and propulsion, behave normally. Within a few hours, each satellite is commanded to fire thrusters to jettison itself to its final altitude of about 1400 km. The satellite injected last is boosted first to minimize the risk of collision.

The Zenit vehicle launches 12 satellites simultaneously, delivering them to an altitude of 920 km. The satellites are held in canisters as with the Delta launch, and ejected within 4 s in rapid succession. Speed is essential to minimize the risk of placing satellites in the incorrect orbit due to movement of the canister. The initial manoeuvres are identical to the Delta launch. But as satellites are at a lower orbit, they travel faster, giving only 10–12 minutes of visibility from ground stations. Moreover, as only six satellites can be placed in each plane, the satellites are grouped and injected in three separate planes. A further consideration is that the satellites experience different radiation and thermal conditions at 920 km to those experienced at 1400 km for which they are optimized. Hence, the satellites must be moved to the higher altitude as soon as possible. Satellite altitude is altered in groups of two or three to minimize the workload on ground controllers. Satellites are allowed to orbit until they reach their respective orbital plane, when they are jettisoned to their final altitude of 1400 km.

The Soyuz launch also sends satellites to the same altitude as Zenit but only four satellites are launched at a time. The eight satellites initially launched from Soyuz were intended to remain partially activated as spares in a 920 km orbit.

After a satellite has acquired its final orbit, the communications payload is checked out, which may last several weeks. As there are no ISLs, each satellite can be checked out independently.

Revision

1. Outline the generic system requirements of an MSS satellite. How do these translate into payload specifications?

2. Compare the advantage and disadvantage of a transparent and regenerative transponder. Develop the relationship between the up-link and down-link carrier-to-noise ratio and demonstrate the trade-offs between the two parts of the link stating assumptions.

3. Illustrate with a block schematic, the core protocols of the RSM-A architecture and its relationship with other functional components of an MSS system.

4. Plot a graph to illustrate the variation in the antenna diameter as a function of altitude in the range 700–1500 km and 10 000 and 14 000 km, given:
   Antenna efficiency = 60%
   E_b/N_o (two cases) = 5 dB-Hz and 10 dB-Hz

5. Compare (i) the launch sequence of Iridium and Globalstar systems and (ii) their manufacturing methodology for mass production satellites.

References

Bellaccini, S. and Rozera, G. (1991) The ITALSAT system. *Journal of the British Interplanetary Society*, **44** (9), 447–452.

Benedicto, J., Rinous, P., Roberts, I. *et al*. (1993) Geostationary payload concepts for personal satellite communications. IMSC, 16–18 June, Pasadena, CA, JPL Publication 93-009, pp. 169–174.

Brugel, E.W. (1994) On-board processing and future satellite communications services. *Space Communication*, **12**, 121–174.

Campanella, J.S., Pontano, B.A. and Chitre, D.M. (1990) A user's perspective of the ACTS hopping beam TDMA system. AIAA Conference Record Paper, AIAA-90-0833-CP, pp. 484–489.

Egami, S. (1995) Applications of multi-port amplifier to personal satellite communications. IMSC '95, Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 67–72.

ESA (2011) Artemis Achievements, last update July 2011, http://telecom.esa.int/telecom/www/area/index.cfm?fareaid=26, (accessed 23 August, 2013).

ESA (2012) Sharing the Satellite www.esa.int/esaTE/SEMWXV0T1PG_index_0.html (accessed 23 August, 2013).

ETSI (2004) Technical Specification. Satellite Earth Stations and Systems, Regenerative Satellite Mesh (RSM-A) Air Interface, TS 102 188 and 102 189.

Foley, T. (1995) Satellite technology – decades of development pay off. *Via Satellite*, **7**, 32–39.

Freeland, R.E., Bilyeu, G.D., Veal, G.R. and Mikulas, M.M. (1998) Inflatable deployable space structures technology summary. 49th International Astronautical Congress, Melbourne, Australia, September 28–October 2, 1998, trs-new.jpl.nasa.gov/dspace/bitstream/2014/20497/1/98-1484.pdf (accessed 23 August, 2013).

Gargione, F. (1990) ACTS hardware. AIAA Conference Record Paper, AIAA-90-0834-CP, pp. 490–496.

Gopal, R. and Arnold, S. (2009) Regenerative satellite mesh system for realtime multi-party multimedia traffic. ICSSC 2009: The 27th AIAA International Communications Satellite Systems Conference, Edinburgh, Scotland, June 1–4, http://www.hughes.com/ProductsAndTechnology/StandardsTechnicalPublications/Pages/TechnicalConferencesPublications.aspx (accessed 23 August, 2013).

Hill, C.A. (2011) Satellite battery technology: a tutorial and overview, *Aerospace and Electronic Systems Magazine, IEEE*, **26** (6), June, 38–43.

IJSC (1988) *International Journal of Satellite Communication (IJSC)*, Special issue on intersatellite links, 1988.

JAXA, (2013) Engineering Test Satellite VIII 'KIKU-No 8 ETS-VIII'', Japanese Aerospace Exploration Agency, Japan, www.jaxa.jp/projects/sat/ets8/index_e.html (accessed 23 August, 2013).

Kato, S., Arita, T. and Morita, K. (1987) Onboard digital signal processing technologies for present and future TDMA and SCPC systems. *IEEE Journal on Selected Areas in Communication (SAC)*, **5** (4), 685–700.

Kawai, M., Tanaka, M. and Ohtomo, I. (1991) ETS-VI multi-beam satellite communication systems. *Acta Astronautica*, **25** (7), 407–414.

Kiernan, V. (1996) Cheaper by the Dozen. *Satellite Communications* (Nov.), 29–32.

Kiernan, V. (1997a) Planting Telephone Poles in the Sky. *Satellite Communications* (Apr.), 38–42.

Kiernan, V. (1997b) Globalstar at the Starting Gate. *Satellite Communications* (Jul.), 36–39.

Leon, R. (2010) Radio Frequency Technologies in Space Applications, JPL Publication 10-11 10/10, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, trs-new.jpl.nasa.gov/dspace/bitstream/2014/41677/1/10-11.pdf (accessed 23 August, 2013).

Loo, C. and Hayes, E.J. (1993) Double hop on-board signal processing using Olympus. IEEE Globecom Telecommunications Conference, pp. 599–603.

Lou, C.M. (2000) Development and application of space inflatable structures. 22nd International Symposium on Space Technology and Science, Morioka, Japan, May 28-June 4, ISTS 2000-C-2 1, Jet Propulsion Laboratory, California Institute of Technology, trs-new.jpl.nasa.gov/dspace/bitstream/2014/14217/1/00-0627.pdf (accessed 23 August, 2013).

Maral, G. and Bousquet, M. (1984) Performance of regenerative/conventional satellite systems. *International Journal of Satellite Communications*, **2**, 199–207.

Morgan, W.L. (1998) Pass it Along. *Satellite Communications* (May), 50–54.

Muri, P. and McNair, J. (2012) A survey of communication sub-systems for intersatellite linked systems and CubeSat missions. *Journal of Communications*, **7** (4), 290–308.

Naderi, M. and Campanella, J. (1988) NASA's Advanced Communications Technology Satellite (ACTS): an overview of the satellite, the network and the underlying technologies. 12th International Communications Satellite Systems Conference, AIAA, March, Paper no. 88–0797, pp. 204–223.

Nelson, A.N. (1999) Spacecraft Battery Technology *Via Satellite*, February 1999, p. 6.

Normand, L., El-Shirbini, E., Neron, J., *et al*. (1988) A versatile array fed reflector antenna Part A: reception, Antennas and Propagation Society International Symposium, 1988. AP-S. Digest, 2, 6–10 June 1988, Syracuse, NY, USA, 502–505.

Oshima, T., Morikawa, H., Oshida, K. *et al*. (1993) Planned Experiments using the Communications and Broadcasting Satellites. 3rd European Conference on Satellite Communications, Manchester, UK, pp. 236–240.

Poivey, C. (2013) Radiation Hardness Assurance for Space Systems, https://nepp.nasa.gov/ (accessed 23 August, 2013).

Richharia, M. (1999) *Satellite Communication Systems: Design Principles*, MacMillan Press, Basingstoke and London.

Richharia, M., Trachtman, E., Westbrook, L., *et al*. (2006) Frequency planning approaches in mobile satellite systems, 3rd Advanced Satellite Mobile System Conference (ASMS 2006), Munich, Germany, 29 May–31 May, 2006.

Taira, S., Takeda, O., Otsu, Y. and Katagiri, H. (1998) Comparative study of switching methods for on-board processor in a mobile satellite communications system. *International Journal of Satellite Communications*, **16**, 131–136.

Takeda, O., Taira, S., Kawakami, Y., *et al*. (1997) A study on on-board processor for advanced mobile satellite communications. Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, CA, June 16–18, Co-sponsored by NASA/JPL and DOC/CRC, Jet Propulsion Laboratory, Pasadena, June 16, 1997, JPL Publication 97-11, pp. 399–404.

Thompson, M.W. (1997) The AstroMesh deployable reflector. IMSC 97, Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, CA, 16–18 June, Co-sponsored by NASA/JPL and DOC/CRC, Jet Propulsion Laboratory, Pasadena, CA, 16 June, JPL Publication 97-11, pp. 393–398.

TIA (2006) TIA-1040.1.07. *Regenerative Satellite Mesh-A (RSM-A) Air Interface*, Telecommunications Industry Association.

Witting, M., Hauschildt, H., Murrell, A., *et al*. (2012) Status of the European data relay satellite system. Proceedings of the International Conference on Space Optical Systems and Applications (ICSOS) 2012, 5-1, Ajaccio, Corsica, France, October 9–12, 2012, p. 8.

7

System Architecture

7.1 Introduction

In this chapter, mobile satellite service (MSS) system architecture is reviewed building on the concepts presented in earlier chapters.

A MSS provides telecommunication services to its user communities over a space segment interfaced to the terrestrial network and, as such, deals with transport of services from source to destination. The terrestrial system(s) can be a propriety network serving a close-user group or a public network such as the Internet, the public switched telephone network (PSTN), or a public switched mobile network. The scope of the MSS in such a network is primarily confined to the space segment. The communication to and from the terrestrial network passes through a MSS gateway where signals are adapted such that transmissions can take place efficiently over the space segment and on reception be reverted back into a terrestrially compatible format. The feeder link(s) (FL) connects gateway earth station(s) to the servicing satellite and the service link(s) connects the satellite to the users. The most critical component of an MSS network is the service link – the cordless tether that provides mobility. The role of such an MSS network is to ensure that a real or virtual connection can be established between users, and once established, the call can be maintained. Connectivity during a call may require handover between spot beams, satellites or fixed stations or point of attachment to the terrestrial node. There is also a desire to minimize the terrestrial tail in order to reduce the propagation delay and additional cost to the end user, attributed to the terrestrial component.

The formulation and development of an MSS system for a given service requirement necessitates an iterative approach involving adjustments to a variety of system parameters and entities – scope of a service, space segment architecture, operational concepts and business plans to cite a few. Since the relationship between these entities is complex and non-linear, it requires the evaluation of a team with diverse expertise. Therefore a diversity of solutions is possible.

The chapter begins with considerations applicable to the air interface, including those of Auxiliary Terrestrial Component (ATC) – also known as, Complementary Ground Component (CGC). We introduce the issues, influences and constraints applied in developing a mobile satellite system followed by an approach to system synthesis. In developing a low

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

earth orbit (LEO) or medium earth orbit (MEO) based system satellite altitude plays an important role in sizing the space segment parameters and user terminals. We introduce, with examples, the sensitivity of various system parameters to altitude. Network considerations, discussed next, include topology, call handling, mobility management (MM) and routing considerations in space-based connectivity architecture.

7.2 Air Interface

We define an air interface in the present context as the radio link and the associated functionalities that assist in the transport of the desired telecommunication services over the satellite system. Consider an MSS network as an extension of the terrestrially available services, for instance, telephony and Internet. In this configuration, the role of the satellite network is to transport the core network's functionalities transparently. Thus a generic protocol architecture comprises a *network independent* layer that deals with radio frequency transmission, which pertains to the satellite network and deals with parameters such as coding, modulation, multiple access, encapsulation of the incoming stream, etc. for transport in the satellite network; and a *network dependent* layer, which deals with transport protocols related to the core network and applications (see Chapter 8). Since a majority of the systems reuse the terrestrial core network, it is usual to deal with the network independent layers primarily when defining the satellite air interface.

Figure 7.1 illustrates the main functional blocks and connectivity architecture of an MSS system.

The service link (SL) provides a radio connection between the user and the satellite constellation; mobile-to-mobile communication is routed through a mobile–mobile service link (MMSL); intersatellite links (ISLs) are used for routing calls in space in preference to terrestrial routing; FLs connect a gateway to the satellite constellation; intra-gateway links (GLN, where N represents link number), inter-gateway links (GGLs) and network management links (NMLs) are essential for exchanging information between system entities to support network management (NM); and terrestrial links (TLs) connect an MSS system to a terrestrial network(s). To reiterate, the feeder and service radio links are crucial to mobility; whereas other links support various network functions but are independent of the transmission medium, and hence use a cost-effective and reliable transport facility such as a private backbone transport network, PSTN lines, leased circuits, a system's own spare space segment capacity, etc.

The air interface comprises an optimized gateway, service radio links and associated transport functionalities to connect the land earth stations, and the user community via a transparent or regenerative transponder. A number of transmission designs have evolved over the years and a number of these have now been standardized and recognized by the International Telecommunications Union (ITU). We will describe a number of prevalent air interface standards in detail Chapter 8.

The service link is the most crucial in MSS, as it imposes fundamental limitations on the capacity and capability of the mobile satellite system due to:

- small, low-gain low-sensitivity mobile terminals;
- power limited radio link;
- limited spectrum;
- a dynamic and harsh RF propagation environment.

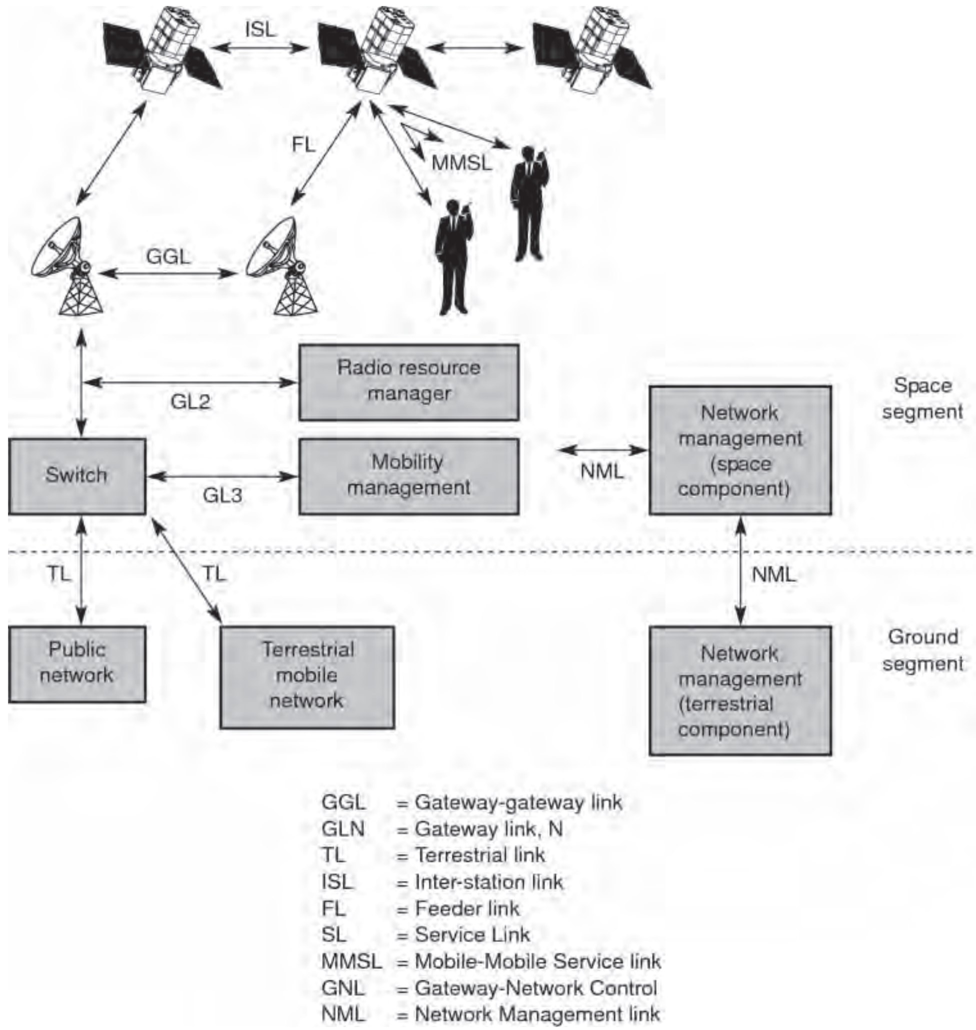


Figure 7.1 Functional blocks and the associated links of an MSS network. (Source: Richharia, 1999. Reproduced by permission of Palgrave Macmillan.)

Figure 7.2 identifies the critical components of the service link and the limitations attributed to each. The space segment capability is governed by the available satellite effective isotropic radiated power (EIRP) and G/T, which are constrained by the state of technology. There have been steady improvements in technology in the past decades, which has enhanced MSS service capabilities remarkably. In the late 1970s, geostationary satellites could transmit EIRP of ~30 dBW; by contrast, satellites being launched three decades later transmit 43 dB higher EIRP level. Similarly, spacecraft G/T continues to improve as spot beam size reduces. In the early 1990s, single beam geostationary satellites were the norm; the present satellites can generate over 300 spot beams; the corresponding increase in service link G/T ranges between −10 and +15 dB/K.

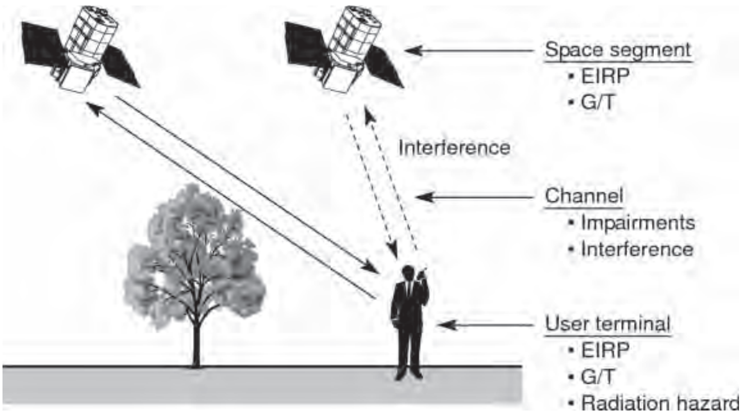


Figure 7.2 Inhibitors of MSS service links

Considerable effort is in progress internationally to alleviate the crucial problem of spectrum shortage attributed to an aggressive growth in the uptake of MSS resulting in an intense commercial competition for provision of mobile communications service. An effective solution for improving spatial spectrum reuse efficiency is to utilize narrow spot beams in a spatial arrangement analogous to the terrestrial cellular systems where each spot beams correspond to a cellular cell. Such spot beams provide enough isolation between non-adjacent beams to reuse frequencies. When compared to a single geostationary satellite, spatial frequency reusability has increased by a factor of ~40–100 within the past two decades assuming respectively 7–3 cell reuse in a satellite deploying 300 spot beams. Similarly a combination of a low altitude satellite and narrow spot beams offers considerable frequency reuse capability assisted by the narrower field-of-view of low altitude satellites. The Iridium system is stated to reuse frequencies 180 times globally with 2150 active spot beams (see Chapter 11). Note that modulation and multiple access methods also influence the system capacity to varying extents (see Chapter 4).

Shadowing and multipath effects restrict the throughput of mobile satellite channels, in particular the land mobile channel. From Chapter 3, we know that antenna directivity, vehicle speed and environmental characteristics influence channel behaviour. Throughput can be increased by deploying directive mobile antennas and operating in clear environments but this is a trade-off against terminal cost/complexity, mobility and link reliability. Effects caused by shadowing/multipath can be partially mitigated by selection of robust modulation, coding and multiple access schemes as discussed in Chapter 3. A more radical solution is to use path diversity, but it requires visibility to more than one satellite in conjunction with network support. A cost-effective solution is to promote user cooperation in selecting a clear satellite view. Because maritime and aeronautical systems suffer less severe shadowing and impose lower constraints on antenna size, a higher throughput can be supported, particularly at high elevation angles. Similarly, high throughputs are achievable for fixed land mobile applications, while for non-real-time applications, store and forward techniques provide a reliable countermeasure.

Radiation health risks impose a bound on system throughput for hand-held applications, as power transmitted from units must be restricted (see Chapter 5), resulting in extremely low received signal levels on satellites. The problem is mitigated by improving satellite G/T and/or path loss reduction by deploying low altitude satellites. Hence, invariably, all the systems that provide a hand-held service use a large number of spot beams and several use LEOs or MEOs to reduce path loss.

7.2.1 Ancillary Terrestrial Component

Satellite links are unreliable in heavily shadowed environments typical of urban areas, tunnels, railway platforms amongst others. This limitation is mitigated in a hybrid terrestrial-satellite architecture by terrestrial re-transmissions in disadvantaged environments of interest. The hybrid architecture is well-established in mobile broadcast systems with the Sirius component of Sirius-XM Radio system operating a terrestrially augmented audio broadcast system in the US region for over a decade to serve millions of listeners (see Chapter 12). This retransmission technique known as ATC and also CGC is based on the rationale that ATC transmissions must result in a more reliable and less expensive ubiquitous system than is feasible with the prevalent (and foreseeable) satellite systems; and that by reusing terrestrial technologies – at levels down to handset processing chips – the system would benefit through economies of scale as the addressable users would far exceed those from a satellite system alone. Such a cooperative network architecture, built on the strengths of respective systems, is envisaged as a potential next generation all-IP (Internet Protocol) universal network to provide seamless ubiquitous availability. Thus it offers wide-area coverage capability of satellite systems and the efficiency, capabilities and mass appeal of terrestrial wireless transport technologies in densely populated areas including the capability to cover satellite-disadvantaged locations. The concept received an impetus (albeit, temporarily) in the US when the regulatory authorities permitted the reuse of satellite frequencies for terrestrial reuse (see Chapter 8).

The coverage areas from terrestrial re-transmitters can be modelled as embedded spot beams wherein frequency reuse would be feasible by avoiding concurrent reuse of frequency between such virtual spot beams and the resident satellite spot beam while keeping the terrestrial transmissions at a low level. To support cellular-sized hybrid phones and maximize spectrum reuse, a high-power space segment with dense reuse capability is a key feature of the system.

Figure 8.22 illustrates pictorially the principle of an ATC system applied to a DVB-RCS+M (DVB, digital video broadcasting; RCS, return channel by satellite) implementation scenario demonstrating augmentation of satellite transmissions by terrestrial retransmissions to extend the DVB broadband service within a tunnel. The terrestrial systems may either repeat the transmissions at the same frequency to boost satellite transmissions or alternatively retransmit on a separate frequency in the same band. Figure 7.3 illustrates such a hybrid architecture that integrates the satellite and terrestrial wireless technologies to provide a seamless coverage. The satellite system is an extended cell sharing the applicable core networks and an identical or similar radio access technology with one or more terrestrial mobile wireless standard supported on a variety of software configurable user terminals.

The gap-filler system can either be a transparent transponder retransmitting the signal after amplification on the same frequency or another frequency in order to avoid regeneration due to transmitter-receiver coupling, and perhaps altering the physical layer if necessary.

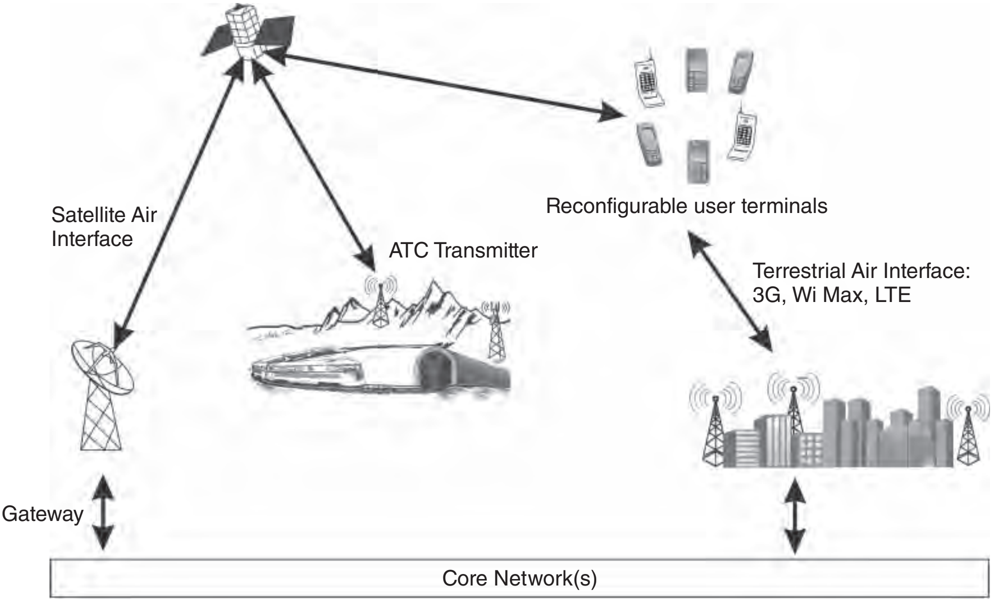


Figure 7.3 An integrated hybrid architecture to provide a range of ubiquitous services on reconfigurable user terminals

Alternatively the gap-filler scheme can extract the packets and retransmit in a suitable terrestrial format such as WiFi; in another approach the gaps are covered directly by terrestrial transmissions without involving satellite signals (Ryu *et al.*, 2007, see also Chapter 8).

Hybrid architectures are under consideration as a candidate for the next generation telecommunication network being defined in the ITU and hence there is considerable research activity in this area undertaken in various international research forums like ITU-R (International Telecommunication Union – Radiocommunication sector) and European Telecommunication Standardization Institute (ETSI). Topics being studied include protocol reference models and analysis, system descriptions and applications, quality of service (QoS) management, network architectures, transport layer issues, standards, services and methodologies such as cross-layer system design and optimization (Kota *et al.*, 2011).

7.3 System Development

In Chapter 2 we discussed constellation design solely in terms of orbital considerations. It must be amply clear by now that such a design cannot be considered in isolation as the constellation has a profound influence on the system architecture. The influence of the orbital characteristics can be perceived differently by prospective operators, leading to a variety of constellations and architectures. For example, an operator familiar with intersatellite communication technology may prefer the use of ISLs for connectivity rather than to use terrestrial connectivity; similarly, an operator with experience of geostationary satellite systems may select a complex geostationary satellite in preference to a system based on a LEO or a MEO constellation.

In this section, we review the factors influencing the system architecture and develop a structured approach to system optimization. The reader can refer to Chapter 9 for a detailed techno-commercial optimization approach to appreciate how intimately technical and financial issues are entwined.

A commercial venture can either be technology driven, wherein innovative products are marketed with the assumption that the product will be appealing to consumers, or products may be developed for real or perceived market needs established through market research. MSS proposals resting on extensive market analysis would be more balanced in view of the huge investments required for MSS. The main factors influencing a commercial MSS system architecture can be summarized as follows:

- communication service type;
- user terminal characteristics;
- service area;
- traffic distribution;
- system capacity;
- QoS;
- network connectivity;
- spacecraft technology;
- frequency band;
- orbital characteristics;
- launch considerations;
- schedule and financial risk;
- cost and revenue goals.

Due to the large number of intra-system dependencies, MSS system development is well visualized as a structured top-down multi-layered iterative optimization, wherein system synthesis cascades down progressively to lower layers iteratively, as design implications are better understood and real-world constraints applied. The concept of such a market-driven system evolution is illustrated in Figure 7.4 (also see Chapter 9).

Research targeted at the addressable market provides a basic set of telecommunication requirements, terminal characteristics and service area, which are used to synthesize a top level system architecture and estimate approximate costs, technology and schedule. Some iterations may be necessary before a realistic solution is achieved. The preliminary design is evolved into a detailed design taking into consideration the state of technology, costs, schedule, and so on. If the detailed design leads to an unrealistic specification, the requirements are modified until an acceptable solution is found; the detailed design specifications are used for developing all system elements, again with a possibility of iterations. The system elements are integrated and tested, leading to the operational phase that incorporates both technical and business elements; finally operational experienced, coupled with market trends is used for developing future strategy.

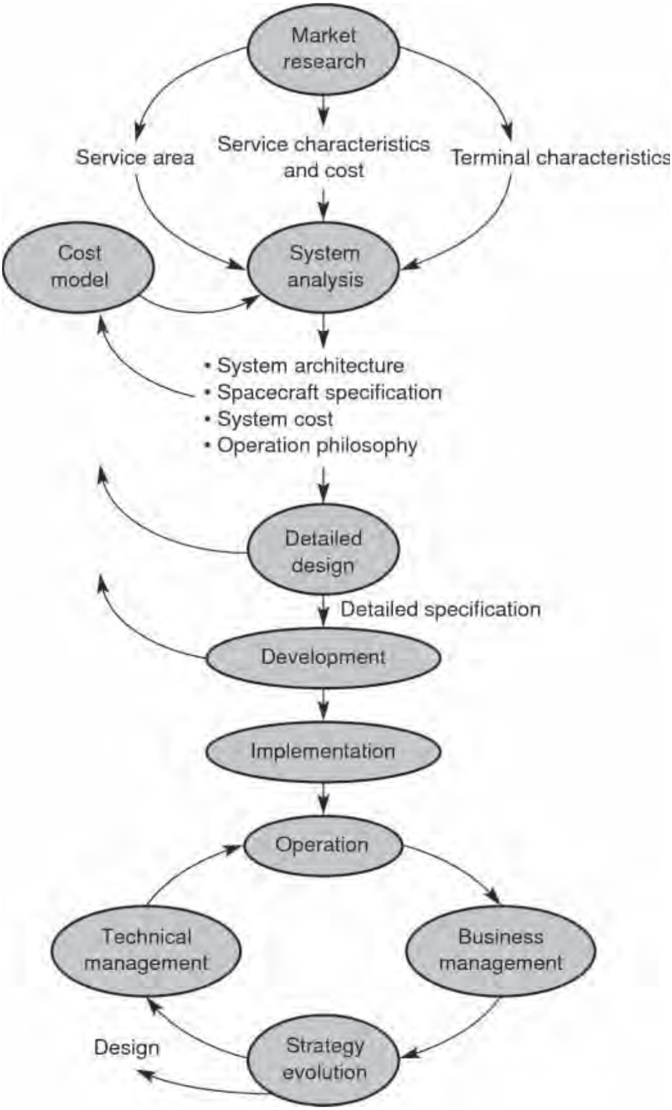


Figure 7.4 Concept of a market driven design methodology

Due to large variability in constraints and assumptions, the architecture is not bound by rigid rules, nor does a set of requirements have a unique solution. The examples that follow should illustrate vast variations in rationale and assumptions used by system developers.

7.3.1 Influences

Communication service, the size of the market, the service area and the cost model form the basis of a preliminary design. Consider each in some detail.

7.3.1.1 Communication Service

The main categories of services are:

- basic rate (a few bit/s to 1 kbps): real-time (e.g. aircraft automatic surveillance) or non-real-time (e-mail, supervisory control and data acquisition (SCADA));
- low bit rate (1–10 kbps): voice, facsimile, e-mail, basic Internet;
- medium bit rate (10–100 kbps): voice, facsimile, data, basic multimedia, medium rate Internet;
- high bit rate (100 kbps to 3 Mbps): multimedia, basic broadband Internet;
- broadband (3–10 Mbps): terrestrial quality broadband;
- broadband plus (> 10 Mbps): fast broadband.

The size and orbital characteristics of a constellation depend on the offered service. For non-real-time applications, intermittent coverage of the service area is adequate; however, for interactive services where time delay is critical, a seamless coverage at sufficiently low orbital altitude is necessary to keep propagation delay within limits. Lower orbital altitude reduces path loss and propagation delay but on the other hand the constellation size increases, the network gets more complex and constellation maintenance becomes more complex. Raising the orbital altitude to its upper bound reduces the constellation size but requires a higher EIRP and a more sensitive front-end receiver to compensate for the additional path loss.

Thus an operator such as ORBCOMM, marketing delay-insensitive products, was able to sustain operation with a partially deployed constellation. Yet others have chosen low orbital altitude to achieve transmission delay comparable to the terrestrial system at the expense of a large constellation and complexity, while several operators have preferred a geostationary orbit because of lower risk and simple network architecture, more stable link quality, etc., losing out on propagation delay and path loss. Iridium system designers chose a LEO, incorporating ISLs for network connectivity to reduce ground segment complexity; Intermediate Circular Orbit (ICO) Global systems achieved similar objectives using a MEO with terrestrial connectivity. Service throughput requirement also has a strong influence on terminal size and cost, spacecraft EIRP and complexity.

7.3.1.2 Market

Market size and scope govern space segment capacity, size and cost of terminals, service cost and the service area. Some of the current marketing opportunities are:

- remote area communications such as rural telephony, or public booths in remote localities or to serve government and aid agencies, and first responders;
- personal communications;
- communication for the transport industry;
- regional or global paging service;
- global/regional messaging such as e-mail, SCADA-type applications;
- mobile broadband services such as Internet access and, mobile television.

The ability of users to pay and the type of use – business, pleasure, distress or remote area communications – govern the portfolio of services, system cost/complexity and service cost.

A system targeted for personal communications must support a low cost hand-held service at the expense of a complex space segment, whereas a system targeting the transport industry, where size and cost of terminals or services are less critical, can deploy a simpler space segment.

Profitability depends on the revenue versus total investment and operating costs of the system. Space segment capacity, cost and anticipated revenue must be matched to break even as quickly as possible. Over-sizing the system capacity of constellations can be expensive and risky; capacity can be enhanced as the constellation is upgraded should demand outstrip capacity. Unforeseen technical problems or complexity may necessitate a revision to constellation size and network complexity later, after a detailed engineering study, as for the Iridium and Teledesic constellations that were each reduced from their original proposals of 77 and 840 to 66 and 288 satellites, respectively.

7.3.1.3 Service Area

Other inputs to architecture design are service area, traffic distribution and special features such as a region-dependent call-barring facility arising from regional licensing/political constraints. The information is used to determine the orbital inclination and eccentricity of non-geostationary constellations, spot beam distribution in the case of geostationary spacecraft and call-barring system features. Examples of such architectural features are:

- spot beam coverage of Inmarsat's fourth generation satellites, which were originally positioned over land masses to cater for land mobile traffic and later expanded to cover the maritime and aeronautical routes;
- coverage of the Globalstar system is optimized for mid-latitude regions and Ellipsat's for mid/high latitude regions;
- Iridium system includes features for region-specific call barring.

7.3.1.4 Cost Model

For developing a preliminary business plan, an approximate cost model is quite adequate, as precise costs can only be obtained after a system has been specified in detail. Planners can investigate the sensitivities of the business case to system attributes such as spacecraft cost, schedule, traffic growth, service cost, etc. on using such a model. Empirical cost models are available in the literature for estimating spacecraft and launch costs from system-level requirements. Established companies possess extensive database and up-to-date market information for accurate cost estimation. Such accurate estimates can only be obtained through a request for proposal (RFP) process as a part of a competitive tender. The RFPs lay out the system requirements as developed in the system design process; the proposal can be evaluated by the operator for the best combination of technology and costs.

7.3.2 Constraints and Considerations

System entities and their relationships must be synthesized applying real-world constraints and considerations. Such constraints may result in solutions, which are, at best, sub-optimal in an academic sense but acceptable in the real world.

7.3.2.1 Business

Financial viability is fundamental to a commercial venture. A number of proposed MSS of the 1990s failed to see the light of day due to financial shortcomings; others (such as Odyssey) merged with competitors; and others ended up selling off the assets to new investors. Consider some important business issues:

- *Financial risk*: In the initial reckoning, a number of assumptions impacting on profitability would have been made. This would include traffic forecast, geographical distribution of traffic, service penetration rate, user requirements, user's ability to pay, impact of competition, etc. Their applicability may become questionable for a variety of reasons such as delay in the introduction of the service, unexpected competition, poor service quality due to technical problems, delay in availability of user terminals, high usage costs, and others. If perceived by potential investors, such risks are detrimental to raising initial capital or cash flow at the start of operation.
- *Capital* for an MSS venture is raised from a number of sources and financial instruments; success in this respect depends on the state of economy, the performance of the satellite industry in general and MSS in particular, the credibility of the proposers and the soundness of the business plan (see Chapter 10).
- *Commercial issues*, such as cost per call, operational and maintenance expenditure and obligations, segmentation of revenue between entities – infrastructure provider, service provider, etc. – must be well understood to avoid perception of exaggerated revenue.

7.3.2.2 Regulatory Considerations

Regulatory issues deal with selection of spectrum and procurement of operating licence(s). The regulatory authority may impose technical and/or financial conditions for granting a licence. In such cases, the authority may require an operator to prove the venture's financial health, modify the proposed design, accept a tentative licence to prove the viability of the proposed design, etc. Some ventures such as LightSquared's proposed ATC venture failed to materialize because of withdrawal of the operating licence due to an interference issue (see Chapter 11). Designers should have included in their plans realistic information on licensing matters – for example, estimates of their share of spectrum in view of competition in the frequency band of interest.

7.3.2.3 Orbital Characteristics

Orbital characteristics are determined by coverage and service requirements. The constraints and detailed optimization techniques are discussed in Chapter 2. Suffice to say here that the impact of the space environment, transmission delay, constellation cost, size and maintenance together with network complexity play a crucial role in determining the altitude, whereas coverage area is influenced by orbital inclination and eccentricity.

7.3.2.4 Call Routing, Connectivity and Mobility Management

Network connectivity requires consideration on the location and management of gateways. System designers may opt for ISLs for increased flexibility in siting gateways. When an ISL is used numerous routing alternatives exist and therefore a suitable routing scheme must be established. Systems such as Iridium, which deployed ISLs, are quite robust in this respect; by contrast, the Globalstar system operates with a large number of gateways to allow regional participation; and the ICO system achieves network through interconnected gateways.

To establish a call, the mobile user must be located, paged and a route established; and when a call is in progress the connection must be retained. This task is achieved by a MM subsystem. In case of an IP connection, an additional task is to maintain the IP connectivity. After a call has been established, the radio connection can be broken when the visibility of the user to the satellite is lost due to the movement of the user or in a non-geostationary orbit (NGSO) by satellite itself. Thus it becomes necessary to handover the on-going call to a radio bearer that can retain the connectivity. Various types of handover can be envisaged:

- One satellite to another (satellite–satellite handover).
- One beam to another beam (beam–beam).
- One earth station to another (earth station–earth station).

The chosen handover scheme can influence spacecraft design, constellation architecture and air-interface design.

When considering connectivity in ISLs, regular handovers are necessary due to constellation dynamics. Numerous algorithms have been proposed based on a various optimization criteria. Since routing algorithms require on-board processing; a more complex algorithm imposes a greater burden on satellite processors.

7.3.2.5 Network Interface

The network must have provisions to interconnect mobile and fixed users with each other. This is connectivity is achieved through appropriate inter and intra network interfaces. The interface with the terrestrial network depends on the interoperability arrangements. For instance Inmarsat's Broadband Global Area Network (BGAN) network provides a seamless connectivity with the terrestrial 3G core network. Similarly, Iridium and Thuraya systems have provisions to interconnect with the GSM (global system for mobile communications) system.

7.3.2.6 Hardware Realization

First generation LEO and MEO systems were based on a number of novel technologies of that period, such as integrated handset design, multiple satellite launch, advanced spot beam technology, on-board processing, ISLs, and others. Depending on factors such as familiarity with technology and risk attitude, a number of constellation architectures have emerged – operators have usually combined mature technologies with novel ones, although the extent of combination differs widely.

7.3.3 System Synthesis

The system synthesis approach depends on the purpose and the scope of the system. A research team may be interested in assessing the sensitivity of system parameters for research purposes; an incumbent operator promotes enhancements to an existing service portfolio with a backward compatibility and, thus, is constrained in certain aspects; a new entrant would build the system based on in-house expertise and experience – thus, the synthesis methodology is not unique.

Here we follow a structured, logical and unbiased approach to system synthesis. The system requirements, influences and constraints are synthesized into a baseline design and cost estimates from a selection of alternatives, resulting in the preliminary definition of the following system elements:

1. space segment architecture;
2. spacecraft specifications;
3. space segment capacity;
4. air interface design;
5. network topology and interfaces;
6. fixed earth station specifications;
7. user terminal specifications;
8. system cost;
9. business plan.

Due to the large number of variables, interdependence and constraints, the synthesis process benefits from a combination of approaches – heuristic decisions may be preferred in some instances, whereas other aspects may benefit from mathematical analysis and simulation such as that developed for planning cost-effective very small aperture technology (VSAT) systems (Dutta and Rama, 1992). This particular model allows a planner to choose key physical and operational parameters and perform important sensitivity analyses like profit/cost versus traffic volume, technical parameters versus traffic, profit/cost versus price, operating parameters versus price and, profit versus demand. The model captures trade-offs between technical variables; estimates capacity and operational parameters through the geometric programming optimization method; and includes physical and regulatory constraints as well as the influence of transmission costs on planning variables. Figures 7.5 and 7.6 respectively demonstrate sensitivity analysis of percentage profit versus change in demand and traffic volume for the system considered by the authors.

7.3.4 Technical Trade-off Analysis

The reader may recall that the groundwork for system trade-offs has already been established in preceding chapters and in earlier parts of this chapter, as summarized in Table 7.1. Further treatment of the topic is available in Chapters 8 and 9.

Let us review the relationship between the air interface and constellation design. The air interface design deals with the radio link, which in addition to the characteristics of the earth stations and frequency, is governed by the characteristics of the propagation path and hence the orbital geometry. A geostationary satellite receiver would require about 24 dB higher G/T in the return direction (Earth–Space) than a LEO satellite system to provide the same received signal quality under identical link conditions. Compare the G/T of the Iridium satellite (-3 to -10 dB/K) with that of the ACes geostationary satellite (18 dB/K).

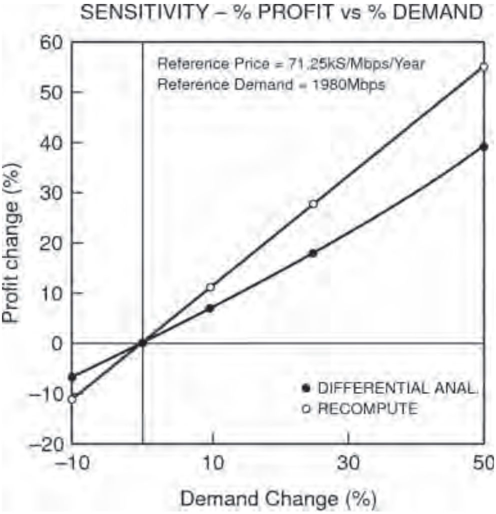


Figure 7.5 Sensitivity of profit change to demand change. (Source: Dutta and Rama, 1992. © 1992 IEEE. Reproduced with permission.)

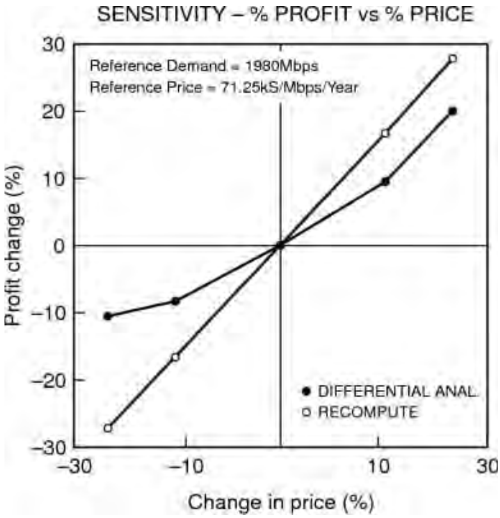


Figure 7.6 Profit versus change in price for the system investigated. (Source: Dutta and Rama, 1992. © 1992 IEEE. Reproduced with permission.)

A critical parameter in the radio link design is the fade margin. System capacity can be traded off against fade margin – a 3 dB drop in link margin can increase the capacity by a factor of 2. The link margin translates into user perception and experience – a higher margin provides a better protection against fading. Cooperative users learn to utilize the system more effectively (e.g. by moving away from obstructions while making a call), as evidenced in the cellular environment.

Table 7.1 Chapters where related system elements have been addressed

System element	Chapter
Constellation size and capacity	2
Multiple access, modulation and coding	4
Radio link	3
Satellite EIRP and G/T	3
Mobile EIRP and G/T	3
Standardization including network aspects	8
Business aspects	10

In order to mitigate fading impairments, the constellation can be designed for multiple satellite visibility from the ground to benefit from a diversity advantage (see Chapters 2 and 3). Systems such as ICO and Globalstar utilize diversity, while the Iridium system relies on a single visibility with considerably high fade margin built into the link design. Satellite-diversity improves the link reliability at the expense of an increase in complexity of the network and the need of additional satellite(s).

Modulation, coding and multiple access schemes were discussed at length in Chapter 4. The system designers compromise between the spectral efficiency, noise resistance, fade margin and cost-effectiveness of the various modulation and coding schemes. Similarly, multiple access schemes are chosen for the highest capacity, noise rejection and other features such as capability to support soft handover.

For a given satellite EIRP per channel and G/T, the capacity required per satellite, transponder type and other requirements such as ISLs, eclipse operation, orbital environment steer the spacecraft design. The estimation of capacity per satellite for geostationary satellite orbit (GSO) satellites is relatively straightforward in comparison to NGSO satellites. An effective method is to estimate the user distribution within the footprint of the satellite and the traffic per terminal; the satellite capacity satellite is then given as the product of the number of active users and the usage per terminal. There is usually an uncertainty in estimating traffic per user terminal particularly when dealing with a new service; temporal variation in traffic profile of the network adds further uncertainty.

It is desirable that radio resources of each satellite be utilized as flexibly as possible. Hence satellites include features for dynamic resource distribution between spot beams and steerable spot beams to direct extra capacity in case of a sudden spurt in demand (see Chapter 6).

Capacity estimation and radio resource management are more demanding for non-geostationary satellite systems because of the dynamically changing service area of each satellite. One approach is to size the capacity of each satellite for the densest service area, which then enables the use of only a single satellite design – resulting in economies of scale, flexibility in constellation deployment and in-orbit spacecraft redundancy needs.

Consider a hypothetical scenario of a global non-geostationary satellite system comprising p satellites serving an area of uniform traffic distribution. If the total global traffic is E Gb, then the capacity per satellite can be approximated as: E/p Gb. In a real situation, traffic has a non-uniform spatial distribution tending to concentrate along highways, ship routes and air corridors, in cities and other population centres with a diurnal variation and hence a more refined model is essential. In Chapter 10, a model to estimate geographic traffic distribution is presented, which may offer a more accurate estimate of geographic traffic dependence. Diurnal variation of traffic is given some attention in the later part of this chapter.

We have observed that performance analysis and optimization require modelling a number of system elements, their interdependence and estimation of performance statistics as a function of a variety of variables. The evaluation criteria in system optimization depend on the scope and intention of the analysis. The investigation may, for example, evaluate the end–end time delay, handover statistics resource optimization, or interference analysis. The associated parameters would be diurnal traffic variation, traffic distribution, constellation dynamics, gateway distribution, ISL connectivity, frequency reuse matrix, etc. Due to the inherent non-linearity and complexity and a need for repetitive analysis with flexibility, computer simulation is commonly used in performance evaluation. Figure 7.7 illustrates a flow chart of a simulation programme for estimating capacity, utilization, propagation delay statistics and length of PSTN lines of LEO or MEO constellations (Böttcher *et al*., 1994).

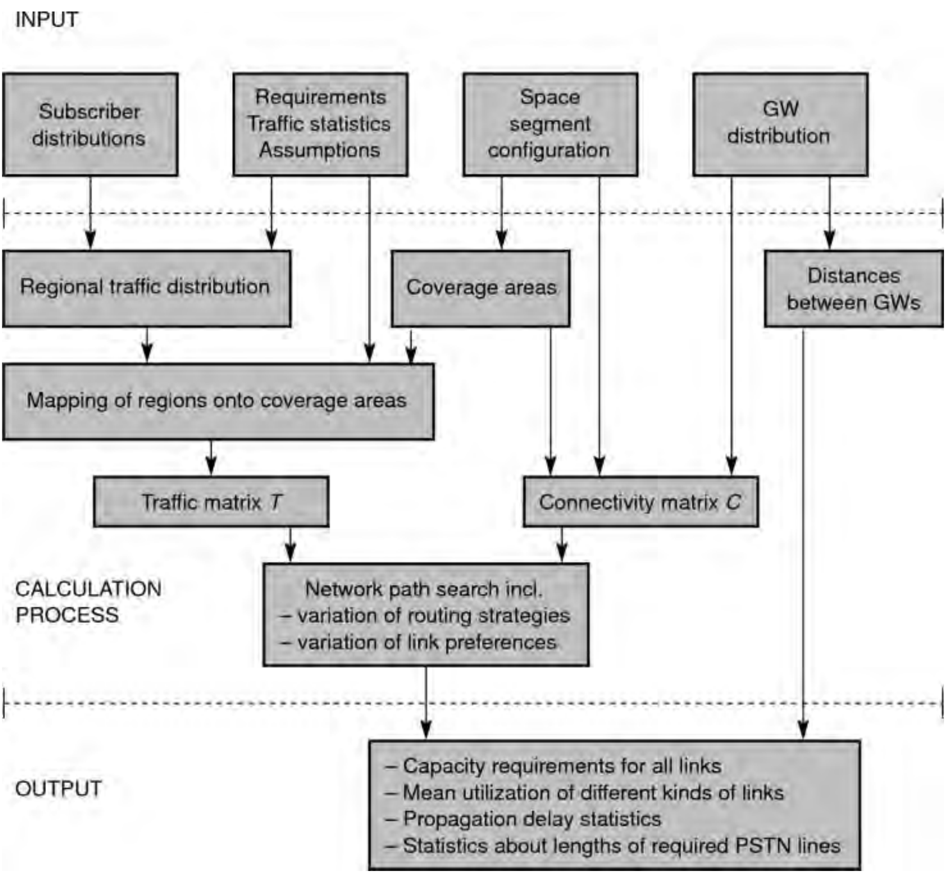


Figure 7.7 A flow chart of a simulation programme for estimating capacity, utilization, propagation delay statistics and length of PSTN lines of LEO or MEO constellations. (Source: Böttcher *et al*., 1994. Reproduced by permission of John Wiley & Sons, Ltd.)

Table 7.2 Regional subscriber distribution used in simulation

Region	Percentage	Number of subscribers – land only (thousands)
North America	25	250
Europe	25	250
Asia	20	200
South America	10	100
Africa	10	100
Australia/New Zealand	10	100

(Source: Böttcher *et al.*, 1994. Reproduced with permission of John Wiley & Sons, Ltd.)

The simulation was conducted with the regional subscriber distribution shown in Table 7.2 (but could be set up with other inputs) and calculations were performed for successive instants of time until the statistics have been estimated for different instants of time.

7.3.5 Impact of Satellite Altitude

In this section we will review the impact of altitude on system design, taking LEO satellite system as a basis. The maximum revenue achievable from an MSS depends on the space segment system capacity capable of maintaining the specified quality in terms of measures such as delay and bit error rate. System capacity is defined as the number of useful channels (circuit-mode) or throughput (packet-mode) available from the satellite constellation and depends on the usable spectrum (including frequency reuses), available EIRP per satellite and the number of satellites in the constellation. While the available spectrum depends on regulatory considerations, satellite EIRP/channel and frequency reuses are influenced by satellite altitude – required satellite power per channel increases with altitude and the spatial spectrum efficiency reduces given the same number and roll-off of spot beams.

Some of the system parameters of interest in determining the orbital altitude are:

- transmission delay;
- spectrum efficiency;
- spacecraft power;
- user terminal EIRP.

Gavish and Kalvenes (1998) studied the influence of satellite altitude lying in range 500–7500 km on these parameters, to produce a number of interesting conclusions. As expected, the study demonstrated that low altitude increases frequency reusability, capacity and power. However, results pertaining to estimated transmission delays were not as obvious. Ground routing exhibited a shorter delay with an increase in altitude, as expected. When using ISLs, however, medium altitude within the range studied proved to have the shortest delay because the number of switching nodes in low altitude increased, which increased the switching delays. It was noted that the delay depended on type of traffic, i.e., local or long distance, and it was noted that with evolution in technology, switching time would reduce. Furthermore it was noted that that frequency reuse increased as altitude is reduced; however, the capacity limit in this case is the power capacity of the spacecraft. As expected, handset power increased with orbital altitude, impacting the size, weight and the available throughput.

Figure 7.8(a) shows the number of frequency reuses possible over the Earth as a function of altitude for a number of spacecraft antenna beam sizes. Reuses at an altitude range from

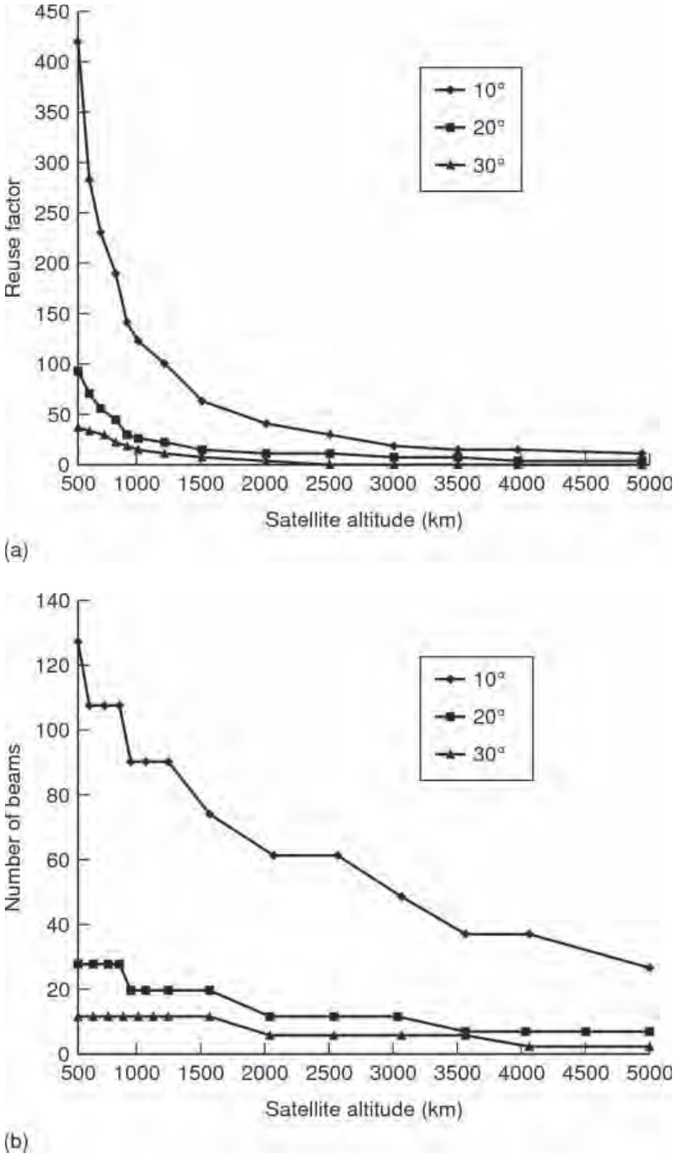
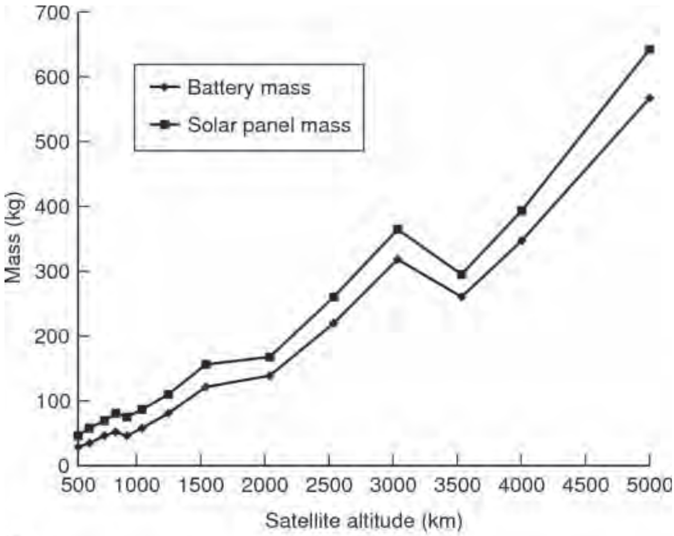


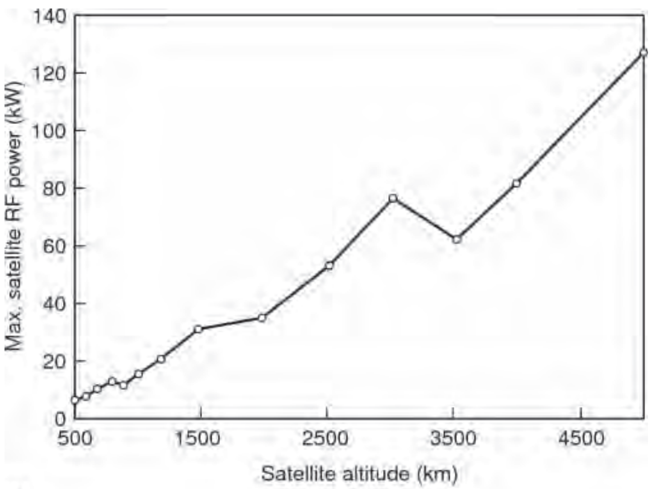
Figure 7.8 (a) Altitude versus possible frequency reuses for spot beam sizes of 10, 20 and 30°. (b) Altitude versus number of beams for spot beam sizes of 10, 20 and 30° (Both parts data source: Gavish and Kalvenes (1998). Graphics AR.)

about 125 for spot beam size of 10° to about 20 for spot beams of 30°. Figure 7.8(b) shows the number of spot beams possible with various spot beam sizes at altitudes up to 5000 km.

Figure 7.9(a) demonstrates the sensitivity of the solar panel size and battery mass with altitude, illustrating that the mass of the power system increases with altitude, even though capacity and eclipse period reduce; this occurs because capacity increases linearly but power increases as the square of the altitude as portrayed in Figure 7.9(b) that shows the increase in satellite power per channel with altitude and Figure 7.9(c) shows the variation in terminal mass with increasing altitude.



(a)



(b)

Figure 7.9 (*continued*)

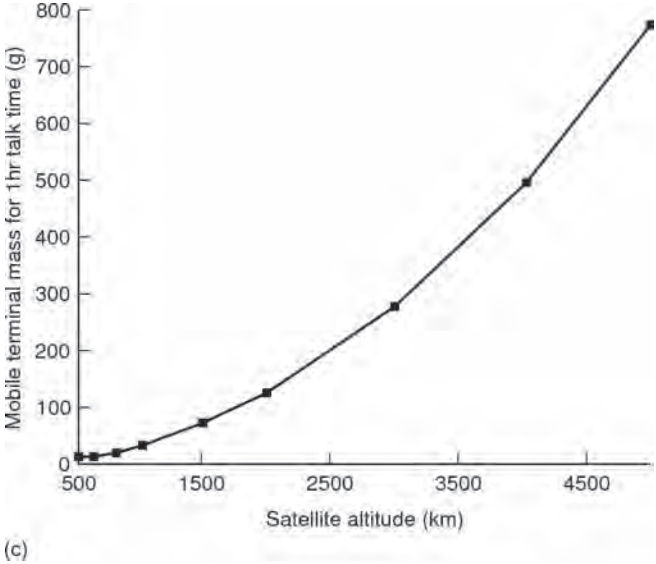


Figure 7.9 (a) Altitude versus spacecraft battery mass, assuming spot beam size of 20°, minimum elevation angle of 10°, a reuse pattern of 10, data rate of 64 kbps and signal-to-noise ratio of 30 dB (b) Altitude versus satellite power requirement using the same assumptions as in (c) Altitude versus mobile terminal mass for 1 hour talk time for the system (Data source: Gavish and Kalvenes (1998). Graphics AR.)

7.4 Network Considerations

7.4.1 General

MSS systems complement terrestrial systems by extending the services to areas outside the terrestrial coverage or where terrestrial infrastructure has collapsed and provide supplementary services unique to satellite systems, such as wide area multicasting. Within the network, there may be a need to create sub-networks – for example *closed user networks*, for management of a fleet; virtual private network (VPN) for corporate network extension. MSS networks are developed around these requirements, and because of differences in services and considerations discussed earlier, there are variations in network architecture.

Conventional network topology consists of mesh and star configurations. *Mesh network* provides full interconnectivity between users whereas in a *star network* a large station communicates with a number of users who are interconnected to each other through the central node. Earlier MSS systems tended to have a star topology because mobile-to-mobile links could not be sustained in power limited satellites but recent systems provide a direct mobile to mobile connectivity. MSS may consist of *close user group network* where the services are confined to operate within a group, or the system may be a part of the *public network*. Within a network, services may be provided as *circuit switched connections* such as used for voice, or *packet switched connections* for data networking. In a circuit switched network, an end-to-end call connection is set up before communication begins and the channel is released at the end of the call. A packet mode network is connectionless and hence a user may begin transmission without a permanent connection to the destination, the network delivers the information on a packet basis to the recipient through a virtual path created on basis of its routing strategy.

Traditionally, voice communication has been dominant in public networks and therefore all MSS networks include a circuit-switched facility; due to recent rapid growth in data traffic and necessity to carry data economically, the use of packet switched networks is increasing. Network configuration depends on the extent of integration between the satellite mobile and terrestrial cellular networks. There was little integration at the network level in the first generation MSS. Dual-mode mobile terminals would operate with satellite and terrestrial networks separately, and were therefore integrated only in packaging, sharing the common hardware. The second and third generation satellite systems integrate quite tightly with the terrestrial networks, allowing a significant commonality at the handset sharing upper layer level protocols and allowing seamless roaming between networks.

A satellite network comprises radio links for user mobility, a user MM system, a mobile switching centre (MSC), and an interface to the public network. Within the MSS itself there are two parts – a component for communication between the fixed parts of MSS such as between gateways, and the mobile component. The signalling associated with the mobile component establishes a stable connection – real or virtual, manages user mobility as well as radio resource and a reliable link, whereas, signalling between the fixed parts of the MSS network establishes communication between the fixed elements for functions such as MM and billing. The signalling for accessing the terrestrial network depends on the nature of the terrestrial service. Examples of terrestrial services, standards and protocols interfaced to MSSs include telephony and facsimile, Integrated Services Digital Network (ISDN), Internet protocol (IP) and ATM. The satellite layer signalling (pertaining to the mobile service itself) is wrapped around those of the fixed network.

Terrestrial systems such as GSM have chosen a well-established signalling method, signalling system number 7 (SS7), as the basis for specifying signalling between fixed elements of their mobile network, for example between the MSC and visitor location registers (VLR). SS7 is a International Telegraph and Telephone Consultative Committee (CCITT) specified signalling standard originally meant for signalling in the fixed terrestrial network between telephone exchanges and later modified to account for user mobility by adding features such as the mobile application part (MAP).

Note 1

The Open System Interconnect (OSI) model is a useful reference for understanding the functionality and architecture of a network. The model was proposed by the Internal Standards Organization (ISO) as a conceptual architecture for developing a telecommunications network. Figure 7.10 illustrates the model.

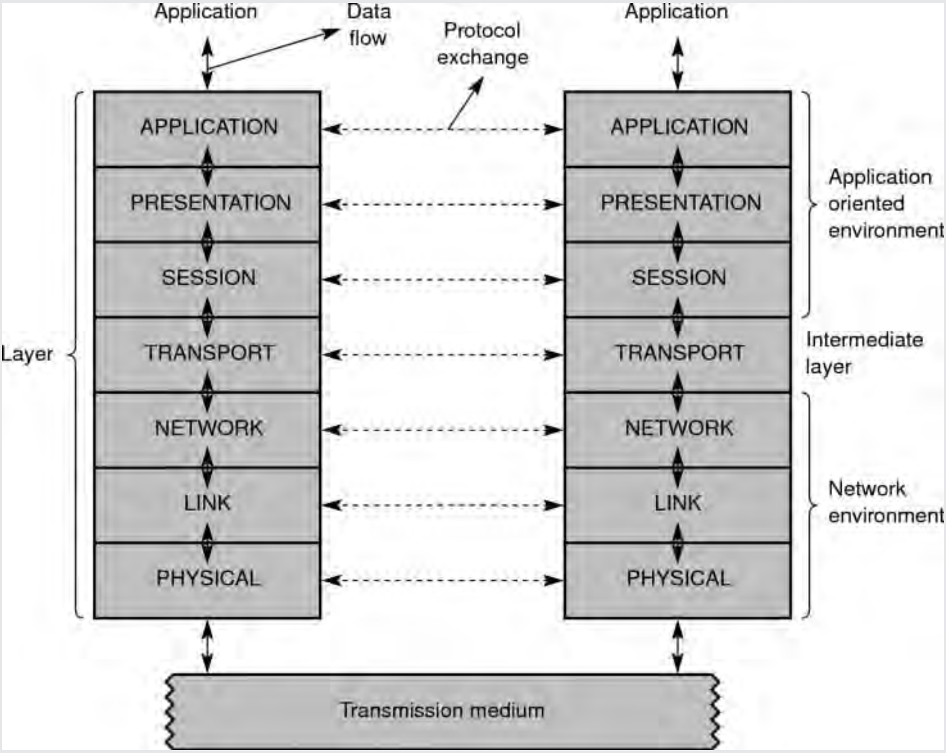


Figure 7.10 OSI seven layer model

OSI 7 Layer Model

The model divides a network node into seven logical layers. Each layer performs a distinct function and communicates vertically. A layer offers service to its next higher layer, which can request specific parameters during the negotiation phase. During transmission, a message from a higher layer is encapsulated by the lower layer, with its descriptive header containing information for its peer layer on another node, as shown in Figure 7.10. The process is continued until the message reaches the physical layer, which transmits the signal through a communication medium, such as a satellite link, to the next node. When in communication with another node, each layer communicates with its peer using a peer-to-peer protocol, as illustrated by the dashed horizontal lines in Figure 7.10. Table 7.3 summarizes the main features of the OSI hierarchy.

Other protocol standards used in distributed systems and computer communications are TCP (Transmission Control Protocol)/IP, now widely used for the Internet; and the X and I series of standards for the telephone network laid down by the ITU. TCP/IP was developed before the OSI model and differs from it in a number of ways. It does not define the data link and physical layer and only partly the network layer; the upper three layers of OSI are compressed into a single application layer. The transport layer of the IP supports two alternative protocols – TCP and User Datagram Protocol (UDP) – and the layer below it, known as the Internet layer, is implemented by the IP.

Table 7.3 Layers and functions of OSI-7 network model

Layer Number	Layer name	Function
7	Application	Provides application protocol, defines interaction between user and communication system; example of services provided: Email, facsimile, file transfer.
6	Presentation	Manages data description and data structure syntax, thereby converting abstract data syntax for applications to transport data syntax for transporting over the network and vice versa; e.g. conversion between different data sets, encryption, data compression.
5	Session	Connects application process that, for example includes synchronization to enable restart in case of failure, remote log-in, etc.
4	Transport	Interfaces physical and logical parts of the network; provides transparent transfer of information between two entities in five service classes depending on application and network quality.
3	Network	Network supervision and flow control; e.g. Monitoring of users and their addresses for maintaining routing table; ordering of received packets in the correct sequence.
2	Link	Detection of transmission errors in frames sent by physical layer; guarantees required error threshold.
1	Physical	Physical transmission; characteristics include transmission channel, type of channel usage such as duplex.

TCP/IP Performance over Satellite

TCP/IP protocol is a widely used transport protocol for reliable data transmission. Typically the IP network is connected to the satellite network via an interworking unit that adapts the IP packets for satellite link transmissions. The TCP/IP protocol suite is optimized for terrestrial transmissions with the premise that congestion is the main source of packet drops in the transmission.

TCP has a mechanism known as slow-start that gradually increases the data rate of the sender by sending one segment and waiting for an acknowledgement; for each acknowledgement, the sender transfers two segments thus increasing the data rate exponentially until the receiver's advertised window is reached or when loss is detected. The congestion avoidance feature probes the network by sending an additional segment for each round trip time and if it does not receive an acknowledgement it drops the sending rate thus reducing the throughput. Throughput loss is attributed to receiver window size limitation, and TCP's slow start and congestion control mechanisms. Moreover, in presence of packet loss TCP's data recovery, the positive acknowledgement mechanism, works poorly over long delay channels.

The problem of TCP performance impairments limitation over satellite channels has been investigated by various researchers and the Internet Engineering Task Force's (IETF's) transmission control protocol over Satellite (TCPSAT) working group. Various types of enhancement have been introduced and demonstrated over satellites' techniques (e.g. Allman *et al*., 1997). Such solutions have continued to evolve and adapted to various types of satellite links over the years.

TCP enhancements technique range from basic improvements to more elaborate data recovery methods (IETF RFC 1072 provides a survey) (RFC, request for comment). Scaled windows and time stamps (RFC 1323), fast retransmit and recovery (RFC 2001) and selective acknowledgement (RFC) are useful enhancements for satellite channels. Other solutions include basic parameter tuning/extensions, improved timer mechanisms, and advanced data recovery procedures such as SACK (Selective ACKnowledgement), FACK (Forward ACKnowledgement) and NewReno. Link-layer interworking techniques or advanced ACK (ACKnowledgement) control algorithms can also bring benefits (Ghani and Dixit, 1999).

A network architecture based on the OSI model can define and form the basis to standardize a mobile system. Standardization offers manufacturers, network operators and service providers the option to develop and market products independently resulting in economies of scale, a wider choice and high quality products due to competition. Consider the example of the terrestrial GSM or code division multiple access (CDMA) system. Its network standards/interfaces are used by operators throughout the world, allowing subscribers to use their phone in partner networks. Due to wide acceptability of the terrestrial standard, many MSS operators adopted the terrestrial network model to facilitate integration with the terrestrial systems. Many terrestrial standards have now been enhanced to include satellite components. Chapter 8 discusses these standards. Figure 7.11 shows a generic network model of an MSS system in an OSI hierarchy.

7.4.2 Functional Entities

A satellite network enables subscribers to make or receive intra-system and inter-system calls. Functions that facilitate communication include call handling, switching, routing, MM and user profile management. Numerous functions are necessary for successful management of the network. The NM functions deals with the management of the network to ensure proper functioning of the network on a daily basis. This involves management of radio resource, QoS, network traffic flow, collection of call data records and dispatch to the *business management* (BM) system, MM for call set-up, traffic trend analysis to assist radio resource management, fault finding/diagnosis and fraud detection and privacy security management. The NM also oversees the management of the space segment. Notice that many of these functions have commonality with terrestrial mobile system and thus benefit by adaptation of a terrestrial system. Chapter 8 demonstrates how these commonalities have been utilized in MSS standards. Chapter 9 addresses the operational aspects of NM.

Figure 7.12 depicts the primary functional entities of a commercial MSS, segmented broadly by their physical association.

Mobile services represent various services on offer to the mobile users including support of associated protocols. Current portfolio of services include, voice, data, facsimile, paging, message delivery, emergency calls; supplementary services such as call transfer, call

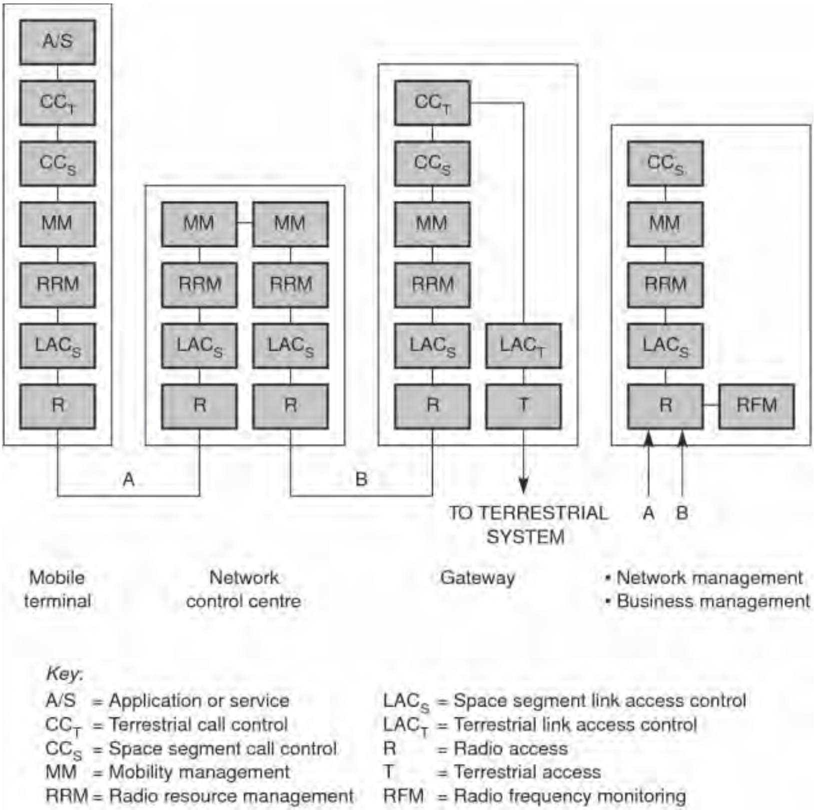


Figure 7.11 Network model of a generic MSS system in an OSI hierarchy

forwarding, call waiting, call hold, conference calls, etc. The *space segment* comprising one or more satellites provides the desired connectivity between the fixed and the mobile segments. Some of the functions shown belonging to the terrestrial part can be either wholly or partially performed by the space segment. For example when regenerative transponders are used some of the NM functions such as call routing can be transferred to the space segment. *Constellation management* involves standard Telemetry and Telecommand (TT&C) functions of satellite health monitoring, ephemeris generation, spacecraft orbit raising, orbital adjustment in case of a spacecraft failure, launch support during initial deployment or replacement of failed spacecraft, etc. The *BM* system constitutes a company's business centre responsible for customer billing for space segment usage, interfacing with the gateways or the NM system to obtain call records, updating user profile of existing subscribers and to introduce new subscribers. The *NM* function deals with real time radio resource management, monitoring radio spectrum, signal quality and network traffic flow, collection and dispatch of call data records to the BM system, interaction with the MM system for call set-up, etc. network traffic trend analysis to assist radio resource management, fault finding/diagnosis and fraud detection. NM functions provide call data records, receive user profiles and other user related information to the BM system. The *MM* system

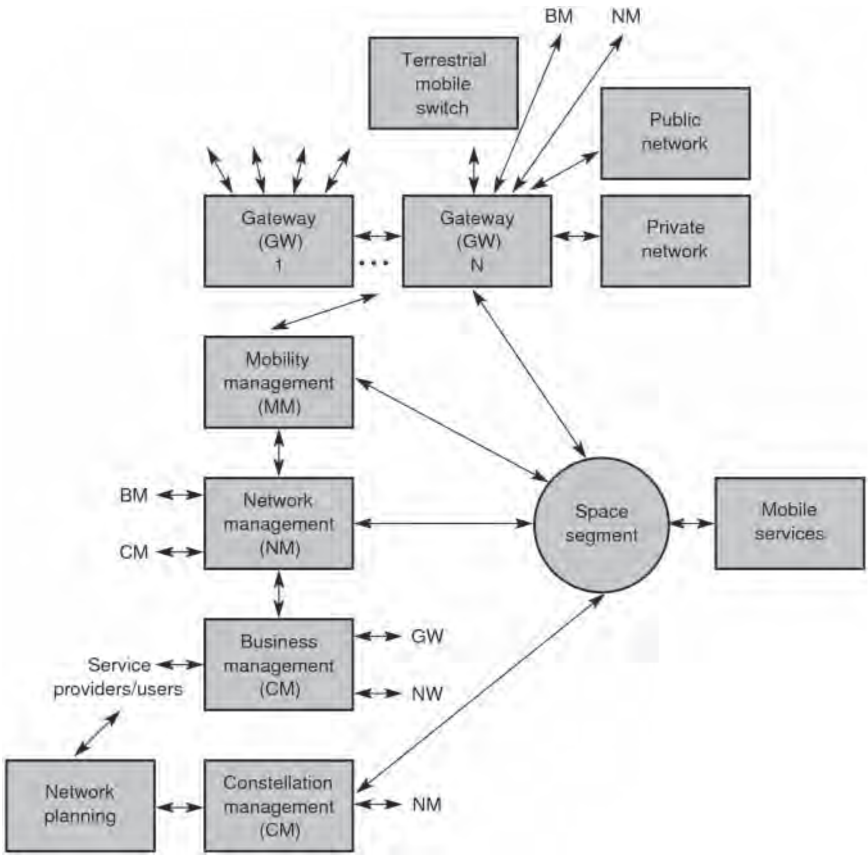


Figure 7.12 Main functional entities of a commercial MSS, segmented broadly by their physical association

maintains user locations in its database and interacts with the NM and gateways for call connection and user authentication, user profile, etc. In a GSM based network the MM involves the VLR and Home Location Register (HLR) with associated protocols for data exchange. In a packet switched network, a route is established for each packet, for example using a pre-calculated routing table. In an IP network, MM involves mobile Internet Protocol (MIP) functionality.

All *gateways* of a network may either belong to the network provider such as in ICO system or owned by individual operators as in Globalstar system. The *network planning centre* oversees NM, capacity and business trend, specific events of interest and develops strategies for changes to the network such as expansion of capacity, introduction of new services, redeployment of satellites in the network, and so on.

MSS systems differ vastly in services, space and ground segment architecture, business arrangement. Figure 7.13 portrays one possible network architecture to illustrate the interaction between the network entities mentioned above. Differences may lie in the techniques

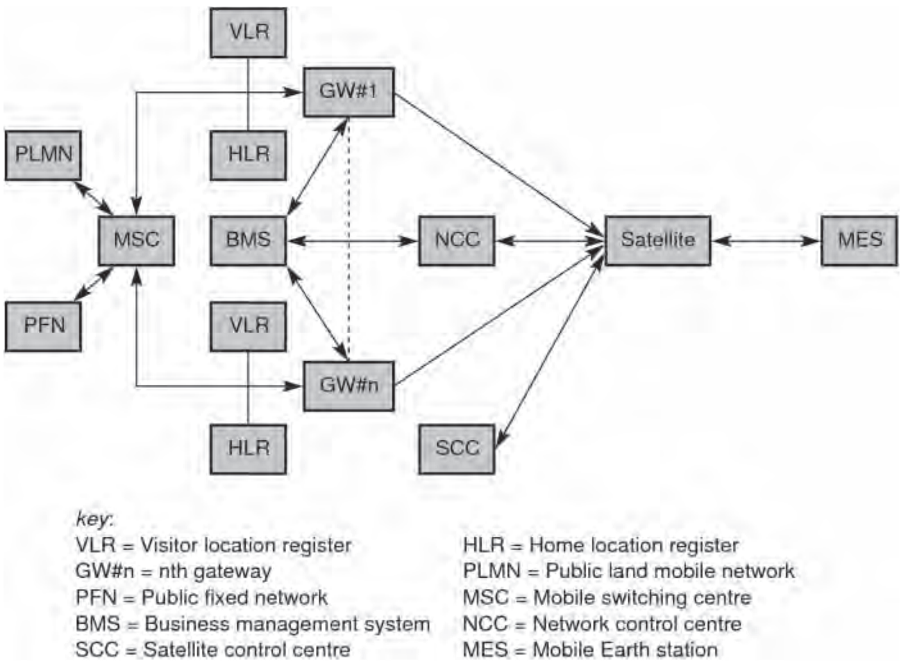


Figure 7.13 A typical MSS network architecture

used for mobility and resource management. For example, radio resource management may be distributed, contrary to the centralized scheme illustrated in the figure.

The application layer of the mobile terminal provides the desired service to the user; this may be through a public or a private network – consider PSTN as an example. Prior to setting up a user alert, a radio frequency path and routing is established by the space segment's call control protocols. When a radio connection has been established, the PSTN invokes the standard call control protocols. This involves sending a ringing tone to the user phone and at end-of-call to tear down the radio path.

7.4.3 Network Connectivity

An MSS system connects with other networks through one or more gateways; the number of gateways in a system depends on the space segment architecture, service area and terrestrial routing arrangements; and their location depends on proximity to the terrestrial network, economics, logistics, political considerations, operator's operational and business plan, etc.

Space segment connectivity is dictated by the architecture of the space segment and the service area, i.e., regional or world-wide. For interactive low delay services, each user requires continuous visibility of satellites, whereas delay-tolerant services can operate with intermittent connections. In a NGSO, connectivity through the space segment changes with time and therefore dynamic routing is necessary to accommodate satellite movement; geostationary satellites have significant advantage in this respect. The length of terrestrial routing is another consideration during the development of network topology. The terrestrial tail adds to the propagation delay and call cost due to involvement of a terrestrial carrier.

Figure 1.3(a) and (b) represents architecture for carrying interactive traffic through terrestrial and intersatellite routing, respectively, while Figure 1.4(a) and (b) portrays two types of store and forward systems – satellite and earth station based systems, respectively.

In a satellite store and forward system, a user transmits messages whenever a satellite appears within visibility. The message is stored in the satellite's buffer and transmitted when the satellite arrives within visibility of the destination, which may be a gateway interfaced to a terrestrial network or a user terminal. System capacity in this case is CN bytes, where C is the storage capacity per satellite and N is the number of satellites in the constellation. Throughput is therefore a trade-off between the storage capacity of a satellite, bound by space-qualified storage technology, and constellation size, which also governs the message transfer delay and the space segment cost.

System throughput can be increased if store and forward functionality is transferred to a ground earth station, where management of message storage and routing is easier and cost-effective. In this type of system, messages are transferred by satellites as soon as an earth station becomes visible; the earth station routes the message to its destination by the most efficient method, either through the space segment or terrestrially.

Interactive services due to their low delay tolerance require near-continuous contact between end users, which implies that the earth station/gateway and the mobile terminal remain in continuous contact. The key difference between architectures of this category lies in the network routing technique.

In the simplest approach, the service area is confined to an area covered by the footprint of satellites visible from a single gateway. Each gateway is connected only to the local/regional fixed network; as the users are confined within a specific area, this arrangement is suited for a regional system. When all gateways are interlinked and a more elaborate MM system introduced, the service can be extended to a wider area. A call from a user in the fixed network can be routed terrestrially (or through a satellite hop) to the gateway covering the mobile user; a route is derived from location information obtained through the MM system, for example HLR/VLR. Similarly, a mobile user can connect to a fixed party through the gateway to which it is connected. A suitable point for connection to the fixed party can be derived by the MM system through the called party's number. Dedicated GGLs can offer lower delay and cost; routing through public networks, on the other hand, removes the burden of installing and maintaining a link but can increase delay and call costs disproportionately.

Figure 7.14 shows connectivity when the service area is confined to a single country. One or more gateways can be deployed, depending on the size of the country. The figure illustrates four gateways covering four areas of a country.

Figure 7.15 represents an example of interconnectivity applicable to a regional MSS deploying a single gateway that serves several countries. It is necessary to interface the gateway to an international switching centre. In this scenario, the TL can be quite lengthy for some participating countries. The call costs increase progressively, with each section of the TL adding costs as the participating entity draws its share of revenue; furthermore, there are security concerns when calls are routed through a hostile country.

We note that the end user charges consist of a number of components (see Chapter 10 for details):

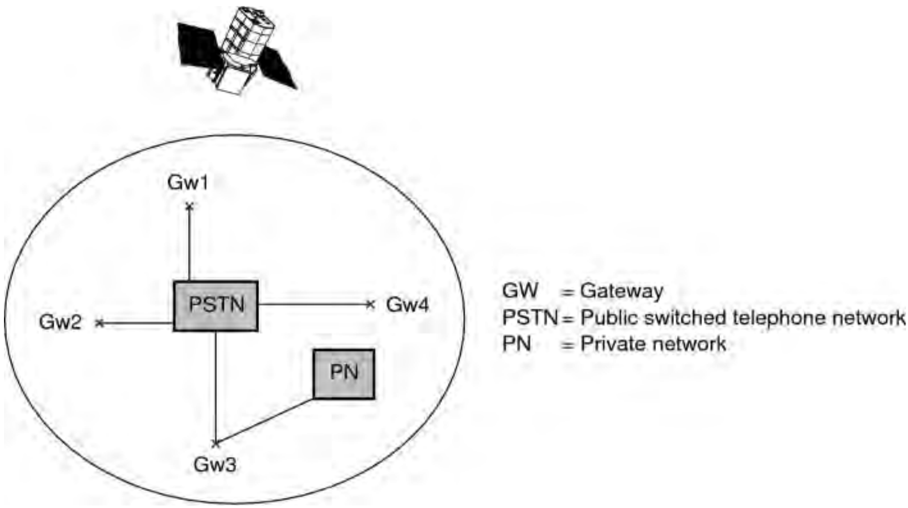


Figure 7.14 Routing in a single country MSS network

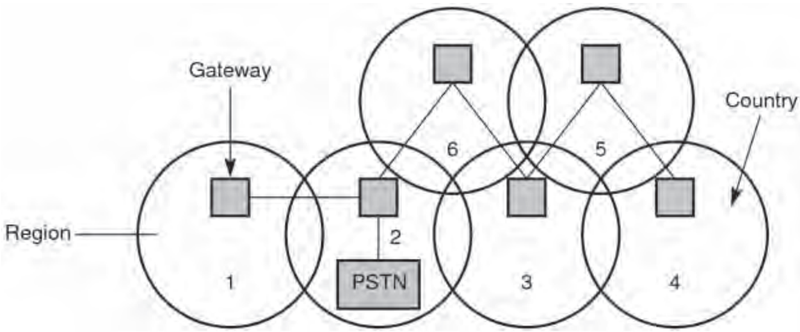


Figure 7.15 Inter-connectivity in a regional/global system

- space segment charge for utilizing spacecraft power and spectrum;
- gateway charge, which may have a non-resident component;
- PSTN/PDN charge and, when applicable, cellular operator's charge;
- service provider's charge;
- lease line charge, if used.

The extent of terrestrial routing can be reduced by deploying ISL as shown in Figure 1.3(b). In this architecture, a major part of the route can be supported on ISLs and therefore the system can operate with a relatively low number of gateways and less dependence on terrestrial routing.

In a centralized routing strategy, a database comprising user and spacecraft locations is maintained centrally, which is interrogated on a call-by-call basis by entities setting up the call for establishing a suitable route. The scheme is susceptible to single point failure and incorporates considerable signalling overheads. In a distributed routing strategy, routing tables are distributed to gateways or satellites, thus the scheme is more resistant to a failure and each node can establish a route on its own. In both the schemes, mobiles transmit their location on a regular basis. In the flooding scheme, packets are sent to all the visible nodes; the message is accepted if a node recognizes the destination address, otherwise it retransmits the message; the process continues until the packet is received at the destination. The scheme is simple and robust as there is no need to maintain a database, but it is wasteful of the network resources. Other schemes depend on the current knowledge of the network and can therefore estimate an efficient route on a per call basis; but this requires processing capabilities at each node.

7.4.3.1 Intersatellite Connectivity

For connectivity on ISLs the routing strategy must consider dynamic nature of ISL. Since constellation geometry is predictable it simplifies route estimates and offers alternatives. The traffic non-uniformity allows a choice of under-utilized route particularly when QoS is an issue. The shortest least-congested path takes into consideration also the propagation delay. Spacecraft technology and its static nature on the other hand imposes power and processing limits and therefore space-borne strategies are less agile compared to terrestrial systems. Routing techniques involve a trade-off between complexity, storage and number of handovers and include measures to:

- Reduce link handovers to minimize call drop probability;
- Minimize ISL link for minimum (propagation + queuing) delay;
- Adapt route based on ISL loading;
- Tailor routes based on QoS requirement;
- Attempt to utilize terrestrial techniques to facilitate integration.

Various algorithms based on the above approaches have been proposed; a sample of such techniques are (Taleb *et al.*, 2005):

- **Finite-state automation:** Topology is divided into fixed number of states within which ISLs remain unchanged; Satellites apply stored tables of the topology;
- **Virtual Node** (VN) **concept:** A fixed topology is superimposed on constellation dynamics so that mobility of satellites remains hidden and satellites utilize routing protocol of the fixed topology;
- **Link handover reduction methods:** (a) Flood and estimate lifetime and select the longest lifetime or (b) Select the longest lifetime path amongst all alternatives that meet QoS;
- **Path-hop minimization schemes:** (a) Path minimization reduces the propagation delay or (b) Hop minimization reduces hops on premises that each hop adds a processing delay;
- **Load balancing algorithms:** Based on distributing traffic judiciously while minimizing path length;
- **Traffic based algorithm:** This scheme differentiates between traffic types in terms of QoS requirements and sends them along appropriate routes.

7.4.4 Gateway Locations

As the number of gateways is increased a network becomes more resistant to single node failures due to path redundancy, and additionally, it is possible for traffic to flow more uniformly, however, an increase in number adversely affects cost and complexity. In practice, the number and location of gateways is dependent on a number of practical considerations such as follows.

1. Position and number of gateways depends on whether the service area covers a country, a few countries (region) or the world.
2. Location of gateway depends on its vicinity to terrestrial traffic as well as logistics. The closer a gateway is to traffic centres, lower the tail-end cost.
3. Political factors have considerable influence in deciding the location of gateways – for example a local operator may be preferred.
4. Each gateway requires signalling capacity and hence consumes (unproductive) radio resource and hence affects the spectrum utilization efficiency.

Consider the proposed routing arrangements of a few NGSO systems – Iridium, Globalstar and ICO. The Iridium system (in its original form) comprised 12 regional gateways, distributed across the world, which were to be shared by 15 operating companies who would own, operate and manage the gateways. The Iridium system has considerable flexibility in choosing the number and site of gateways, as it deploys ISLs. Theoretically, a single gateway can support this type of network topology, but this would lead to an increase in size of the terrestrial component, thus increasing terrestrial line costs.

The Globalstar system has been designed to support single country as well as a region. Several inter-connected gateways are used to extend the coverage area. The decentralized gateway architecture, gives the operators independence in operating their part of the network, which also allows service areas to expand as necessary. Initially, 50–60 gateways were planned, with gateways to be introduced incrementally as the system evolved.

The proposed ICO system comprised 12 self-owned regional gateways, called satellite access nodes (SAN), interconnected to each other and interfaced with PSTN/PDN (Public Data Network), and two NM centres. The SAN and the NM centres were to be connected through a dedicated high capacity optical-fibre link to minimize dependence on public networks.

7.4.5 Call Handling

Call establishment techniques, standardized in the GSM terrestrial system, were emulated in many mobile satellite system architectures including ICO, Iridium and ELLIPSO™, thereby simplifying integration with GSM terrestrial systems and maximizing benefits of the mature terrestrial technology. A satellite standard called Geostationary Earth Orbit Mobile Radio (GMR) interface has been based on this architecture (see Chapter 8).

In the GSM system, each mobile registered to the network is assigned a home area. User service profile, location, billing and other data are maintained in the HLR in the home area of each mobile. When a mobile migrates outside its home MSC, it registers itself with the visited MSC. The mobile identification is entered in the VLR of the visited MSC and the information is communicated to the HLR, which updates the mobile's location details. Whenever the mobile is called, the gateway where the call is placed interrogates the mobile's HLR for location, authentication details and routing to establish the call.

Typical call handling sequences for terrestrial and mobile-originated calls at mobile and network level are shown in Figure 7.16(a–d). Note that there may be variations in the details, but the process should essentially be similar. A GSM-type network architecture is assumed, as many satellite systems have modelled their architecture around it (Tisal, 1997).

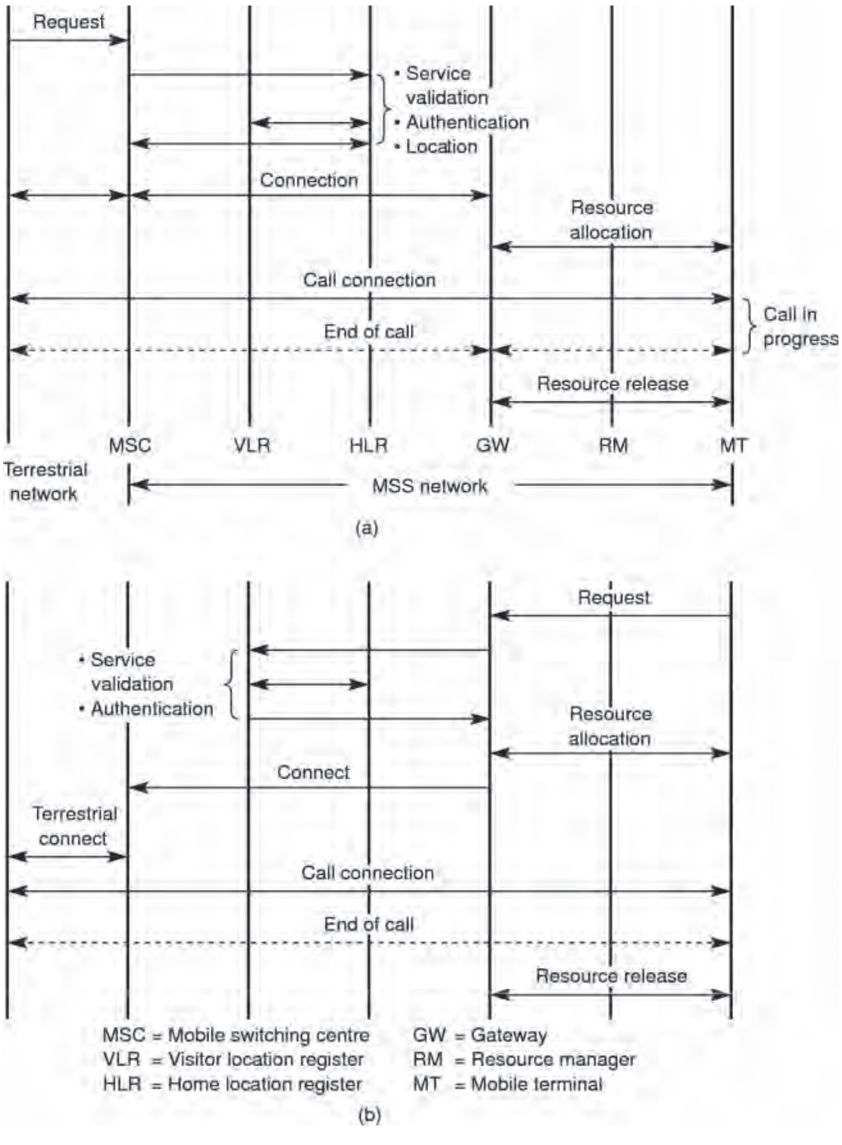


Figure 7.16 (*continued*)

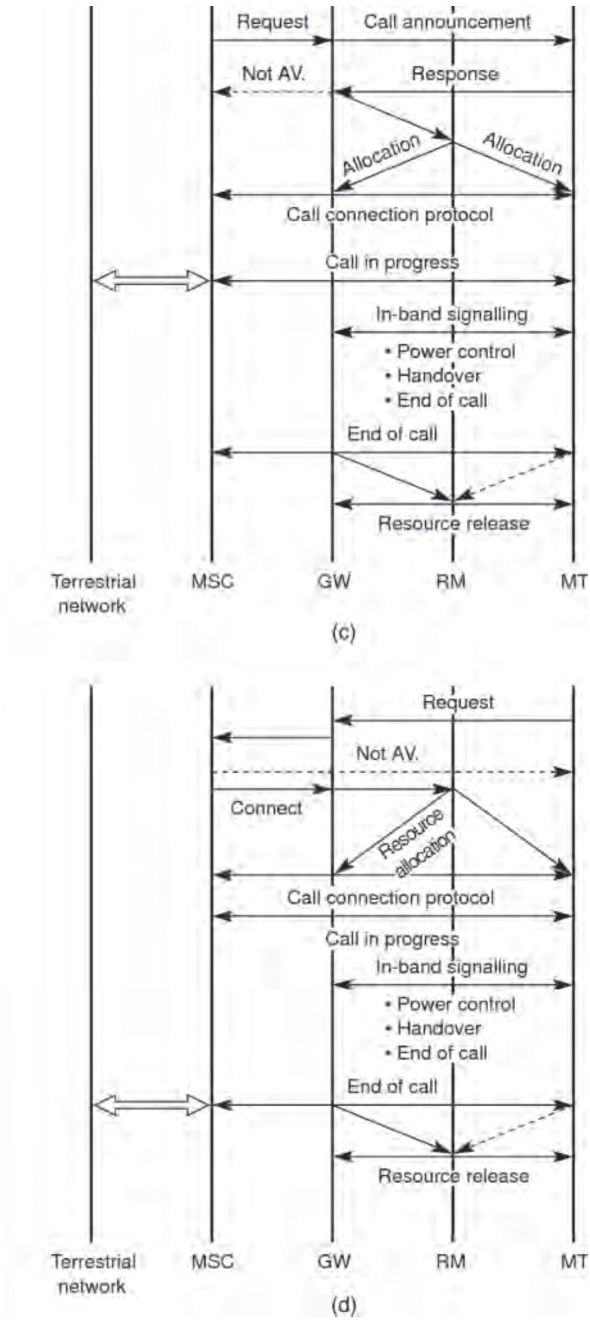


Figure 7.16 Signalling sequence for (a) terrestrial originated call; (b) mobile originated call; (c) network functions for terrestrial originated call and (d) network functions for mobile originated call

A call originated by a terrestrial party is received by the MSC; the MSC interrogates the HLR for service validation, authentication, location of the mobile and mobile status – busy or not (see Figure 7.16(a) and (c)). Assuming that the mobile is authorized and available, the interrogation provides the location of the mobile, and if the mobile has roamed to another area, the address of the VLR, which may be the address of a gateway. The radio resource manager sends a call announcement on the broadcast channel at the appropriate location of the network. If there is no response by the mobile within a specific time, the terrestrial party gets an 'unobtainable' message by return signalling message. When the mobile responds to the call announcement, the radio resource manager assigns resources to the gateway and the mobile. Following allocation, the mobile receives a notification such as a ring; when the device is off-hook, a handshake protocol ensures radio connectivity. The call then proceeds, after exchange of application-specific protocols. Any further exchange of signalling for functions such as power control, handover, end of call, etc. are conducted 'in-band', that is within the assigned channel. When the device is kept on-hook at the end of a call by either party, depending on which party has terminated the call, the gateway or the mobile informs the resource manager, and the resource manager then returns the radio resources to the common pool.

When a call originates at a mobile, a request is sent on a shared pre-assigned channel; the mobile's identity and service request are authenticated by the network by interrogation of the mobile's HLR, which maintains its service profile (see Figure 7.16(b) and (d)). After authentication, the radio resource manager proceeds to allocate the appropriate resources to a gateway which may, depending on the network arrangement, be identified by the mobile or the network on the basis of some criterion, for example closest gateway to the called party. Establishment of radio connectivity between the gateway and the mobile is confirmed by a simple technique such as a 'loop back'. Simultaneously, the gateway sends a call request to the public network, which initiates a ring at the called party's number using the network's standard procedures. When the called party has answered, the applicable end–end protocols are exchanged, followed by communications.

7.4.6 Mobility Management

The primary goal of MSS systems is unrestricted mobility in the service region. This network feature demands continuity of service both in the short term during calls, and in the long term involving large geographical mobility, which may necessitate migration across networks on a single number or address. Encapsulated within these lie the concepts of wireless access, terminal mobility, personal mobility and management of user service profiles.

A variety of wireless systems exist, each matched to specific environments; therefore, a universal system cannot be based on satellite system(s) alone. This leads us to the concept of a universal personal communication network, a topic that has received considerable attention from the international community. The concept of a unified system is introduced in Chapter 14.

MM of mobile systems requires a number of unique features, summarized as follows:

- **Unrestricted connectivity**, whereby calls are established with a mobile anywhere within the service area. The functionality is generally achieved through location registration, paging and call establishment. Location registration involves logging-on and regular

location reporting by a terminal to facilitate call establishment. Paging involves transmitting a message to verify the current location of a mobile prior to the establishment of a call. In general, a call connection comprises terrestrial and satellite components. We will explore in a following section the wide variations in routing strategies.
- **Roaming** is a network function that allows users to migrate to other networks or to a MSC other than its home MSC.
- **Handover** involves handing live calls from one beam (or cell) to another or from one satellite to another. Handover may be necessary at a mobile, a gateway or a satellite.

The architecture of a MM system for MSS depends on various factors:

- **Number of spot beams**: The number of handovers increases as the number of spot beams increases that may demand a more complex management scheme.
- **Satellite motion**: MM for LEO/MEO systems, where satellites themselves are moving, requires a more agile MM.
- **NGSO satellites with ISLs**: In these systems, handovers are also necessary on the ISLs due to dynamic variation in routing.
- **Multiple operators**: When gateways belong to different operators, the MM system must provide suitable provision for roaming between them.
- **Extent of integration with terrestrial system**: Depending on the level of integration, architectural features to manage inter-network mobility have to be introduced.

7.4.6.1 Handover

Handover is the process of changing beams, satellite or gateway during a call. It is necessary for a number of reasons:

- the communicating satellite may have moved below the specified elevation;
- a better satellite path may be available when path diversity is used;
- the user may have moved in to a new spot beam;
- the radio resource manager may need to divert capacity to a less congested beam;
- satellites may have moved out of visibility of feeder stations;
- when ISLs are used, a satellite–satellite handover can occur due to a change in routing because of constellation dynamics.

Due to a static link geometry, handover management is simpler for a GSO satellite system than for a NGSO satellite system. When considering an IP network a MIP protocol is necessary to ensure continuity with the point of attachment. To minimize or eliminate handover between spot beams, non-geostationary satellites may deploy a quasi-stationary spot beam system, wherein the spot beams of each satellite remain fixed to predefined points on the surface of the Earth until the satellite moves below visibility, at which time a new satellite replaces the area vacated by the old beam. This type of Earth-fixed arrangement simplifies frequency planning but requires rapid, accurate and synchronized steering of spot beams on satellites. Figure 7.17 shows various types of handover.

Beam–beam or satellite–satellite handover need not be applied to geostationary satellite systems deploying only a one beam or a few spot beams, as coverage boundaries are

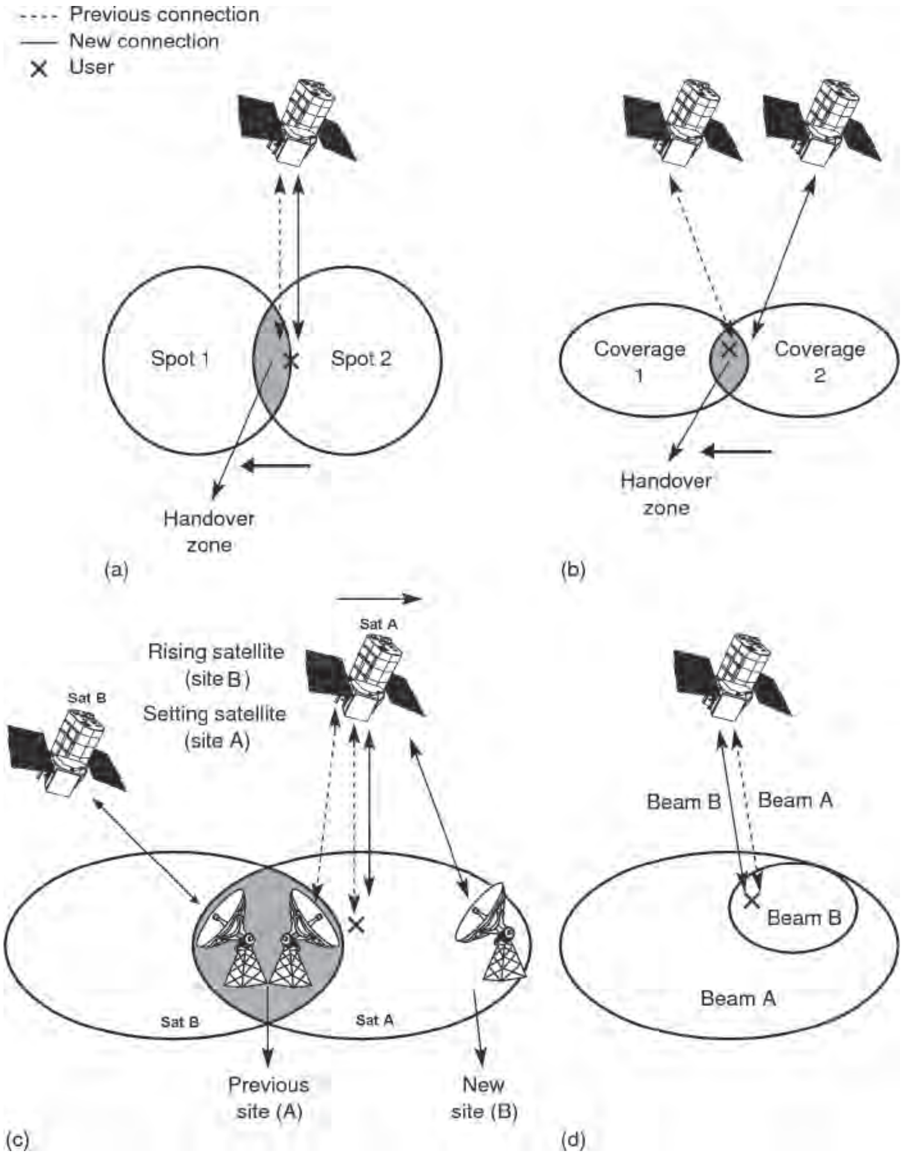


Figure 7.17 Various types of handover: (a) spot–spot, (b) satellite–satellite, (c) feeder link–feeder link and (d) congested beam-non-congested beam

fuzzy over long distances and it may take hours for slow-moving mobiles – ships, land vehicles or pedestrians – to move across spot beams. Inmarsat third generation systems operating with four to five spot beams do not use handovers whereas in the BGAN system that operates over Inmarsat's fourth generation satellites deploying some 200 spot beams, handovers are used in only in the aeronautical system but not for land and maritime systems.

Handovers can be mobile initiated with network assistance or may be network-initiated. The handover protocols ensure that the process is seamless by checking availability of resource and transferring call details such as routing to maintain connectivity between the source and destination before a call is handed over. Specific examples of handover mechanisms are available in Chapter 8.

Next generation satellite systems are evolving towards IP solutions to leverage established technology and easier integration with terrestrial. A handover in an IP network requires a different approach than used in a circuit mode handover. MM involves location management for reachability and hand-offs with data delivery for continuity. It involves sending a binding update that associates each node's unique identifier, Reachability IDentity (Reach.ID); a Routing identity (Route.ID) that is a location dependent variable sent by each node regularly; and data delivery that deals with transport of data. IETF has designed MIP and MIP version 6 (MIPv6) for MM of hosts in terrestrial networks and network mobility (NEMO). Satellite handover protocols are extensions of terrestrial mobility protocols such as MIP, MIPv6 and Network mobility Basic Support Protocol (NEMO BSP) for satellite extensions (Zafar *et al.*, 2008).

Consider as an example an application of MIP over a non-geostationary satellite acting as a mobile host (MH) (see Figure 7.18a); the home agent (HA) is co-located with Earth station 1; the foreign agent (FA) is co-located with Earth station B. The communicating node, denoted as CN, exchanges data with the satellite through the HA. When the satellite comes in contact with Earth station B, it requests a care of address (CoA) from B, and passes the assigned CoA to the HA. The HA establishes a tunnel to the FA and any subsequent communication with the mobile node is established through the tunnel. On the return path the FA sends the data through standard Internet routing protocols (i.e. not through the tunnel). In order to improve performance of MIP over satellite, improved techniques such as seamless IP diversity-based general mobility architecture (SIGMA) have been proposed. MIP suffers from performance degradation such as high handover latency, packet drop-offs and low throughput. There is no concept of HA and FA (see Figure 7.18b). The mobile remote host obtains a new IP address (IP-2) when it comes in contact with a new access router (satellite-2) while maintaining communication through the old router's (satellite-1) IP address (IP-1) and switches to the new IP address (IP-2) after assignment; the new address is also communicated by the MH to a location manager so that routing information is available to the network to begin a new communication.

7.4.6.2 Mobility Management Examples

In this section, we will highlight some of the issues and techniques discussed in the preceding sections with specific examples. Chapter 8 includes solutions proposed in various ITU recommendations covering MSS air interface.

Geostationary Satellite Systems

Satellites using the MSAT (Mobile SATellite) first generation system consist of six spot beams (see Chapter 11). Four beams cover contiguous territories – East, East/Central, West Central and Western regions of the USA and Canada, the remaining two cover the states of Alaska and Hawaii, as well as southern reaches of the USA and Caribbean. A mobile can make or receive calls from anywhere within these regions.

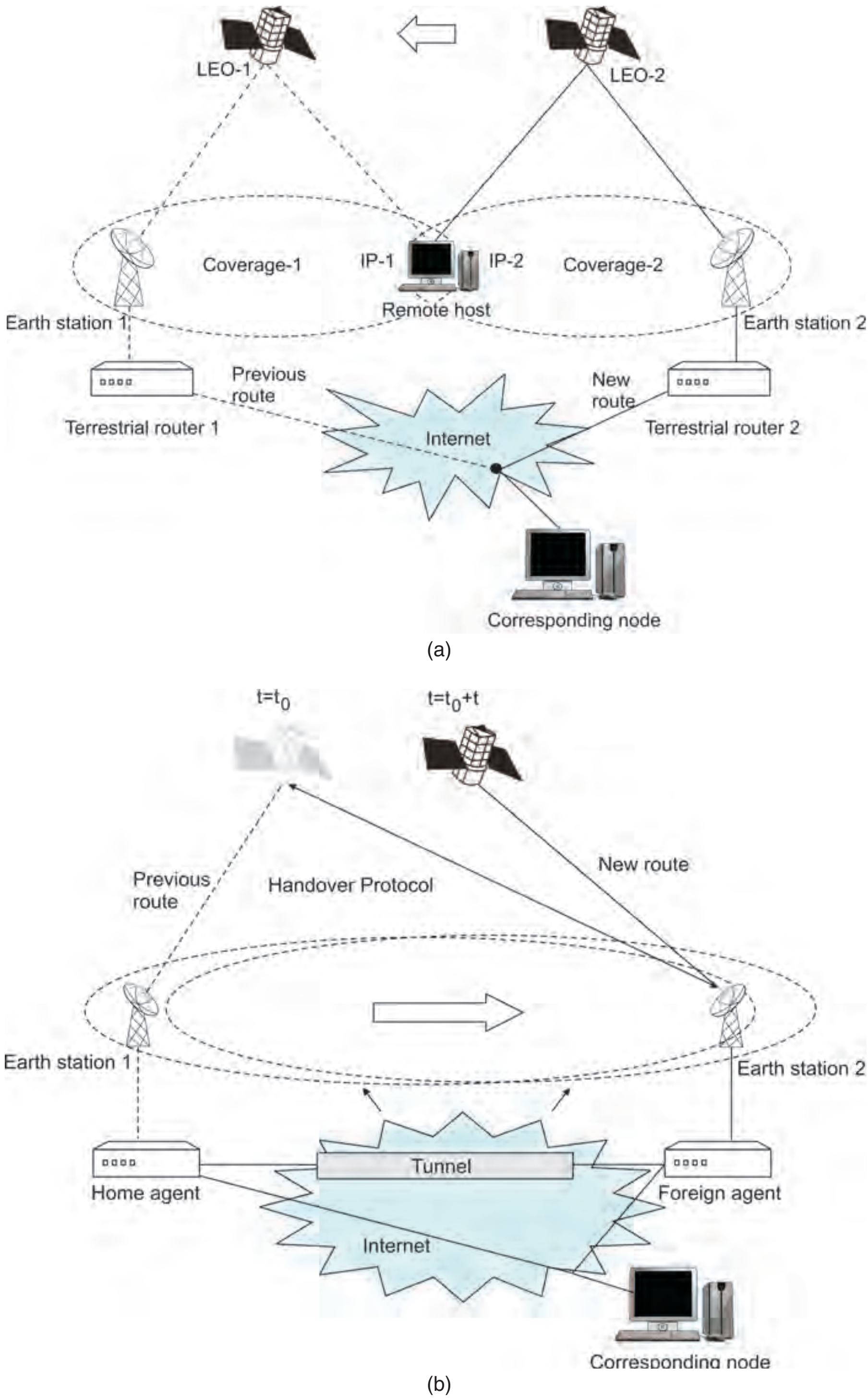


Figure 7.18 (a) Application of MIP in satellite networks where satellite is the mobile host. (b) Appli-

Each mobile is assigned a home region, a beam at the time of initial registration and is provided with details of other beams. At the same time information regarding other beams is introduced in to mobile memory. When a mobile is switched on, it searches for the home broadcast channel; locks to it; extracts relevant network information; and is then ready to make or receive calls. If the home channel is absent the mobile determines identity of the broadcast channel being received from its database; this information is transmitted to the network control station (NCS); thereafter this beam becomes the new home beam of the mobile until it moves to another beam.

A MM system was introduced in Inmarsat's legacy systems to enable use of a single number in all regions. The system requires each mobile to indicate its location. The information is received and maintained in databases located within designated land earth stations for call routing purposes.

The third generation Inmarsat system uses four to five spot beams per satellite. Therefore it becomes essential to locate a mobile for placing a call. Land and maritime systems use similar techniques whereas the aeronautical system, which has a distributed architecture (i.e. each ground earth station operates individually with minimal central control), uses a different scheme. In the land and maritime systems (Inmarsat B and M systems), a system bulletin board is transmitted in the global beam from which mobiles read frequencies of spot beam identifiers. Frequency of each broadcast channel is hard-coded in to mobiles (and hence is fixed); whereas changes to the spot beam identifying frequencies are possible through updates to the bulletin board. For fixed-to-mobile calls, the mobile is paged by the NCS on the global broadcast channel to which all idle mobiles are tuned. On receiving a call announcement, the mobile responds with its spot beam identity, which it derives from the spot beam identifiers. The NCS then assigns a channel in the mobile's spot beam. The size of a spot beam is governed by the level of spot beam identifier. When a mobile makes a call, it requests a channel in its present spot beam. The NCS then proceeds to establish the call in the selected beam. If all the channels for the spot in the channel are busy, a channel is borrowed from the global beam, provided the service can be supported in the global beam. There is no beam–beam or satellite–satellite handover during a call, as the spot beam roll-off is gradual.

The aeronautical system is based on an earth-fixed spot beam system in which spot beam contours are broadcast on a global beam broadcast channel. An aeronautical mobile establishes its current spot beam by checking the beam within which it lies. The preferred spot beam is signalled to the communicating earth station for call establishment. There is no beam–beam or satellite–satellite hand over. A timer is used to drop a call if the spot beam crosses a spot beam boundary during a call.

Let us consider some techniques applied to more recent GSO proposals. In contrast to the first generation systems that typically deployed only a few spot beams, these *super-geostationary systems* use hundreds of spot beams. We will also consider systems that operate in non-stationary orbits where MM must include the impact of constellation dynamics.

Super-Geostationary Systems

Super-geostationary satellites generate several hundred spot beams, and although in comparison to the first generation spot beam systems, the spot beam sizes are vastly narrower the dimension of these narrow spot beams extends hundreds of kilometres on the Earth's surface depending on their relative position with respect to the satellite location. Therefore, beam–beam handoffs are not necessary for slow moving mobiles as the signal level change at the beam edges is gradual and thus unlikely to cause noticeable affects to the quality of a call in progress typically lasting only a few minutes; however, handoffs become essential for high-speed mobiles in which case signal level at receivers deteriorate rapidly near spot beam boundaries.

Inmarsat's BGAN system supports handover in the aeronautical systems but not in maritime or land systems. A handover can be initiated when aircraft migrates to a new spot beam judged on basis of the knowledge of spot beam boundary with respect aircraft's own position and estimating.

An approach similar to the one discussed for Multifunctional Transport Satellite (MTSAT) has also been proposed in the literature, with the difference that spot beam broadcast frequencies are reusable rather than unique to avoid the need for a large number of frequencies and consequent spectrum loss (Johanson, 1995). This necessitates the mobile extracting spot beam/region information from the broadcast messages of the new beam when migrating outside its home beam rather than on the basis of frequency (as the broadcast frequencies may be reused). When a mobile migrates outside a beam, it sends an update of the new location to the network. In a system that serves a region partially, that is use of terminals is disallowed in certain countries, it becomes necessary for the mobile to send more precise location information, for example by using a global positioning system (GPS) receiver. It would then be possible to disallow calls when the mobile migrates to forbidden territories. When this type of system comprises a number of gateways and/or operators, the GSM concept of VLR and HLR can be used for call management, provided the operators incorporate a cooperative arrangement.

Non-Geostationary Satellite Systems

In non-geostationary systems, the MM system requires management of satellite movement in addition to that of the mobile. Satellite movement is predictable and therefore routing to a mobile can be established through the appropriate satellite with relatively straightforward routing calculations, when satellite orbital parameters are available at each gateway. In a GSM-type architecture, the mobile periodically transmits its location over a passing satellite that relays it to the HLR through the gateway connected to the satellite at the time, or via an ISL if used. The beam and/or the satellite handover mechanism is/are similar to the cell–cell and base station–base station handover experienced in terrestrial cellular systems.

Beam–beam handover can be initiated either by the fixed earth station or by a mobile. In many systems, the service links operate in different up and down link bands, for example, the L and S bands, which have different transmission characteristics. Consequently, signal monitoring at a ground station for the purpose of beam handover is unreliable due to differences in propagation behaviour in the two bands. In these cases, mobile-initiated handover is preferred. One proposed solution is to use a dual receiver at the mobile (Johanson, 1995). While a call progresses through one receiver, the second receiver monitors other spot beams. When a better quality of signal is received from another beam, the mobile requests a beam handover from the gateway, which assigns radio resources to the new beam and transfers the call in synchronization with the mobile transceiver. Satellite–satellite handover requires the mobile transceiver to compensate for differences in path loss and Doppler between the old and new radio connections, which is not necessary for beam–beam hand-over of the same satellite.

Let us consider examples of non-geostationary satellite systems with regard to the handover methods. Although the Odyssey system has now been shelved, it pioneered the Earth-fixed frequency reuse technique. The service area is divided into predefined regions, each of which are assigned a block of frequencies; satellite antennas utilize a steering antenna system that locks each of its spot beams to these predefined Earth-fixed contours until the elevation angle moves below a predefined threshold, when the Earth-fixed contours are handed over to a new satellite. This system had no provision for spot beam handover because satellites in MEO have a long visibility of about 2 h, during which spots remain fixed. The signal quality near the beam edges undergoes a graceful degradation, much like geo-stationary satellite systems, and hence inter-beam handover is not necessary – the size of each spot is in the order of a 800 km diameter for this MEO system. The probability of a high number of calls remaining active at the instant of satellite handover is low because, during the coverage overlap of about 10 min between satellites, most existing calls, which typically last 2–3 min, terminate normally; new calls during this interval are established via the new spot beam. Furthermore, the network was capable of handing over calls between satellites, thereby offering the capability of re-establishing a call on a new satellite.

The Globalstar system incorporates a soft handover technique enabled by the system's CDMA scheme. Feeder-link handovers are not used. When a gateway receives signals transmitted by a mobile from two beams of either the same or an adjacent satellite, the gateway begins transmission of a time-shifted version of the code though the new beam. The mobile's rake receiver is able to track both signals, combining them to provide diversity advantage, and finally drops the signal from the old beam when the signal is unusable; following this handover, the gateway receiver begins to prepare for the next handover. In this technique, handover is not instantaneous or 'hard', as the receivers are able to exercise an autonomous decision to initiate the handover at the most appropriate instant. By contrast, in a 'hard' handover where the network commands a mobile to switch to a new gateway, there are occurrences of dropped calls and hence loss of information.

By eliminating the need for feeder-link station handover, Globalstar system coverage is limited to the area around the earth station visibility arc – a cell of about 650 km. Therefore, a honeycomb of earth stations is necessary for global coverage.

An earlier proposal was to reduce the number of spot beam handovers by elongating the beam shape in the direction of satellite motion, though this would have been achieved through loss in antenna gain. The arrangement allows calls to be carried for longer, thereby minimizing the need for spot beam handover.

Finally, the Iridium system uses all types of handover – spot beam, FL and, by virtue of ISL with on-board processors, ISL handover.

Revision

1. With the help of a block schematic, discuss various network functions and components of an MSS system, including the interaction between the functional entities.

2. Suggest the desirable features of an MSS service link to support: (a) hand-held services (b) broadband services on nomadic user terminals.

3. A prospective MSS operator intends to provide a seamless service aiming to address the mass-market that covers a wide region encompassing large uninhabited open areas,

motorways, a large railway network and several densely populated metropolitan cities. Suggest an efficient network architecture, stating the rationale for the selection of each element of the network.

4. In comparison to a technology driven, a market-led approach in developing the requirements of an MSS reduces the financial risks to the system. Describe the process in evolving a market-led system development process including the influences and constraints that must be taken into consideration.

5. Suggest the main elements and processes of a simulation scheme meant to derive system parameters such as system capacity, propagation delay, link utilization statistics and the length of the PSTN lines, given the subscriber distribution, the traffic statistics, the space segment configuration and a gateway distribution.

6. In evaluating LEO and MEO satellite systems, orbital altitude has a profound effect on a number of system parameters; list such parameters and outline with reasoning the impact of altitude on these parameters. Suggest a favourable range of altitudes to support a handheld voice-medium bit rate communication service.

References

Allman, M., Hayes, C., Kruse, H. and Ostermann, S. (1997) TCP performance over satellite links. International Conference on Telecommunications, March, www.isi.edu/tcpsat/papers.html (accessed 23 August, 2013).

Böttcher, A., Jahn, A., Lutz, E. and Werner, M. (1994) Analysis of basic system parameters of communication networks based on low earth orbit satellites. *International Journal of Satellite Communications*, **12**, 85–93.

Dutta, A. and Rama, D.V. (1992) An optimisation model of communications satellite planning. *IEEE Transactions on Communications*, **40** (9), 1463–73.

Gavish, B. and Kalvenes, J. (1998) The impact of satellite altitude on the performance of LEOS based communication systems. *Wireless Networks*, **4**, 199–213.

Ghani, N. and Dixit, S. (1999) TCP/IP enhancements for satellite networks. *IEEE Communications Magazine*, **37** (7), 64–72.

Johanson, G.A. (1995) Mobility management in satellite networks. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 29–34.

Kota, S.L., Giambene, G. and Kim, S. (2011) Editorial special issue on Hybrid/integrated satellite–terrestrial networks. *International Journal of Satellite Communications and Networking*, **29**, 187–190.

Ryu, J., Han, S.M., Shin, M., *et al*. (2007) The gap filler technology for mobile satellite system, Antennas and Propagation Society International Symposium, 2007 IEEE, June 9–15, 2007, pp. 3033–3036.

Taleb, T., Kato, N. and Nemoto, Y. (2005) Recent trends in IP/NGEO satellite communication systems: transport, routing, and mobility management concerns. *IEEE Wireless Communications*, **12**, 63–69.

Tisal, J. (1997) *GSM Cellular Radio Telephony*, John Wiley & Sons, Inc., New York, Section 7.3.

UMTS Forum (1998) Cost Impacts, Report Number 3.

Zafar, A., Shahriar, M., Atiquzzaman, M. and Rahman, S. (2008) Mobility management protocols for next generation all-IP satellite network. *IEEE Wireless Communications*, **15** (2), 46–54.

8

Satellite Radio Interface Standards

8.1 Introduction

In this chapter we intend to introduce various mobile satellite service (MSS) radio interface standards to provide a sound understanding of how radio interface technologies are utilized in practice, their evolution and recent technical advances that have reached maturity. We have covered many of these technologies elsewhere in the book; and here we will see how these have been woven into operational systems. More specifically, we will cover the following:

- the Geostationary Earth Orbit Mobile Radio Interface (GMR) family;
- International Mobile Telecommunication-2000 (IMT-2000) satellite radio interfaces (SRIs) recommended by the ITU;
- Digital Video Broadcast/Return channel via satellite + mobility (DVB-S/RCS + M) developed in the DVB project forum.

Each standard and recommendation has been developed over several years of concentrated effort by a large pool of international experts and specified in great detail. To keep within the scope of the book we will only introduce the main features of each with cross-references to the original text for the benefit of those who require more details. It is emphasized that the tabulated or other data are included only for illustrative purposes and the interested reader should refer to the most recent source material for details or to clarify design and implementation issues. Radio interface is also referred as air interface in the literature. We will use these nomenclatures interchangeably. The satellite air interface standards tend to reuse the terrestrial 'core' network to facilitate interworking and maximize benefits through economies of scale enabled by terrestrial systems. We will only address layers associated to and affected by the SRI with the premise that the remaining layers are transparent to satellite systems.

A generic protocol architecture used in the European Telecommunication Standard Institute (ETSI) standards (and others) is presented in Figure 8.1. It comprises a *network independent* layer that deals with radio frequency transmission comprising synchronization within the satellite network, coding, modulation radio frequency conversion/filtering/amplification,

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

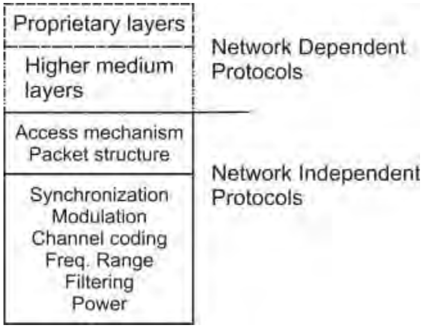


Figure 8.1 A generic protocol architecture applicable to the satellite radio interface standards. (Source: DVB-1, 2012. Reproduced with permission of ETSI.)

multiple access and encapsulation of the incoming stream for transport in the satellite network and a *network dependent* layer that deals with transport protocols related to the core network and applications. The figure lists the critical parameters of the lower physical layers addressed in many standards.

Since a majority of the standards and recommendations use the same generic architecture, there is an underlying similarity across these systems; the differences lie in the solutions applied in achieving respective goals. The general approach has been to provide as high a throughput as possible within practical limitations. A majority of the proposals target L and S band implementation but are not necessarily limited to these bands; however, the DVB/RCS + M standard mainly targets K_u/K_a bands to provide high throughputs (through directive antennas) difficult to achieve in the L/S bands due to spectrum scarcity and preference to provide service to small non-directive user terminals (UTs) with high mobility.

The GMR family (GMR-1 and GMR-2) specifications extend the reach of the well-established terrestrial Global System for Mobile communication (GSM) standard to a geostationary earth orbit (GEO) mobile satellite system by provisioning the access of the MSS to the GSM core network, incorporating alterations to the GSM specifications where necessary. Thus the GMR system specifications are a combination of satellite-specific, GSM-shared and evolved GSM specifications. These enhancements were generated in the third generation partnership project (3GPP).The 3GPP is a global partnership 'to prepare, approve and maintain globally applicable Technical Specifications (TSs) and Technical Reports that includes GSM, an evolved third Generation (3G) and beyond Mobile System and an evolved IP Multimedia System' (3GPP, 2007).

The *satellite component of the Universal Mobile Telecommunication System* (S-UMTS) constitutes a family of standards developed to complement the terrestrial Universal Mobile Telecommunication System (T-UMTS) 3G mobile services to the extent feasible in a satellite environment utilizing low earth orbit (LEO), medium earth orbit (MEO) or GEO satellite systems with a capability to inter-work with other IMT-2000 family members through the UMTS core network. ITU recommendation ITU-R M.1850 (ITU-1, 2010) defines IMT-2000 as follows:

> IMT-2000s are third generation mobile systems which provide access, by means of one or more radio links, to a wide range of telecommunications services supported by the fixed telecommunication networks (e.g. PSTN/ISDN/Internet protocol (IP)), and to other services which are specific to mobile users.

The ITU recommendation M.1850 identifies those SRIs that satisfy IMT-2000 features and design parameters that include worldwide compatibility, international roaming and access to high speed data services. These specifications continue to evolve in line with technology, user expectations, business goals, operational aspects and other influences. Parts of GMR-1 specifications are included within this family.

The *DVB* Project is a multi-organization industry-led consortium committed to develop global standards for the delivery of digital television and data services to foster market driven solutions for the needs and economic circumstances of the broadcast industry stakeholders, applicable VSAT (very small aperture terminal) segment, mobile communication stakeholders and consumers. The DVB standards cover terrestrial and satellite domains and cover transmission, interfacing, conditional access and interactivity for digital video, audio and data delivery. The evolution of the standards follows an agreed open framework. The standard is produced by the Joint Technical Committee (JTC) Broadcast of the European Broadcasting Union (EBU), Comité Européen de Normalisation ELECtro technique (CENELEC) and the ETSI (DVB-2, 2009).

8.2 Satellite Radio Interface Standards

8.2.1 GMR

GMR radio interface specifications are prepared by ETSI Technical Committee-Satellite Earth Stations and Systems (TC-SES). The evolution of these specifications has continually tracked those of the GSM system, as illustrated in Figure 8.2 (adapted from ETSI-1, 2009).

Table 8.1 summarizes the salient system, services and specific features of the three major GMR releases so far (ETSI-1, 2009).

GMR-1 (ETSI TS 101 376), released in 2001, supports circuit-switched voice and data up to 9.6 kbps, short messaging service (SMS), cell-broadcast and location-based services over GSM core network interface-A. GMR-1 Release 2 known as global mobile radio services (GMPRS) comprising three sub-releases over the period 2003–2008, incorporates general packet radio service (GPRS) capabilities to GMR-1 with progressively new terminal classes. GMR Release 2.1 (2003) incorporates packet switched (PS) GPRS-like services up to 144 kbps based on GPRS release 97 (r97) to provide mobile services up to 144 kbps over GPRS core network interface, Gb. GMPRS Release 2.2 (2005) extends services to hand-held terminals at a throughput of 64 kbps in the forward direction and 28.8 kbps in the return direction. Release 2.3 (2008) extends the data rates to 444 kbps units in the forward direction and up to 400 kbps in the return direction. GMR Release 3, known as GMR-1 Third Generation, a satellite adaptation of Enhanced Data Rate for GSM Evolution (EDGE), enhances the throughput to 512 kbps via the core network interface, Iu. EDGE, standardized by 3GPP, is compliant to ITU's IMT-2000 recommendations. It is based on 3G release six protocols interfacing with the core network on Iu-PS interface (Interface between base station system and core network) to provide voice and broadband services including Internet Protocol (IP) multimedia services up to 592 kbps depending on terminal types, which range from small hand-held terminals to large high gain fixed or transportable terminals.

Due to the differences in terrestrial and satellite operational environment, the physical layer of GSM system was modified for the satellite channel. A significant number of upper protocol layer attributes could be shared either unchanged or with modifications. Accordingly, some specifications are unique to GMR-1 and others are either the same or

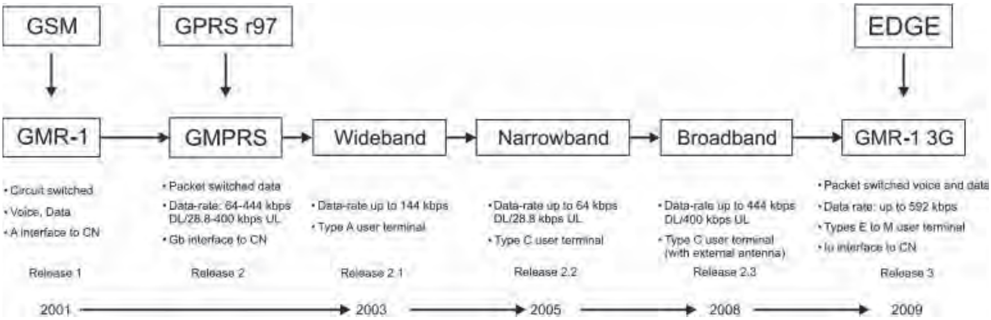


Figure 8.2 Evolution of the GMR system

Table 8.1 Features of GMR-1 releases

GMR standard	Salient system characteristics	Services and specific features
GMR-1	Satellite orbit: Geostationary L and S band operation Circuit mode support GSM infrastructure reuse Dual-mode terminal support Roaming support between terrestrial and satellite networks Standardized interface (Interface-A) with the core network	Voice, data up to 9.6 kbps, short message service (SMS), cell broadcast services, position-based services, high penetration alert
GMPRS-1 (3 versions: 2.1.1, 2.2.1, 2.3.1)	Satellite orbit: Geostationary L and S band operation Dynamic link adaptation IP data services GPRS infrastructure reuse Transportable user terminal QoS support Roaming support between terrestrial and satellite networks Standardized interface (Gb interface) with the core network	Version 2.1.1 – Bidirectional packet data up to 144 kbps Version 2.2.1 – Packet data services support to handheld terminals: 64 kbps forward/28.8 kbps return Version 2.3.1 – Packet data services to A5 sized terminals: 444 kbps forward/202 kbps return (400 kbps with an external antenna); enhanced air Interface
GMR-1 3G	Satellite adaptation of ETSI TDMA EDGE radio interface (ITU-1, 2010) Protocol basis: 3GPP release 6 Air interface basis: GMR-1 Standardized interface (Iu-PS interface) with the core network	Voice, broadband services up to 592 kbps to mobile and nomadic user terminals including IP multimedia services

(Adapted from ETSI-1, 2009.)

a derivative of the GSM system specifications. To retain consistency, GMR specifications follow the same numbering scheme as those of the GSM system. The specifications are typically structured in a multi-part deliverable and each part can have sub-parts to address sub-systems (ETSI-2, 2003).

8.2.1.1 GMR-1 (Release 1)

System

The GMR-1 family utilizes a geostationary orbit to operate in L and S MSS bands for telecommunication services to a variety of mobile earth stations (MES). Figure 8.3 (ETSI-3, 2001) illustrates an architecture of the GMR-1 system (described later in detail). It comprises a space segment, which inter-connects with the required public and private fixed network and the GSM core network through standardized interfaces. Integration with the terrestrial system at the network level provides a tight integration of functionalities at dual-mode UTs. Such facilities include terrestrial-satellite roaming.

The system comprises gateway stations (GSs) connected to public, private and mobile terrestrial network as applicable, one or more multi-beam geostationary satellite, an operational control centre (OCC – not shown in the figure), satellite operations centre (SOC) and MES with a capability to support GMR-1 service throughout the satellite footprint. The satellite(s) and the SOC, are outside the scope of the GMR-1 specifications.

The gateway comprises one or more gateway transceiver system (GTS) with each GTS serving a single spot beam, a gateway station controller (GSC) to manage the GTSs, one or more mobile switching centres (MSCs) and a traffic control system (TCS) responsible for optimal routing and supporting various services such as position based services. The MSC interfaces with the gateway station subsystem (GSS) comprising the GTS, and GSC over

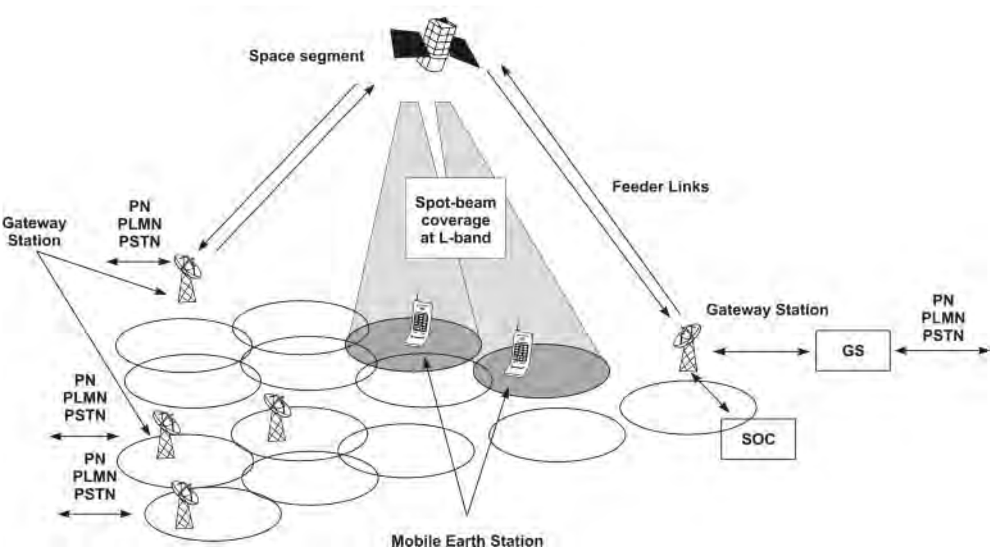


Figure 8.3 Architecture of the GMR-1 system. (Source: ETSI-3, 2001. Reproduced with permission of ETSI.)

an A-interface to connect with the MESs served by the GS. Gateways are also interfaced to GSM's mobility management (MM) system to manage MES mobility in the service area.

The TCS manages the resource allocated by the OCC and GMR-specific service support that are not available in the terrestrial GSM systems – for example direct terminal-to-terminal calls, optimal routing, high penetration alerts and position based services.

The OCC performs various centralized functions that include satellite control, resource management of GSs, network monitoring and network data collection.

Mobile Earth Stations

MESs typically operate in the L band MSS allocation 1525–1559 MHz downlink and 1626.5–1660.5 uplink; MES portfolio includes hand-held units, vehicular units and fixed terminals. A typical unit comprises a Mobile Equipment (ME) and a Subscriber Identity Module (SIM). The ME consists of mobile termination (MT) unit that supports various combinations of Terminal Adapter (TA) and terminal equipment (TE) groups as described in (ETSI-4, 2001). Table 8.2 lists a summary of salient MES characteristics arranged by power class showing the average power, G/T and MES category. MES EIRP (Effective Isotropic Radiated Power) ranges from 5 dBW for the handheld to 11 dBW for vehicular terminal with adjustable antenna and G/T ranges from −24 dB/K for handheld and vehicular (fixed) to −18 dB/K for nomadic MES used in fixed installations.

The receiver characteristics are specified for Gaussian (strong direct signal), and Ricean channels corresponding to walking individual (Carrier to multipath ratio (C/M) = 9 dB, multipath bandwidth = 10 Hz) and a vehicle travelling at 60 kmph (C/M = 12 dB, multipath bandwidth = 200 Hz).

Services

GMR-1 standard set of services, based on GSM phase 2 services, are formally described as the bearer capabilities, the teleservices and the supplementary services as described in the GSM specifications. The GMR-1 system supports subscribers roaming to partner GSM networks with their SIM card provided the participating networks can support GSM MAP (Mobile Application Part) protocols as described in GSM specifications (ETSI-5, 1996). It

Table 8.2 Example features of GMR-1 MESs (frequency of operation (MHz), forward/return: 1525–1559/1626.5–1660.5)

Power class	Average transmit power for normal operation: min-max (dBW)	G/T (dB/K)	MES category
1	5–7	−24	Handheld
2	5–7	−24	Vehicular (fixed)
3	Reserved	–	–
4	7–11	−22	Vehicular (adjustable antenna)
5	Reserved	–	–
6	5–9	−18	Fixed
7	Reserved	–	–

(Adapted from ETSI TS 101 376-5-7 V1.3.1 (2005-02).)

supports dual-mode (i.e. terrestrial-satellite) GSM stations with the capability to support routing to a single user number. The system supports the same addressing and numbering as GSM with the addition of location based information (Location Area Identification – LAI) used to identify an associated spot beam and GS. GMR-1 retains the security features of the GSM system.

Optimal routing of MES originated call allows routing to a GS other than the MESs registered GS. The routing criteria may include the subscriber's service provider, the called party number and/or the MES position. High penetration alerts are used to alert a user who may be in a disadvantaged location. Position-based services depend on GPS-derived (global positioning system) position fix. The changes to the protocols relative to GSM are described in a number of GMR specifications.

Layered Architecture

Lower Layers

The physical layer characteristics are specified in part 5 of the specifications. Here we summarize salient features of the lower layers including modulation/coding, power control, multiple access with examples of representative physical and logical channels. A logical channel contains data related to a specific function whereas a physical channel is responsible for physical transport. Several logical channels can be multiplexed on a single physical channel.

Coding and Modulation

The applied channel coding rate depends on factors such as the channel impairments on per call basis, information block size and the desired protection level. It comprises an inner coding scheme in tandem to an outer code. The outer code comprises Cyclic Redundancy Check (CRC) (8, 12 or 16 bits parity, depending on the channel) whereas the inner code comprises convolutional coding with a typical constraint length of 5 and rates of half, one third, one quarter and one fifth, as required.

The power control messages use the (24, 12) systematic Golay encoder with a soft-decision Golay decoder.

Reed–Solomon code is used for a number of channels. For example, the basic alerting channel (BACH) uses a systematic (15, 9) Reed–Solomon code.

Various puncture masks are used to fit the coded bits into the channel's bit capacity. Channel-dependent interleaving is used and can be applied intra-burst or inter-burst and is based on block interleaving methods with pseudorandom permutations. A scrambler adds a binary pseudo-noise sequence to randomize the incoming bit stream. Certain channels such as the traffic channel include data encryption to prevent eavesdropping.

The modulating symbol rate is fixed at 23.4 ksps. A majority of traffic and control channels use $\pi/4$-CQPSK (coherent quadrature phase shift keying) modulation shaped by a root raised cosine filter of 0.35 roll off factor; $\pi/4$-CBPSK (coherent binary phase shift keying) modulation with identical filter shape and roll off factor is also used. A keep-alive burst is transmitted during periods of speech inactivity to save the battery life, satellite power, reduce co-channel interference, add-comfort noise and maintain the power control and timing/frequency synchronization. The bursts uses $\pi/4$-DBPSK (differential binary phase shift keying), with the same (i.e. 0.35) roll-off root-raised cosine filter. The BACH uses pulse shaped 6-PSK (Phase Shift Keying) modulation and the frequency correction channel (FCCH) burst is a real-chirp signal spanning three slots.

Power Control

Power control is exercised in both the directions to minimize interference, conserve power at satellite and MES, and maintain signal quality. The closed-loop control utilizes a feedback from the remote receiver for power adjustment; in the open-loop method the transmitter adjusts power based on the behaviour of the received power, with the premise that receive and transmit channel behaviours are correlated. The open loop system has a faster response as it avoids the delay incurred in a feedback loop from the distant transmitter and hence suits events of rapid fading.

In the idle mode, the MES is required to lock to the broadcast control channel (BCCH) to extract the system information in readiness for communication.

Multiple Access

A time division multiplexed/time division multiple access (TDM/TDMA) accessing scheme is used in the service link. The forward channel consists of a continuous TDM channel and the return link consists of TDMA bursts, synchronized in time and frequency to avoid interference. Message synchronization is also necessary to ensure that the start of each message sequence is identifiable. Procedures for these types of synchronization have been specified. Table 8.3 summarizes the timing features of the TDMA.

Physical and Logical Channels

Frequencies are allocated in steps of 31.25 kHz in pairs separated by 101.5 MHz. There are 1087 pairs in the 34 MHz MSS allocation. The minimum allocation of channels per spot beam is five.

The physical channels can be used to carry traffic or to support various control functions to ensure proper functioning of the network. The features and uses of these physical and logical channels are summarized in Table 8.4.

Upper Layers

GMR-1 upper layer protocol structure, illustrated in Figure 8.4 (ETSI-3, 2001), is similar to that of the GSM. Differences exist in the lower layers due to differences in the physical characteristics between the satellite and terrestrial systems, that is the propagation environment (Ricean versus Rayleigh), propagation delay (270 ms versus a few ms), cell size (hundreds of km versus a few km) and addition of satellite-specific services. Moreover, spectral efficiency in satellite systems is crucial due to the large reuse distances of hundreds of kilometres in satellite systems.

Table 8.3 TDMA timing characteristics

Timeslot	5/3 ms – 78 bits
Frame	40 ms (24 timeslots)
Multiframe duration	640 ms (16 TDMA frames)
Superframe	2.56 sec (4 multiframes or 64 TDMA frames)
Hyperframe	3 hr 28 min 53 sec 760 ms (4896 superframes, 19 584 multiframes or 313 344 TDMA frames)

(Adapted from ETSI TS 101 376-5-7 V1.3.1 (2005-02).)

Table 8.4 Example features and uses of various traffic and control logical channels

Channel type	Category/Id	Sub-category	Features and uses
Traffic	Traffic channel 3 (TCH3)	–	Carries voice at gross rate of 5.8 kbps; 3 time-slots wide
	Traffic channel 6 (TCH6)	–	Carries user data at 2.4 and 4.8 kbps; 6 time-slots wide
	Traffic channel 9 (TCH9)	–	Carries user data at 2.4, 4.8 and 9.6 kbps; 9 time-slots wide
Control	Broadcast	–	There are various types of forward broadcast transmissions
		Frequency correction channel (FCCH)	Used for MES system synchronization and frequency correction
		GPS broadcast control channel (GBCH)	Carries Global Positioning System (GPS) time information and GPS satellite ephemeris information
		Broadcast control channel (BCCH)	Broadcasts system information and system timing
		Cell broadcast channel (CBCH)	Demand assigned channel to broadcast short message service (SMS) and deliver cell broadcast information to the MESs on a spot beam basis
	Common control channel (CCCH)	–	There are various types of control channels
		Paging channel (PCH)	Downlink channel to page MESs
		Random access channel (RACH)	Uplink request channel for traffic and standalone dedicated control channel (SDCCH)
		Access grant channel (AGCH)	Downlink channel to alert MES
		Basic alerting channel (BACH)	Downlink channel to alert MES, transmitted with greater power and coding than normal paging channel for disadvantaged MES locations following failure of normal paging
		Common idle channel (CICH)	Downlink channel for calibration measurements at MES
		Dedicated control CHannel (DCCH)	Bidirectional control channel dedicated to a MES (Note: Terminal-to-Terminal Associated Control CHannel (TACCH) is downlink only)
		Power control sub-channel	To maintain power control during a call, the information bits of the power control sub-channel are multiplexed into six consecutive bursts during TCH3/6/9 calls

(Adapted from ETSI TS 101 376-5-2 V1.1.1 (2001-03).)

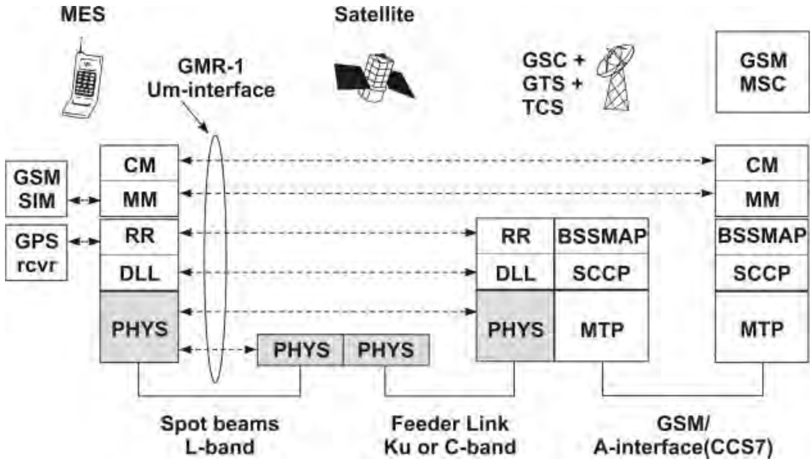


Figure 8.4 GMR-1 protocol architecture. (Source: ETSI-3, 2001. Reproduced with permission of ETSI.)

GMR-1 data link layer (DLL) protocol – Link Access Protocol for SATellite (LAPSAT) is a modified version of the GSM protocol that incorporates the use of a selective reject and repeat protocol called group reject, inclusion of larger window size and new timer values. The radio link protocol (RLP) for data and telemetric services includes some phase 2+ features of GSM. The DLL protocols are described in part 5 of the specifications.

The radio resource (RR) management sub-layer was modified to accommodate the physical differences with respect to the GSM radio interface. For example, the GMR-1 random access channel (RACH) procedure can accommodate the large differential path delay within a spot beam as well as the enhanced features of GMR-1 with regards to position based services, optimal routing and terminal-to-terminal call. The RR sub-layer protocols are described in part 4 of the GMR specifications.

The GMR-1 MM sub-layer in the MSC remains identical to that of GSM. Its peer in the MES is modified to accommodate the added functionalities of position based services and optimal routing. These protocol differences are described in part 4 of the GMR specifications. The connection management (CM) sub-layer manages connection-oriented calls performing tasks such as call setup, release.

8.2.1.2 GMPRS-1

The GMPRS-1 is release 2 of the GMR-1 specifications, appending to release 1, packet data service based on the GPRS. These specifications enable packet data services at data rates ranging from 64 kbps up to 444 kbps in the forward link, depending on the capability of mobile terminals. Version 2.1.1 (2003) enables bidirectional packet data rates up to 144 kbps supporting quality of service (QoS) differentiation across users with dynamic link adaptation; Version 2.2.1 (2005) specifies packet data services to handheld terminals at up to 64 kbps on the forward link and 28.8 kbps on the return link. Version 2.3.1 (2008) supports broadband packet services at data rates up to 444 kbps on the forward link and 202 kbps on the return link on A5 size transportable terminals with a provision to support data rates up to 400 kbps on the return link through an external antenna. Version 2.3.1 utilizes advanced techniques such as low density parity check (LDPC) and 32-amplitude phase shift keying (APSK) modulation to enhance spectral efficiency and provides bi-directional streaming services.

Differences in radio transmission and reception characteristics with respect to GMR-1 release 1 specifications include:

- Support of unpaired frequencies.
- Differences in test conditions due to different energy per symbol to noise power density ratio (E_s/N_o) requirements in packet mode.
- Modifications in the switching time requirements.
- Addition of four power classes of UTs (C, A and D – two types) and associated specifications such as radiation pattern, receiver G/T, frame error rate at receiver.
- A few modifications applicable to all classes of UTs – for example transmit requirements (burst ramp-up and ramp-down time, unwanted emissions).
- Updates to packet burst data structure and modulation scheme.
- Radio link failure detection, capability to identify BCCH type (Temporary-broadcast control channel (T-BCCH) related to a temporary spot beam or Anchored broadcast control channel (A-BCCH)) in idle mode and idle mode loss of T-BCCH when camped on it.
- Link adaptation procedures.
- Some clarifications with regards to the reporting by MESs of link quality.

Figure 8.5(a) and (b) illustrate the control plane and user plane protocol architecture of the system (ETSI-1, 2009). Familiarity with GSM protocols is assumed.

While a number of layer 3 (signalling) procedures are identical to GSM (Phase 2+), various modifications have been incorporated for the satellite environment – these are specified in (ETSI-6, 2003). The layer 3 comprises four sub-layers: Radio Resource Management (RRM), Mobility Management functions (GMM), Logical Link Control (LLC), CM functions and Session Management (SM) functions to activate, modify and delete the contexts for Packet Data Protocols (PDPs) and supporting functions for SMS control.

The RR sub-layer contains RR management protocol and Dual-Tone Multiple Frequency (DTMF) transmission and Dual-Tone Reception Service (DTRS) management.

The CM manages Call Control (CC) for non-GMPRS services, SMS Support for non-GMPRS services, GMPRS Short Message Service Support (GSMS) for GMPRS services supported on Class C MESs, SM for GMPRS services supporting Class C MESs and Supplementary Services Support for non-GMPRS services.

8.2.1.3 GMR-1 3GPP

System Description

GMR-1 3GPP air interface specification, released in 2009 as a satellite component of the 3G mobile communication system is an extension of the EDGE specified in ETSI TDMA EDGE (ITU-1, 2010). EDGE is a terrestrial 3G IP based broadband service envisioned as an IP multimedia system (IMS). The GMR-1 3GPP air interface is an evolution of the GMPRS-1 system using new developments in the satellite physical layer air interface technology. Its protocol architecture is a derivative of 3GPP Release 6. The satellite base-station (gateway in GMR-1 terminology) is the equivalent of GPRS Edge Radio Access Network (GERAN).

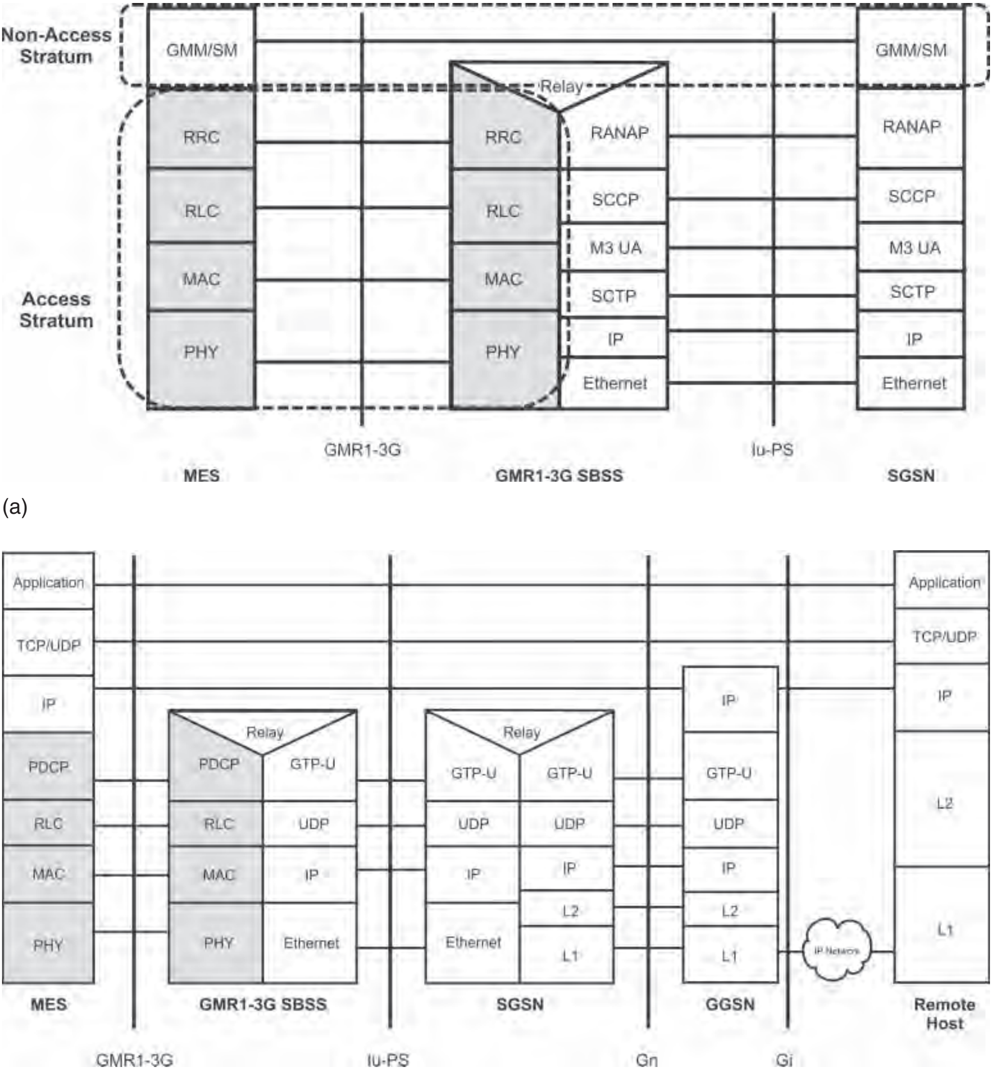


Figure 8.5 Protocol architecture of GMPRS (a) control plane and (b) user plane. (Both parts source: ETSI-1, 2009. Reproduced with permission of ETSI.)

The air interface is applicable in L and S MSS bands and connects with the core network over an interface known as I_u-PS. The interface has been defined in the document GMR-1 3G 45.005 (ETSI-7, 2005).

Key features of the air interface include (ETSI-3, 2001):

- Up to ~592 kbps mobile-user throughput;
- Spectrally efficient multi-rate Voice over Internet protocol (VoIP) with 0 byte header compression;

- IP Multimedia Services;
- IPv6 compatibility;
- Differentiated QoS across users and applications;
- Robust waveforms to improve link closure reliability;
- Dynamic link adaptation;
- Multiple carrier bandwidth operation;
- Multiple terminal types – Hand-held terminals, Personal Digital Assistant (PDA), vehicular, portable and fixed;
- Terrestrial-satellite handovers;
- Unmodified Non-Access Stratum (NAS) protocols and core network.

Figure 8.6 illustrates the end-to-end architecture of GMR-1 3G interfaced with different core network interfaces, which offers an operator the flexibility to select the most appropriate interface and backward compatibility. Packet data are services through Gb and Iu core-network interfaces. Figure 8.7 illustrates the protocol architecture of air interface for the user plane.

Mobile Earth Stations

A variety of MES are supported – hand-held, fixed, transportable and nomadic. In addition to provide IP data traffic at data rates commensurate with MES capability, MESs support voice at 2.45 and 4 kbps using 0-byte header compression. For signalling purposes, an MES type is identifiable by a type identifier, which is based on its operational frequency band (L or S) and technical features. Technical features necessitating MES distinction include TDMA burst processing capability, antenna type and EIRP, communication support in terms of half/full duplex, usability as handheld or fixed, core network capability in terms of A, Gb or Iu mode. Table 8.5 illustrates parameters of a subset of terminal types.

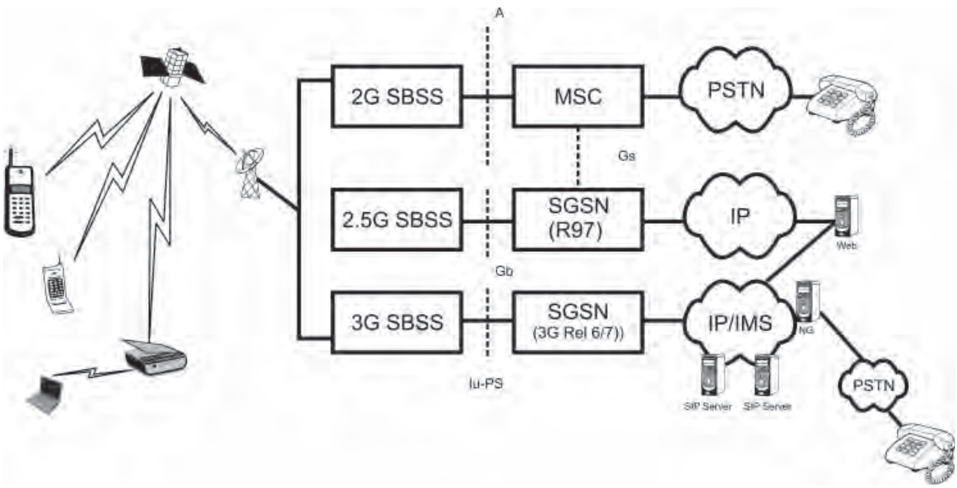


Figure 8.6 End-to-end architecture of GMR-1 interfaced with different core network. (Source: ETSI-1, 2009. Reproduced with permission of ETSI.)

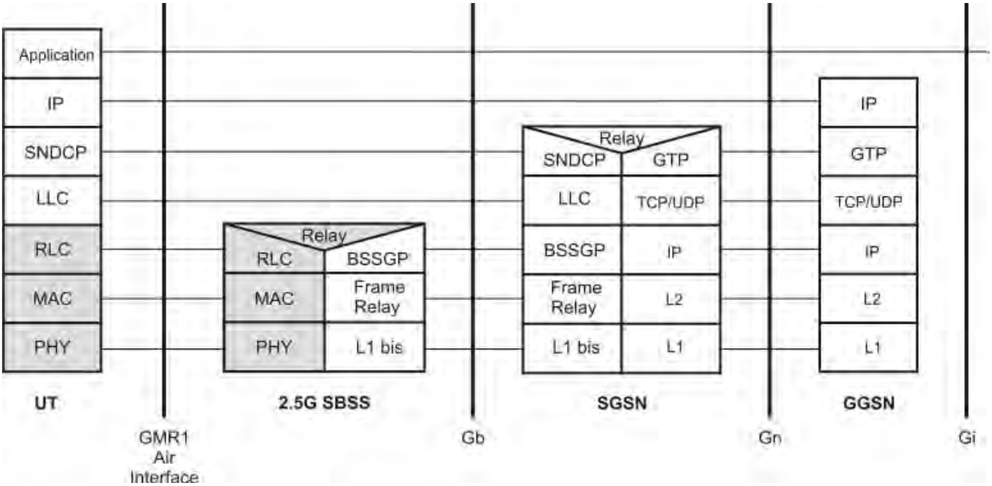


Figure 8.7 User plane protocol architecture of GMPRS. (Source: ETSI-1, 2009. Reproduced with permission of ETSI.)

Table 8.5 Example parameters of a few GMR-1 3G MESs (LHC = Left hand circular) (Note: The average EIRP is obtained as an average of 200 bursts with each burst averaged for 90% of the active part of the burst)

Terminal type identifier (Hex)	Terminal type	Operating band	Transmit/receive antenna gain (dBi) – minimum	Average nominal EIRP (dBW) (normal condition) – min-max	G/T (dB/K)	Transmit/receive polarization
48	A	L	12	12.1–14.9	−16.2	LHC/circular
09, 0A, 0B, 0C	C	L	0 (ideal case)	5–7	−24	LHC/circular
0D	D	L	8.5	8–10.8	−18	LHC/circular
0E, 0F	D	L	15	15–17.8	−18	LHC/circular
1A	G	S	2	1–3.8	−27	Circular/LHC
1F, 20	H	S	2	4–6.8	−27	Circular/LHC
24, 25	I	S	13	12–14.8	−17	Circular/LHC
38	M	L	12	11–13.8	−17	Circular/circular

(Adapted from ETSI TS 101 376-5-5 V3.2.1 (2011-02).)

The L band terminal receive and transmit frequencies are 1525–1559 and 1626.5–1660.5 MHz respectively, and the S band frequency bands are 2170–2200 and 1980–2020 MHz respectively, with tuning steps of 31.25 kHz in each band. Thus there are 1087 paired channels in the L band and 960/1280 in the S band – note the asymmetric allocation of S band. The L band duplex channels are paired with a separation of 101.5 MHz.

Layered Architecture

The system uses a frequency division multiplexed TDM/TDMA scheme forward/return link – with the same frame structure in each part. The TDMA frame consists of 24 slots; 16 TDMA frames constitute a multi-frame (MF) and 4 MFs make a super-frame.

Two categories of channels are used – traffic and control channels. Circuit mode traffic channels (A-mode) use convolutionally coded $\pi/4$ CQPSK to provide encoded voice and data rates upto 9.6 kbps. Packet data traffic channels (PDTCHs) carry packet mode traffic in Gb or Iu mode at data rates between 8.8 and 587.2 kbps using $\pi/2$ BPSK, $\pi/4$ QPSK, 16-APSK or 32-APSK modulation, the modulation selection being dependent on the chosen data rate. (Note: Iu = Interface between base station system and core network).

The RF bandwidths range from 31.25 to 312.5 kHz. The modulated signals are filtered by a 0.35 roll-off root raised cosine filter. In addition to traffic and control channels, a constant envelope frequency-modulated (called dual-chirp) channel is transmitted for initial time and frequency acquisition at the UTs (see below for a description of the control channels). Depending on the transmission bit rate a variety of coding schemes are used – convolution code, LDPC and turbo-code. Table 8.6 lists the forward error correction (FEC) codes supported by GMR-1 3G.

The return PDTCHs are defined as PDTHC (m, n) where m denotes the bandwidth as a multiple of 31.25 kHz and n the number of TDMA slots assigned to it. For example, PDTHC (1, 3) denotes a physical channel occupying a bandwidth of 31.25 kHz comprising three slots in its TDMA frame. A dedicated PDTCH channel can be assigned for carrying voice traffic. The bearer service ranges from 1.2–592 kbps.

The transmission rate depends on a combination of the modulation scheme and the coding rate. For example, the peak rate of 590.4 kbps for PDTCH3 (10, 3) can be achieved by 16-APSK at a code rate of 2/3 or, $\pi/4$ QPSK at a coded rate 5/6.

Table 8.7 gives a sample of PDTCHs.

Control channels provide signalling and synchronization information. Three types of logical channels are defined – broadcast, common and dedicated. Each logical channel type has a defined frame structure and performs a specific function. Each logical channel is mapped on to a physical channel for transmission. Each category of signalling channels can be assigned to A, Gb or Iu communication modes together or individually (see Figure 8.6).

Table 8.6 Code rates used in GMR-1 3G

Code type	Explanation
Convolutional	Constraint length K = 5, 6, 7 and 9. Base code rates = 1/4, 1/3 and 1/2. Various rates can be derived by puncturing or for small FEC block by tail biting
Turbo	Based on 3GPP/3GPP2 turbo code. Various code rates are derived by puncturing
Reed–Solomon	Systematic (15,9)
Low density parity check (LDPC)	DVB-S2 LDPC based; optimized further for small FEC block size
Extended Golay	Extended (12, 24)
Cyclic redundancy check	3, 5, 8, 12 or 16 bit CRC for error detection

(Adapted from ETSI TS 101 376-1-3 V3.1.1 (2009-07).)

Table 8.7 Examples of packet data traffic channels

PDTCH Id	Direction	Modulation	Coding	Transmission bandwidth (kHz)	Peak data rate
PDTCH (4, 3)	Up/down	π/4-QPSK	Convolution	125	116.8
PDTCH2 (5, 3)	Up/down	32-APSK	LDPC	156.25	382.4
PDTCH3 (10, 3)	Down	16-APSK	Turbo	312.5	590.4

(Adapted from ETSI TS 101 376-1-3 V3.1.1 (2009-07).)

Broadcast Channel Examples

FCCH or FCCH3, transmitted in the forward direction, is used by mobile for frequency correction and synchronization of system information cycle. FCCH is a chirp signal (mentioned above) covering three time-slots whereas FCCH3 covers 12 time-slots.

The *GPS broadcast control channel* (GBCH or GBCH3), transmitted in the forward direction, provides time and GPS satellite ephemeris information to the MESs. GBCH is formatted for transmissions on two bursts whereas GBCH3 is formatted for 12-slot transmission.

The BCCH broadcasts system information to the mobiles on BCCH burst comprising three slots or a 12-slot structure.

Common Control Channels (CCCH) Examples

The *paging channel* (PCH) is transmitted on a six or 12 slot convolutionally encoded π/4 CQPSK modulated channel in the forward direction to page UTs.

The RACH is transmitted on six or 12-slots in the return direction to request the network for traffic channel.

The *access grant channel* (AGCH), allocates traffic channel to the MES in a 6- or 12-slot burst.

The BACH is transmitted in two-slots 6-PSK modulated burst to alert MESs of a call.

The system utilizes *modulation and coding adaptation* in both the forward and the return links to match the user's prevailing channel conditions to promote effective utilization of satellite RR and improve link reliability. The network matches the modulation and coding combination to the prevailing channel conditions on the basis of the measured signal quality at the gateway or as reported by the UT. The UT extracts the modulation and coding advised by the gateway from the traffic burst.

Power control is applied at the UT to provide a relatively uniform received power flux density from the receiver population at the spacecraft inputs thereby minimizing the probability of co-channel interference, while reducing unwarranted battery power drainage at the UT. The power control mechanism provides up to 24 dB EIRP reduction in steps of 0.4 dB. In a closed loop power control, the MES EIRP is controlled under the direction of the gateway on the basis of MES's received signal quality, whereas in an open loop power control, the MES takes an autonomous decision whether to make a transmit power adjustment by measuring the received signal and assuming a positive correlation between the receive and transmit signals. The closed loop adjustment has a slower time response due to the propagation delay involved in the process.

The *Medium Access Control* (MAC) *and Radio Link Control* (RLC) layer design are derivatives of the 3GPP GPRS/EDGE but include features to mitigate the impact of long propagation delay of GEO satellites on throughput efficiency by minimizing protocol exchanges ('chatting'), while maximizing RF bandwidth utilization. The *MAC layer* is responsible for functions that include mapping of logical channels to the physical channels, selection of logical channels for signalling and traffic, MES measurement reporting, broadcasting and listening of BCCH and CCCH and identification of different traffic flows of one or more MESs on the shared channel. The *RLC layer* performs link control functions such as segmentation of upper layer protocol data units (PDUs) into RLC data blocks, concatenation of upper layer PDUs into RLC data blocks, padding to fill out RLC data block, reassembly of RLC data blocks into upper layer PDU, in-sequence delivery of upper layer PDUs, link adaptation, ciphering and deciphering in Iu Mode and sequence number check to detect lost RLC blocks.

The *Radio Resource Control* (RRC) layer design is based on the 3GPP GERAN Iu mode. The RRC specifications have been modified to account for the delay of GEO satellites and improve spectral efficiency. RRC functions include: assignment, reconfiguration and release of RRs for the RRC connection, establishment, reconfiguration and release of radio bearers, release of signalling connections, paging, routing of higher layer PDUs, control of requested QoS, control of ciphering and integrity protection, integrity protection, support for location services, timing advance control. Satellite specific enhancements in RRC layer includes enhancements to cell update procedure to reduce number of round-trips, fast RRC connection setup using RACH, fast GERAN Registration Area update using the RACHs and fast RRC connection reject/connection release using AGCH.

The Packet Data Convergence Protocol (PDCP) layer design is also based on 3GPP TSs with modifications incorporated for satellite channels (figure 8.5(b)). It performs functions such as header compression and decompression of IP data streams, transfer of user data with satellite optimizations including early context establishment procedures and 0 byte header compression. PDCP layer in general improves spectral efficiency, reduces satellite power usage, reduces packet loss rate and improves – capacity, MES battery life and interactive response time.

8.2.1.4 GMR-2

System Description

The GMR-2 system provides wide-area voice, data, fax and supplemental communication services between mobile and fixed/mobile users through public and private switched telecommunication network via a geostationary satellite.

System Features

The system supports roaming to partner networks using SIM cards and supports roaming of non-GMR users in to the GMR-2 network when such network agreements are in place and the given network supports MAP protocols of the signalling system number 7 (SS7).

The system supports communication on a single number for a dual-mode (satellite-terrestrial) UT. It supports optimal routing such that the shortest terrestrial route is made available to each call. To minimize propagation delay the system can provide direct mobile-to-mobile calls by switching a spacecraft transponder that inter-connects service-links directly (e.g. L-L or L-S). The system does not support inter-beam handover on the premise that spot-beam boundaries are coarse allowing adequate time for the user to complete a call; the system supports handover in-between frequencies, code rate and time slot on the same beam to contend quality degradation by local interference, propagation or other impairments. It utilizes power control in both the directions.

Typically, the space segment consists of a geostationary satellite comprising a large number of spot beams, a number of gateways operating at C band and communicating to the MES on L band MSS allocation. The TS of GMR-2 have been produced by ETSI TC-SES. In principle, its architecture and services are similar to GMR-1 (Release 1) and as such can be represented by the architecture illustrated in Figure 8.3. The naming conventions are identical to those used in the GMR-1 system. However, there are differences in details – for instance, the physical layers parameters are markedly different. The TC detailed the similarities and differences between the GSM circuit mode system and a geostationary satellite system, summarized in Table 8.8 (ETSI-8, 2001) in order to optimize the end-end system, retaining the GSM features as far as feasible.

Figure 8.8 (ETSI-8, 2001) shows the system elements proposed in GMR-2. It comprises a geostationary satellite, Network Control Centre (NCC), Satellite Control Facility (SCF), Customer Management Information Centre (CMIS), a number of Gateways and a population of MES. The UT types include handsets, vehicle-mounted terminals and fixed terminals. The Gateways have external interfaces to existing fixed telecommunications infrastructure, namely public switched telephone network (PSTN), private network (PN) and PLMN (public land mobile network).

The gateways implement the radio modem functions of the terrestrial BTS (Base Transceiver Station), the RR/traffic channel management, call setup functions of (terrestrial) base station controller (BSC) and switching functions of MSC, along with databases for subscriber data (see Figure 8.9(a)). An example of traffic and signalling channel radio connectivity between gateways, network and satellite control station and UTs is presented in Figure 8.9(b). Broadcast and common control signalling channels, provided by the NCC, are used during the initial call set up. Note that single hop user-to-user traffic link through the satellite uses L band links. Call control for user-to-user circuits are performed by the NCC, Gateways and by switching on the satellite, to achieve single-hop connectivity.

The MES comprises ME consisting of a MT unit that can include an interfaces for TAs and TE. A SIM provides a unique user identity and profile. The MT is the main unit of the MES responsible for all major radio and communication functions:

- Reception of broadcasts, paging, access grant information, making call requests, network authentication and encryption functions, SIM support;
- Transmission and reception of two-way voice communication, transmission of DTMF messages from handheld and vehicle phones over an established voice link;
- Interfacing data and fax;
- Support of commissioning procedures.

Functions performed by the network management centre include system resource management, congestion control and network synchronization; day-to-day operations and maintenance; inter-station signalling management; user terminal and gateway commissioning support, common channel signalling management; defining and managing payload configurations.

The Gateway sub-system (GWS) comprises transceivers, controllers and other units interfaced to the MSC through a single A-interface The Gateway Sub-system Controller (GSC) controls the functioning of Gateway Transceiver Sub-system (GTS). Each GTS is responsible for call management in a given spot beam.

Gateway functions include management of dedicated signalling and traffic channels; connectivity to external fixed networks; MM; interoperability functions related with partner land

Table 8.8 Similarities and differences between a satellite and terrestrial system used as the basis in the development of GMR-2 specifications

System parameter	Conceptual similarity	Differences	Comment
Spot beams	Similar to terrestrial cell	Significantly larger distance between MES and gateway in comparison to terrestrial base station-user terminal distance leads to issues such as a more difficult time synchronization requirement and exerts an adverse impact on protocols	–
Frequency reuse between spot beams	Similar to cell reuse	Reuse distance is much larger (hundreds of kilometres) in satellite hence there is a greater need of a higher spectral efficiency	–
Physical layer	–	Different propagation environment – Ricean in satellite, Rayleigh in terrestrial leading to different physical layer optimization criteria; Doppler effects caused in satellite systems due to movements of satellites and fast-moving mobiles such as aircraft have to be countered	–
Upper layers	Similar functions	–	Core network is reused
MES-Gateway propagation distances	–	Large distances in satellite systems leads to much greater attenuation and propagation delay; significant distance variability within satellite coverage leads to stringent time synchronization requirement	–
Satellite/base station EIRP	–	Spacecraft power is severely limited necessitating Ricean operation due to lower link margin and necessity of features like power control	–
Gateway/base station routing	–	Satellite systems can utilize any gateway in the footprint whereas cellular system communicate only with the base station covering the cell	–
User cooperation	–	Satellite systems require user cooperation due to severe impact of shadowing – leading to the need of high-power alert signals for shadowed mobiles	–

(Adapted from TS 101 377-1-3 V1.1.1.)

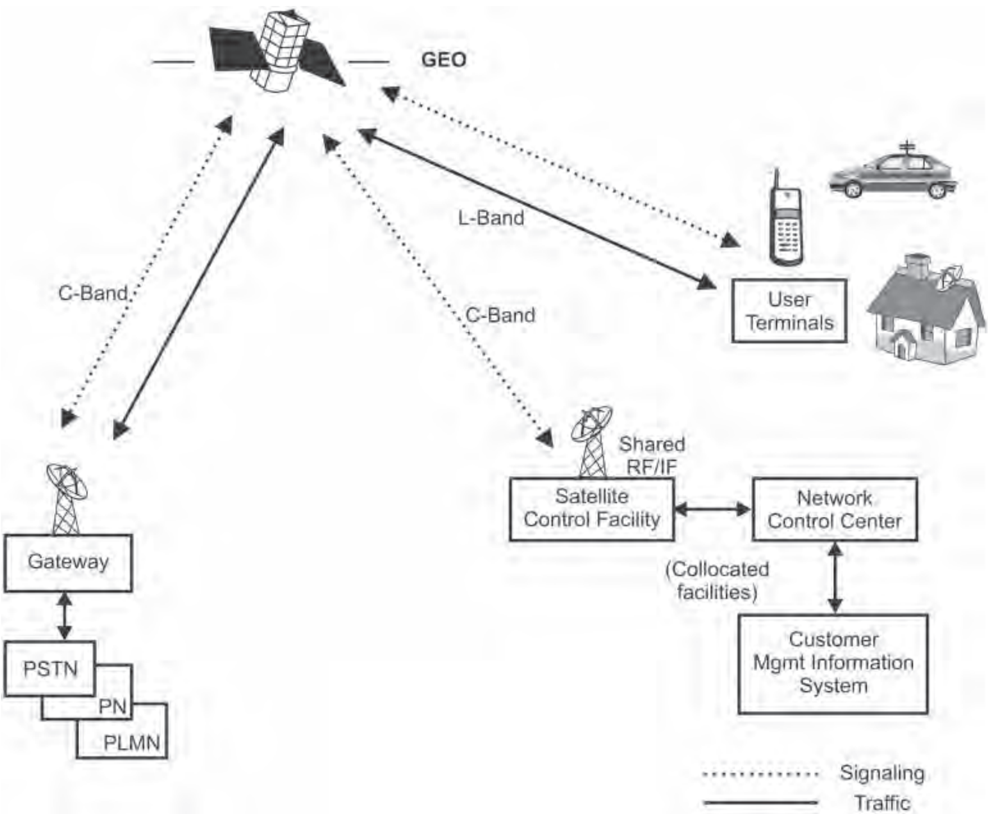


Figure 8.8 GMR-2 system element. (Source: ETSI-8, 2001. Reproduced with permission of ETSI.)

mobile networks; authentication and encryption services; RRM, support for user terminal commissioning, and others.

The SCF is responsible for spacecraft monitoring, managing payload configuration, generation and distribution of satellite ephemeris, telemetry and commands processing for spacecraft management, real-time range and range rate measurement for satellite ephemeris generation, real-time payload switching, etc.

The CMIS manages gateway configuration, billing and accounting and processing of call records.

The NCC functions include call management comprising network parameters broadcast, payload configuration through the SCF for mobile-mobile connections, traffic monitoring, RRM and network timing synchronization.

Network functions to support basic services operation involve: call handling, subscriber identity authentication, support of emergency calls, supplementary services support, SMS support (ETSI 9, 2001). Network support for satellite operations entails location registration to keep track of mobile locations, intra-spot beam handover, high penetration alerting and centralized call management from the NCC.

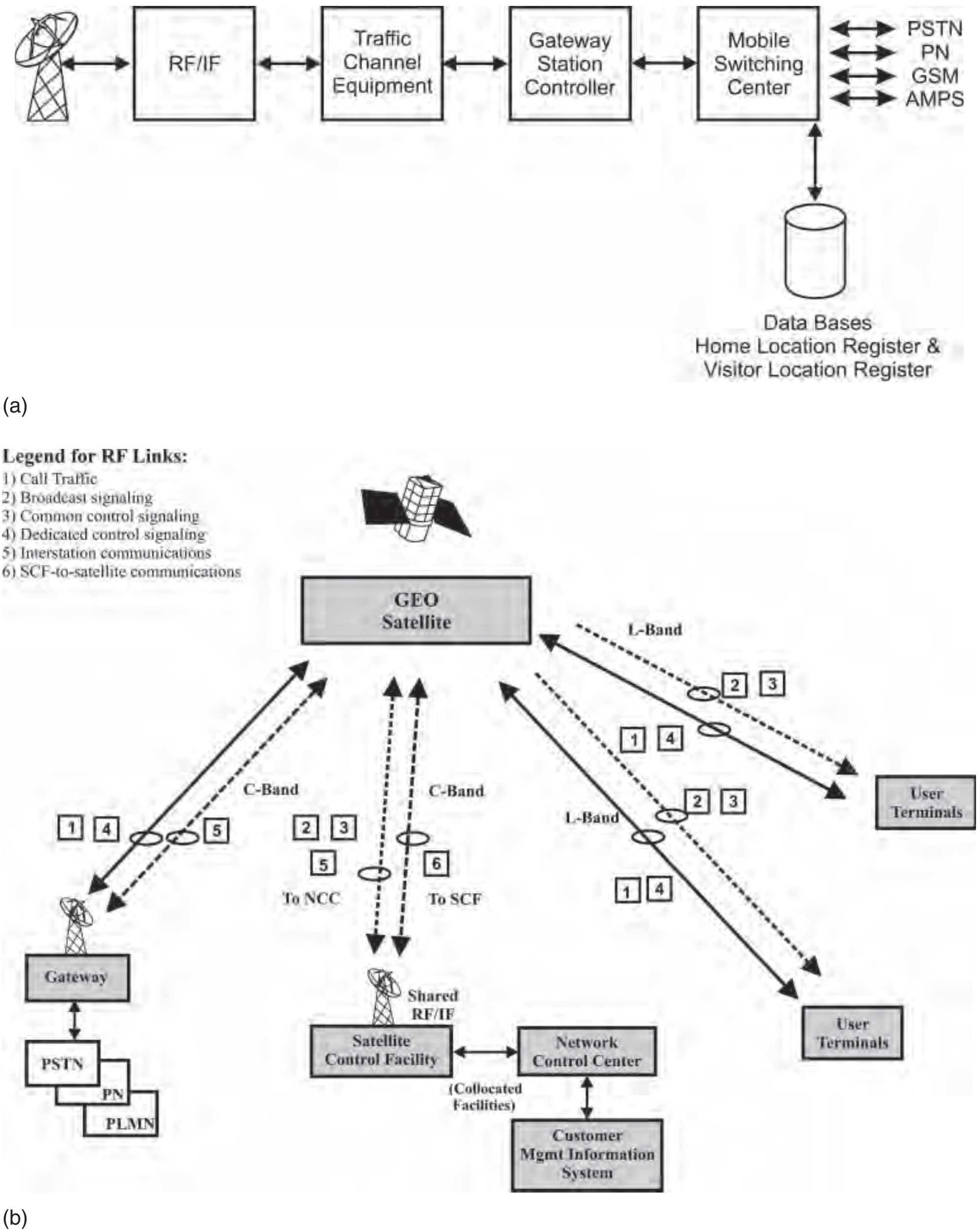


Figure 8.9 (a) Gateway block schematic (b) Example of traffic and signalling channel connectivity. (Both parts source: ETSI-8, 2001. Reproduced with permission of ETSI.)

The specifications outline procedures for additional call handling that perform functions such as: queuing mobile originating or terminating calls at GWS, signalling, user data security, discontinuous transmissions/reception to minimize MES power drainage and off-air call set-up.

MESs perform all essential network and call functions to support voice, data and facsimile calls, including support of features essential in commissioning and network control directives. MES functions cover transmissions of DTMF signalling message for external applications, automated signalling, and adherence to terminal power-up/down and dynamic power control procedures.

Lower Layers

Service Link

The service and feeder link frequencies are as follows:

- *Service link*
  Space–Earth: 1525.0–1559.0 MHz, Right Hand Circular Polarization (RHCP)
  Earth–Space: 1626.5–1660.5 MHz, RHCP
- *Feeder link*
  Earth–Space: 6425.0–6725.0 MHz, Linear Horizontal Polarization
  Space–Earth: 3400.0–3700.0 MHz, Linear Vertical Polarization

Channel Types

There are two categories of physical channels – Satellite traffic channel (S-TCH) and satellite control channel (S-CCH), each sub-categorized into logical channels consistent with their respective function.

There are four types of physical traffic channels to carry voice or user data – Full rate traffic channel (24 kbps), half rate traffic channel (12 kbps), quarter rate channel (6 kbps) and one eighth rate traffic channel (3 kbps). Table 8.9 lists the capability of each channel type.

Signalling channels are categorized as: Satellite BCCH, Satellite Common Control Channel and Satellite Dedicated Control Channel, and are listed in Table 8.10.

Multiple Access

The FDM/TDMA scheme is used with 200 kHz carrier spacing in the forward link comprising eight slots and 50 kHz in the return link comprising two slots.

The forward time-slot lasts 15/26 ms (~0.58 ms) at a burst-rate of about 270.833 kbps. The return time-slot lasts 60/26 ms (~2.3 ms) at a burst-rate of ~67.708 (1625/24 kbps).

Table 8.9 Traffic channel and application

Traffic channel type	Speech quality	User data rate (kbps)
Full rate	Enhanced	9.6
Half rate	Robust	4.8
Quarter rate	Basic	2.4
Eight rate	Robust	–

(Adapted from ETSI TS 101 377-5-1 V1.1.1 (2001-03).)

Table 8.10 Control channels and functions

Type of control channel	Main functions
Broadcast	To broadcast network functions and assist receiver synchronization
Control	To page and alert mobile users and to enable call set-up
Dedicated control	Dedicated use during a call

(Adapted from ETSI TS 101 377-5-1 V1.1.1 (2001-03).)

The TDMA frame structure hierarchy comprises a hyperframe of 3 28 min 53 s 760 ms duration, chosen to support GMR-2's cryptographic mechanism. Each hyperframe comprises 2048 superframes of 6.12 s, which is further sub-divided into MFs. There are two types of MFs, partitioned to carry specific logical channels −120 ms MF of 26 frames and ~235.4 ms MF of 51 frames. Twenty-four MFs of the 26-frame MF are used for traffic and two are used for in band signalling. The 51-frame MF is used for control channels.

Six types of TDMA bursts are used, depending on the functions – to carry traffic, access the network, MES time synchronization, frequency and time correction, etc.

Coding and Modulation

Various block and convolution codes are used with interleaving – the selection depending on the function of the given logical channel. For example, the basic voice channel uses a half-rate, 64-state punctured convolutional code; the associated and dedicated signalling channels uses a half-rate, 64-state convolutional code along with a Fire Code. The information bits in some control channels utilize a half-rate, 16-state Convolutional code and a Fire Code. Yet another coding scheme utilizes a one-third-rate, 64-state convolutional code. Interleaving dimensions for the communications channels vary from 3−81 bursts depending on the required protection.

The forward link uses OQPSK modulation followed by a 0.35 roll off square root raised cosine filter. The return link modulation (including the single hop MES-MES link) for the traffic and in-band signalling utilizes Gaussian Minimum Shift Keying (GMSK) – pre-coded in the same manner as specified in the standard GSM.

Other physical layer functions in GMR-2 are – management of dynamic power control mechanisms of the MES and the gateway; synchronization of time and frequency at the MES receivers and measurement procedures for the selection and reselection of spot beam.

Transmit and Receive Characteristics Examples

Various transmitter and receiver parameters have been specified for the MES, Gateway, the NCC and the spacecraft (ETSI-10, 2001). A few illustrative examples are included here.

MES power is specified at the antenna connector of the equipment and for user TE with integral antenna, a reference antenna with 0 dBi gain has been assumed. For all types of fixed terminals, the requirements are specified in terms of EIRP where transmit power refers to the power when averaged over the useful part of the burst. Peak power refers to the maximum taken over a sufficient time such that the power level has settled. Table 8.11 lists the specified transmit power of MESs.

Nominal L band EIRP from satellite is specified as 50.7 dBW and the C band EIRP as 5.6 dBW. Spacecraft is channelized with a granularity of 200 kHz. The gateway terminals

Table 8.11 Transmitter power of MESs (tolerance levels not included)

Power class	Maximum output (dBm) – except power class 1	Comment
1	35	
2	39	–
3	37	–
4	33	Hand-held

(Adapted from ETSI TS 101 377-5-5 V1.1.1 (2001-03).)

Table 8.12 Receiver G/T for satellite and two types of user terminals: gateway and NCC

Receiver	G/T (dBi/K)	Comment
L band satellite	15.3	Minimum over full coverage
C band satellite	−6.5	Edge of coverage
Handset	−26	–
Fixed user terminal	−14	–
C band gateway	30.5	Minimum in clear sky conditions
C band NCC	32.7	Minimum in clear sky conditions

(Adapted from ETSI TS 101 377-5-5 V1.1.1 (2001-03).)

are specified to be able to adjust the EIRP of each burst individually, compensate rain fades and support the power control mechanism. Various other parameters related to MES RF are specified, for example the frequency tolerance, harmonics and spurious transmissions.

Table 8.12 summarizes the G/T of various types of terminals used in the network.

Bit error rates are specified for various types of channels under quantifiable measurement condition. For example, under nominal conditions – defined as a channel without interference and an input level ranging from 10–20 dB above the reference sensitivity level in an additive white noise Gaussian channel (AWGN) – the bit error rate prior to error correction shall be $<10^{-4}$ (for voice) and that of data traffic shall be $<10^{-7}$.

8.2.2 Satellite Component of UMTS/IMT-2000

8.2.2.1 Introduction

Figure 8.10 shows a generic IMT-2000 satellite network demonstrating various interfaces (ITU-1, 2010). Some of these interfaces are excluded from the IMT-2000 radio interface specifications since their implementation depends on operators' design and optimization criteria. Such interfaces include satellite–feeder link, satellite–satellite link, internal interface between terrestrial and satellite elements within a MES and the interface between satellite and the core network. However, it is recommended that the satellite component should interface with the terrestrial core network in a similar manner as the terrestrial system, incorporating satellite-specific alterations such that all the key IMT-2000 service requirements

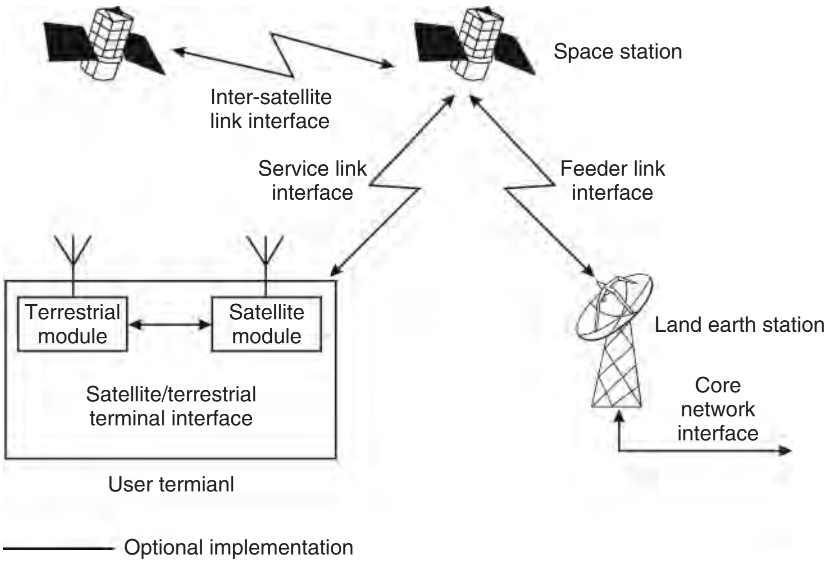


Figure 8.10 Generic IMT-2000 satellite network. (Source: ITU-1, 2010. Reproduced by permission of ITU.)

applicable to the given technical and marketing considerations can be implemented. In a dual mode, satellite-terrestrial, user terminal, it should be possible for the terminal to select the appropriate mode either automatically or under user control. The user terminal should provide bearer services in both terrestrial and satellite modes with a roaming facility and be aligned to IMT-2000 service management and provisioning. Handover between terrestrial and satellite networks is not mandatory but can be applied at an operator's discretion. In absence of this type of handover, switching to the appropriate network can be done by the user. The specifications recommend – location and roaming procedures between terrestrial and satellite networks to be as those used for updating locations between different PLMNs; a single number access regardless of which component (terrestrial or satellite) the terminal is currently using; and provisions to select the desired terrestrial network when this is the set preference.

The ITU recognizes eight IMT-2000 compatible radio interfaces – termed, SRI A, B, etc. These interfaces deal with the radio part of *the service links* only. Architectural and system description are included where appropriate due to their strong relationship with the radio interface. We will introduce high level descriptions of the architecture, services and the lower layers. The reader will note that the IMT-2000 recommendations uses the term 'radio interface', while we have used the air interface elsewhere in the book.

8.2.2.2 Satellite Radio Interface A

System Description

Satellite Wideband-Code Division Multiple Access (SW-CDMA), adapted from the terrestrial IMT-2000 radio interface – Universal Terrestrial Radio Access (UTRA) Frequency

Division Duplex (FDD) or Wideband code division multiple access (WCDMA) – was examined by ETS-TC SES TC as an open standard. The system intends to reuse the terrestrial core network and the radio interface specifications for the Iu (UTRAN-CN) and Cu (USIM (Universal Subscriber Identity module) smartcard-ME) interface to the extent feasible. The adapted satellite service radio link operates in FDD mode with RF channel bandwidth of either 2.350 or 4.700 MHz, corresponding respectively to 3.840 Mchip/s (full chip-rate) and 1.920 Mchip/s (half chip-rate) for each transmission direction to support numerous packet and circuit bearer services in the range 1.2–144 kbps and bit error ratio (BER) between 10^{-3} and 10^{-6}.

The half-rate transmissions allow a tighter packing of carriers for finer granularity in sharing. Satellite-specific features include soft handover with satellite diversity, Doppler pre-compensation in the forward direction, half chip-rate support, inclusion of a high power PCH in the forward direction, a reduced power control rate to contend long propagation delays and changes to scrambling and preamble sequences.

The system can utilize any constellation design – LEO, MEO, HEO, GEO, etc. Multiple satellite diversity is supported, but remains optional. There are no restrictions on the type of satellite transponders except for the necessity of a transparent transponder if return link path diversity is used. The maximum permissible delay is 400 ms to ensure compatibility with each constellation type.

Transmission Scheme Outline

Forward link

Three types of logical channels are used; broadcast, control and traffic, each mapped to a suitable physical channel for radio transmission. Broadcast and control channel functions include:

- Transferring network and beam specific information.
- Transferring control information to an identified MES when its position is known.
- High penetration paging.
- Signalling and access to the network for resource allocations.

The period of each frame is 10 ms for the 3.840 Mchip/s option and 20 ms for the 1.920 Mchip/s. The forward frame comprises 15 time-slots containing TDM data bits for various types of logical channels – for example dedicated control and data channels. A MF lasting 80 ms consists of eight frames (full-rate option) or four frames (half-rate option) and one super-frame consists of nine MFs. A close-loop power control scheme is used in conjunction with an open-loop scheme that supports packet mode transmissions.

Three different QoSs are supported, controlled by the appropriate code and interleaving scheme. The standard services with inner convolution (one-third rate) coding and interleaving targets a BER of 10^{-3}. The high quality services with inner coding and interleaving plus outer RS (Reed–Solomon) coding and interleaving (or optionally turbo-coding) targets a BER of 10^{-6}. Service-specific FEC coding can be applied as needed (e.g. at a higher layer). Table 8.13 lists the coding requirements for various QoS.

Each in-phase and quadrature-phase channel of the QPSK modulation scheme is spread to the chip rate with the same code, followed by data scrambling with a beam-specific complex scrambling code. The BPSK modulation is used for higher data rates (> 4.8) kbps to reduce receiver sensitivity to phase errors.

Table 8.13 FEC and interleaving characteristics

FEC	Standard quality: convolutional coding with code rate 1/3 or ½, constraint length $k = 9$, Variable puncturing to match the required information rate High quality concatenated RS code over GF(2^8), concatenated with inner convolutional code with rate 1/3 or 1/2, constraint length $k = 9$. Turbo coder is an option
Interleaving	Interleaving on a single frame basis (default) Interleaving on a multiple frame basis (optional)

(Adapted from ITU-1, 2010.)

The scrambling code is a complex quaternary sequence of length 2560 chips, and by staggering a few chips, can be reused in each beam. If a spacecraft is accessed by different LESs (land earth stations) on the same frequency slot, they must be either mutually synchronized or use different scrambling codes. The channel codes belong to the orthogonal variable spreading factor (OVSF) family, which can preserve orthogonality between forward link channels of different rates and spreading factors (SFs).

To assist initial acquisition of the transmissions at mobiles, LESs transmit a pilot channel of known code and modulating pattern characteristic. MESs search for the pilot's scrambling code. When locked to the pilot, the MES can de-spread the common control physical channel (CCPCH) and extract network information from the logical broadcast channel. The convergence length of the initial acquisition depends on the number of codes to be searched and therefore depends on MES's knowledge of the candidate satellites.

Return Link

The dedicated physical channel frame structure is identical to that of the forward channel but the accessing scheme is code division multiplexed (not TDM).

Data modulation of the return link is BPSK, where the dedicated physical data and traffic channels are mapped to the carrier I and Q branches respectively. These branches are also spread to the chip rate with two different channel codes – the same OVSF code as the forward link, and complex-scrambled by a MES specific complex quadri-phase scrambling code assigned to the MES on a semi-permanent basis by the LES.

Table 8.14 summarizes the salient baseband features of both forward and return links.

Mobile Earth Stations

SW-CDMA supports hand-held (H), vehicular (V), transportable (T) and fixed (F) user terminals to cater for various types of user population. The service capability depends on the size of user terminal. All classes of service and QoS (defined in terms of bit error rate) can be supported by all the terminals type in the range 1.2–16 kbps; 32–64 kbps services are supported by all terminals except the hand-held. 144 kbps is supported only on fixed and transportable terminals. Table 8.15 lists examples of key RF parameters of the hand-held, vehicular and transportable MESs.

Mobility – Handover

The system supports four types of soft handovers (or 'hand-offs') – beam, satellite, LES and frequency.

Table 8.14 Salient baseband features

Multiple access	Direct sequence CDMA
Chip rate (where appropriate)	1.920 Mchip/s or 3.840 Mchip/s
Time slots (where appropriate)	15 time-slots per frame
Modulation type	Dual-code BPSK in the return link QPSK or BPSK in the forward link
Dynamic channel allocation	No
Duplex method	FDD
Synchronization between satellites required	Synchronization between BSs working on different satellites is not required Synchronization between BSs working on the same satellite is required

(Adapted from ITU-1, 2010.)

Table 8.15 Typical parameters of MESs

RF parameter	MES class		
	Hand-held	Vehicular	Transportable
Channel bandwidth (kHz)	2350[a], 4700[b]	2350[a], 4700[b]	2350[a], 4700[b]
Maximum EIRP (dBW)	3.0	16.0	16.0
Antenna gain (dBi)	−1.0	2.0[c], 8.0[d]	4.0[c], 25.0[d]
Power control range (dB)	20.0	20.0	20.0
Transmit/receive isolation (dB)	> 169	> 169	> 169
G/T (dB/K)	−23.0[c], −23.0[d]	−23.5[c], −20.0[d]	−23.5[c], −20.0[d]
User caused Doppler shift compensation	Yes	Yes	Not applicable
Maximum mobile speed (km/h)	250[a], 500[b]	250[a], 500[b]	Not applicable

[a]Half rate option (1.920 Mchip/s).
[b]Full rate option (3.840 Mchip/s).
[c]Typical value for LEO constellation.
[d]Typical value for GEO constellation.
(Adapted from ITU-1, 2010.)

Beam handover – MESs regularly report pilot signal power to noise ratio (C/N) of adjacent beams to the communicating LES; the LES initiates a beam handoff procedure at a set threshold by transmitting the signals to the operating and the candidate beams and instructs MES to start demodulating both the signals. When the LES receives reports of successful reception of the candidate signal at the MES, it drops transmissions to the operating beam.

Intersatellite handover – The MES measures the pilot C/N and scrambling code of candidate beams. When a new scrambling code is detected, the measurements are reported to the LES, which may exploit satellite diversity through maximal ratio combining by transmitting the same signal through both satellites and invoking a handoff at an appropriate time following further network protocol exchanges.

Inter-LES handover – Inter-LES hand-off is achieved by negotiations between the present and the new LES. The new LES starts transmitting a carrier towards the affected mobile, which is simultaneously instructed by the current LES to search for the carrier of the new LES. When the MES confirms to the current LES, satisfactory reception of the new carrier, the current LES stops transmission thus handing over the communication to the new LES.

Inter-frequency handoff – This is a hard hand-off and is either intra-gateway or inter-gateway.

8.2.2.3 Satellite Radio Interface B

System Description

Wideband-code/time division multiple access (W-C/TDMA) radio interface called satellite radio interface B is being examined by ETS TC-SES as a voluntary standard. The system utilizes the same physical channel structure as the terrestrial radio interface. The interface is compliant with the core network and the Iu and Cu interface.

It is based on a ‘hybrid’ CDMA/TDMA technique occupying RF channel bandwidth of either 2.350 or 4.700 MHz in each transmission direction to provide bearer services globally at 1.2–144 kbps and BER ranging from 10^{-3} to 10^{-6} depending on UT capability. The system includes a high penetration PCH and incorporates a means to provide user location. It supports FDD/TDD (time division duplex) operation that allows the use of lower complexity diplexers. The half-rate option provides a finer granularity grid in frequency planning and adds robustness to chip synchronization and tracking in a channel with high Doppler shift. The use of a code division multiplex allows full frequency reuse between spot beams.

The system can be built on all types of well-known constellations without imposing any restriction on the type of spacecraft transponder, that is, transparent or regenerative. The maximum tolerable propagation delay is 400 ms. It incorporates a capability to utilize satellite diversity at the LES.

Transmission Scheme Outline

Forward/Return Links

Closed-loop power control is used in both forward and return links. An open loop power control is available to support packet transmission and for initial power setting during a call set-up.

To manage the QoS, the system supports three service classes, each controlled by an appropriate code and interleaving scheme. The standard services with inner convolution coding (one-third rate) and interleaving targets to achieve a BER of 10^{-3}.The high quality services with inner coding and interleaving plus outer RS coding and interleaving (or optionally turbo-coding) targets a BER of 10^{-6}; for service-specific coding FEC can be applied as needed (e.g. at a higher layer).

In the forward direction, an orthogonal CTDM is adopted and in the return link quasi-synchronous W-C/TDMA is adopted.

The transmission is organized in frames of 20 ms subdivided in eight time-slots and transmitted in bursts of one or more time-slots. MFs consist of eight frames plus one extra frame. Frame 0 of each MF in the forward direction is dedicated for broadcast functions like paging, high penetration messaging channel, synchronization, etc. and frame 0 of the return direction is reserved for asynchronous traffic such as the request burst. User data is sent on dedicated bursts.

There are two physical broadcast channels in the forward direction – primary and secondary CCPCH. Other physical channels transport general control information and traffic.

Transmission schemes include two-way stream mode when a dedicated channel is made available in both the directions. Various types of logical channels are used, each mapped appropriately to a physical channel. For example, consider the primary and secondary CCPCH. The primary CCPCH transports the broadcast control logical channel (BCCH) that carries system and beam specific information to the MESs. The secondary CCPCH transports control information to an identified MES on two forward access logical channels (FACH) when MES position is known. In the return direction the physical return access channel (PRACH) transports the logical return access channel (RACH), carrying control information, and the Return channel (RTCH), carrying short user packets. Example of a bi-direction channel includes dedicated data physical channel (DDPCH) used to transport control information such as higher layer signalling, the logical dedicated control channel (DCCH) and bidirectional user data to carry the dedicated traffic channel (DTCH).

As mentioned, the system utilizes two spreading rates in each direction – 3.840 Mchip/s (full chip rate) and 1.920 Mchip/s (half chip rate). The proposed spreading codes are identical to those used in the terrestrial system – the OVSF codes based on length 128 bits Walsh-Hadamard code set for the 1.920 Mchip/s option and length 256 bits Walsh-Hadamard code set for the 3.840 Mchip/s option.

Let us briefly consider the transmission sequence. The incoming multiplexed data stream is spread, randomized and modulated. The data stream is split into two bi-polar data streams (I and Q streams). These data, (clocked at symbol rate), are multiplied with the bi-polar components of the spreading code vector (clocked at chip rate). The spread sequences are further randomized using bi-polar PN-sequences, called randomization codes such that the transmitted signal appears noise-like to an unsynchronized receiver and modulated.

Dual-BPSK and QPSK are used in the forward link, depending on the physical channel. In the return link the proposal is to use either $\pi/4$-QPSK or dual BPSK data modulation both with complex randomization for DPCH and pre-compensated frequency modulation (PFM) is an option as a trade-off.

The system time and frequency reference signals emitted by the satellite must correspond to the nominal frequencies and timing. In case of a transparent transponder, the LES can pre-compensate the transmission times, frequencies, chip rates of its uplink as necessary. In the return link, the LES controls timing of the individual MES such that signals arrive at the intended satellite in quasi-synchronism with the nominal system time and frequency.

The system must keep the transmit frame structure in all the beams of a satellite aligned at fixed time offsets of a few chip periods to permit reuse of the same randomization code in all beams of the same satellite. Similarly, time offsets must be applied in the return link frame structure of signals arriving at the satellite from different beams, if the same randomization code is to be used in all the beams. The LES controls the MESs in a manner such that these time offsets occur at the LES receiver.

Furthermore, the system provides for a network-wide time synchronism, covering all the satellites to facilitate handover and diversity advantage.

MES acquisition and synchronization process involves initial acquisition using the synchronization word of the common channel, selection of the highest correlation peak to obtain the correct satellite/beam, extracting the randomization code from the common pilot channel, further improvements to the process using the pilot, and finally, locking to the BCCH to extract relevant high level synchronization and system information.

Following the forward link acquisition, the MES uses the information gathered from the broadcasted information to pre-compensate the frequency and burst time. On successful reception from a mobile, the LES updates the frequency and burst time correction estimated for the specific mobile, following which the MES can begin transmission, while tracking/updating the forward link frequency and time. A reacquisition procedure is initiated by the LES in case of loss in MES's acquisition. The return link synchronization is more demanding for non-geostationary satellites due to dynamic path length variations.

Table 8.16 lists the main transmission features.

Mobile Earth Stations

The system supports various classes of UT – a fixed class and three classes of mobile terminals, namely, hand-held, vehicular and transportable. All the terminals can provide services at the specified QoS (defined in terms of bit error rate) in the range 1.2–16 kbps; whereas

Table 8.16 Transmission characteristics of satellite radio interface B

Multiple access	Forward link Hybrid wideband Orthogonal CDM/TDM (W-O-C/TDM) Return link Hybrid wideband Quasi-synchronous quasi-orthogonal CDMA/TDMA (W-QS-QO-C/TDMA)
Chip rate	3.840 Mchip/s or 1.920 Mchip/s
Time slots	Eight time-slots per frame
Modulation type	QPSK or dual-code BPSK in the up-link QPSK or BPSK (low data rate) in the down-link
Dynamic channel allocation	No
Duplex method	FDD or FDD/TDD
Forward error correction	Standard quality: convolution code with code rate 1/3 or ½, constraint length $k = 9$. Variable puncturing to match the required information rate High quality: RS code concatenated with inner convolutional code with rate 1/3 or 1/2, constraint length = 9, optionally – turbo-coder
Interleaving	Interleaving on a single burst (default) Interleaving on multiple bursts (optional)
Synchronization between satellites	Synchronization between LESs working on the same channel of different satellites is required Synchronization between LESs working on different channels of the same satellite is not required

(Adapted from ITU-1, 2010.)

Table 8.17 Examples of key RF parameters of mobile user terminals

RF parameter	MES class		
	Handheld	Vehicular	Transportable
Channel bandwidth (kHz)	2350[a], 4700[b]	2350[a], 4700[b]	2350[a], 4700[b]
Maximum EIRP (dBW)	8.0[c], 12.0[d]	11.0[c], 18.0[d]	20.0[c], 20.0[d]
Antenna gain (dBi)	2.0	2.0[f], 8.0[e]	4.0[f], 25.0[e]
Power control range (dB)	20.0	20.0	20.0
G/*T* (dB/K)	−23.0[f], −22.0[e]	−23.5[f], −20.0[e]	−23.5[f], −20.0[e]
Doppler compensation	Yes	Yes	Not applicable
Maximum permissible mobile speed (km/h)	250[a], 500[b]	250[a], 500[b]	Not applicable

[a] At 1.920 Mchip/s.
[b] At 3.840 Mchip/s.
[c] FDD/TDD mode.
[d] FDD mode.
[e] Typical value for GEO constellation.
[f] Typical value for LEO constellation.
(Adapted from ITU-1, 2010.)

all the terminals excepting the hand-held support 32–64 kbps and only the fixed and transportable terminals support 144 kbps. Table 8.17 lists examples of key RF parameter of the mobile UTs.

8.2.2.4 Satellite Radio Interface C

System Description

The SAT-CDMA is a LEO and GEO SIR to offer various IMT-2000 mobile services at a user data rate of 144 kbit/s for LEO and 384 kbitps for GEO using W-CDMA at a chip rate of 3.84 Mchip/s. The characteristics of the proposed baseline LEO constellation are summarized in Table 8.18.

Table 8.18 Characteristics of the proposed baseline configuration

Orbit configuration	LEO
Orbit altitude (km)	1 600
Orbit inclination (°)	54
Number of orbit planes	8
Number of satellites per orbit plane	6
Phase offset between adjacent orbit satellite (°)	7.5
Orbit period (min)	118.2
Service area latitude (15° elevation)	± 69°
Intersatellite link	Provisioned

(Adapted from ITU-1, 2010.)

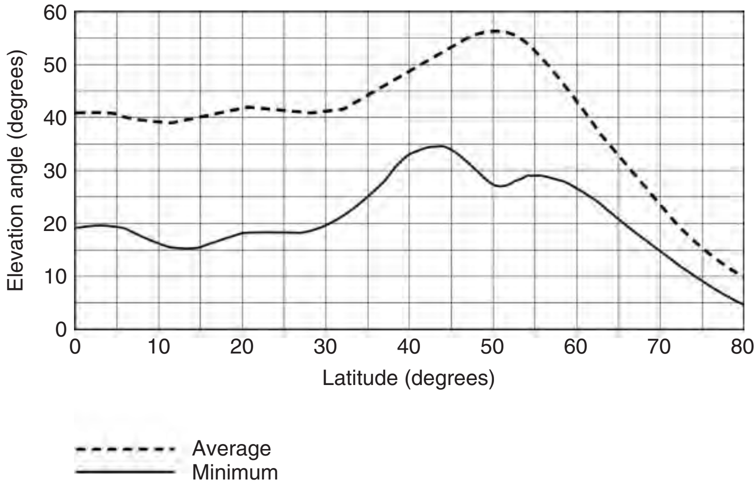


Figure 8.11 Average and minimum elevation angle as a function of latitude. (Source: ITU-1, 2010. Reproduced by permission of ITU.)

Figure 8.11 (ITU-1, 2010) shows the average and minimum elevation angle as a function of latitude to illustrate that the minimum elevation angle of 15° can be sustained up to about 69° latitude for the LEO constellation. The average elevation angle in this range is about 25°. In the densely populated mid-latitude 30–60° regions the minimum elevation angles range between ~20 and 27°, averaging, ~41 and 43°. It was demonstrated that the probability of simultaneous two-satellite and three-satellite visibility exceeds 98 and 40% respectively in the latitude range 30–50°. The reader will recall the advantages of diversity and propagation aspects at high elevation angles.

The satellite comprises 37 spot beams – the foot-print spanning about 2721 km and a satellite ground pass of about 16 min. The elongation in spot beam span from the beam centre to the edge of foot print ranges from about 520 to 1654 km due to the Earth's curvature.

The GEO satellites use multiple spot beams and a global beam. The system may comprise a single satellite or multiple satellites for extended coverage.

Services include:

- Basic bearer service at data rates 2.4–64 kbps.
- Packet data services at 144 kbps for LEO system and 384 kbps for the GEO system.
- Teleservices such as emergency calls, facsimile, SMSs, paging, and so on.
- High penetration paging for users in disadvantaged locations.
- Multimedia Broadcast and Multimedia Service (MBMS).
- Point to multi-point broadcasts including facility of a return channel.

The space segments comprise satellite with on-board processing units. Service links operate at 2.5 GHz band, the feeder links at 4/6 GHz band and the intersatellite links at 60 GHz band. Table 8.19 gives a sample of the main characteristics of the satellites in LEO and GEO.

Table 8.19 Main characteristics of satellites

	LEO	GEO/with global	GEO/with multi-spot beams
Nominal EIRP (dBW)	9.6	64	64–74
Service link receive antenna gain (dBi)	20	30	36–39
Noise temperature (K)	50	550	550
Service link *G*/*T* (dB/K)	−7.0	2.6	8.6–11.6

(Adapted from ITU-1, 2010.)

Table 8.20 User terminal characteristics for the LEO space segment

Terminal type	Data rate (kbps)	Nominal speed restriction (kmph)	Maximum EIRP (dBW)	Antenna gain (dBi)	Receive temperature (K)	G/T (dB/K)
Hand-held	2.4–16	500	2	2	300	−22.8
Vehicular	2.4–32	500 (maximum 1000)	15.8	2	300	−22.8
Transportable	2.4–64	0	21	4	300	−20.8
Fixed	2.4–144	0	36	23	500	−4

(Adapted from ITU-1, 2010.)

The UTs for the LEO space segment are handheld, vehicular, transportable and fixed. Table 8.20 presents the characteristics of UTs including the achievable user data rate (kbps) and mobile speed limits. For GEO systems additional classes of mobile terminals namely portable and aeronautical are available and the system can provide higher throughput for transportable terminals as listed in Table 8.21.

The system supports various types of mobile-originated network-assisted handovers.

Table 8.21 User terminal characteristics for the GEO space segment (NA = not available)

Terminal type	Data rate (kbps)	Nominal speed restriction (km/h)	Maximum EIRP (dBW)	Antenna gain (dBi)	Receive temperature (K)	G/T (dB/K)
Hand-held Class 1, 2, 3	2.4–32	500	3, −3, −6	0, 0, 0	290, 290, 290	−33.6, −33.6, −33.6
Portable	2.4–64	500	5	2	200	−26
Vehicular	2.4–144	500 (maximum 1000)	13	4	250	−25
Transportable	2.4–384	0	17	14	200	−14
Aeronautical	2.4–64	1000	6	3	NA	NA

(Adapted from ITU-1, 2010.)

Inter-Beam Handover

Each MES monitors the pilot level of adjacent beams during a call, and reports to the network. When the pilot level of the present beam falls below a predefined threshold, the network may decide to proceed with an inter-beam handover based on the present satellite's ephemeris and the prevailing operational conditions. The network (LES) begins transmission of the call through the present and the candidate beam and instructs the mobile to receive the additional transmissions. Mobiles incorporate a capability to combine signals received from both the beams. As the mobile combines the two signals it informs the LES that the new signal is being received correctly; following this the previous channel is released by the LES.

Intersatellite Handover

Intersatellite handover can be instigated when two satellites are simultaneously visible from an MES and the communicating LES and the call is likely to fail unless transferred to a new satellite; the scenario also offers path diversity. The MES monitors the pilot signal levels from adjacent satellites and reports the estimates to the network when the pilot level of the present satellite falls below a predefined threshold. Based on the satellite ephemeris and other operational conditions, if the network decides to instigate a handover it transmits the call through the new satellites and instructs the MES to receive the additional signal. The MES can utilize satellite path diversity during this phase. The present channel is released when the visibility with the satellite is lost and the call continues over the new satellite.

Inter-LES Handover

Consider a satellite handover scenario when the new satellite is not visible to the LES handling the call. It is then necessary for the LES to handover the call to a new LES with visibility of the new satellite thus instigating an LES-LES handover. In this scenario, after an LES-LES negotiation phase involving resource allocation and signal rerouting the new LES begins transmission of the call and the LES handling the call instructs the MES to receive the new signal. The MES informs the LES when the new signal has been acquired following which the LES drops the signal and the communication is established with the new LES.

The channel bandwidth is 5 MHz. Power control uses up to 20 dB control with steps in the order of 0.25 dB. Doppler compensation is effected by pre-compensating transmission frequency at the LES based on satellite ephemeris in conjunction with post-processing at the receiver based on signal processing of the received signal involving a two-dimensional search of frequency and time followed by a closed-loop fine control using an FFT-based (fast Fourier transform) frequency detection algorithm.

A general guidance is available regarding propagation effects for LEO system at low elevation angles of $10-20°$as tabulated in Table 8.22.

Transmission System

Channel Types

Logical channels perform broadcast, control and traffic functions and are mapped to an appropriate physical channel for transmission. Table 8.23 tabulates a few representative examples.

Table 8.22 Propagation impairments guidelines at 10–20° elevation angles for LEO systems

Fade level	General: −7 to + 4 dB Occasional: −10 dB at very low elevation, particularly in a suburban environment, where specular multipath dominates
Fade duration	Moving car: Typically, 100–200 ms Fixed user: 10–20 s in a suburban environment, where specular multipath dominates

(Adapted from ITU-1, 2010.)

Table 8.23 Representative examples of logical channels

Logical channel	Function
Broadcast control channel	Broadcasts system control information
Paging control channel	To page, when MES location is unknown
Common control channel	Bi-directional channel for exchanging control information
MBMS control channel	Transfers control information related to MBMS
Dedicated traffic channel	Bidirectional channel to exchange information to a specific MES
Common traffic channel	Forward point to multipoint channel for transferring information to all or a particular user group
MBMS traffic channel	Transfers MBMS traffic

(Adapted from ITU-1, 2010.)

Physical Channel

A total of 10 physical channels are used in the forward direction and three in the return direction. Table 8.24 lists representative examples.

The transmissions chain involves a number of steps broadly outlined as follows. Parity bits are attached to each transport block for CRC. The transport blocks are then segmented into channel coding blocks and concatenated in preparation for channel coding. Convolutional or turbo-coding is applied, depending on the requirement of the logical channel. Convolution codes use a constraint length of 9 with coding rates of one-third and one-half. The turbo-coder comprises the classic arrangement of a parallel concatenated convolution code of two encoders and an internal interleaver at the input of one of the arms to achieve a code rate of one third. The type of code is selected and signalled by the upper protocol layers. The resulting symbols are interleaved to randomize error bursts. The blocks are rate-matched to ensure compatibility with the multiplexed physical channel and segmented into 10 ms radio frames. The individual transport blocks (corresponding to each logical channel) are multiplexed and segmented, interleaved a second time, and mapped to the corresponding physical channel ready for modulation. Some physical channels (e.g. dedicated physical data channel) include a transport-format combination indicator field encoded using a (32, 10) sub-code of the second order Reed–Muller code to indicate the format of the physical channel. Where needed, the transmit power control (TPC) bits are included and encoded by repetition such that they are mapped to the assigned number of TPC bits on the physical channel.

The primary synchronization code for LEO is constructed as two generalized hierarchical Golay sequences and for the GEO satellites from a generalized hierarchical Golay sequence chosen to have good aperiodic autocorrelation properties.

Table 8.24 Representative examples of physical channels

Direction	Physical channel	Description
Forward	Common pilot channel	30 kbps channel to carry a predefined symbol sequence to provide a phase reference to various physical channels
Forward	Primary common control channel	30 kbps fixed-rate channel to carry the broadcast logical channel; also used as a timing reference for all forward/return physical channels
Forward	Secondary common control channel	Carries paging channel and forward access channel
Forward	Synchronization channel	Used by MESs for beam search
Return	Random access channel	Carries random access logical channel used by the MESs on an Aloha basis
Return	Common control channel	Carries common control logical channel
Return	Dedicated physical channel	Comprises a multiplexed set of logical channel carrying user data and control information to the network

(Adapted from ITU-1, 2010.)

Spreading

The forward direction spreading comprises a short code for channelization and a long code for scrambling. A periodic direct sequence spreading code is used in the long code with a period of 38 400 chips (equal to frame length of 10 ms). The code is known as orthogonal variable spreading factor code and it uses orthogonal complex dual-channel quadrature phase shift keying (OCQPSK) modulation at 3.84 Mchip/s with a roll-off of 0.22. The code is constructed from two binary m-sequences each derived from specified generator polynomials.

The return link uses the same OVSF code with OCQPSK modulation scheme as the forward direction.

Power Control

Power control is used in both the forward and the return direction to resolve the near-far problem, conserve power and minimize self-interference. In absence of feedback information an open loop power control is used.

General

Procedures for beam selection, random access, open and closed loop control, beam selection/diversity reception have been laid out in the specifications.

8.2.2.5 Satellite Radio Interface D

System Description

The radio interface specifies a global MEO MSS system for services to hand-held, vehicular, aeronautical, maritime mobile and semi-fixed UTs at data rate up to 38.4 kbps on satellite-terrestrial dual-mode UTs. The network architecture provides robust reliability against satellite or LES failure by including redundancy in both the segments of the network.

The MEO constellation positioned at an orbital altitude of 10 390 km inclined at 45° comprises 10–12 satellites in two 180° phased planes with intra-plane phasing of 72° and 60° respectively for 10- and 12-satellite constellations respectively. Figure 8.12 illustrates the visibility of one to four satellite(s) as a function of latitude.

At least two satellites are visible simultaneously throughout the world for about 95% of time and up to four satellites are likely to be available for about 15% of time at the worst visibility location. Analysis of minimum-maximum visibility elevation angle as a function of latitude demonstrates that a minimum elevation of > 10° (Average 20°) is available throughout the world and the minimum elevation exceeds 25° (equivalent to an average of 50°) within a 20–50° latitude (see Figure 8.13).

Essential features of the satellite to comply the requirements include a transparent transponder with 163 actively generated spot beams using 127 element direct-radiating array with a capability to continuously switch any of the 490 170 kHz wide, satellite filter channels on a 150 kHz grid across 30 MHz band and a capability to maintain antenna gain and reuse performance throughout its lifetime. Depending on the angular distance from the nadir, there are 19 beam types, within each of which, the range and Doppler remain almost unchanged. A four-cell satellite-fixed frequency plan is used.

The ground segment comprises 12 globally-dispersed terrestrially-connected LESs with network management and billing centre, each interfaced to terrestrial networks that include second generation and third generation (IMT-2000) infrastructure (Figure 8.14). The Interworking functions (IWFs) available at each LES provide roaming capabilities with terrestrial

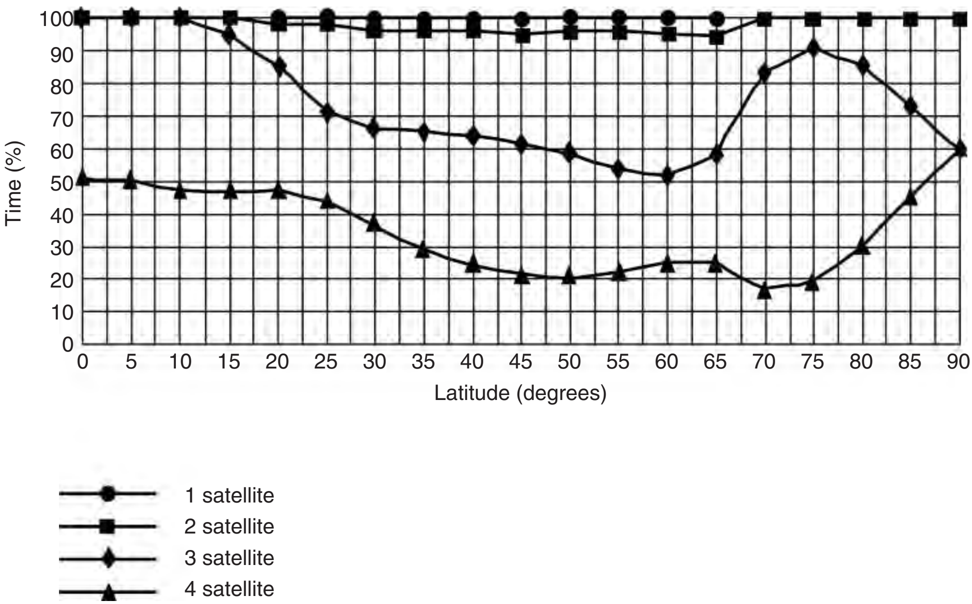


Figure 8.12 Visibility of satellites as a function of latitude. (Source: ITU-1, 2010. Reproduced by permission of ITU.)

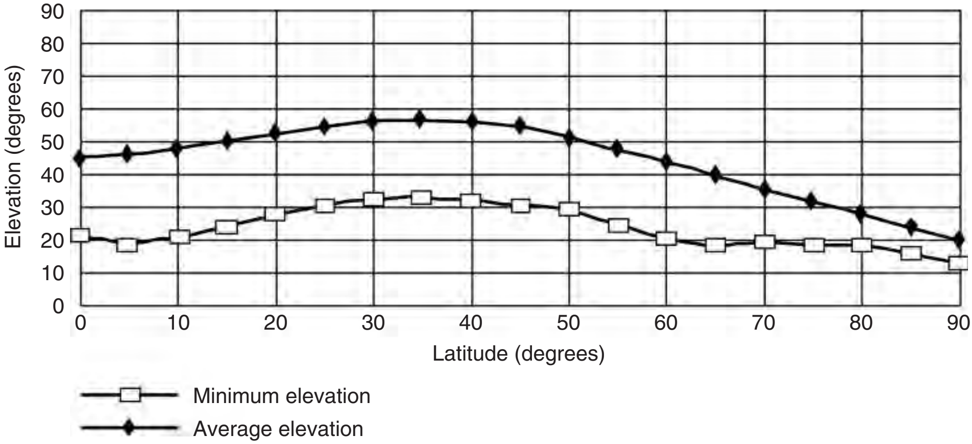


Figure 8.13 Minimum-maximum visibility elevation angle as a function of latitude. (Source: ITU-1, 2010. Reproduced by permission of ITU.)

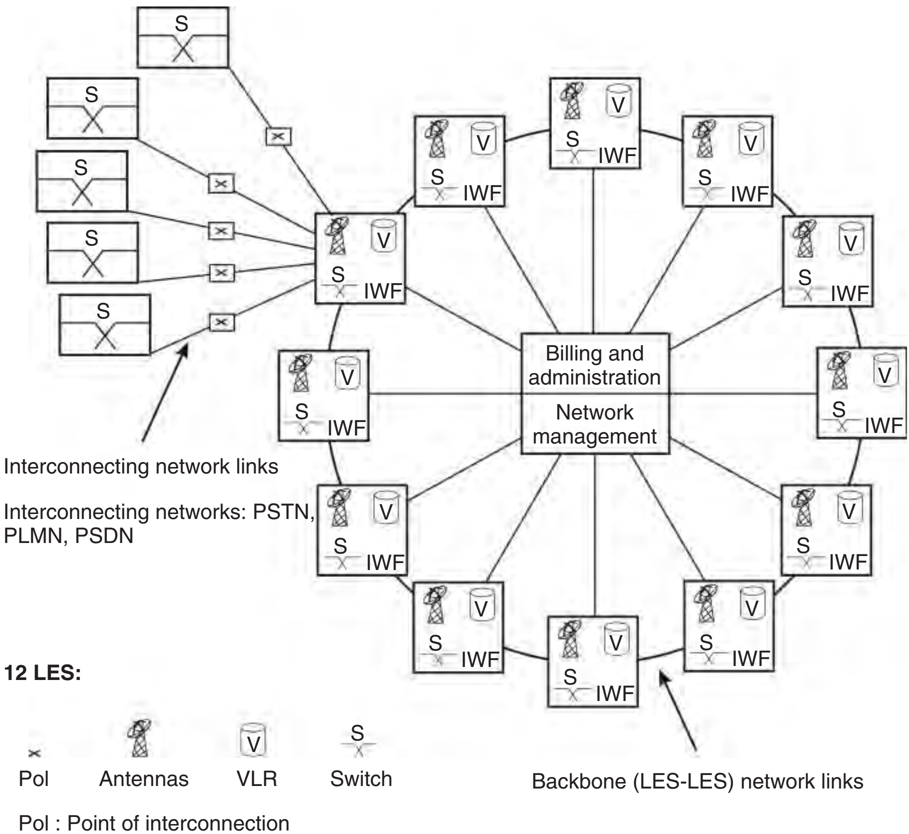


Figure 8.14 Ground network infrastructure of SRI–D. (Source: ITU-1, 2010. Reproduced by permission of ITU.)

mobile networks. The visitor location registers (VLRs) facilitate inter and intra-network movement of users. Each LES is connected to one or more terrestrial network through point of interconnections (POIs) and switches. The mobility support in the satellite component is provided through VLR and an HLR (home location register) allocated to each mobile.

Table 8.25 lists examples of the telecommunication services supported.

The system provisions UT assisted, hard (make before break) and soft handovers (preferred option) for beam–beam and satellite–satellite handovers and supports LES-LES handover. Open-loop Doppler compensation, based on knowledge of satellite ephemeris, is used to limit uncertainty to 1.1 and 40 Hz in the forward and return links, respectively. The system exploits space and time diversity for traffic and signalling and frequency diversity for broadcast and common control channel.

Table 8.26 lists salient characteristics of the UTs. Five classes of UTs are envisaged with maximum capability of up to 38.4 kbps for larger UTs. The hand-held unit supports voice

Table 8.25 Examples of supported services

Service type	Service example	Comments
Teleservices	Telephony, emergency calls, DTMF support	Voice codec rate: 4.8 kbps; voice activation is used for resource conservation in forward and return links
Bearer	Asynchronous and synchronous transparent and non-transparent circuit-switched data up to 38.4 kbps; packet switched data	Implementation of packet switched data is under review; Radio resource allocation depends on user requirements and availability
Supplementary	Forwarding, call waiting, multi-party service, call restriction service, location services, charge advice	–
Messaging	Voice messaging, fax messaging, SMS	–

(Adapted from ITU-1, 2010.)

Table 8.26 Characteristics of user terminals (BER for data = 10^{-5}, BER for voice before error correction = 4%)

Terminal	Service	Bit rate (kbps)	Gain (dBi)	G/T (dB/K)	Peak EIRP (dBW)
Hand-held	Voice, Data	4.8, 2.4–9.6	2	−23.8	≤ 7
Ruggedized transportable	Voice, Data	4.8, 2.4–9.6	3.5	−21.5	≤ 7
Private vehicle	Voice, Data	4.8, 8.0–38.4	3.5	−21.5	≤ 10
Commercial vehicle	Voice, Data	4.8, 8.0–38.4	6.5	−18.0	≤ 10
Semi-fixed	Voice, Data	4.8, 8.0–38.4	10.5	−14.0	≤ 10

(Adapted from ITU-1, 2010.)

and data up to 9.6 kbps. G/T ranges from −23.8 dB/K for the hand-held to −14 dB/K for the semi-fixed UT.

Satellite EIRP is distributed across all the spot beams, on basis of traffic distribution. The maximum available EIRP per beam (assuming all the power is directed into it) varies from 55.4 dBW in the outermost beams to 58.1 dBW in the inner-most beam; the corresponding service link G/Ts are respectively 1.5 and 2.6 dB/K. LESs are time-synchronized to each other; RHCP is used both in the forward and the return service links.

The system uses TDMA and FDMA (frequency division multiple accessing) combination with frequency spacing of 25 kHz. TDMA frames are 40 ms wide, each supporting six time slots of 40/6 ms duration with two guard symbols at either ends. Each slot supports 6 kbps and thus 36 kbps per frame.

The system uses QPSK in the forward service link and GMSK in the return service link for voice and data and BPSK for signalling channel with convolution code, interleaving and soft decoding applied in each case. Voice and data channels use code rates respectively of 1/one-third and one-half and signalling channel code rates vary between one-half, one-fourth, and one-sixth depending on the required robustness.

8.2.2.6 Satellite Radio Interface E

System Description

The SRI-E is based on a constellation of three multi-spot beam (e.g. 300 spot beams) geostationary satellite system (inclination $\leq 3°$) to provide world-wide IMT-2000 compatible multimedia services at bit rates up to 512 kbps to fixed, portable and mobile UTs, including the provision of higher data rates on specialized UTs through carrier aggregation. The primary goal of the interface is to deliver, support and provide interoperability with UMTS type applications using a directional antenna connected to a notebook or laptop computer. The main objective is to provide data services and applications such as voice and web browsing over the Internet and private intranet. The radio interface layer is sufficiently distanced from the services carried such that the satellite system can transport a variety of services and traffic. Although the interface targets 1–3 GHz as the operational frequency band, it is applicable to any frequency band taking into consideration the propagation environment.

Table 8.27 lists features of a geostationary satellite that can meet the target.

The system operates as a packet data system compatible with ATM (Asynchronous Transfer Mode) and IP transport mechanisms to facilitate full access to the Internet. The forward link is shared between users through a number of TDM physical bearers while the return link uses a TDMA allowing connection with individual UTs. The slots are either of 5 or 20 ms durations, provided to the MESs in a return schedule transmitted on a forward bearer. FDD scheme is used with unpaired carriers.

The satellite includes a provision to assign frequencies to a spot beam dynamically in case the radio resources (RRs) become insufficient. The shared mode of forward transmission implies that data rates fluctuate with traffic load. Circuit mode traffic can be carried through the system's QoS mechanism (virtual connections) negotiated at the call set-up. The interface intends to adopt 4 kbps adaptive multi-band excitation (AMBE + 2^{TM}) codec for voice.

Table 8.27 Representative characteristics of a satellite suitable for SRI-E

Number of spot beams per satellite	Up to 300, depending on desired coverage
Configuration of spot beams	Reconfigurable in response to evolving traffic patterns
Spot beam size	Approximately 1° beam width (~ 800 km diameter at the sub-satellite point)
Frequency reuse	7-beam clusters
Service link G/T of satellite beam	Average: 10 dB/K Minimum: 9.5 dB/K
Service link saturation EIRP of each beam	Minimum: 38 dBW Maximum: 53 dBW
Service link total saturation EIRP	67 dBW
Satellite EIRP per RF carrier	43 dBW (Maximum) 42 dBW (Average)

(Adapted from ITU-1, 2010.)

The burstiness and asymmetrical nature of Internet traffic is managed through statistical multiplexing and dynamic resource allocation allowing an ‘always on’ Internet connection. Dynamic resource allocation allows efficient use of RRs by reallocating the RRs of an inactive user until the user becomes active again. The system supports multiple concurrent independent call connections with different requirements and supports position location report dispatch to the network on a signalling channel for scenarios where users’ position reports are essential. Dynamic RR handover allows efficient use of RRs by expanding or compacting RR dynamically. The block error rate of 10^{-3} is considered adequate for multimedia applications but upper layer protocols can tailor higher quality when needed. Both background and interactive modes of PS traffic can operate in an acknowledged mode so that the lost packets are retransmitted; circuit and streaming mode operation is unacknowledged. An adaptive coding arrangement counters the fluctuations of the received signal quality to maintain the desired QoS.

At least one gateway is required for each satellite. Radio network systems (RNSs) of the gateway provide radio connectivity to spot beams. GSM/UMTS core network architecture is used for MM and each spot beam is considered as a ‘location area’. The system supports all types of terrestrial interface transparently and all satellite-specific functions including MM remains hidden from the terrestrial network.

The handovers supported include:

- Beam–beam handover of same beam type serviced by the same radio network controller (RNC).
- Beam–beam handover of the same beam type on the same satellite serviced by a different RNC.
- Beam–beam handover of the same beam type on a different satellite serviced by the same RNC.

Handover is initiated by a radio resource management (RRM) process and conducted by the system’s bearer control layer. After the target bearer control process has been reconfigured, the old connection is dropped. There is some loss of data for streaming traffic such as voice; however, for data an ARQ (automatic repeat request) mechanism recovers the lost data.

The bearer control layer also manages the timing corrections of mobiles so that each mobile transmits on its assigned time slot. MES transmission frequencies are locked to the forward transmissions to ensure correct transmissions. The system does not support space diversity.

To support a variety of UTs and variable link conditions a number of modulation and coding rates are used – 16-QAM (quadrature amplitude modulation) and QPSK in the forward direction and 16-QAM and 4-ary schemes in the return direction. Turbo codes with puncturing provide an efficient way of link-dependent code adaptation. The code rates are chosen by the network such that each rate provides 1 dB change in the carrier to noise density ratio (C/N_o). Return schedules transmitted in the forward broadcast carrier provides the symbol rate and modulation to each MES. The basic spectral efficiency for data is estimated as 2.4 bit/s/Hz, increasing to 3–7 bit/s/Hz due to advantages gained by statistical multiplexing. Moreover, voice activation is estimated to double the spectral efficiency. MES EIRPs range between 10 and 20 dBW. The system does not include any recommendation for Doppler compensation considering receiver's automatic frequency control scheme as adequate. Multipath effects are accounted through provision of link margin and the channel is considered non-dispersive. The data rate supported in the forward link ranges from 21.6 to 512 kbps, adjustable on a burst basis; on the return link the data rate ranges from 19.2 to 512 kbitps.

Table 8.28 lists the forward link frame duration, symbol rate, modulation scheme and FEC blocks per frame of a sample of bearers. The frame duration is 80 ms and FEC block size vary from 80 ms for the lowest symbol rate bearer to 10 ms for the highest symbol rate bearer. The short block size of the 151.2 ksps bearer improves data latency.

Table 8.29 lists the main characteristics of a sample of return bearers. Two burst durations are used – 5 and 20 ms with 16-QAM and $\pi/4$ QPSK modulation scheme; 5 ms burst is used for low latency applications.

Table 8.28 Characteristics of forward bearers. The identifier gives carrier description in shorthand.

Identifier	Frame duration (ms)	Symbol rate (ksps)	Modulation	FEC blocks per frame	Occupancy (kHz)	Primary use
F80T0.25Q1B	80	0.25×33.6	QPSK	1	10.5	Global beam signalling
F80T1X4B	80	33.6	16-QAM	4	42	Signalling and traffic for small user terminals
F80T4.5X8B	80	4.5×33.6	16-QAM	8	189	Primary traffic
F80T1Q4B	80	33.6	QPSK	4	42	Signalling and traffic for small user terminals

(Adapted from ITU-1, 2010.)

Table 8.29 Example characteristics of a sample of return bearers. Carriers with 5-ms bursts carry short messages including signalling whereas 20-ms bursts carry longer messages. $\pi/4$ QPSK modulation is more robust in presence of channel impairments than 16 QAM carriers. Modulation and FEC are variable and are matched to channel conditions and terminal type

Identifier	Burst duration (ms)	Symbol rate (ksps)	Modulation	FEC blocks per burst
R5T1X	5	33.6	16-QAM	1
R5T4.5X	5	4.5×33.6	16-QAM	1
R20T1X	20	33.6	16-QAM	1
R20T4.5X	20	4.5×33.6	16-QAM	2
R5T4.5Q	5	4.5×33.6	$\pi/4$ QPSK	1
R20T1Q	20	33.6	$\pi/4$ QPSK	1
R20T4.5Q	20	4.5×33.6	$\pi/4$ QPSK	1

(Adapted from ITU-1, 2010.)

The available code rates are 0.34, 0.4, 0.5, 0.6, 0.7, 0.8 and 0.84; the applicable code rate is signalled on a burst basis by the unique word, depending on the channel conditions. Code generation involves puncturing of the turbo-coded parity streams using one of a number of pre-defined puncturing matrices described algorithmically. The functions for control encoder and decoder and the channel interleaver matrices are also conveyed algorithmically (rather than in tables) to minimize the probability of implementation errors.

8.2.2.7 Satellite Radio Interface F

System Description

SIR F known as SATCOM 2000 is targeted as a global LEO mobile satellite system to provide telecommunication services at data rates up to 144 kbps to fixed, nomadic, portable, mobile, maritime and aeronautical terminals on a single contact number with roaming capability to terrestrial mobile systems. The satellite constellation consists of 96 regenerative LEO satellites interconnected in space and a ground infrastructure interfaced to IMT-2000, PSTN, PSDN (Public Switched Data Network), PLMN and other terrestrial networks. The system is capable of providing voice and data services, 'including a combination of voice, data, facsimile transfer, Internet access, e-mail, voice-mail, paging and messaging applications'. The system supports symmetric, asymmetric and asynchronous transmissions.

The architecture consists of a satellite radio access network (SRAN), comprising gateways and a regenerative satellite constellation in a meshed configuration connected to the core network through an interface known as I_{us} and to mobiles over an interface called U_{us}. The feeder and intersatellite links are considered internal to SRAN and hence are not specified. The SRAN is responsible for network functions including call set-up, MM, network monitoring/management, RRM, statistics, etc. The system is capable of operating in all the IMT-2000 frequency bands. Two gigahertz was opted as a baseline to specify the RF characteristics.

The satellite constellation consists of 96 satellites in eight 98.6° inclined slightly eccentric orbital planes, each with 12 equi-spaced satellites. The orbital design takes into consideration the combined needs to provide global coverage, minimize overall cost of constellation and minimize complexity of intersatellite geometry. The average altitude close to 860 km is chosen to minimize spacecraft hardware complexity – a lower altitude (than chosen) would require a tighter station-keeping (hence more on-board fuel) and a higher altitude would be subject to larger radiation dose requiring on-board component hardening that would increase the hardware cost. The minimum elevation angle in the service area would be 15°.

The proposed satellite would have 228 reconfigurable, switchable spot beams with a dynamic resource allocation capability. The spot beams would be generated by an array of separate transmit and receive satellite-fixed antennas.

The system comprises five segments: the space segment, SCF/constellation control facility, the fixed ground infrastructure, the user segment and business/customer support segment that includes billing.

The satellite supports a hybrid multiple access architecture of a combined TDMA-CDMA to meet the diverse nature of the anticipated traffic and provide optimum use of satellite resource on the basis of service demands. Both TDD and FDD duplex modes are supported. The CDMA scheme was targeted for the fixed/transportable UTs where signals are relatively stable – a condition that minimizes the 'near-far' problem of long-delay satellite channels. The CDMA scheme provides a good spectral efficiency in this environment. In a mobile environment where signals undergo rapid variations, power control efficiency degrades and hence TDMA provides a better spectral efficiency and service quality.

The voice channels in FDMA/TDMA mode are QPSK-modulated, transmitted at a burst rate of 34.545 kbitps, occupying 27.17 kHz to provide 147 voice channels per MHz. Each TDMA frame supports four bursts and the voice codec incorporates 2/3 rate FEC. The frame duration is 40 ms. The channel coding scheme consists of an RS outer code to contend bursty error concatenated with an inner convolution code for contending random errors. In addition, an interleaving scheme disperses bursty errors. Two types of ARQ schemes are anticipated for delay tolerant, negligible-BER applications such as executable file transfer. These schemes are – selective-to-repeat and go-back-N.

The CDMA scheme operates on 1.25 MHz sub-bands at spreading rates in the range 1.228–4.096 Mbitps. The FEC coding rate is half in the forward link and one third in the return link to provide a basic user rate of 9.6–144 kbps by aggregating multiple channels where necessary. The modulation schemes in the forward and return directions are 16-QAM and QPSK respectively. The CDMA scheme allows frequency reuse in every beam. A power control logical channel is added to each link.

The system supports diversity (space, time, etc.) and intra/inter handover with a capability of operating in a variety of satellite operating environments including urban, rural, fixed-mounted and indoor, as recommended by the ITU (ITU-R M.1034).

The maximum velocity of users for operating hand-held terminals is set to 500 kmph and those of aeronautical terminals to 5000 kmph. Table 8.30 lists suggested RF characteristics of the satellite and hand-held UTs at 2 GHz. Details of other classes of UTs are anticipated to be market driven.

Table 8.30 Representative RF characteristics at 2 GHz

Maximum/average EIRP for hand-held terminal	(−2 to 4) dBW/(−8 to −2) dBW
G/*T* for hand-held terminal	−24.8 dB/K
Antenna gain for hand-held terminal	2 dBi for hand-held
Maximum satellite EIRP/channel	29.6 dBW
Maximum satellite *G*/*T*	0.1 dB/K
Channel bandwidth	TDMA: 27.17 kHz CDMA: 1.25–5 MHz
Multiple channel capability	Yes
Power control	Yes
Doppler compensation	Yes
Maximum fade margins for each service type	Voice: 15–25 dB Messaging/paging: 45 dB

(Adapted from ITU-1, 2010.)

8.2.2.8 Satellite Radio Interface G

System Description

This interface promotes a tight integration of a mobile satellite system with the terrestrial IMT-2000 CDMA standard such that the satellite component can be easily integrated with terrestrial handsets. Such an approach would minimize the cost of dual-mode 3G handsets and encourage market penetration. This is achieved by reusing the terrestrial waveform in a neighbouring MSS frequency band.

The user throughput ranges from 1.2–382 kbps, depending on the terminal type; and there is a provision to support multiple services over a single connection. A number of recent advances in technology were incorporated – such as, capacity and coverage enhancement technologies (e.g. adaptive antennas, advanced receiver architecture, transmitter diversity, etc.), handover provisions for operation with hierarchical spot beam architecture, and handover to other systems such as GSM.

The terrestrial standard called 3GPP Universal Mobile Telecommunication System Terrestrial Radio Access Frequency Division Duplex Wideband-Code Division Multiple Access (3GPP UTRA FDD W-CDMA) forms the genesis of the satellite air interface G, incorporating modifications to manage the differences in the operational environments. The interface uses a modified but compatible version of the terrestrial W-CDMA technology aiming to maximize the economies of scale for dual-mode UTs and network infrastructure, while keeping the technology compatible with the evolution of the terrestrial 3G standard. The W-CDMA technology is well developed and thus offers the potential of a relatively straightforward implementation of hybrid networks through a complementary ground component (CGC). The interface was developed by TC SES, standardised by the ETSI (ETSI-13, 2008) and registered by the ITU-R as IMT-2000 SRI G.

Figure 8.15 illustrates an overview of the system architecture including the standardised interfaces. The space segment comprises a single or multiple satellites, each deploying a single or multiple spot beam in the preferred constellation arrangement (i.e. LEO, MEO, HEO (highly elliptical orbit) and GSO (geostationary satellite orbit)). The ground segment comprises gateways to inter-connect the core network and the remote users. The RF part of

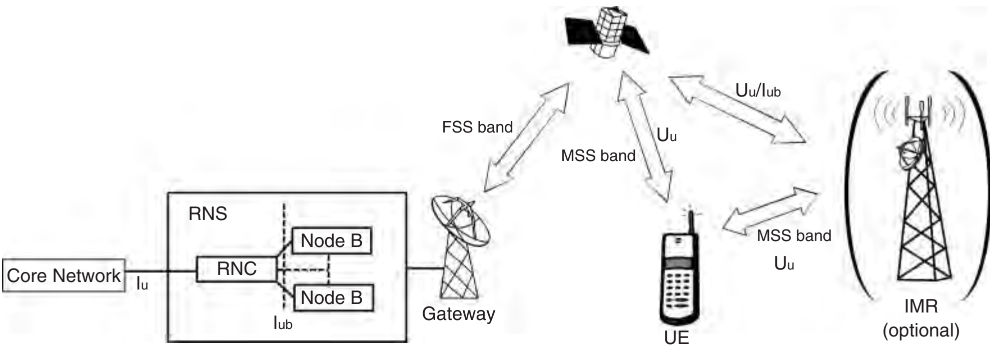


Figure 8.15 An example of extended multi-beam coverage of Europe. (Source: ITU-1, 2010. Reproduced by permission of ITU.)

the gateway comprises RNS including an antenna and a RNC that controls various traffic nodes.

Figure 8.16 shows an example of a multi-satellite/multi-spot beam/multiple visibility coverage of Europe. The network divides the service region into 'location areas' each comprising a single or multiple spot beams. A GEO satellite network can be configured in a multi-beam, extended multi-beam or multi-satellite configuration, in compliance to the size of the service area and its partitioning (by region, on linguistic basis, etc.). The gateways connect with the core network over an interface designated as Iu. Each node in the gateway is connected to the RNC over an interface known as Iub. The service link uses

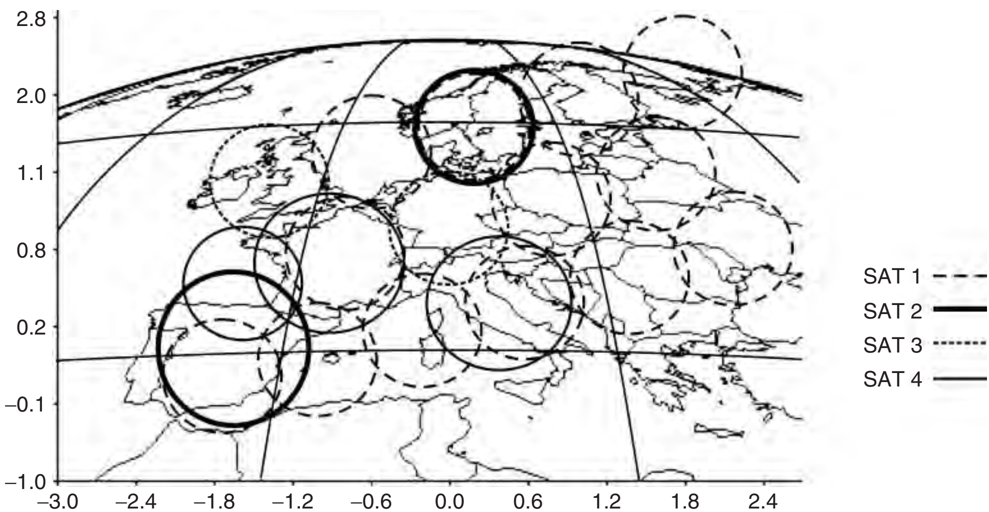


Figure 8.16 Satellite radio interface G system architecture. (Source: ITU-1, 2010. Reproduced by permission of ITU.)

Table 8.31 RF characteristics of user terminals supported by SRI-G UE.

UE type	Data rate (kbps) (Forward/return)	Mobility (km/h)	Maximum transmit power (dBm)	Typical reference antenna gain (dBi)	Maximum EIRP (dBW)	Antenna temperature (°K)	G/T (dB/K)
Handset (three classes)	1.2–12.2/1.2–384	500	24–33	0	−6–3		−33.6
Portable	1.2–384/1.2–384	500	33	2	5	200	−26
Vehicular	1.2–384/1.2–384	Up to 1000	39	4	13	250	−25 dB/K
Transportable	1.2–384/1.2–384	Static	33	14	17	200	−14 dB/K
Aeronautical	1.2–384/1.2–384	5000	33	3	6	–	–

(Adapted from ITU-1, 2010.)

an interface called Uu, which can optionally be connected to a complementary ground terrestrial repeater known as Intermediate Module Repeater (IMR); IMRs are excluded from the specifications.

Services include, basic bearer services (voice and data up to 384 kbps), packet data services (data rates 1.2–384 kbps), teleservices (emergency calls, SMS, facsimile transmission, video telephony, paging, etc.), deep paging service (paging directed to users in disadvantaged positions), multicasting services over MBMS at bit rates from 1.2 kbps to $n \times 384$ kbps ($n = 2$, 3, etc.).

The system uses DS-CDMA as the multiple access scheme with FDD at a chip rate of 3.84 Mchip/s with a carrier spacing of 5 MHz. The system uses variable convolutional code rates of half to one-third and turbo-code at one-third rate and a frame length of 10 ms.

The interface envisages five types of user equipments (UEs). The *3G standardized handset* is a dual-mode handset where frequency has been extended to the MSS band. These sets use omni-directional antennas and are sub-divided into three power classes 1, 2 and 3. The *Portable set* is integrated with a notebook PC interfaced to an external antenna. The *vehicular UE* utilizes roof-mounted antenna. In the transportable units a patch antenna is built on notebook PC that has to be pointed to the satellite for connectivity. The *aeronautical UE* uses antennas mounted on the fuselage with the user electronics housed inside the cockpit/cabin. Table 8.31 lists the salient service features and RF characteristics of each UE.

Handover Scenarios

Beam-Beam Handover

UEs report carrier to noise plus interference [C/(N + I)] of the resident and adjacent beams regularly during a call; the network evaluates the necessity of a handover based on factors such as congestion and selects a suitable beam when a hand-over becomes essential; it begins simultaneous transmissions on the candidate beams, and instructs the UT to receive the new carrier. Finally, the existing connection is dropped when the new carrier is received satisfactorily at the UE. Path diversity cannot be introduced in this type of handover.

Intersatellite Handover

When a new satellite appears in view of a UE's antenna the terminal extracts its scrambling code and after synchronization, reports the signal quality to the network. The network may decide to proceed with the handover based on criteria such as suitability of the new satellite's ephemeris. The handover necessitates an LES handover. The network negotiates routing and resource availability with the 'new' LES and on success, instructs the new LES to transmit. Diversity can be exercised by the network in this type of handover due to the different paths of the satellites prior to the handover.

Inter-Frequency Handover

Inter-frequency handoff is a hard handoff. It can be invoked by a LESs without UE involvement either by the same gateway or inter-gateway during handover.

Satellite Diversity

Satellite diversity is utilized in the system to reduce link margin by augmenting signal quality in situations where the line-of-sight signal is obstructed, and during satellite–satellite handover to enable a smooth transition. The UE can be commanded to a diversity mode when it receives signals from a second source. Since scrambling vectors between satellites differ, the UE must initially acquire the scrambling code of the new satellite. In the return link diversity combination can be initiated at the gateway on receipt of a new signal by combining the diversity signal received on different arms of the rake receiver.

Figure 8.16 illustrates a multi-satellite and multi-spot beam scenario to illustrate a possible space segment configuration amenable to diversity and, as an adjunct, add coverage redundancy for safety-critical applications such as aircraft distress and safety communication. Simulations indicate that the maximum space diversity advantage of several decibels occurs in a non-line-of-sight scenario when signals from both the satellites are obstructed; the advantage is of the order of a decibel when both the satellites are in view; and it is advantageous to switch off the obstructed satellite channel to avoid unnecessary power drainage at the satellite when signals from one of the satellites is obstructed as the obstructed signal provides little diversity advantage.

The system utilizes a number of control and dedicated logical channels to perform air interface functions. Each logical channel is mapped to a corresponding physical channel. A CRC is used for baseband error detection. Channel coding is used as counter-measure against radio channel impairments.

Common logical channels carry network and control information to UEs and receive information from each UE. For example, the broadcast channel (BCH) in each beam broadcasts network and control information to UEs. The PCH carries control information to locate MESs and send paging indicators to support efficient sleep-mode procedures. The Forward Access Channel (FACH) carries user or control information to MES when its location is known. The RACH is used by the UEs to send to the gateway user or control information. The common packet channel (CPCH) is used to carry information from UEs to the network. In addition, there is a dedicated channel for communicating with each active UE.

A number of forward and return physical channels are used for transportation of logical channels. The common pilot channels transmitted at 30 kbps carries a known sequence of bits as a phase reference to aid reception. The primary control physical channel carries the broadcast channel. The secondary control physical channel carries FACH and PCHs. Other physical channels include the synchronization channel, downlink dedicated physical channel, return link RACH, amongst others. The CCPCH is used as a time reference for all the other forward and return channels.

Transmission Format

Forward Link

The forward link transmission process involves CRC attachment to the transport block, coding of each block using convolution or turbo code – the selection being signalled by the upper layer the blocks, rate-matching, interleaving to randomize bursty errors, and framing. The coding robustness depends on protection needs of the channel. For example, BCH, PCH and FACH are convolutionally coded and CPCH and others are turbo coded. Convolutional codes utilize a constraint length of 9 at one-third and half rate; turbo-codes are coded at one-third rate. The framed blocks of 10 ms each are then multiplexed with other compatible transport blocks. Some physical channels (e.g. dedicated physical data channel) include a transport-format combination indicator field encoded using a (32, 10) sub-code of the second order Reed–Muller code to indicate the format of the physical channel. Where needed, the TPC bits are included and encoded by repetition so as to map to the assigned number of TPC bits on the assigned physical channel.

The multiplexed blocks are segmented, interleaved a second time and mapped on to physical channel ready for spreading and modulation.

The spreading comprises channelization and scrambling. The direct sequence spreading used in the scrambling code is periodic with a period of 38 400 chips (equal to frame length of 10 ms). The scrambling code is constructed from two binary m-sequences each derived from specified generator polynomials. The chip rate is 3.84 Mchip/s. QPSK Modulation is used for transmission.

Return Link

The modulation applied for transmission is dual-channel QPSK where the filter used for pulse shaping is a root-raised cosine shaped filter with a roll-off of 0.22. The spreading process comprises two steps – channelization and scrambling. Channelization transforms every data symbol into a number of chips – the increase in the number of bits is called the Spreading Factor. Data symbols on I and Q branches are independently multiplied with an OVSF code. The modulating chip rate at is 3.84 Mchip/s. Next, the resultant signals on the I- and Q-branches are each multiplied by a complex-valued scrambling code.

Power control is used in forward and return directions to resolve the near-far problem, conserve power and minimize self-interference. The open loop power control is used in absence of any feedback information. Procedures for beam selection, random access, open- and closed-loop control, beam selection/diversity reception are available.

8.2.2.9 Satellite Radio Interface H

This interface constitutes the 3G version of GMR-1 already described in Section 8.2.1.3.

8.3 Interactive Mobile Broadband Broadcast Standard

The DVB-S system was developed to deliver digital television. Systems deployed in various parts of the world serve over a 100 million receivers (DVB-3, 2011). To meet the demands of a more efficient delivery and leverage technical evolution, the standard was enhanced to DVB-Second generation satellite (DVB-S2) by using advancements in physical layer technologies. DVB-S2 utilizes more efficient channel coding, modulation and error correction techniques and combined with the most recent video compression technology, offers a commercially viable solution for the transmission of High Definition Television Services (HDTV) than its predecessor.

In response to requests from the VSAT industry, a return channel via satellite (RCS) component was incorporated into the DVB standard to provide interactivity with the remote user to facilitate construction of broadband VSAT networks at affordable costs.

To meet a growing demand of mobile broadband services, a mobility element was next added to DVB-RCS. The system is called DVB-RCS + M to indicate the addition of mobility. This standard provides support for mobile and nomadic terminals and supports terminal-to-terminal (mesh) connectivity.

These widely adapted DVB standards have facilitated remarkable economies of scale and enabled introduction of satellite communication products to the consumer mainstream.

We introduce features of DVB-S2, RCS, and RCS + M standards with an emphasis on mobility. Our interest here is therefore primarily in DVB-S2 enabled RCS + M standard.

8.3.1 DVB-S2/RCS + M

8.3.1.1 Introduction

DVB-S2 is a forward-only broadcast system for fixed installations where the radio propagation is stable (DVB-1, 2012). Similarly, the RCS component deals with fixed installations. The economies of scale for DVB-S2/RCS + M standard were obtained by retaining the structure of the DVB-S2/RCS transmission schemes and building changes over these.

DVB-S2

The DVB-S2 standard, enshrined as EN 302 307 (DVB-2, 2009) enhances the DVB-S standard (EN 300 421), by improving its modulation and coding/error correcting schemes. DVB-S2 utilizes QPSK and 8-PSK modulation for broadcast applications to provide efficient operation in non-linear satellite transponders and supplementing the schemes with 16-APSK and 32-APSK for professional applications such as news gathering, which can accommodate higher C/N required by these spectrally-efficient modulation schemes. The coding comprises concatenation of BCH (Bose–Chaudhuri–Hcquengham) with LDPC (Low Density Parity Check) inner coding to provide capacity close to the theoretical limit. In order to contend propagation impairments due to rain, etc. the system utilizes adaptive coding and modulation (ACM) whereby transmission parameters can be altered frame by frame for each user. There is a provision for backward compatibility to support legacy DVB-S receivers. These enhanced features provide an increase in throughput typically of more than

Table 8.32 Example performance of DVB-S2

Satellite EIRP (dBW)	51	53.7
Modulation; Code rate	QPSK; ¾	8-PSK; 2/3
C/N (dB) in 27 MHz	5.1	7.8
User bit rate (Mbps)	46	58.8
Per cent improvement over DVB-S	36	32
Number of SDTV channels	10 MPEG-2	13 MPEG-2
Number of HDTV channels	2 MPEG-2	3, MPEG-2

(Source: Morello and Mignone, 2004.)

30% over DVB-S. Table 8.32 illustrates typical performance scenarios of DVB-S2 (DVB-4, 2011; Morello and Mignone, 2004).

The variable modulation scheme combined with a large choice of code rates provides spectral efficiencies spanning 0.5–4.5 b/s/Hz and satisfies a variety of users. The system offers continuous coding and modulation (CCM) for constant protection, variable coding and modulation (VCM) for service-based differentiated protection, and ACM to contend dynamic fading. The use of pilot symbols provides a robust carrier recovery scheme. The system supports different transport mechanisms based on Transport Stream (TS) or Generic Stream (GS) (explained later).

RCS Standard

The DVB-RCS standard released by ETSI in 2000 applies to interactive TV and VSAT networks arranged in a star configuration (DVB-5, 2011).

The 2004 revision of this standard integrates the DVB-S2 forward link transmission to the RCS standard. The 2008 revision enhances the system to support mobile and nomadic terminals by including countermeasures to contend mobile channel impairments and include mobility management.

DVB-RCS systems are in wide use and have been deployed in various frequency bands – K_u, K_a, C including the extra high frequency (EHF) band.

Several features were introduced to meet the inherently different radio and network environments of a mobile system. These features include live handovers between spot beams, a spread-spectrum wrapper in both the directions to comply regulatory constraints with regards to inter-system interference and a quasi-continuous carrier transmission feature for shared terminals to benefit from user-traffic aggregation. Features such as link-layer FEC were introduced to counter breaks in radio connectivity due to shadowing and multipath.

Figure 8.17 (DVB-6, 2009) depicts the basic RCS system model broadly divided into a satellite network dependent and independent segments. The service provider(s), as depicted, communicate with the remote communication satellite terminal (RCST – a term used in this standard). The applications and the content are network independent while the transport segment consisting of packetization and physical transfer constitutes the network segment. The broadcast signals are received at the RCST in the broadcast interface module and transferred to the set-top unit. The user interacts with the network on interactive channels and an interaction network.

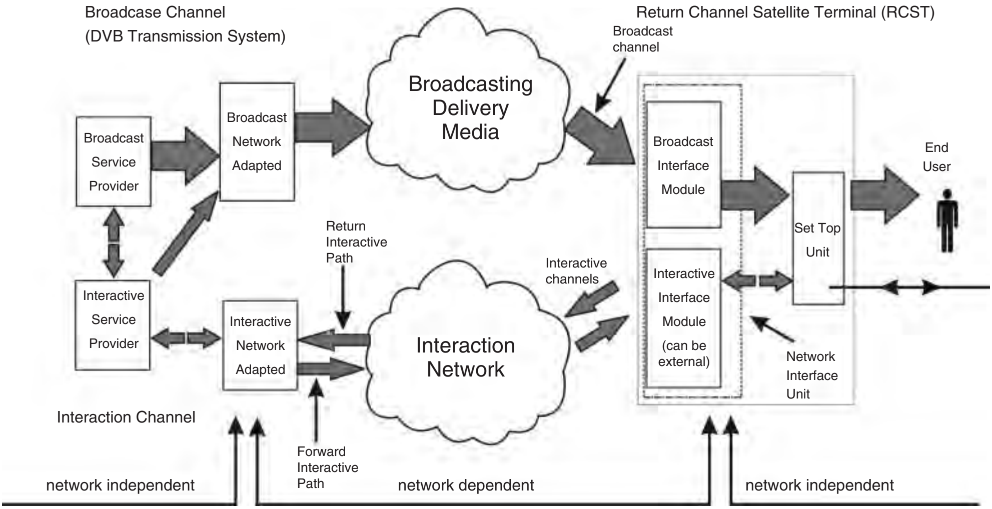


Figure 8.17 DVB RCS system model (DVB-6, 2009). (© European Telecommunications Standard Institute 2009. Further use, modification, copy and/or distribution are strictly prohibited. ETSI standards are available from http://pda.etsi.org/.)

Since the launch of the RCS standard in 2001 the physical layer technology has evolved and the IP platform has become more stable and its use widespread. Therefore, a second generation RCS system known as RCS-2 has been developed (DVB-1, 2012). The changes have not dealt with mobile option as such with provision to address it in future

8.3.1.2 RCS + M Standard

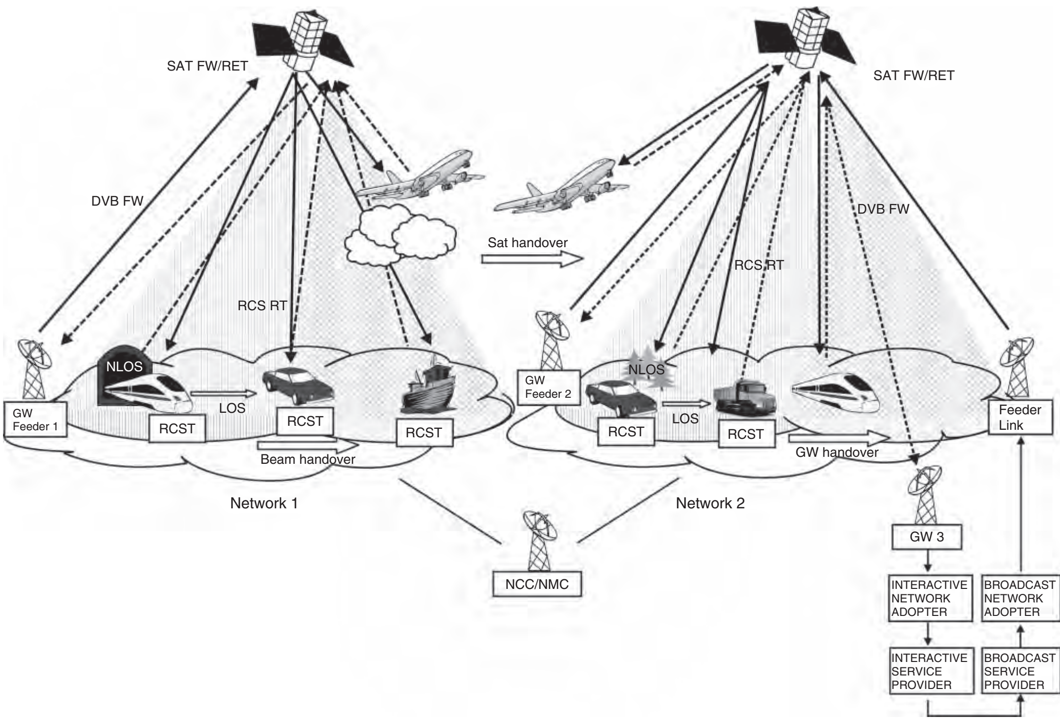
Figure 8.18 (DVB-6, 2009) illustrates the reference model used in DVB-S2/RCS + M standard.

It comprises

- A space segment of multi-spot beam geostationary satellite;
- DVB-RCS + M user terminals mounted on ships, land vehicles and ships, which service UEs such as PC over a local area network;
- A fixed ground segment comprising hubs/gateways interfaced to the terrestrial core network;
- A network operations centre (NOC) to manage the network including a SCF.

Service can be regional or global covering vast expanse of the size of a continent.

The NCC provides network control, timing and monitoring functions. Traffic *gateways* (GW) receive and route the RCST return signals; provide accounting functions and user interactivity with the interactive service provider; and broadcast the forward channel over DVB-S or DVB-S2 uplink containing multiplexed user data, network control and timing signals. The system supports a variety of shared RCSTs in land, maritime and aeronautical captive environments such as busses, aircrafts and railways.



re 8.18 Reference model of RCS + M (DVB-6, 2009). (© European Telecommunications Standard Institute 2009. Further use, modif and/or distribution are strictly prohibited. ETSI standards are available from http://pda.etsi.org/.)

The DVB-RCS + M radio interface comprises an enhanced DVB-S2 forward link that incorporates spectrum spreading over the standard DVB-S2 bearer and a Link Layer FEC feature, and the return link enhancement includes spectrum spreading, burst repetition and a continuous carrier mode for traffic aggregation in shared mobile UTs.

The standard includes handover procedures to maintain continuity of service in mobiles during a call. Handovers are supported between spot beams, satellites and gateways including the possibility of a handover to a terrestrial gap-filler system in non-LOS situations such as tunnels, railway stations, cities with dense buildings, etc.

Other enhanced features include off-axis EIRP control and UT exclusion zones around specific areas to avoid interference-prone areas.

To counter propagation impairments and uncertainties in frequency and time synchronization robust techniques are available for system access, log-on procedures and rapid receiver re-synchronization after signal interruptions.

The standard proposes a proactive physical layer retransmission technique in the return link to overcome the unreliability (packet loss) caused by short breaks in communication – the underlying premises being that the forward and return signal fades are correlated. Thus the UT predicts a fading event in the return link by measuring the S/(N + I) ratio of the forward carrier and withholds transmission of the return carrier until the fading event is over. It is presumed that a safety margin would be applied in making the decision as the short-term correlation in multipath situation can be weak and the forward carrier can be reacquired quickly enough and therefore not requiring the return-withhold countermeasure. The proactive retransmission scheme can also offer the advantage of avoiding a reduction in throughput due to IP protocol recovery mechanisms such as the transmission control protocol (TCP).

IP encapsulation process over ATM cell can cause fragmentation of the packet with the possibility that a part of the IP packet may be corrupted during transmission. To facilitate transmission of IP packet over ATM/AAL5 IP, the retransmission strategy ensures that each corrupt IP packet is retransmitted in entirety since the ATM transport mechanism does not support packet reassembly in case of partial retransmission.

The following sections discuss the implications of mobility on the RCS standard in detail and outline the countermeasures techniques mentioned earlier as specified in the standard. The system is mainly targeted towards K_u and K_a satellite frequency bands as they are the most likely candidates for broadband mobile services in the foreseeable future. The following enhancements in the RCS standard for the mobile option are discussed:

- Spectrum spreading in forward direction.
- Log-on requirements in presence of large timing uncertainty due to variability in locations of mobile RCST.
- MM.
- Interference avoidance.
- Continuous carrier operation for traffic aggregation in shared mobile RCSTs.

Considerations

It is the basis that the mobile broadband system is likely to benefit travellers in a shared mobile environment primarily – that is aeronautical, maritime, railway and bus. Other users would include a limited set of nomadic users and individuals travelling on land vehicles. A mobile environment entails RF propagation impairments that adversely impact the conventional DVB-S2 (and DVB-S) forward and RCS return signal quality. As the allocations in the L and S MSS frequency bands are limited and congested, it is necessary to utilize K_u or K_a bands where adequate spectrum is available. MSS allocations in the K_u band have a secondary status whereas 1200 MHz of primary allocation is available worldwide in the K_a band (20.1–21.3 GHz).

Regulatory compliance necessitates adherence to transmission power spectral density limits; and additionally interference tolerance from the transmissions of the primary and shared satellite services is necessary. Antenna size and profile of mobile/nomadic UTs are restricted in order to promote mobility and thus the antenna side lobe levels are higher than the conventional VSAT UTs causing high off-axis transmissions levels. Thus specific provisions are necessary for mobility to ensure regulatory compliance, particularly for the Ku band.

Furthermore, considering the inherent shadowing and multipath effects on mobiles, the conventional DVB/RCS radio link design necessitates enhancements to meet the QoS. For example, the user should not get logged off when the terminal gets shadowed or suffer unacceptable loss of quality during multi-path events. The requirement of high power per bit to counter the mobile environment in the forward link increases the probability of interference into adjacent region and off-axis EIRP transmissions caused by low side lobe discrimination in the return direction can cause interference to adjacent satellite systems. Thus, the radio link design and UT antenna performance exert a strong influence on DVB/RCS + M performance.

Forward Link Power Spectral Density

Studies conducted in the DVB technical forum estimate that the minimum antenna aperture for service provision in the regional and global beams respectively would be 30 and 50 cm with an unmodified DVB-RCS UT and such antenna sizes would restrict mobility. It is further estimated that a spreading of the DVB carriers by a factor of 4 provides the necessary reduction in the transmit power density in the forward direction, whereas the spreading factor (SF) should be at least 16 in the return direction to accommodate UT antenna diameters of 30 cm for the global beam (Morlet *et al.*, 2007/2008). The increased bandwidth would have to be traded off with the user data rate.

Accordingly, the RCS + M standard utilizes channel spreading as a countermeasure towards interference mitigation in both forward and return directions.

Forward Link Spreading Arrangement

The forward link transmission is processed in two stages – spreading and scrambling. Spreading is achieved by multiplying each incoming (I + jQ) symbol by a sequence of chips. The SF is given as the number of chips per symbol, that is when $SF = 1$ the signal is a conventional DVB signal. The spread signal is then scrambled spreading the energy across the transmission band to avoid spectral spikes. The concept is illustrated in Figure 8.19.

Consider an (I + jQ) symbol sequence of a physical layer frame PLFRAME:

$$\{d[k]\}, k = 0, 1, \ldots, NPLFRAME - 1 \tag{8.1}$$

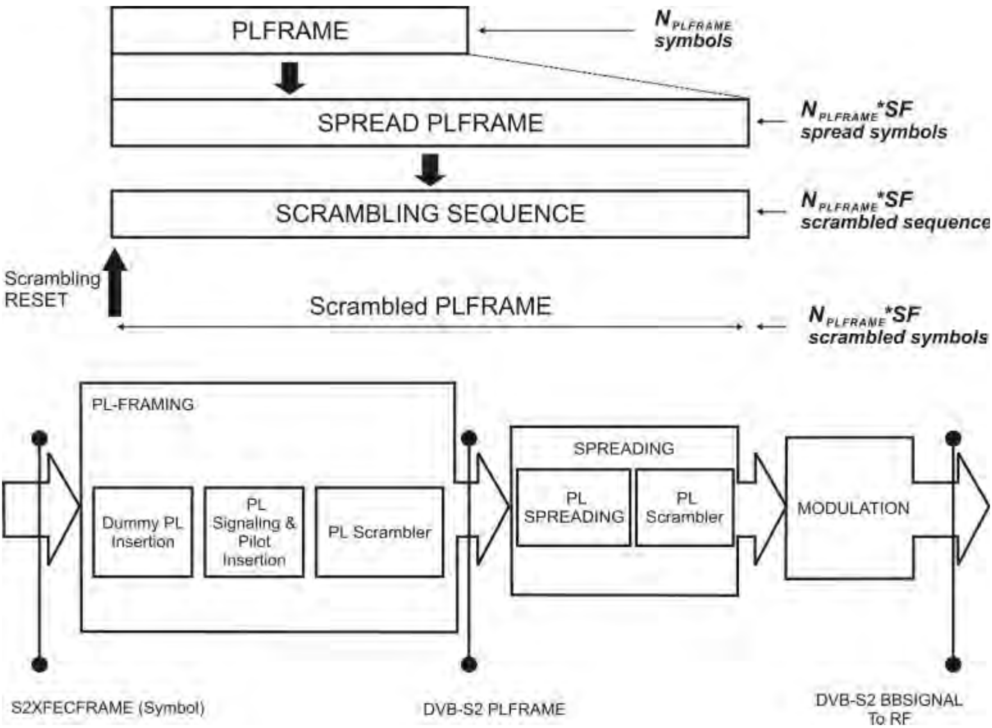


Figure 8.19 Forward link spreading sequence (© European Telecommunications Standard Institute 2009. Further use, modification, copy and/or distribution are strictly prohibitated. ETSI standards are available from http://pda.etsi.org/.)

where *NPLFRAME* = number of symbols in one PLFRAME. The spreading operation yields the spread sequence $s(k)$ given as:

$$\{s[k]\}, k = 0, 1, \ldots, (NPLFRAME \times SF) - 1 \tag{8.2}$$

$$\mathrm{S(i) = d(\lfloor i/SF \rfloor)C(mod(i, SF)\ for\ i = 0, 1, \ldots, (N_{PLFRAME} \times SF) - 1} \tag{8.3}$$

Where SF = spreading factor.

The standard defines spreading codes $C(i)$ for SF = 1, 2, 3 and 4. The return code type is signalled to the mobile in the forward link.

The second operation, scrambling, is applied to the spread signal; the resultant signal is square root raised cosine filtered and passed onwards for further processing.

Countermeasures for Non-line of Sight Propagation Impairment

In Ku/Ka band mobile environment shadowing, multipath and channel dispersion under certain conditions cause severe degradation to the received signals beyond the impairments caused by the Earth's atmosphere.

The situation is quite severe in a land mobile environment. For example, studies conducted in Italy for a railway environment in K_u band indicate 2–3 dB periodic attenuation due to electrical cables and 15–20 dB due to electrical trellises around the railway tracks, superimposed on environment dependent statistical fluctuations (Scalise, Schena and Ceprani, 2004) (see also Section 3.3). The coding at the physical layer alone is inadequate in such a fading environment. The concept of upper-layer coding has been used in DVB/RCS + M to improve performance in a fading environment such that the DVB-RCS transmission structure remains intact. The standard does not support ARQ schemes but these can be implemented above the standardized components with the provision of buffering at both ends. A more robust solution optionally supported is the use of terrestrial retransmitters that could cover prolonged shadowing such as inside a tunnel.

Demodulator synchronization/resynchronization can be severely tested in presence of Doppler fluctuations; a receiver can lose lock during a shadowing or multipath event and hence requires rapid synchronization following such an event. In the RCS + M system an interleaver is used to disperse bursty errors to counteract short-term blockages and multipath fades. Another countermeasure is to use a frame structure containing a large number of pilot symbols. Solutions to minimize Doppler effects include Doppler pre-compensation at the transmitter and the use of enhanced frequency acquisition algorithm and capture range.

Link Layer Forward Error Correction (LL-FEC) allows the use of the DVB-RCS FEC while applying FEC at an upper layer. The LL-FEC is only directed to those RCSTs that have signalled ability to support the countermeasure in the common signal channel burst.

The technique can also be applied in the return direction by those terminals that opt for continuous return link carrier transmissions.

The use of LL-FEC is defined separately for each elementary stream in the TS. LL-FEC carried over Generic Stream Encapsulation (GSE) is defined separately from that over Multi-Protocol Encapsulation (MPE). LL-FEC use the Raptor code for LL-FEC frame application data table (ADT) sizes up to 12 Mbytes or the MPE-FEC Reed–Solomon code ADT sizes up to 191 Kbytes. The code selection is signalled in the forward link.

Figure 8.20 shows a LL-FEC frame. It consists of ADT and FEC Data Table (FDT). The ADT contains layer-3 datagram such as IP packets and the FDT comprises the corresponding parity bits depending on the code selected based on factors such as required level of protection, bit rate, and so on. The size of the frame is variable on a frame by frame basis.

A descriptor for LL-FEC specifies the elementary streams and the GSE-FEC streams that are applying LL-FEC, together with definition of LL-FEC – for example type of code, frame size.

Spectrum Spreading in Return Direction

The return link MF-TDMA transmission is spread in bandwidth in order to be compliant to the regulatory limits on transmitted power spectral density, particularly when using the secondary allocations. The provision is also available for continuous-carrier return link transmissions, although the process is different to that of the MF-TDMA mode.

MF-TDMA

Two methods for transmission of the return link bursts are available – $\pi/2$-BPSK modulation, equivalent to spreading a QPSK modulated signal by a factor 2, and burst repetition so that the effective bandwidth of the signal corresponds to that of the $\pi/2$-BPSK signal. Each

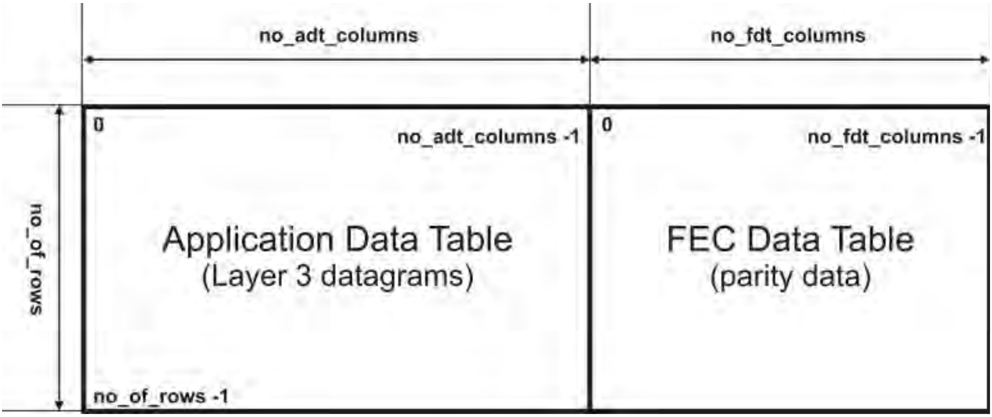


Figure 8.20 LL-FEC frame (DVB-6, 2009). (© European Telecommunications Standard Institute 2009. Further use, modification, copy and/or distribution are strictly prohibited. ETSI standards are available from http://pda.etsi.org/.)

burst is repeated F times within the duration of the burst, which results in a symbol rate F times the unrepeated burst.

Logging in Presence of Large Timing Uncertainty

The mobiles are dispersed throughout the coverage area causing time dispersion in the received bursts. The burst timings of all the mobiles must be aligned for the bursts to arrive without overlaps. At the time of initial log-on the system timing is unknown to an RCST. The RCST then searches the largest continuous number of available slots and to minimize the probability of a collision, transmits at the centre of these slots if the number is uneven and in either one of the two centre slots when the number of available channels is even. Thereafter, the network advises the RCST of the reception time of the received signal in response so that the mobile corrects its timing.

Mobility Management

Figure 8.21 shows various types of handover supported in DVB RCS + M system. A *gateway to gateway handover*, HO 1, occurs when a mobile crosses the area covered by gateway 1 and moves to the area covered by gateway 2; a *satellite-satellite handover* (HO 2) occurs when a mobile crosses from the area covered by gateway 2 operating via satellite 1 to the area covered by gateway 3 operating through satellite 2; a *spot beam to spot beam* handover occurs when a mobile moves from spot beam 1 of satellite 2 to spot beam n of the same satellite; a *satellite to a gap-filler system* handover occurs when the mobile moves into an area where a prolonged shadowing is known to occur.

DVB-S2/RCS + M adopts a distributed handover management scheme wherein the mobiles initiate the process, the controlling NCC evaluates the request and, if appropriate, proceeds with execution under its control.

The beam handover is performed by the NCC at a lower protocol layer. The gateway and satellite handover require involvement of upper layers, such as Mobile Internet Protocol

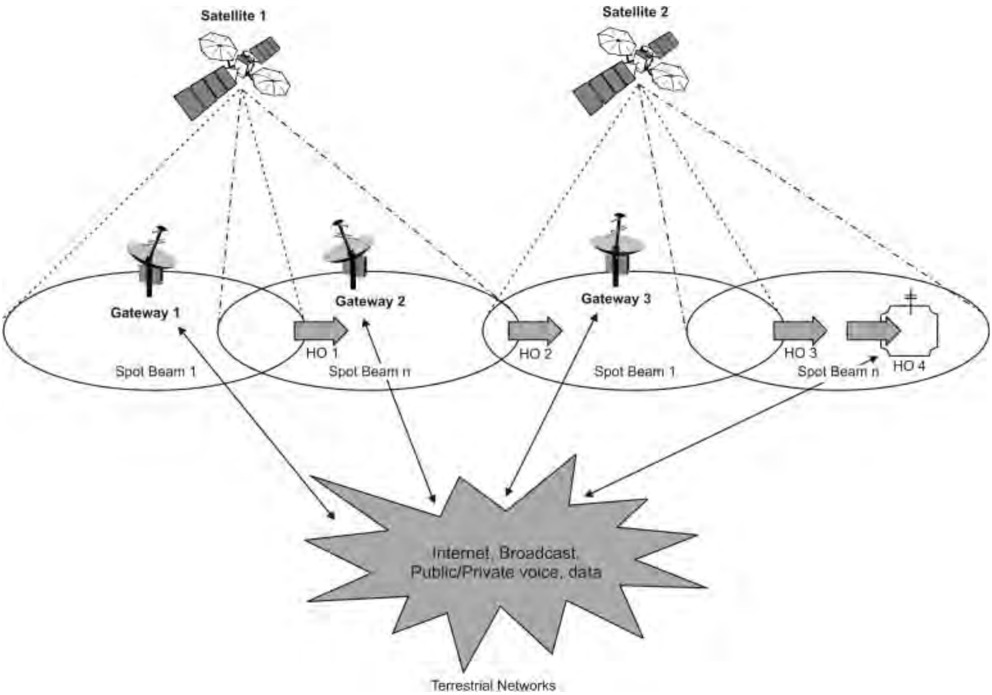


Figure 8.21 RCS + M handovers

(MIP) and Session Initiation Protocol (SIP). The reader will recall that MIP is a Layer 3 solution that requires upgrade of routers at gateways, and edge routers serving as home agents also need modifications. SIP implements mobility at the application layer. Other schemes such as the one based on SNMP (Simple Network Management Protocol) have also been proposed (Alamanac *et al*., 2010).

Beam–Beam Handover

Each mobile stores 'beam profile' maps of its present area; when it detects that it is moving into a new beam it alerts the network requesting a forward, return or both beam handover. Alternatively, the network decides that a handover is necessary when a mobile provides its location.

The NCC checks the resource availability in each of the candidate beam. When a target beam has been found, a handover command is issued in a forward signalling Terminal Information Message (TIM). The mobile is synchronized to the new beam with the new composition tables through the current beam in order to reduce the handover execution time.

The NCC redirects the forward traffic and signalling to the target beam and updates mobile's status. The handover includes the new DVB general Service Information (SI) table and carrier synchronization. On receiving a handover TIM the mobile re-tunes to the new beam and starts to use the return link resources to complete the handover.

Satellite–Satellite Handover

It is understood that a beam-beam handover is a prerequisite during a satellite handover. The handover includes functionalities to support transfer of the connection to a new satellite including the possibility of a network handover when the new satellite serves a different network. In case of a network handover the activity is coordinated by a NOC that oversees the overall network operations. Note that here NCC refers to the entity that is responsible for intra-network operations. Thus, the NCC manages the satellite handover for intra-system satellite handover and the NOC's participation is essential for an inter-network handover. The handover process is similar to gateway handover (assuming that a different gateway services the new satellite), when the candidate beam is on a satellite belonging to the same network. For a network handover the NCC enquires from the NOC the network status of the candidate network and, if favourable whether it is able to accept the mobile into its network. If the response is positive and when the target network is ready to accept the handover the NCC sends the handover command to the mobile. On receiving the command, the mobile synchronizes to the new beam and logs to the new network. The new NCC re-routes the forward traffic to the mobile through a new gateway and the mobile begins sending the return traffic to the new gateway. There is the provision of invoking a handover independently in each link direction.

Gateway Handover

A gateway handover is necessary when the target beam is served by a different gateway. It is assumed that a gateway handover is associated with a beam handover and thus the beam-beam and gateway handover procedures are closely knitted. On a handover request from a mobile the NCC selects a target beam (in forward, return or both directions, as required) and checks whether the chosen beam is associated to a new gateway. When a gateway handover is necessary the NCC communicates with the target gateway and instructs it to prepare for a handover. The NCC then sends a gateway handover command to the mobile. The mobile retunes to the new beam and begins to receive traffic and signalling in the new beam from the new gateway. The mobile begins to send the return traffic and signalling on the newly established link thus completing the handover.

Satellite System to Complementary Terrestrial Component Handover

In areas of known heavy shadowing such as railway or road tunnels and cities with dense building cover the satellite link becomes unreliable or unachievable. Terrestrial re-transmitters (called gap-fillers) are considered to be viable in maintaining the coverage continuity in these situations. Handovers applicable to two types of gap-fillers are proposed – *integral gap-filler* and *terrestrial gap-filler* (Alamanac *et al*., 2010).

Figure 8.22 shows the concept of the integral gap-filler (adapted from Alamanac *et al*., 2010). An integral gap filler consists of a fixed transparent terrestrial relay placed in a location of direct satellite visibility, which takes over the satellite links from blocked mobiles. These mobiles remain connected to the repeater via a local radio link that is linked to the satellite on the same satellite frequencies as before. Thus the satellite RR assignment remains unchanged. At the end of the blockage the repeater hands over the communication to the mobile. If the mobile has moved to a new beam then a beam-beam handover is invoked. The

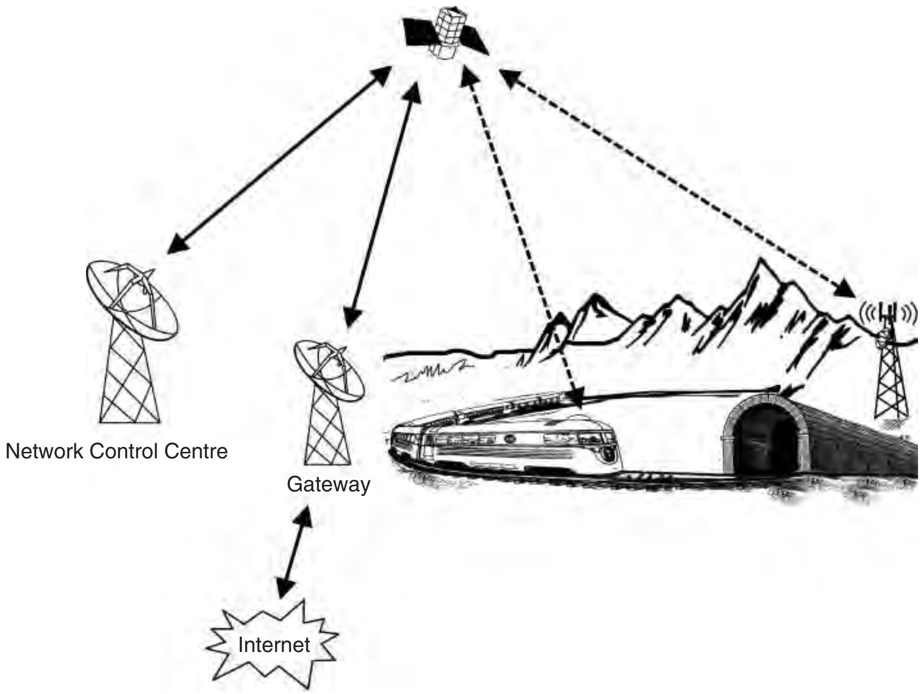


Figure 8.22 An integral gap-filler architecture

mobile should be aware of the presence of the prolonged shadowing area and accordingly search the signals of the terrestrial transmitter and when acquired, hand over the satellite link after retuning to the local transmitter frequency. The entire procedure is managed by the terrestrial gap-filler.

A terrestrial gap-filler can be used by mobiles disadvantaged locations in situations where it is impractical to install an integral gap-filler and rather a terrestrial system such as a cellular system or another wireless network can be accessed by the mobile. This inter-system handover may occur, for example when a vehicle enters a dense city or when a train enters a railway station. Due to the differences in air interface technologies between the networks the mobile should support a multi-mode capability. The handover requires a dedicated network protocol between the terrestrial and satellite networks in addition to the mutual support of the on-going communication protocols to ensure a faultless handover (DVB-7, 2009).

Interference Avoidance in Mobiles

A number of interference avoidance mechanisms dealing with aeronautical, maritime and land mobiles have been standardized by ETSI in K_u and K_a bands as a part of terminals' control and monitoring functions to comply European Commission's (EC's) radio equipment and telecommunications terminal equipment (R & TTE) conformity directive. For example, ETSI EN 302 340 deals with satellite earth stations on board Vessels operating in the 11/12/14 GHz frequency bands allocated to the Fixed Satellite Service (FSS). Article 3.2 of the EC's directive concerns the use of spectrum.

The control and monitoring functions of the mobile terminals dealing in interference avoidance address:

- off-axis EIRP emission density into adjacent FSS satellites;
- power flux density of received interference signal at specified terrestrial stations;
- fault conditions.

The NCC is assigned the responsibility to control the level of *off-axis EIRP* emission density to comply the ETSI regulatory requirements by adjusting the corresponding on-axis EIRP density level.

Further, the NCC also controls the mobile terminal EIRP density to ensure that the *received power flux density* at the earth stations of other services does not exceed the relevant regulatory limits.

The RCST has the task to predict when such interference may occur in the immediate future and send an ‘exclusion zone’ (a prohibited area in terms of interference compliance) entry form at the earliest opportunity. The mobile may request a remedial action from the NCC to avoid harmful interference such as:

- Log off.
- Change frequency to a non-interfering band.
- Change transmission parameters to ensure compliance.

To minimize the probability of interference caused by equipment malfunction the NCC is designated the task to monitor transmission to detect faults that result in harmful interference and take action. The general requirements of such fault conditions have been specified in ETSI specifications of MES.

Continuous Carrier Operation in Return Link

The RCS + M standard addresses broadband services in a shared environment such as ships or aircrafts. Simulations indicate that continuous transmission instead of a demand assigned mode (MF-TDMA) could be more efficient from terminals that produce net traffic so as to appear quasi-continuous. Thus, there is a provision for this class of RCST to support a continuous transmission mode. The RCST signals its ability to support this mode in the common signalling burst. An RCST that supports continuous carrier operation can operate either in a continuous carrier or an MF-TDMA signal, but not in both the modes simultaneously; however enhanced continuous carrier operation allows simultaneous transmissions of both carrier types. The return continuous carrier can be transmitted using $\pi/2$-BPSK modulation.

Carrier Spreading for Return Link Continuous Carrier

Continuous carriers are spread to mitigate interference to adjacent satellite systems. The baseline transmission scheme is that of DVB-S2/RCS so that the carrier can use constant code modulation (CCM) or adaptable code modulation (ACM) modes. The characteristics of the return carrier can be signalled by the RCST or managed by the NCC. The corresponding forward carrier can be transmitted as a standard DVB carrier with or without spreading.

The process of spreading is similar to that of the forward direction, as described. It involves a spreading process, followed by scrambling of the spread carrier. Each (I + jQ) symbol is multiplied by a sequence of chips to spread the bandwidth (see Figure 8.19). The SF is signalled in the time slot composition table transmitted in the forward link. In the scrambling operation, a scrambling code redistributes the bits to spread the spectral density and ensure presence of sufficient number of transitions to facilitate synchronization at the receiver. The scrambling code can either be signalled by the RCST or a default code can be used.

The system defines other features of continuous carrier mode including stream format, return link signalling and procedures for carrier assignment and release. Implementation aspects of the RCS + M standard have been detailed in DVB-7 (2009).

Note 1 DVB Generic Stream Encapsulation (DVB-GSE)

DVB-GSE is an ETSI standard (TS 102 606) that allows IP-based content transport in its native form on DVB physical layers to provide an efficient alternative to the traditional Motion Picture Expert Group Transport Stream (MPEG-TS) to promote convergence of the DVB system with IP-enabled applications (DVB-8, 2011).

The MPEG-TS transport allows a means of encapsulating IP datagrams on MPEG-TS packets for DVB transmissions, but the process incurs overheads.

GSE supports encapsulation of multiple protocols, for example IPv4, IPv6, MPEG, ATM, Ethernet, etc. and several addressing modes. If necessary IP packets can be fragmented transparently within the GSE for mapping to the baseband frame in preparation for the physical layer transmission, that is the fragmentation resides within the GSE. Furthermore, the hardware implementation complexity is low. Figure 8.23 demonstrates the DVB-GSE encapsulation process.

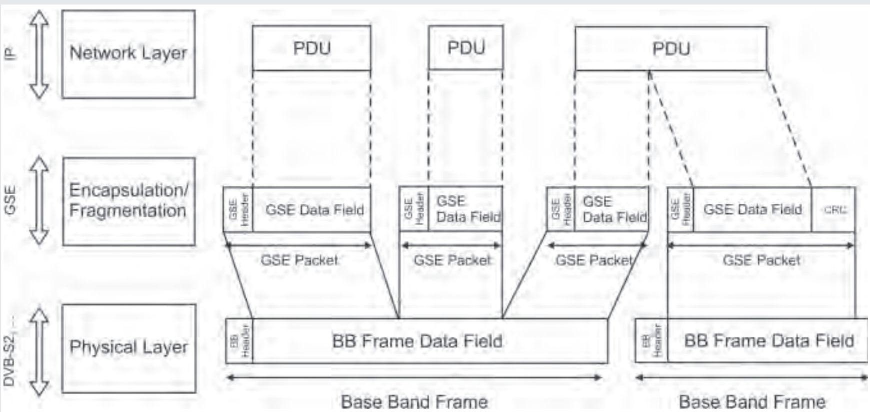


Figure 8.23 DVB-GSE encapsulation process. (Source: DVB-8, 2011. Reproduced by permission of the DVB Project.)

The network layer PDUs are encapsulated in GSE packet adding control information such as type of packet protocol, address, and so on to each packet and if necessary fragmenting the incoming PDU. A CRC can be included for error check, if needed. The reassembly of the IP packets at the receiver is performed within the GSE layer.

Fixed Broadband Standard

For completeness we mention ETSI's (fixed) broadband satellite multimedia (BSM) standard (ETSI-12, 2007) developed with the intention of providing 'high performance, quick to set up, competitive' multimedia and other broadband alternatives including Internet and those not yet defined, to wire-based access systems at comparable QoS targeting geographical areas where deployment of terrestrial systems is uneconomic. The BSM systems can inter-work with several core networks – MPLS (MultiProtocol Label Switching), ATM or IP and others, which may not have been defined yet. The systems operate in FSS frequency bands and provide an efficient use of satellite network resources and facilitate the provision of simple and low cost UT. Figure 8.24 shows the protocol structure of the standard and Table 8.33 summarizes the functions of each layer.

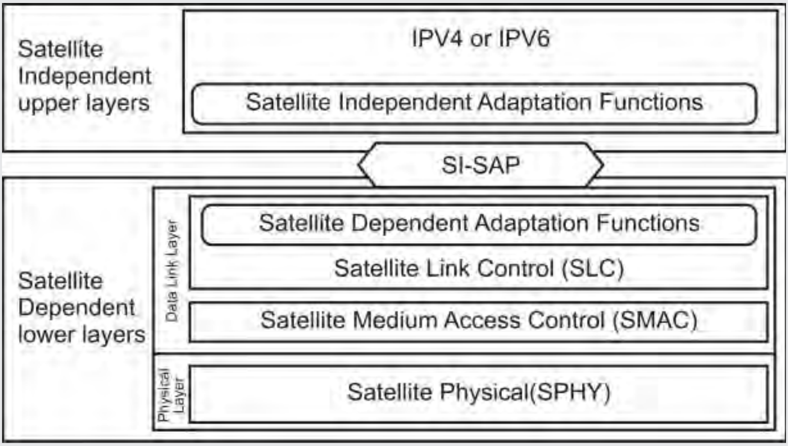


Figure 8.24 Representation of the BSM protocol architecture (see Table 8.33 for a description). (Source: ETSI-12, 2007. Reproduced by permission of ETSI.)

Table 8.33 Layers of the BSM protocol architecture (see Figure 8.25)

Layer	Description	Comments
1	PHYSICAL LAYER Satellite Physical layer	Satellite dependent layer conforms to ETSI harmonized standards
2	DATA LINK LAYER Satellite Link Control (SLC) and Satellite Medium Access Control (SMAC)	Satellite dependent layer – SLC and SMAC layers may be combined or remain separated.
3+	Higher layer, for example IP	Satellite independent layers

(Adapted from ETSI, 12, 2007.)

One of the strengths of the BSM protocol is to facilitate interworking of satellite systems with a variety of broadband core network by demarcation of satellite dependent and independent parts at the *Satellite Independent Service Access Point* (SI-SAP).

Based on the use of SI-SAP a standard known as Internet protocol over satellite (IPOS) has been developed. It was initially approved by various standards organizations in 2005, that is the Telecommunications Industry Association (TIA) in North America, and ETSI and ITU. In 2006, ETSI approved the IPoS.v2 air interface standard, which incorporates the DVB-S2 standard (see ETSI document BSM 1503 and TIA standard TIA-1008).

The IPoS system was developed to deliver IP services at residents and small offices. In addition to (primarily) broadband Internet access the system can provide IP Multicast services. A number of differences exist between IPoS and DVB-RCS systems. For example, the DVB-RCS standard does not follow the SI-SAP specifications implying that DVB-RCS upper layers need not work with other (than DVB-RCS) types satellite systems. DVB-RCS operates with QPSK in the return link whereas IPoS uses OQPSK – the difference has implications on the UT transmit operating point. There is also a difference in the size of the return burst – DVB-RCS operates on 1, 2 or 4 ATM cells (each 53 bytes long) whereas IPoS allows bursts that vary in multiples of 7, 8 or 9 bytes up to lengths that are considerably longer than 4 ATM cells. These differences are advantageous for IPoS compared to the RCS.

Revision

1. Describe a generic protocol architecture that can be used in standardisation of MSS radio interface. What are the advantages in using such an approach?
2. Describe the evolution of the GMR standard, mentioning the incremental advances in radio interface technology at each stage.
3. Compare the lower layer technologies of the three GMR releases.
4. What are the similarities and differences between a terrestrial cellular system (such as the GSM system) and a geostationary satellite system, which can be used to assist in optimising the air interface of a satellite system, retaining features of the terrestrial system as far as possible?
5. Compare the lower layer features of GMR-1 release 1 and GMR-2 systems.
6. With the help of a schematic describe the functional entities and their interaction of a typical MSS system such as GMR-1 and GMR-2.
7. With the help of a generic IMT-2000 satellite network schematic identify various interfaces including those that are excluded from the IMT-2000 radio interface specifications. List ITU recommendations pertaining to the integration of satellite systems with terrestrial IMT-2000 systems.
8. Several types of handovers are used in MSS. Handover algorithms are system-specific but broadly follow a similar sequence. List various handover types and outline the algorithm used in any one of the ITU-recommended satellite radio interface, mentioning the air interface.

9. Compare the following characteristics of the satellite radio interfaces recommended by the ITU:
- Orbit
- Frequency band
- Mobile terminal types and throughput
- Modulation and coding scheme
- Multiple access scheme
- Diversity
- Power control

10. What are the constraints in incorporating mobility to the DVBS2/RCS standard? What are the techniques used in the DVBS2/RCS + M standard to circumvent these constraints?

References

3GPP (2007) 3GPP Scope and Objectives, www.3gpp.org/ftp/Inbox/2008_web_files/3GPP_Scopeando310807.pdf (accessed 5 September, 2013).

Alamanac, B., Pauline, A., Chan, M. *et al*. (2010) DVB-RCS goes mobile: challenges and technical solutions, *International Journal of Satellite Communications and Networking*, **28**(3-4), 137–155.

DVB-1 (2012) Digital Video Broadcasting (DVB); Second Generation DVB Interactive Satellite System; Part 1: Overview and System Level Specification, ETSI TS 101 545-1.

DVB-2 (2009) Digital Video Broadcasting (DVB); Second Generation Framing Structure, Channel Coding and Modulation Systems for Broadcasting, Interactive Services, News Gathering and other Broadband Satellite Applications, ETSI EN 302 307.

DVB-3 (2011) DVB Project–An Introduction Factsheet, www.dvb.org/technology/fact_sheets/DVB-Project_Factsheet.pdf (accessed 5 September, 2013).

DVB-4 (2011) DVB Fact Sheet-Second Generation Satellite, /www.dvb.org/technology/fact_sheets/DVB-S2_Factsheet.pdf (accessed 5 September, 2013).

DVB-5 (2011) DVB Factsheet-Return Channel Satellite, www.dvb.org/technology/fact_sheets/DVB-RCS_Factsheet.pdf (accessed 5 September, 2013).

DVB-6 (2009) Digital Video Broadcasting (DVB); Interaction Channel for Satellite Distribution Systems, ETSI EN 301 790 V1.5.1.

DVB-7 (2009) Digital Video Broadcasting (DVB): Interaction Channel for Satellite Distribution Systems; Guidelines for the Use of EN 301 790 in Mobile Scenarios. Technical Report ETSI TR 102 768 V1.1.1, European Telecommunications Standards Institute.

DVB-8 (2011) DVB Fact Sheet, Generic Stream Encapsulation, www.dvb.org/technology/fact_sheets/DVB-GSE_Factsheet.pdf (accessed 5 September, 2013).

ETSI-1 (2009) Technical Specification, GEO-Mobile Radio Interface Specifications (Release 3);Third Generation Satellite Packet Radio Service; Part 1: General specifications; Sub-part 3: General System Description, GMR-1 3G 41.202, ETSI TS 101 376-1-3 V3.1.1.

ETSI-2 (2003) Technical Specification, GEO-Mobile Radio Interface Specifications (Release 2); General Packet Radio Service; Part 1: General Specifications; Sub-part 2: Introduction to the GMR-1 Family, GMPRS-1 01.201, ETSI TS 101 376-1-2 V2.1.1.

ETSI-3 (2001) GEO-Mobile Radio Interface Specifications; Part 1: General Specifications; Sub-part 3: General System Description, GMR-1 01.202, ETSI TS 101 376-1-3 V1.1.1.

ETSI-4 (2001) GEO-Mobile Radio Interface Specifications; Part 4: Radio Interface Protocol Specifications; Sub-part 2: GMR-1 Satellite Network Access Reference Configuration, ETSI TS 101 376-4-2: GMR-1 04.002.

ETSI-5 (1996) Digital Cellular Telecommunications System (Phase 2+); Mobile Application Part (MAP) Specification. (V5.10.1), GSM 09.02 (ETSI ETS 300 974).

ETSI-6 (2003) Technical Specification, GEO-Mobile Radio Interface, Specifications (Release 2); General Packet Radio Service; Part 4: Radio Interface Protocol Specifications; Sub-part 7: Mobile Radio Interface Signalling Layer 3, General Aspects, GMPRS-1 04.007, ETSI TS 101 376-4-7 V2.1.1 (2003-03).

ETSI-7 (2005) GEO-Mobile Radio Interface Specifications (Release 3); Third Generation Satellite Packet Radio Service; Part 5: Radio Interface Physical Layer Specifications; Sub-part 5: Radio Transmission and Reception, GMR-1 3G 45.005, ETSI TS 101 376-5-5.

ETSI-8 (2001) GMR-2 General System Requirements GMR-2 01.202 15, TS 101 377-1-3 V1.1.1.

ETSI-9 (2001) Technical Specification: GEO-Mobile Radio Interface Specifications; Part 3: Network specifications; Sub-part 2: Network Architecture, GMR-2 03.002, ETSI TS 101 377-3-2 V1.1.1.

ETSI-10, (2001) Technical Specification GEO-Mobile Radio Interface Specifications; Part 5: Radio interface physical layer specifications; Sub-part 5: Radio Transmission and Reception, GMR-2 05.005, ETSI TS 101 377-5-5 V1.1.1.

ETSI-11(2001) Technical Specification, GEO-Mobile Radio Interface Specifications; Part 5: Radio Interface Physical Layer Specifications; Sub-part 1: Physical Layer on the Radio Path, GMR-2 05.001, ETSI TS 101 377-5-1 V1.1.1.

ETSI-12(2007) Satellite Earth Stations and Systems (SES); Broadband Satellite Multimedia (BSM); Services and Architectures, ETSI TR 101 984 V1.2.1.

ETSI-13 (2008) Satellite Earth Stations and Systems (SES); Satellite Component of UMTS/IMT2000; Parts 1 to 6; Sub-part 1: G-Family, S-UMTS-G 25.211, ETSI TS 101 851.

ITU-1 (2010) Detailed Specifications of the Radio Interfaces for the Satellite Component of International Mobile Telecommunications-2000 (IMT-2000), Recommendation ITU-R M.1850, (01/2010).

Morello, A. and Mignone, V. (2004) DVBS2 – Ready for lift-off EBU Technical Review, http://tech.ebu.ch/docs/techreview/trev_300-morello.pdf (accessed 5 September, 2013).

Morlet, C., Alamañac, A.B., Gallinaro, G. *et al*. (2007/2008) Introduction of mobility aspects for DVB-S2/RCS broadband systems. *Space Communications*, **21**, 5–17.

Scalise, S., Schena, V. and Ceprani, F. (2004) Multimedia service provision on-board high speed train: demonstration and validation of the satellite-based FIFTH solution. 22nd AIAA International Communications Satellite Systems Conference and Exhibit, Monterey, CA, May 2004.

9

Operational Considerations

9.1 Introduction

A mobile satellite network comprises a number of entities that operate in concert to provide mobile telecommunication and/or broadcast services. This chapter addresses the mobile system in an operational context emphasizing the technical issues. The management goals are to:

- guarantee that the system provides the negotiated quality of service (QoS) to the user by monitoring direct or inferred measurements of the performance and taking corrective action when necessary;
- guarantee the specified reliability by taking appropriate counter-measures in case of failure of network components;
- supervise the functioning of equipment, collect and analyse anomaly reports and take corrective action for fault repairs;
- maximize network productivity, for example by allocating resource where and when necessary;
- Provide support to the commercial and business framework.

The chapter begins by presenting a perspective of the network with an introduction to various facets of the network management, followed by a discussion on the technical issues in relation to the network management. It then introduces issues and tasks related to radio resource management, including radio frequency monitoring and interference assessment, QoS management and spectrum licensing.

9.2 Perspective

An efficient network management system is fundamental to the success of a commercial satellite communication system. It comprises operational, commercial, planning, administrative and liaison tasks. This chapter deals with the operational management of the system.

An mobile satellite service (MSS) network comprises numerous functional units that interact with each other to deliver the desired services to the end user consistent with the requisite

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

quality. The task is to monitor the performance of each unit, apply configuration changes as and when needed, and take corrective action to recover from anomalies.

The operational framework consists of the following logical elements:

A physical layer, consisting of various pieces of equipment and associated software; for example high power amplifier (HPA) in an earth station;

Transmission layer, comprising a group of physical elements performing a network function – for example a gateway;

Service layer, which provides the desired circuit- or packet-based transport: Management of this layer involves service performance monitoring for QoS compliance;

Commercial management layer that deals with billing, commissioning, marketing, customer support, and so on.

External interfaces that deal with other organizations, vendors, manufacturers, service providers, and so on.

The transmission network used for the management system is an independent telecommunications network that interfaces with various elements of the managed network as necessary.

Figure 9.1 represents a global view of a network management system depicting the entities that constitute the operational MSS network lifecycle. *Constellation management* involves monitoring and controlling the performance of the constellation in its entirety. The health of each satellite is monitored and its orbital location and operational parameters adjusted as necessary. *Gateway management* involves monitoring and managing the performance and functioning of each gateway sub-system. The *switching system*, resident within a gateway, performs switching with respective core network on a call-by-call basis. The *mobility management* system deals with ensuring connectivity with a mobile population. The *security sub-system* ensures on-air security as well as physical security of vital elements of the network. Security measures are defined by each network operator according to internal

Figure 9.1 A telecommunications management network structure of a MSS

operational philosophy. Example of security features include maintaining subscriber confidentiality, subscriber or/and mobile authentication and data and signalling security.

Maintenance involves regular check-out and calibration of various equipments across the network. *Operations* refers to the real-time monitoring of traffic-flow, radio-frequency integrity (i.e. effective isotropic radiated power (EIRP) and radio frequency) of each element of the network, that is satellite, gateways and other feeder stations, mobile terminals and interference, interfacing with other operators and QoS monitoring. Radio resource management deals with effective utilisation of spectrum and spacecraft EIRP. *Configuration management* takes care of sub-system/system reconfiguration – for example spacecraft on-board switching or system reconfiguration in case of a satellite failure. It involves configuring a number of system parameters depending on the network requirement – constellation reconfiguration in case of a satellite failure or reconfiguration of spot beams. *Network management control systems* deal with issues such as radio resource management in real time or non-real time. Network control function also includes mobility management to ensure connectivity during a call or when the user changes spot beams or satellites. It also involves network information broadcasts to mobile and gateways – such details include frequency of signalling channels, gateway identification flag, changes to space segment configuration, and others.

Information flow is managed so that information is available to the network manager on time so that timely corrective actions can be applied.

In addition there are numerous supporting tasks. The *commercial support system* deals with customer support on real-time basis, this includes direct support or indirect support through service providers, billing, provision of traffic trends for forward planning or revenue prediction, etc. Administrative and commercial management functions (including aspects of configuration management) include subscriber management – commissioning of new subscribers, modification of existing user service profiles and billing; user equipment identity database management; user terminal status list management for issues such as checking fraudulent use; traffic statistics generation for network provisioning, planning, and so on; report generation, including incident, fault and network change reports.

Long-term planning tasks deal with strategic planning in terms of procurement of new satellites, development of new products, and spectrum forecast. Interface with external bodies would involve liaising with operators, service providers and regulatory bodies in support of issues such as interference management, spectrum negotiations, inter-system coordination, collaborative endeavours, publications, and so on. The research and development management ensures participation in the industrial collaborative efforts such as standardization, development of new-generation products, tracking recent development to remain competitive.

The *operational* tasks, the main theme of this chapter, involves monitoring, interpretation and control of a vast number of parameters in various domains – for example EIRP, carrier-to-noise ratio, bit error rate (BER) in a non-IP environment and packet loss and latency in an IP domain. Such data lie scattered throughout the network. Thus data collection, abstraction, analysis and presentation on powerful graphical user interface facilitate rapid and intuitive interpretation. Desirable features of such a monitoring system are:

centralized monitoring, in preference to a number of scattered facilities;

display and monitoring of the network topology so that faults may be isolated easily and rapidly;

capability to display functioning of both top and lower-level entities to provide a quicker fault diagnosis;

statistical processing capability to indicate trends and assist fault analysis; for example network fault trends, variations of operational parameters, information to refine alarm limits, and others.

By maintaining the performance standards of critical network parameters, the network operator is able to comply with the promised QoS. Such parameters include satellite EIRP variations with traffic load, frequency of each radio carrier, carrier-to-noise ratio, network congestion, numbers of repeat requests in an automatic repeat request (ARQ) scheme, message delivery time, throughput analysis, failed call analysis, holding time trend, and others. The MSS network provider often has no control over the QoS of another network from where a call may originate or for which it is destined. Thus, it is necessary to monitor end-to-end call quality (periodically) and keep a vigil on user feedback to gain an insight into the network behaviour.

Measuring the users' received signal quality during regions' busy hours when the network is fully loaded provides a realistic representation of the users' perception. During these periods, satellites are heavily loaded, resulting in increased inter-modulation noise, a possible degradation in satellite EIRP due to excessive loading, increased probability of interference and congestion, and a reduction in users' throughput in packet switched networks. QoS measure includes grade-of-service, end-to-end connection time, BER, packet loss and latency. The congestion observed by a user can be caused either in the terrestrial network, in the gateway or in the space segment resource management system. The system has to be provisioned to ensure that congestion is managed at all the levels.

9.3 Subscriber and Gateway Commissioning

In order to minimize the probability of interference and ensure the desired QoS, operators set compliance guidelines on the equipment performance for the manufacturers and participating service providers. The operator checks out the compliance guidelines before a gateway, a feeder earth station or a mobile earth station genre is allowed an entry into the network.

9.3.1 Gateways

To preserve the radio frequency and network integrity it is essential that, prior to its introduction into a network, performance of each earth station is checked-out. These checks include radio frequency compatibility tests and functional checks of the supported services. Many of the mandatory tests are witnessed by representatives of the operator to assess compliance. The earth station is certified for introduction into the network when all tests have passed successfully. Example radio frequency parameters used in these tests include: earth station G/T, antenna main-lobe and side-lobe performance, frequency and EIRP stability, phase noise and spurious emissions. Examples of functional performance include signalling, end-to-end call establishment and QoS compliance.

9.3.2 Mobile Earth Stations

Each mobile station model must pass rigorous tests before the operator permits its introduction into the network. The network operator certifies the mobile station model for production after the terminal has qualified for the tests. Tests check that such terminals respond correctly to network control signals and protocols, their ability to log on, quality of voice facsimile and data transmission, distress/safety-related functions, satisfactory reception of system management messages, end-to-end connection with public/private networks, transmitted and spurious transmission power compliance, regulatory requirements such as compliance with distress messages, and so on. After the type approval process, the model is allowed production.

When a new subscriber starts service, the mobile terminal is formally commissioned to ensure that the system has been configured for terminal identification, routing, billing, accounting, service, for the purpose of validation and billing. The commissioning procedures depend on the operator and the service. Aeronautical or maritime terminals required to comply with international safety standards would require a more rigorous commissioning procedure, whereas a satellite telephone for personal use has simple requirements.

Note 1 Voice quality in circuit mode

To maintain end-to-end voice signal quality, the following considerations apply in the transmission link of mobile systems:

- mobile earth station speech codec performance;
- impact of cascaded echo control devices in a transmission path, as they can interact with each other;
- performance of the fixed link, which includes the fixed link of the mobile system as well as the connected fixed network.

Transmission parameters affecting end-to-end quality include propagation delay, speech level at transmitter and receiver, quantization distortion introduced by the speech codec, thermal noise, multipath noise, co-channel interference and group delay. It is necessary to apportion degradation amongst various components of the link, including other network and international connections. An agreed planning standard, and if necessary, contractual agreement is necessary for guaranteed performance. In many countries, such as the UK, guidelines to facilitate link design and interconnect agreement between operators have been laid down by the operators' interest group.

Since mobile terminals usually operate under conditions of high ambient noise they should therefore exhibit good acoustic and echo characteristics. Send loudness rating (SLR) and receive loudness rating (RLR), determined by acoustic transducers and the analogue–digital/digital–analogue converters, specify user loudness requirements under operational conditions. Delay gets added during voice codec processing, channel coding and transmissions causing degradation to speech quality. Long propagation delays incurred in mobile networks require good echo control because echoes from such links cause a further degradation to the voice quality. Acoustic loss between the mouth and microphone must be low, and good contact must be possible between the ear and the earpiece of the handset. The shape and size of the handset also affect echo level. Significant echo is caused by coupling of the received signal to the microphone. The permitted echo coupling within a network is specified by terminal coupling loss; the smaller handsets preferred by users therefore cause difficulty in meeting the echo coupling loss specifications. ITU-T G.131 (ITU, 1987) specifies echo protection versus transmission delay of up to 300 ms.

Echo cancellers connected in tandem tend to reduce the net advantage and therefore intelligent cancellers are preferred. In their interoperating arrangements, UK network operators specify a particular set of rules for disabling/enabling echo cancellers. The scheme was implemented using CCITT No. 7 signalling system.

Fixed networks usually incorporate speech codecs at various points of their link, such as for rate adaptation, voice messaging and call forwarding, causing degradation to end-to-end quality. Quality can also be degraded by digital speech interpolation (DSI) systems along the transmission route. Tandeming an MSS service with a terrestrial mobile system may stretch the propagation delay to its limit. For example, introducing a Global System for Mobiles (GSM) circuit to a geostationary MSS circuit would result in a delay of the order of 355 ms, which when added to non-linear speech coding, echo and link bit errors can degrade signal quality quite severely (ITU, 1989). Therefore some terrestrial systems prefer cable routes for GSM–GSM links rather than satellite routes. A careful assessment of degradation along the route is therefore essential.

Some practical measures can be adopted to achieve high end-to-end voice quality – a good terminal type approval test ensures good voice quality/echo performance; apportioning of link quality with other operators, using measures such as a standard contract, can guarantee the specified quality; following sound end-to-end link design practices based on ITU-T recommendation, and identified above, ensures a reliable basis.

Note 2 Voice over IP on satellite channels

Voice over Internet protocol (VoIP) has become a dominant application for users with the availability of IP network and particularly the Internet as it provides a cost-effective solution. VoIP applications include voice calls, support for call centres, and directory services over telephones, IP video conferencing, radio broadcasting, etc.

A number of parameters inherent to satellite systems impact the quality of voice when transporting VoIP streams over satellite channels, necessitating countermeasures. The delay associated with satellite channels, particularly geostationary earth orbit (GEO) systems, can cause difficulty when coupled to the latency of the network and echo. To minimize IP transport delay VoIPs are carried over the User datagram Protocol (UDP) suite, which delivers packets without waiting for acknowledgements thus compromising packet loss in favour of low latency. Real-Time Protocol (RTP) is used over UDP for streaming applications such as voice and video. These protocols add overheads to the voice packets, which increases the transmission bandwidth, and to minimize their impact on bandwidth these headers are compressed by using techniques such as CRTP.

The Internet engineering task force (IETF) has also specified protocols that attempt to overcome the limitations of UDP, that is lack of congestion control and fairness control over multiple incoming streams. The Datagram Congestion Control Protocol (DCCP) addresses the fairness issue and is suitable for carrying multimedia streaming traffic including voice and video (Kohler *et al*., 2006). The VoIP performance of the DCCP protocols has in fact been evaluated over satellite channels using a number of voice coders and improvements proposed (Sarwar *et al*., 2012).

Packet switched networks carry a variety of traffic usually with a provision of QoS management; this mechanism can be utilized to prioritize VoIP data streams on basis of the negotiated QoS. A mechanism referred as call admission control allows admission of calls subject to availability of resources; proposed solutions include reservation mechanism using Resource Reservation Protocol (RSVP), Differentiated Service (DiffServ), MultiProtocol Label Switching (MPLS) and End-to-end Measurement Based Admission Control (EMBAC).

Jitter caused due to irregular arrival of packets can cause 'choppiness' in speech quality affecting comprehension. The average inter-arrival packet delay should approach that at the transmit end for a good quality speech. A smooth flow can, for example be maintained by including a smoothening buffer, which stores the incoming packets and delivers them at a uniform rate similar to the transmit end.

In order to provide good speech quality the packet error rate should be minimized that suggests the need a high quality radio connection.

VoIP schemes utilize the standardized voice codecs, such as G.729, which require 8 kbps that increase to about 16–18 kbps due to protocol headers; techniques such as CRTP reduce the throughput to about 10 kbps. Other standards such as G.723 can encode speed at 5.3 kbps (idirect, 2008).

9.4 Radio Resource Management

The maximum revenue earning capability of an operator is bound by the operator's space segment resources, that is, satellites' available EIRP and radio spectrum available to the operator (see Chapter 10). The spectrum efficiency depends on transmission scheme and frequency plan, that is, arrangement of the RF channels in a satellite and in the network in entirety (Richharia, 1992). Another practical consideration is the management of transmitted power from each satellite, which is limited by the technology available at the time of satellite manufacture. MSS systems require transmission of large amount of power per channel and hence it is essential for the satellite EIRP to be managed efficiently.

9.4.1 Spectrum Management

Operational spectrum management involves frequency planning, frequency list maintenance, real-time monitoring of frequency usage and spectrum assessment for future needs. Figure 9.2 (Richharia, 1992) illustrates the process of spectrum management.

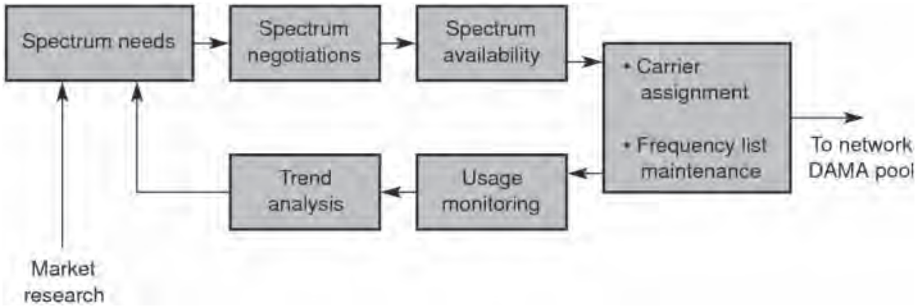


Figure 9.2 The main entities of the frequency planning process

MSS operators assess their spectrum requirements through market research and, when historic data are available, through trend analysis or a combination of these two processes. The spectrum requirement is presented at operators' coordination meetings or to the regulatory authority within the operators' jurisdiction. Depending on a number of factors, as outlined in Chapter 3, an operator may be granted access to the spectrum, though not necessarily the entire demand.

The transmission scheme (multiple access/modulation/coding/power control mechanisms) is optimized during the design phase and thereafter remains fixed. Spectrum management at this stage then becomes an operational issue, requiring detailed assignment of carriers wherein frequencies are assigned to signalling and traffic as efficiently as possible in response to short term traffic needs, applying constraints, rules and optimization. The resulting frequency plan is disseminated to various network entities, such as network control stations (NCSs), the network control centre, traffic monitoring systems, system analysts, and so on. Real-time management of the frequency plan on a per call basis is the responsibility of the radio resource manager located at the NCS.

Frequency plans are generated by a judicious utilization of software using dedicated computers that interface with radio resource manager to transfer frequency plans in a timely manner. The radio resource management system, for instance, may be a demand assigned multiple access pool in a circuit or a packet mode transmission system. In a packet mode transmission the incoming data streams are packed optimally onto each physical channel: the packing algorithm is constrained by factors such as QoS compliance to each incoming stream that requires a further level of optimization.

The operational tasks related to spectrum management are dynamic, due to rapidly changing requirements. A related task is monitoring spectrum usage and its trend analysis for planning. Note that circuit-mode MSS systems invariably use single channel per carrier (SCPC) or single slot per carrier assignment due to their thin communication needs.

The remainder of the section discusses the process of generating frequency plans – one of the technically challenging parts of spectrum management. Figure 9.3(a–c) identifies processes used for managing spectrum in an operational system. Figure 9.3(a) captures the process at the top level. Channel requirements are generated by trend analysis of recent traffic usage, including events that may distort the trend. Traffic usage is monitored as call holding time, data volumes and call failures due to congestion and other reasons. This data is analysed for trends, resulting in a set of detailed requirements in terms of the number

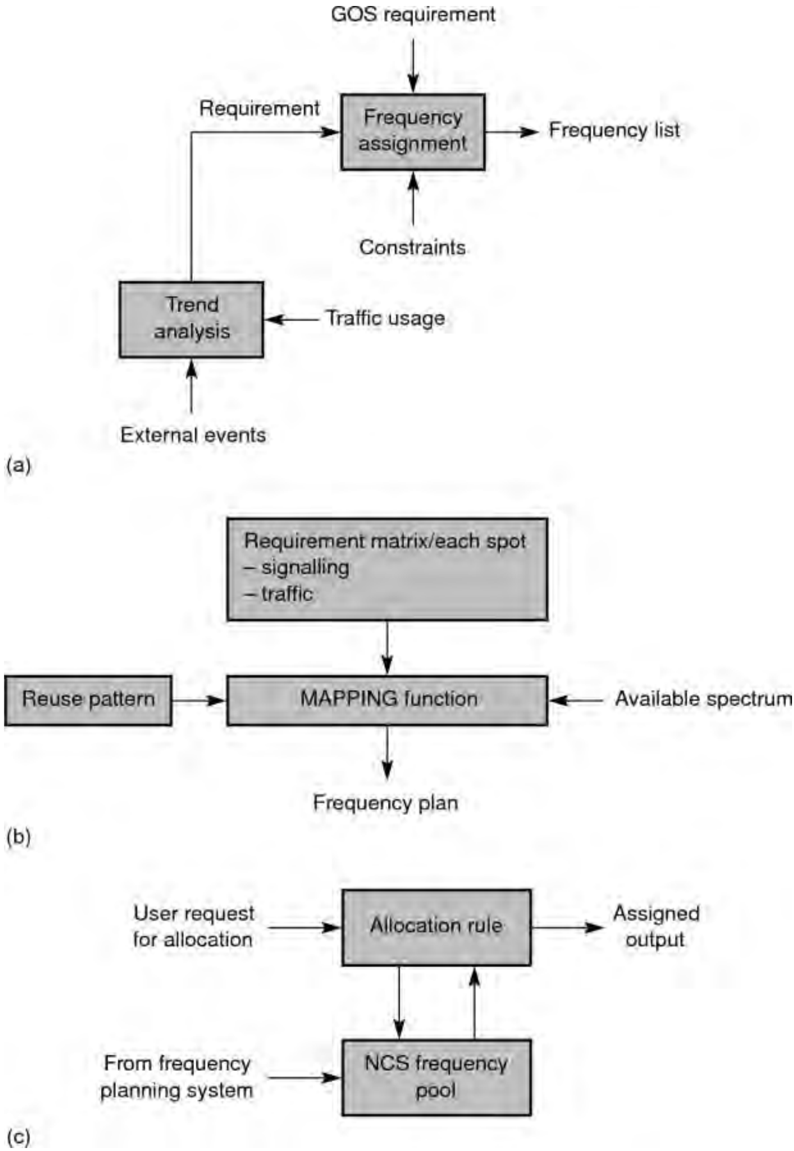


Figure 9.3 Processes used for managing spectrum resource: (a) overall process, (b) process for generating frequency list and (c) real-time channel allocation

of channels for traffic carrying and signalling carriers. The frequency assignment process generates a frequency list respecting the constraints and complying with the desired quality/grade of service (GOS). Figure 9.3(b) presents the frequency assignment process, which converts the raw data to the number of required RF channels and maps the requirements to a frequency plan, taking available spectrum, reuse matrix, optimization rules and constraints as the basis. The mapping function, shown in Figure 9.3(b), creates a frequency plan applicable at time t by taking into consideration the relative positions of satellites and spot beams. The constraints include practical considerations, such as receiver tuning range, signalling carriers and location-specific interference. The process can repeat for time t_1, if necessary. Figure 9.3(c) shows the real-time frequency assignment process per call or session. On request for a channel (in circuit mode) or capacity (in an IP system), the NCS allocates a channel or capacity on basis of a pre-assigned allocation rule. When dealing with packet switched network where assignments are multiplexed in time the radio resource manager assigns slots on basis of the QoS negotiated by the user at the time of assignment. For example, a streaming service would get an uninterrupted assignment whereas a delay tolerant service would be assigned slots when higher priority assignments have been served.

The actual performance of each carrier in the frequency plan can be evaluated by analysing the statistical behaviour of the performance of each carrier. For example, if the call holding time of a circuit mode SCPC channel is significantly below the average then this would indicate that users are unable to complete the call to satisfaction, or if an assignment remains unused this would indicate an anomaly in real-time resource manager database. Such performance measures can be introduced into the frequency planning process. In the earlier example, the corrupt slots will be removed from the pool and an attempt would be made to reassign the slot to a robust carrier or the slot could be assigned to another region. An anomaly investigation on this slot would isolate the fault – for example the frequency slot may have been subject to an inter-modulation noise spike that requires readjustments of the offending carriers.

Different carrier arrangements are possible, depending on the chosen optimization method, constraints and criteria. Sources of RF noise include transponder inter-modulation noise, spurious or out-of-band emissions from other satellite operators, terrestrial interference within a specific region, intra- and inter-system interference. A system becomes interference limited when intra-system frequency reuse is very intense (e.g. four-cell versus seven-cell reuse).

A rather complex frequency planning problem is that of a non-geostationary satellite constellation comprising satellites with a large number of spot beams due to the additional complexity of spatial and temporal variations in interference due to constellation dynamics.

In earlier years inter-modulation noise minimization was the primary optimization criterion; channels would be numbered in an ascending order, beginning from the least-affected channel. Thus if the real-time radio resource manager assigned carriers in an ascending order early-choice assignments would offer a better signal quality even under heavily loaded condition. Moreover, as the system is not fully loaded throughout the day, only the better-quality slots would be used for a majority of the day. Voice-activated carriers provide a further 2–3 dB advantage in inter-modulation noise. This type of arrangement does not include all the impairments, and since inter-modulation noise is only a small component of the noise budget, can lead to an unsatisfactory solution.

Several techniques have been investigated for frequency assignments in cellular radio, such as generalized graph colouring (Cozzens and Wang, 1984; Hale, 1980). There has also been interest in applying heuristic optimization methods to frequency assignments (Castelino, Hurley and Stephens, 1996; Holland, 1992, 1975; Goldberg, 1989; Kirkpatrick, Gelatt and Vecchi, 1983). In particular, genetic algorithms (GAs) based on Darwin's theory of evolution, simulated annealing derived from the concept of minimal entropy, and

Tabu search have been investigated for their applicability to mobile satellite communications (Jahn, 1999; Sammut, 1999; Pujante and Haro, 1997). Their work demonstrates the potential application of such schemes to operational satellite systems.

For geostationary satellite systems, antenna patterns remain static on the surface of the Earth and hence interference geometry is stationary, whereas in non-geostationary satellite systems interference geometry is dynamic. Thus, frequency management is simpler for geostationary systems in comparison to non-geostationary satellite systems. In geostationary satellite systems, frequency plans are generally static with adjustments required periodically in response to traffic variations – except when a dynamic frequency plans is in use (discussed later). In non-geostationary satellite systems, frequency plans tend to be quasi-static implying that they remain static over parts of the orbit where the traffic and interference profiles are quasi-static. Figure 9.4 represents coverage at two instances t_0 and t_1 of a low earth orbit (LEO) satellite system, assuming satellite-fixed reusability as illustrated. Note that the reuse matrix applicable at t_0 does not apply at instant t_1 (frequency sets 1, 3 and 4 interfere) necessitating a change to the frequency plan. Moreover, the use of certain frequencies may be forbidden in parts of an orbit due to inter-system or other reasons such as licensing. The use of a dynamic frequency plan on non-geostationary orbit, particularly LEO, would appear to be complex but cannot be ruled out.

Geostationary mobile satellite systems tend to use fixed frequency plan, that is frequency lists remain unchanged for prolonged periods depending on traffic dynamics (typically, months – unless the traffic load is skewed by an unusual event such as a natural calamity or war) during which the real-time radio resource manager assigns frequencies from the existing list.

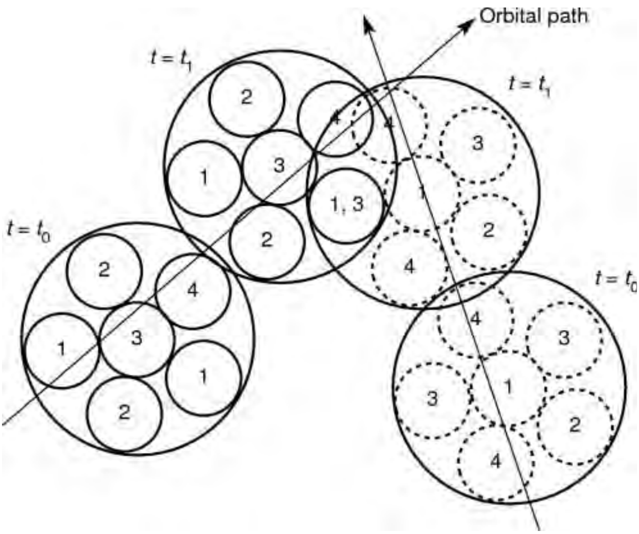


Figure 9.4 Satellite reference frequency plan demonstrating the need of frequency plan change. The figure represents the coverage patterns of satellites in two LEO planes at two instances t_0 and t_1. Spot beams that reuse frequencies are marked with identical numbers. At instant t_1, co-channel frequencies in patterns 1, 3 and 4 interfere, requiring changes to frequency plan

It is well known that traffic load on mobile satellite systems varies during the day with the busiest period coinciding with the region's business hours. In a wide-area network covering multiple time-zones, the load would ripple from the east spot beams to the west over a day. This has led to the concept of time-of-day dependent frequency plans wherein time-of-day dependent frequency plans are uploaded to comply with temporal variations.

Static frequency plans are suitable when the traffic usage profile is well-behaved. Each spot beam is allocated channels to satisfy its peak demands, and as stated, frequency lists are updated to meet changes to traffic demands over relatively long periods. Such schemes can also be applied to non-geostationary satellite systems. But fixed assignment schemes become increasingly inefficient when the dynamics of traffic distributions change rapidly.

An example of an event-driven real-time change in frequency refers to the change infrequency plans by Inmarsat during at the turn of millennium (i.e. 31 December 1999). Due to the (expected) unprecedented increase in traffic around the midnight of each region – the operational staff (lead by Dr Michael Wasse and supported by the author) continued to change the frequency plan of each ocean region in real-time as the traffic rippled from the east-most region (Pacific) to the western-most region (Atlantic – West).

With the advent of on-board processing satellites with agile frequency-beam switching capability, it is now possible to redirect frequency blocks to any spot beam thus creating the paradigm of *dynamic frequency plans* in which blocks of spectrum are moved dynamically in response to the traffic demands, ensuring that interference is kept within limits. An alternative is to manage the traffic dynamics on the ground at the input to the radio resource manager layer. In a *hybrid frequency plan* a part of the plan is static while the remaining is dynamic.

Richharia and Trachtman (2006) present the results of various techniques applicable to a MSS geostationary network.

9.4.1.1 Spectrum Management Architecture

A spectrum management architecture can be centrally controlled or can be distributed. In a *centrally managed architecture*, frequency lists are prepared centrally and transferred to each radio resource manager of the network. The rate of transfer depends on various system variables – traffic dynamics, real-time optimization capability, acceptable number of call handovers for reliable operation, and so on. In a *distributed architecture*, frequency plans are prepared by land earth stations (LES) autonomously, with feedback from other earth stations or by real-time measurements. Various possibilities exist.

- Frequency plans are prepared centrally so each gateway can implement its plan without communicating with others. The architecture is suitable for geostationary satellite systems. It applies to medium earth orbit (MEO) systems but becomes difficult for LEO satellite systems due to the rapidly changing interference and traffic conditions.
- Frequency plans are prepared such that gateways borrow channels from each other in real-time that do not infringe intra- and inter-system interference limits;
- Frequency plans are prepared in each gateway autonomously and verified through real-time interference measurements, feedback from other gateways and network management systems, etc.

- Frequency plans are prepared on the ground and uploaded regularly to the resource manager resident on each spacecraft; an alternative with coarser granularity would be to switch frequency blocks on satellite to spot beams in response to traffic demands.
- Satellites compute frequency plans autonomously based on rules/algorithms downloaded from the ground.
- Fixed and dynamic allocations are combined to obtain optimum performance. The fixed pool is sized for a specific traffic threshold, with the dynamic pool serving traffic beyond the threshold and shared by contention or rules, such as mentioned next.

Several schemes, studied extensively for terrestrial cellular systems, have been investigated for satellite systems. For example, the following schemes were investigated by Sammut (1999).

- *Satellite-based decision* (Bjelajac, 1996): In this distributed scheme, assignments are made autonomously on satellites, based on carrier to interference ratio assessment on candidate channels and supplementing this information with measurements on other satellites. An Iridium-type system was evaluated through a computer simulation. We note that in the ORBCOMM system this scheme is simplified by removing intersatellite links. This approach can be associated with the concept of cognitive radio (see Chapter 14).
- *Ground station based autonomous decision* (Finean, 1996): In this scheme, channels are managed at ground stations instead of satellites. Channel assignments are based on interference measurements of candidate channels at the ground station as well as the called mobile. The scheme was evaluated for an Iridium-type LEO system. In this scheme, the number of attempts to find an unoccupied channel can become large in high load conditions, causing delay in call set-up and a reduction in handover reliability. This approach also can be associated with the concept of cognitive radio.
- *Cost function minimization* (Del Re *et al*., 1994): In this scheme, satellites assign channels autonomously without any coordination with other satellites or gateways, through minimization of a GOS cost function. The optimum channel at any instant is chosen to be the one that minimizes the number of spot beams where the channel will be unusable because of interference caused by it. As proposed, the scheme does not consider intersatellite interference and therefore requires refinement for practical purposes where the constellation comprises more than one satellite – but, recall that regional geostationary satellite systems can operate with a single satellite.

One approach towards simplification in frequency plan of non-geostationary satellite systems is to use an Earth-fixed reuse pattern as illustrated in Figure 9.5.

In this scheme, spacecraft antenna beams are steered to fixed regions on the Earth in synchronism ensuring seamless coverage, thereby emulating a geostationary satellite system. In addition to simplifying the frequency plan, the satellite and user terminal handover numbers are reduced. Figure 9.6 portrays an Earth-fixed seven-cell frequency reuse pattern over which satellite beams would dwell as described. In this simplified representation, each satellite could, for example, be pointing at cells marked A. The (defunct) Odyssey and Teledesic systems proposed the uses of such a scheme.

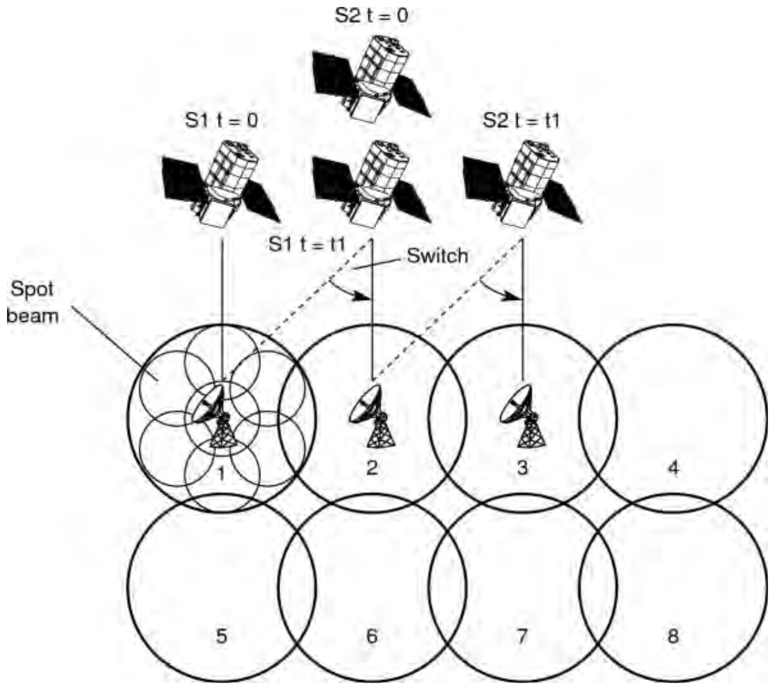


Figure 9.5 Earth-referenced reuse scheme. The service region is divided into a grid and each grid unit into a cell cluster; blocks of spectrum are assigned to each cluster and further subdivided into cells ensuring reusability; Each satellite footprint dwells on the cluster lying within its field-of-view and switches to the next in synchronism when moving out of the cluster

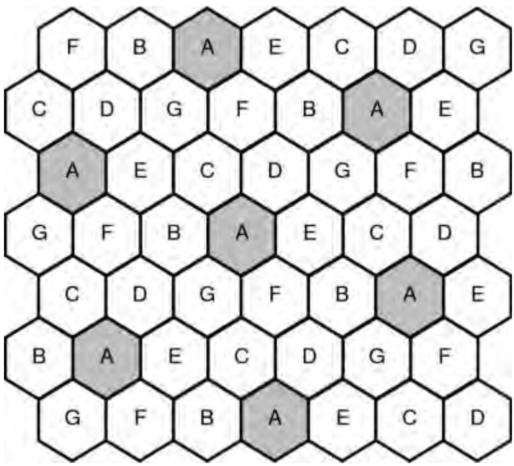


Figure 9.6 Seven-cell reuse pattern

9.4.1.2 Frequency Plan Examples

We present the frequency planning strategy of Intermediate Circular Orbit (ICO) system as an example. We briefly review the Iridium and Globalstar strategies as known and highlight results of investigations conducted by Inmarsat for enhancing the efficiency of real-time and off-line frequency planning of next generation satellite systems.

Iridium satellites consist of 48 spot beams per satellite (see Chapter 11) resulting in 2150 active beams on the surface of the Earth. The system uses frequency division multiplexed time division multiple accessing (TDMA) with seven-cell reuse, enabling an effective frequency reuse about 180 times around the world. The frequency allocation process is not available in open literature as it is the subject of a US patent filing. The constellation being polar, satellites converge as they approach the pole, causing an increase in coverage overlap. The reuse pattern is maintained for the entire orbit by progressively switching off overlapping spot beams; this is possible due to the predictability of satellite paths. As an aside, we note that this system avoids diversity in favour of spectrum efficiency.

The Globalstar system uses frequency division access with wideband code division multiple access (CDMA) and transparent transponders (see Chapter 11). This scheme minimizes restrictions in frequency reuse because of the interference rejection property of spread spectrum modulation. A direct sequence frequency division spread spectrum CDMA scheme with a 1.23 MHz spread and quadrature phase shift keying (QPSK) modulation is used and the transmission band is spread over 13 1.23 MHz bands. Each gateway uses a unique code, allowing signal isolation between different ground stations, beams and satellites. To avoid code overlapping in adjacent beams, gateways apply a different time off-set to each beam within its view. In the return direction, mobiles use the same binary sequence as the forward channel but with user-specific time-offsets to allow identification.

ICO proposed a 10-satellite MEO system placed in two inclined planes (see Chapter 11). Frequency plans would be generated centrally, taking into consideration the orbital and spot beam characteristics, daily traffic variability on each satellite and other factors such as interference. Channels would be assigned such that inter and intra-satellite reused channels had sufficient isolation. The spectrum would be divided into non-overlapping blocks, which would be divided between two orbital planes, thus avoiding spectral overlap between orbital planes. Within a plane, each block would be further subdivided into sets; within each satellite, frequencies would be assigned so that they do not cause interference to each other. Intersatellite interference would be managed by ensuring that spectrum assigned to leading spot beams does not overlap with the set assigned to the trailing spot beams of the satellite in front, and similarly, the spectrum assigned to the trailing edge of the satellite does not overlap with the set assigned to the leading edge of the satellite at the rear.

Figure 9.7 (based on ITU, 2010) shows an example of such a frequency plan for 10 and 12 satellite constellation. The constellation comprises 10 or 12 satellites distributed in two 45° inclined orbital plane at a nominal altitude of 10 390 km (similar to the ICO system). The available spectrum is divided into 16 blocks with eight blocks attached to each orbital plane. Within each plane each satellite is assigned eight blocks – divided in two sets each of four blocks such that the leading edge is assigned blocks 1–4 and the trailing set 4–8. As the leading edge does not overlap with the trailing edge of the satellite in front the blocks can be reused safely.

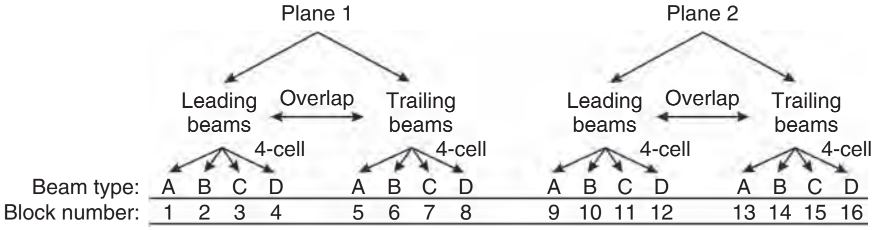


Figure 9.7 Block-wise arrangement of frequency plan

Inmarsat investigated methods in refining frequency planning techniques for real and non-real time applications to support Broadband Global Area Network (BGAN) and legacy systems over a heterogeneous mix of three satellite generations (Richharia *et al.*, 2006). Inmarsat-2 and Inmarsat-3 frequency plans are rigid due to constraints on satellite channelization and switching. Although Inmarsat-3 has a channel switching capability, for practical reasons it can only be exercised occasionally for specific operational purposes. The difficulty arises because of relatively large bandwidths channels, a rigid spectrum allocation regime, difficulty of clearing out spectrum from a filter quickly due to resident mix of carrier types (e.g. signalling) and a manual (rather than electronic) interface between network and satellite control systems. On Inmarsat-4 satellites, filters are switchable rapidly between beams on a 200 kHz grid via an interface between the network traffic management system and spacecraft's payload control systems.

Three approaches were considered in the study – static, dynamic and hybrid. Static frequency plan deal with off-line formulation of the plan, dynamic frequency plan deal with real-time management of frequency plan and the hybrid technique considers a mix of a core fixed plan over which a dynamic element resides to service temporal traffic variations. The study was primarily directed at the physical layer, although the results could be applied to the real-time resource management algorithm of packet switched services at the link layer.

Algorithms evaluated for static frequency planning included sequential allocation, Tabu search, GA and adaptation of a commercial off-the-shelf optimization engine. An agile interference calculation methodology eliminating redundancies in interference analysis, and storing results of repetitive calculations was developed permitting a computationally-efficient interference calculation, which is central to dynamic frequency planning in an interference limited environment of the type under consideration (other constraints such as spectrum coordination compliance, tuning steps of mobiles, and so on, being quasi-static). Due to the uncertainty associated to the position of a mobile within the coverage area (essentially a non-uniform distribution) and the fact that call initiation and holding time is random in a demand assigned SCPC system, a statistical approach involving a probability-weighted location dispersed over the service area was preferred as opposed to the traditional worst-case single-point interference analysis. Thus, it would be possible to satisfy the QoS criteria to any arbitrary percentage locations. It was concluded that due to the complexity and variability of the global MSS operational environment (a mix of spot beam types, a large mix of carrier types exceeding 100, a variety of user terminals), a single algorithm would not be optimal and hence a library allowing selection of the most appropriate algorithm offered a flexible alternative. It was observed that for demand assigned systems operating on

Inmarsat-4, a dynamic plan involving real-time spacecraft switching generally outperformed other schemes for the frequency planning requirements and assumptions chosen in the study. The hybrid scheme approaches the effectiveness of a dynamic plan when the mix of fixed and dynamic components is chosen judiciously.

Two classes of optimization algorithms were evaluated – those based on rules and constraints belonging mostly to sequential algorithm category and others not following rigid rules belonging to the meta-heuristic class such as Tabu search and GA that can provide near-optimal results within a reasonable time. After initial trials, GA was preferred over the Tabu search algorithm. The GA scheme is used to order the carrier in a semi-optimal manner; a sequential algorithm is then applied to the ordered set. The sequential algorithm assigns carrier on a first- come-first-served basis respecting all the rules and constraints including that of aggregate interference analysis. Sequential algorithms are very quick heuristic algorithms, but their performance relies heavily on the ordering of the carriers in the selection process. The GA presents a mechanism for giving a good ordering of carriers to the sequential assignment process. GAs are based on the mechanics of genetics (Davis, 1991). The string of numbers (gene, or carrier in this analogy) that encode the problem are known as chromosomes (carrier assignment list in the present context); associated with each chromosome is its fitness – that is its performance. A group of chromosomes forming a population reproduces to produce children, and the fittest chromosomes have greater likelihood of becoming parents. In reproduction, two parents are recombined where a different part of each parent is given to the child. Mutation may also occur where a gene within the chromosome is randomly changed. The position of the gene in the chromosome represents its sequential order. By using the operators of recombination and mutation the GA population converges to find an ordering that attempts to give a robust solution. Initial population is generated by random combinations of the input carrier list; fitness of each chromosome is achieved by evaluating the performance of the frequency plan generated using a sequential algorithm; parents are selected from the population by using roulette wheel selection with normalized fitness. The population is updated using steady-state reproduction without duplicates. At each generation, a small number of children are generated and introduced into the population provided their fitness is greater than the fitness of the worst member of the population. Towards the end of the GA the population converges to a set of chromosomes having similar fitness. Either recombination or a mutation operator is chosen to form a child based on a crossover probability. Mutation is used to allow a greater diversity and is useful in the final stages when the solution begin to converge. Two types of recombination operators were used – Uniform Order Crossover (UOX) (Davis, 1991) and Cycle Crossover (CX) (Oliver *et al*., 1987). In UOX a random-sized mask is chosen and gene of one parent is chosen to be within and the other outside the mask retaining the ordering. The CX operator cycles between the two parents selecting a gene on each cycle for the child until all the genes are filled. For mutation a swap operator was used that takes two genes at random and swaps them. Figure 9.8 demonstrates the improvement in the frequency plan as the evolution progresses. It illustrates that the majority of improvement occurs in the first 1000 generations. It also shows that UOX operator performs better than the CX operator.

An off-the-shelf package (ILOG, 2006) was adapted to automate the frequency planning process. A feature of the ILOG packages is a facility to search a large solution space for a solution to a problem. It is possible to define searches in a non-procedural way by means of goals. In order to speed up the otherwise slow frequency planning process based on ILOG

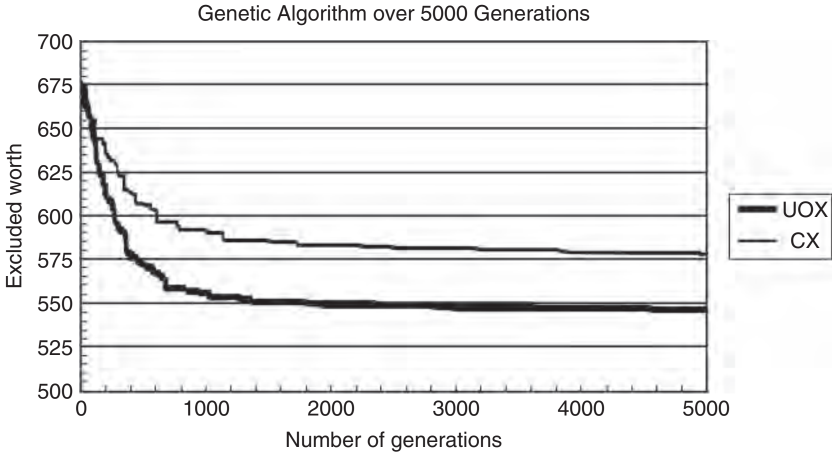


Figure 9.8 Performance of UOX and CX over 5000 generations. (Source: Richharia *et al.*, 2006.)

alone, a design concept called *Metamodel* was developed in which an intermediate representation of the allocation problem called Metamodel is built as a one-time activity. From the Metamodel it is possible to extract representation of any sub-problem quickly so that the same can be presented to ILOG within a processing time of a fraction of a second. The Metamodel is built on the knowledge of the frequency planning regime such as reuse matrix, forbidden frequencies, and so on.

The modelling objects are extracted from the database and built in to representative structures. The model for obtaining solution to a problem can be arrived by selecting appropriate objects in a Metamodel format and linking them to create a model suited to ILOG. The Metamodels can be stored and retrieved using Metamodel IDs in the database. The Metamodel operates in three iterative steps. In iteration 1 an algorithm termed the *Load Reduction Algorithm* (LRA) attempts to find a solution beginning from no solution. If a complete solution is not found in iteration 1, the process proceeds to iteration 2, termed Incremental Allocation Algorithm (IAA) where an attempt is made to assign the remaining carrier by adding the load in small manageable steps. If a solution is not complete in iteration 2, the process proceeds to add the remaining carrier by relaxing the reuse distance (thus exceeding the interference budget) in a controlled manner using an algorithm called Heuristic Distance Reduction (HDR). Iterations 1 and 2 provide a clean frequency plan, whereas iteration 3 allows a controlled relaxation of the rigid constraints at the discretion of the operator.

The process was evaluated by developing frequency plans of Inmarsat's heavily loaded 3F1 satellite (Indian Ocean Region) and 3F3 satellite (Atlantic Ocean Region-West) using as benchmark the carrier count of the existing frequency plans, which evolved over several years incrementally using computer-assisted manual procedures. As illustrated in Figure 9.9(a) and (b), the results demonstrated that for the 3F1 satellite, the ILOG solution populated 98% of the carrier, out of which 2% belonged to the HDR category. For the lightly loaded 3F3 satellite (Atlantic Ocean Region-West) satellite, all the carriers were allocated successfully with 1% belonging to the HDR category. The retained carrier category represents carriers that had to be retained at their assigned frequencies in the

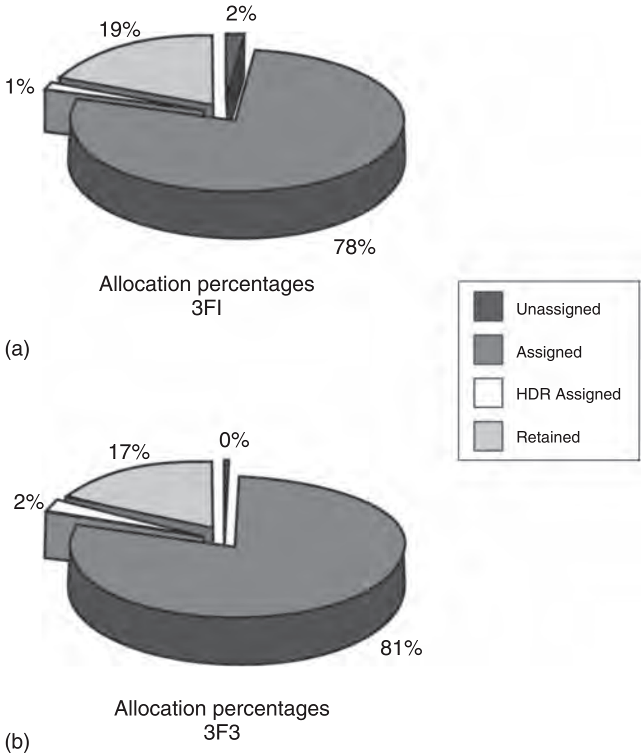


Figure 9.9 Performance of the frequency planning process using an off-the-shelf optimization engine on two Inmarsat satellites (a) heavily loaded 3F1 satellite and (b) lightly loaded 3F3 satellite. (Source: Richharia *et al*., 2006.)

frequency plan for operational reasons. The existing plans were developed gradually over several years, whereas the software achieved the fully automated solution in a few hours.

Real-time frequency allocation (as opposed to off-line) studied included the following configurations limited to Inmarsat-4 satellites (I4):

- Fixed carrier assignments with fixed filters (Note: A filter denotes a segment of spectrum 200 kHz wide).
- Fixed filter assignments but dynamically assigned carriers within those filters.
- Dynamic variation of filters.

When a sequential or similar algorithm is used to create a static plan, then simulations show that the use of dynamic assignment always achieves lower blocking than fixed assignment. Dynamic assignment with borrowing of filters out-performs other schemes, depending on the number of whole (200 kHz) filters that can usefully be accommodated in each beam. However, the advantage at high loading becomes lower when the efficiency of the fixed frequency plan is improved (e.g. using the GA scheme as suggested earlier).

Figure 9.10 compares the performance in terms of blocking for fixed, dynamic and hybrid frequency assignment schemes. Dynamic assignment without filter borrowing is seen to be

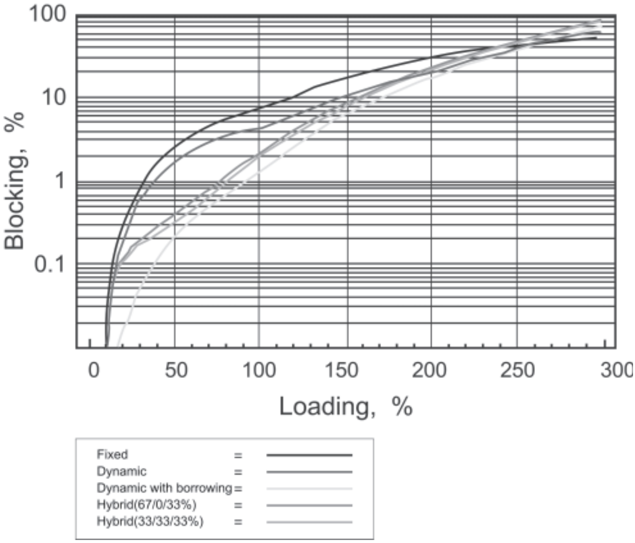


Figure 9.10 Blocking performance of fixed, hybrid and dynamic (with and without filter borrowing) frequency plans over an Inmarsat-4 satellite. (Source: Richharia *et al*., 2006.)

only marginally better than fixed assignment. Dynamic assignment with filter borrowing results in almost an order of magnitude less blocking over a significant range of loadings. The hybrid assignment (with filter borrowing) schemes produce very similar results for the two different filter pool sizes (67/0/33% and 33/33/33%) improving rapidly from the fixed assignment values as the loading is increased and tending towards the dynamic assignment performance. The test bed allows a large number of situations to be evaluated depending on the evolution of the traffic on I4s.

9.4.1.3 Traffic Analysis

Frequency lists used by the resource manager should match the traffic needs of the network to comply the specified quality/GOS. In an MSS network, traffic varies in the short term over minutes or hours, with an underlying trend that manifests over weeks and months. These variations are modelled in different ways due to differences in their mechanisms and characteristics.

In a circuit-mode configuration, the measured traffic (Erlangs) can be converted to the number of circuits using a Poisson or Erlang model (see technical note in the next section). Estimates of the number of RF circuits for packet-mode transmissions involves assessment of the total traffic, the traffic mix, the capacity of RF channel, and the QoS (compliant to the traffic mix) including packet delivery latency and the throughput perceived by users.

Short-Term Traffic Trend

It is well known that the traffic load carried over mobile satellite networks varies diurnally, with a peak during the business hours (e.g. Morgan and Gordon, 1989; Richharia, 1992).

When the service area encompasses several time zones, diurnal traffic profile of each time zone is offset proportionally; hence, the peak traffic carried over the satellite migrates from the spot beam covering the easterly time zone towards the western spot beams. Figure 9.11(a) portrays a hypothetical scenario where a geostationary satellite is located at 60°E, and provides coverage over regions lying within a 5° elevation contour. Dominant traffic-generating regions, their peak traffic in Erlangs and time zones are also marked on the figure. The beams

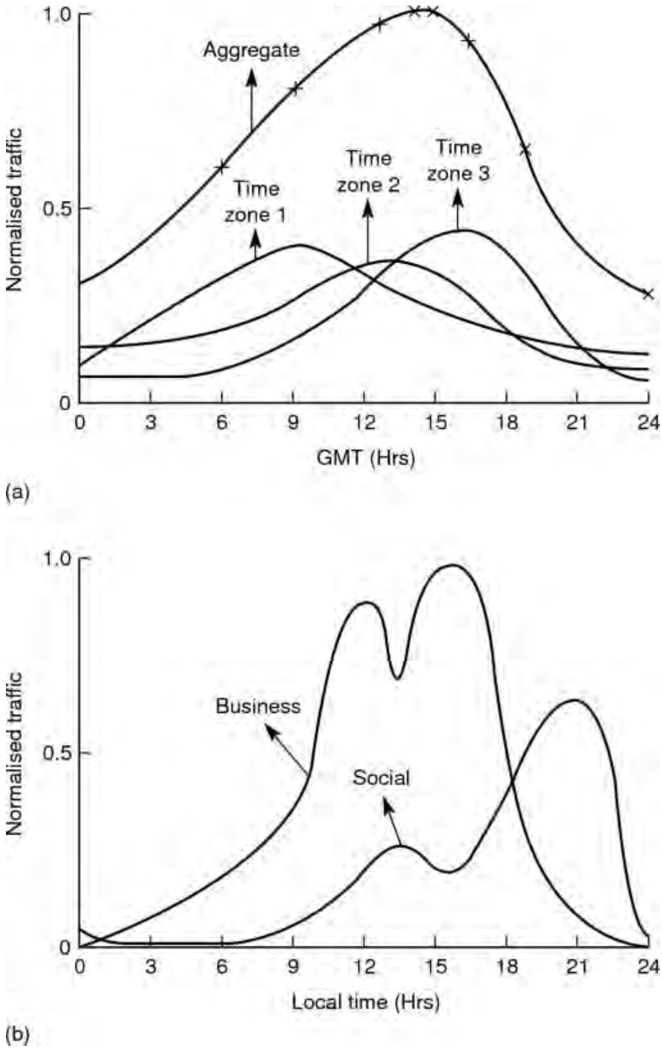


Figure 9.11 (a) A hypothetical model for obtaining weekday diurnal variation in traffic load in a system servicing multiple time zones. Such a trend was routinely observed over Inmarsat satellites in all the four ocean regions for its legacy services. (b) A hypothetical model of business and social traffic over a satellite for generating diurnal traffic profile on satellite. The trough in the mid-afternoon is caused by traffic dispersion during lunch breaks

covering time-zone 1 peak at about 9:00 h GMT; the beams carrying time-zone 2, peak at about 14:00 h GMT; and the beams covering time-zone 3 peaks at about 17:00 h GMT. The illustrated time zones are not adjacent and hence the figure demonstrates that the traffic peaks are concentrated in non-adjacent regions, for example Indian sub-continent, Europe and the US. Notice that the aggregate load on the satellite consists of the sum of spot beam loads, peaking at about 15:00 h.

Figure 9.11(b) represents a 24-h traffic profile within a single time zone for typical weekday business traffic, assuming a normal working day between 9:00 and 17:00 h with about an hour's lunch break. Traffic rises rapidly at the start of the business; there is a trough in mid-afternoon as people disperse for lunch over an hour, followed by another surge in the traffic. The traffic subsides rapidly after the end of the business day. Weekday traffic for social needs would be expected to rise gradually over the day, with an increase between 6:00 and 22:00 h and then tapering off. A weekend traffic profile would show a small back-ground business traffic whereas the social traffic would be higher. A refinement to the model would be to segment traffic by market sector, user terminal and traffic category as exemplified in Table 9.1.

Temporal and location dependence would depend on the environment, geography, traffic type and the time of day. Business traffic for land, maritime and aeronautical services would continue to peak in business hours. Traffic for the aeronautical environment would, in addition, be influenced by aircraft route, flight time and duration; land traffic distribution would demonstrate a strong relationship to geography (cities, motorway, railway routes, etc.) and the local telecommunications infrastructure; and spatial distribution of maritime traffic would exhibit a strong correlation to sea routes. This type of segmentation can allow an effective service/time of day/spot beam/location dependent radio resource management.

Since business traffic drops to low levels at night, during weekends and public holidays, satellite capacity remains under-utilized during such duration. To encourage the use of unutilized transponder capacity, operators tend to lower the tariffs. Inmarsat has developed a system to utilize the spare capacity that provides off-peak lease services that hand over the circuits to regular traffic if an unexpected surge in demand occurs (Richharia and Trachtman, 2006). The technique provides a low-cost framework for leasing spare satellite circuits without affecting the grade of the regular service. The arrangement is flexible without the need of operator intervention, unlike the conventional leases in which the leased spectrum is unavailable to the regular traffic. The operator can utilize existing NCS and signalling,

Table 9.1 Market segmentation by sector

Sector	Variants
Land	Terminal type: vehicle mounted, hand-held, paging, broadband, rural telephony, pay booth, railway
	Traffic type: social, business, tourism, event coverage, circuit, packet
Maritime	Terminal type: installed on cargo ships, yachts, cruise ship
	Traffic type: business, social, pleasure, event coverage, circuit, packet
Aeronautical	Terminal type: small, medium, large
	Traffic source/type: Air traffic control, cockpit communication, passenger communication
Distress and safety	Aeronautical, maritime, land

while the lessee can make use of off-the-shelf mobile Earth stations (with minor alteration). Due to a finite probability of pre-emption at a short notice, the capacity is attractive for non-critical applications such as caching, multicasting and web-surfing. The lease is extendable to busy hours, provided the lessee tolerates a larger pre-emption probability and the operator accepts an increase in signalling overheads during the busy hours. To ensure minimal impact to the regular users, off-peak lease circuits are pre-empted and returned when the number of channels in the regular pool falls below a threshold. To achieve the functionality, the system incorporates a real-time handover mechanism of leased channels to the regular pool. To avoid call set-up latency to the regular service user during the handover, the NCS assigns a channel from a buffer pool before initiating a handover. At the end of a handover, the recovered channel is deposited into the buffer ensuring that the buffer remains full. Thereafter channels released by regular users are released into the lease pool.

In summary, traffic carried by a satellite is influenced by:

1. Spatial distribution of users and their respective time zones;
2. Day of week (working day or not);
3. Temporal distribution of traffic – short (minutes) and long (hours) term;
4. User base, that is business or social;
5. Service type, that is voice, data (Internet);
6. Operational environment – aeronautical, land, maritime;
7. Anomalous events such as New Year's Eve, war, natural calamity.

Medium-Term Traffic Trend

For medium-term planning extending months and years, the underlying trend variation and external influences are of primary interest (see Chapter 10 for long-term forecast methodologies). Trend analysis through curve fitting is a simple and effective technique under stable conditions (i.e. periods without anomalous events), commonly used in telecommunication networks. Its accuracy can be improved by superimposing seasonal trends such as the effects of a holiday season, or a traffic surge such on the New Year's Day, knowledge of new commissioning, etc. Figure 9.12 illustrates the trending approach using an exponential (a), linear (b) and negative exponential (c) curve fitted to a set of hypothetical data. We note by inspection that the error resulting from estimates (a) and (b) can become significant whereas those from curve (c) are accurate demonstrating that unless the fitting is chosen judiciously the error in forecast can become significant. Similarly, unexpected deviation in the usage can distort the well-behaved trend causing errors in this type of forecast.

The main limitations of this approach are

1. **Need of historic database:** Historic data are not always available. Data are non-existent when a new operator introduces a service or an incumbent one launches a new product. In such cases, theoretical assumptions have to be estimated based on heuristics, logic, etc. (see Chapter 10 for such methodologies).
2. **Absence of external influences:** The impact of external influences is absent; while applicable for short-term forecasts, the steady-state assumption may not apply in the longer term for a number of reasons – for example introduction of new and/or improved services, unexpected increase in competition, societal change, and so on. Refinements to

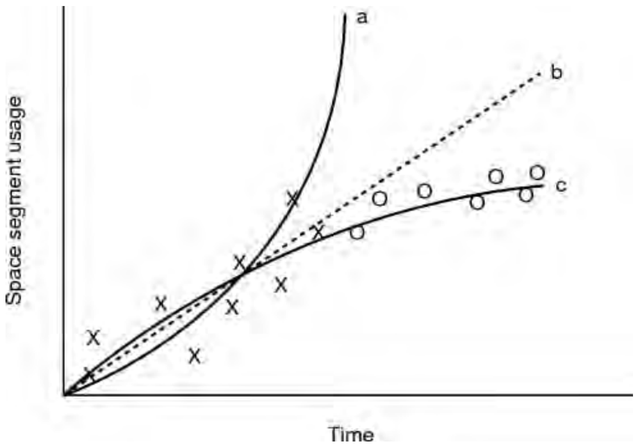


Figure 9.12 Trend analysis using curve fitting. Various types of curves have been fitted to a hypothetical set of data and extrapolated for forecast. Note the sensitivity of the results to type of curve (a = exponential fit, b = linear fit and c = negative exponential fit; x = historic data; O = actual growth)

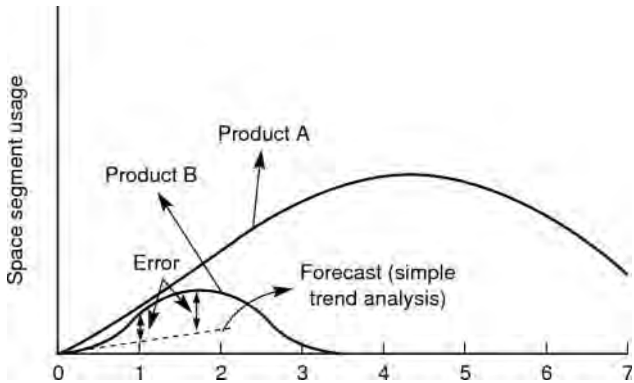


Figure 9.13 Hypothetical lifecycles of two products – A and B. Note the forecast error for product B when a simple linear extrapolation is used. The linear forecast model could be improved by incorporating additional knowledge. Note the difference in the lifecycle of the products

the model can be made by including the impact of external influences, such as estimates of traffic lost to a competing product, a change in the usage trend, and others.

Product lifecycles follow the well-known S curve whose timescales are influenced by a number of factors – evidently, the accuracy of long-term projections depends on how well a product lifecycle has been modelled, as illustrated in Figure 9.13. For example, the trend in the decline of analogue systems was much slower than anticipated – operators of analogue systems were able to retain customers by call cost reduction, aided by the availability of low-cost handsets enabled by mature technology.

The accuracy of simple trend analysis should generally be adequate for periods of a few months to a year, depending on the stage of the product cycle, particularly if external influences are factored in.

Forecast models should have the necessary granularity as resources have to be planned for each beam or beam cluster when beam sizes are smaller. Note that spot beams in regions within the same time zone would have identical time profiles but spatial distribution may vary; the largest time-offset in the profile will be experienced between beams that are longitudinally furthest apart (see Figure 9.11a).

In an operational environment, the GOS of the space segment can degrade under unusual network loading conditions. Some possibilities include:

1. **Unusual events:** To reiterate, events such as a natural disaster, war, an election, a sporting or social event result in unusual flow of traffic through an MSS space segment. This type of traffic surge is also experienced in terrestrial mobile systems in conditions such as a road traffic jam.
2. **System malfunction:** A variety of system malfunctions cause anomalous loading condition. For example, a faulty beam-forming network on a satellite can distort the shape of the beam resulting in an unexpected change to the traffic pattern; a NCS fault can cause anomalous congestion. System architectures using signal strength detection techniques for spot beam identification are sensitive to variations in the signal strength of the cell identifier and hence a fault in the spot beam identifier transmission can result in an anomaly.

Figures 9.14 and 9.15 portray the scenario when the signal strength of the cell identifier in spot beam f either increases (case 1) or reduces (case 2) by, say, 1 dB due to a system malfunction. In case 1, traffic from adjacent beams is flows into beam f, as a number of mobile stations originally locked to other beams are ‘captured’ by beam f due to its higher signal level, causing an increase in traffic carried by the beam f (Figure 9.14). A traffic

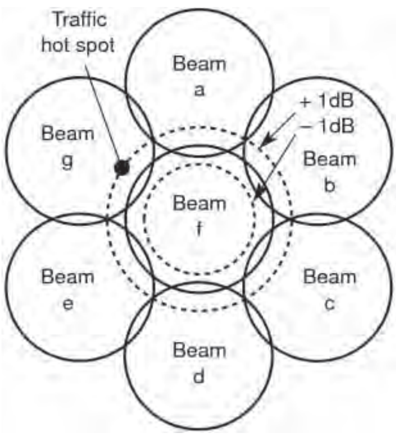


Figure 9.14 Spot beam size of beam f is changed by +1 dB (case 1) and −1 dB (case 2) due to change in level of spot beam identifier

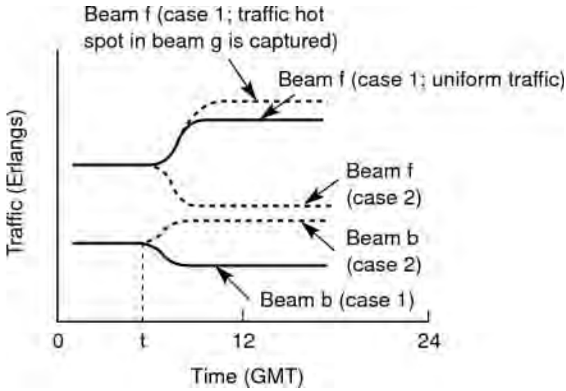


Figure 9.15 Change in traffic captured by beams f and b for cases shown in Figure 9.14 including a situation when beam f captures a hot spot in beam g

hot spot, such as marked in spot beam g (see Figure 9.15), can further increase the traffic captured by beam f relative to the scenario of uniformly distributed traffic, depicted by the dotted line in Figure 9.14. For case 2, the spot beam f shrinks in size, resulting in migration of its traffic to other beams and a corresponding increase in traffic carried by other beams – as demonstrated for beam b in Figure 9.15.

Figure 9.16 depicts a scenario when the spot beam identifier of the central spot beam has failed. Traffic of the beam in common areas migrates to the other beams whereas the central dark area is left unserved by spot beams and its traffic migrates to an umbrella global or regional beam when such a beam is available, but only supportable services can be transferred to the global beam. Note that the capability of the global (or regional) beam will be limited with lower G/T and EIRP. Such a scheme is available in Inmarsat's BGAN system

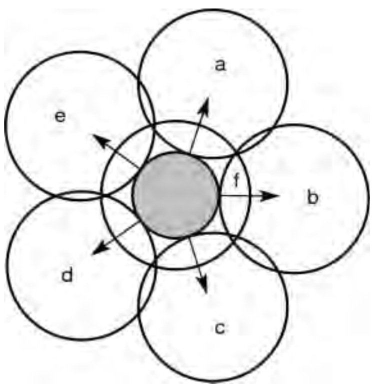


Figure 9.16 Migration of traffic from beam f to other beams in case of failure of its cell identifier. The mobiles in the central hatched area will have either no service or migrate to an umbrella beam (if available) provided that the service can be supported on such a beam

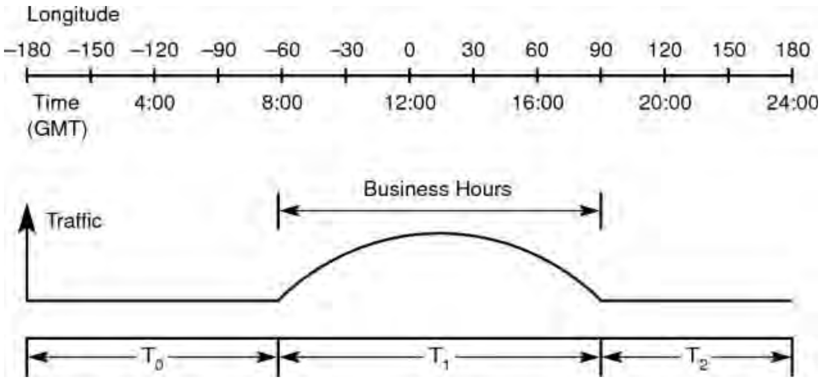


Figure 9.17 Hypothetical traffic flow as a function of longitude on a low earth equatorial orbit satellite for a mid-afternoon pass over a traffic zone of high intensity

where services with limited throughput are supported on the regional spot beams when a narrow spot beam is unavailable.

The scenarios depicted previously become time and spatially variant for non-geostationary satellite systems. Figure 9.17 (described in the next section) represents a hypothetical traffic flow as a function of longitude on a low earth equatorial orbit satellite for one pass, assuming high traffic intensity within latitudes of about −60 to +75°.

When we consider systems based on earth-fixed spot beams, spot beam identification is made by comparing the spot beam boundary with the location determined by each mobile. Here the impact and type of system malfunction differ – an anomaly occurs due to a malfunction in the user terminal's position determination system or a mismatch between real and broadcasted spot beam maps. The consequence in either case is an error in spot beam identification in mismatched areas and the associated impact (e.g. assignment of incorrect frequencies for a call); however, the problem of traffic migration is absent. In these systems, the operator has an option to change the size of the spot beam by broadcasting a spot-beam periphery update.

Applicability of Traffic Forecast to Non-Geostationary Satellite Systems

The coverage pattern of satellites in a non-geostationary satellite system is time variant and illuminates a smaller area (except satellite systems operating in high altitude elliptical orbits), implying lower traffic per satellite and a composite diurnal and spatial variation traffic profile through each satellite. A simple model is developed here to illustrate the concept (see Figure 9.17). The surface of the Earth is divided into 15° segments (in the equatorial region), which gives a difference of an hour between local time at the centre of lateral neighbouring cells. Hypothetical business traffic at 12:00 GMT for different time zones are shown (ignoring the lunch hour). Traffic is assumed to peak around mid-day, tapering on either side. Using this model, it is possible to estimate the traffic on each satellite of an equatorial constellation such as ECCO as it orbits the Earth. Traffic will be maximum for T1/(To + T1 + T2) of the orbital period on the satellite. In reality, traffic distribution is likely to be non-uniform and constellations may be in an inclined, polar or hybrid orbit

9.4.1.4 Signalling Channel Requirements

The previous discussions dealt with the traffic channels. A mobile network requires a number of signalling channels to transfer network information to mobiles and for mobiles to initiate and maintain communication. The number of signalling channels depends on the network topology. The numbers are higher for distributed architecture than for a centralized architecture. Typically, it is necessary to provide a broadcast channel for each spot beam and one or more signalling circuits in the return link for the mobiles. As signalling channels do not earn revenue and consume satellite resources, an attempt is, therefore, made to minimize their number but make them robust in presence of noise and fading due to their significance.

Note 3 Congestion theory

Congestion theory is a branch of traffic engineering dealing with a telecommunication system's ability to manage traffic offered to a network. Traffic in this context is defined as an aggregate of messages or calls that may pass through the network's shared facilities, such as trunks. Traffic intensity is measured as the product of the number of calls C and their average duration or holding time during a specified time. Thus if C calls pass through the circuit switched network in T hours, the traffic carried is CT Erlangs. If a call lasts for 1 hour, the traffic carried by the network for the call is 1 Erlang. In the USA and Canada, traffic intensity is measured as hundreds of call-seconds or CCS (centum call second) per hour; thus CCS/hour = C.H, where C is the average call rate per hour and H the average call duration in hundreds of seconds. The measure of a call's duration is holding time, that is, the time when a circuit is occupied by a call. Other definitions of holding time may be used by operators for calculating revenue, taking into consideration signalling overheads during call set-up.

A network is usually sized to offer an acceptable QoS at the busiest hour of the network. There are a number of definitions of busiest-hour traffic; for example CCITT recommendations Q80 and Q87 define it as the average of busy hour traffic for the five busiest days of the year. GOS is a measure of congestion in a network specified for the busiest period. Grade of service is defined as the probability that a call offered to the network fails to obtain a circuit at its first attempt in the busiest hour. A 2% GOS implies that out of 100 call attempts made, two were blocked due to congestion. A 0% GOS is ideal but would require a large number of circuits and is therefore wasteful of network resources. Operators allow some calls to be lost due to congestion. Mathematical statistics are used to estimate the number of circuits required to achieve a specified GOS, and therefore their applicability to networks depends on the accuracy of the model compared to real traffic.

The Poisson model assumes that there are infinite sources, lost calls are held, calls arrive randomly and holding-time distribution is negative exponential. In practice, the condition of infinite sources is satisfied approximately when the ratio of number of sources to number of circuits is more than ~ 20. Lost calls held implies that when a call is not satisfied at the first attempt, it is held for a duration equal to the holding time, and if it is satisfied subsequently, the call is cleared at the end of the remaining of the holding time however, the call is considered lost in any case. In practice, it has been observed that the call holding time of telephone conversations follows a negative exponential distribution. The probability P of exactly N circuits being busy for a mean offered load of A is given as

$$P = \frac{A^N}{N!}e^{-A} \tag{9.1}$$

where $N =$ number of trunks and $A =$ offered traffic (Erlangs).

The Erlang B model assumptions are identical to the Poisson model assumptions except that here lost calls are cleared. If a call is not satisfied at the first attempt it is cleared and does not appear again during the busy hour. The probability P of exactly N circuits being busy for a mean offered load of A P is given as:

$$P = \frac{A^N/N!}{1 + A + A^2/2! + A^3/3! + ... + A^N/N!} \tag{9.2}$$

The Erlang C model assumptions are also identical to the Poisson model assumptions except that here the lost calls are delayed. If a call is not satisfied at the first attempt, the source continues to demand a circuit until a circuit becomes available, at which time the idle circuit is seized for the duration of the holding time. The probability of exactly N circuits being busy for a mean offered in this model is given as:

$$P = \frac{(A^N/N!)(N/(N - A))}{1 + A + A^2/2! + A^3/3! + ... + (A^N/N!)(N/(N - A))} \tag{9.3}$$

Figure 9.18 compares the number of channels versus GOS for various levels of traffic for these three models. Notice the sensitivity of the models to the number of circuits. The Poisson model gives the most conservative estimate of the traffic; thus to carry say 100

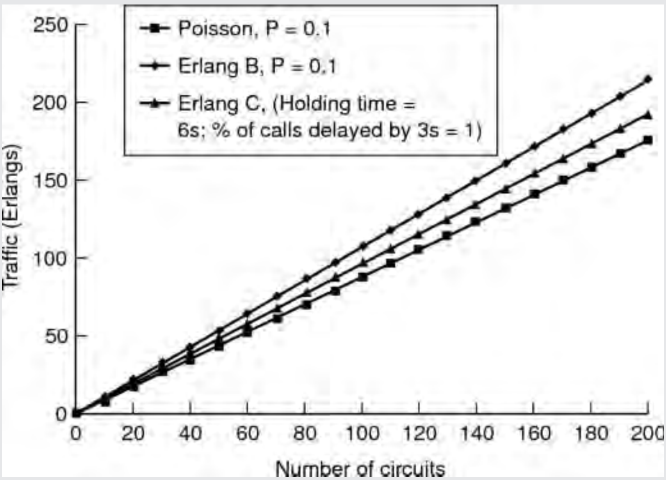


Figure 9.18 A comparison of Poisson, Erlang B and Erlang C traffic models

Erlangs at 10% GOS the Poisson model predicts a circuit of about 110 circuits, whereas, the Erlang model predicts only 90 circuits.

Voice over IP

When considering bandwidth of VoIP traffic similar principles can be applied, since VoIP uses real-time transport protocol. However, the bandwidth calculation must take others into consideration, which include voice codec bit rate, sampling rate, voice activity, RTP header consideration and the VoIP network topology (point-to-point versus point-to-multipoint) (Cisco, 2013).

9.4.2 EIRP Management

We have observed that EIRP demands on MSS service links are high to compensate for the low sensitivity of mobile receivers and offset propagation impairments. This requirement constrains the space segment capacity – particularly for geostationary satellites – due to limitations on spacecraft antenna gain, power-generation and HPA technologies. A typical average forward link power requirement of an L-band geostationary satellite to support medium bit rate communications on a large mobile terminal is of the order of 25 dBW. Thus a spacecraft capable of transmitting a total of 60 dBW in the service link, assuming 10% EIRP overhead for signalling and network management functions, would carry about 2850 radio carriers or 285 000 users, assuming an average traffic intensity of 10 milli-Erlangs per user at the peak load. Now consider further the scenario when the forward service link EIRP is not maintained within limits, resulting in an overdrive of 1 dB in the satellite EIRP (i.e. carriers are transmitted on an average at a higher than the nominal EIRP); this would cause a reduction in carrier capacity by about 20%! In practice, spacecraft power is not hard-limited and therefore the same capacity could be served with a signal quality reduced due to increased inter-modulation noise and carrier suppression – the extent of each depending on whether an automatic-level-control sub-system is incorporated in the transponder. Furthermore, a higher carrier level than nominal, results in harmful interference to other users, whereas a lower level makes the carrier susceptible to harmful interference.

It has been observed in practice that a common source of forward EIRP power loss is due to inadvertent transmission of excessive high power by one or more of the participating fixed earth stations. The network operations centre (NOC) keeps a vigil on such anomalous transmissions to identify, prevent and resolve such instances. Although undesirable but power stealing of a transponder by unscrupulous operators cannot be ruled out and again operators employ countermeasures to identify and eliminate such sources.

The return link satellite EIRP requirements are significantly lower due to the low EIRP transmissions of user terminals – for the example considered previously, depending on the ground earth station size, the requirement would be lower by 20–40 dB. However, transmissions emanating from terminals operating through other satellites can be received at relatively high levels (particularly if there is a coverage overlap) because of the low directivity of MSS user terminals. Such transmissions are retransmitted in the return link by the wanted satellite causing 'power stealing'. This type of power loss can be minimized by: switching off radio-frequency channels used by adjacent satellites; deploying regenerative transponders that can reject unwanted signals; when possible, tailoring satellite antenna patterns such as to reject signals emanating from unwanted directions.

From the preceding discussions, we infer the necessity to utilize, monitor and maintain EIRP efficiently to maximize spacecraft resource, maintain signal quality and minimize RF interference.

A number of techniques are incorporated in radio link design to maximize forward link EIRP utilization. Dynamic power control ensures optimum satellite EIRP utilization to achieve the desired signal quality, although the latency caused by feedback can limit the advantage in applications intolerant to delay. The advantage gained through power control depends on signal fade distribution amongst users. If a majority of users operate under fading signal conditions, then a larger than nominal EIRP is drawn from the satellites that could lead to an overdrive; but when the majority of users operate in clear sky conditions the satellite EIRP is distributed equitably. Voice (or data) activated carriers reduce average power and the average inter-modulation noise. Signals are transmitted only when information is present; since the carrier remains switched off during the remaining time, the average EIRP per carrier is reduced. Other EIRP reduction techniques include coding (at the physical or upper layers), and store and forward communication; each reduces the EIRP by introducing redundancy in the signal to make it more tolerant to noise. Packet-switched networks using multiplexed RF streams, however, do not incorporate the voice activity advantage directly; the advantage can instead be accrued by reallocating the capacity to other users during pauses.

Diurnal satellite EIRP variation provides a measure of spacecraft load profile (see Figure 9.11a). Trending the peak EIRP provides information necessary to predict the date of EIRP saturation to assist operational planning. Typically, this sets a basis to procure higher capacity (next-generation) satellites. Temporary mitigation techniques in such conditions utilize operational techniques such dual-satellite operation where two or more satellites are collocated sharing their EIRP within the allocated spectrum.

EIRP management is also essential in ground earth stations. The NOC may budget an EIRP to each earth station, according to its requirement. A level-monitoring feedback loop at the ground station would ensure compliance to the budgeted EIRP.

9.4.2.1 Power and Bandwidth Limitation

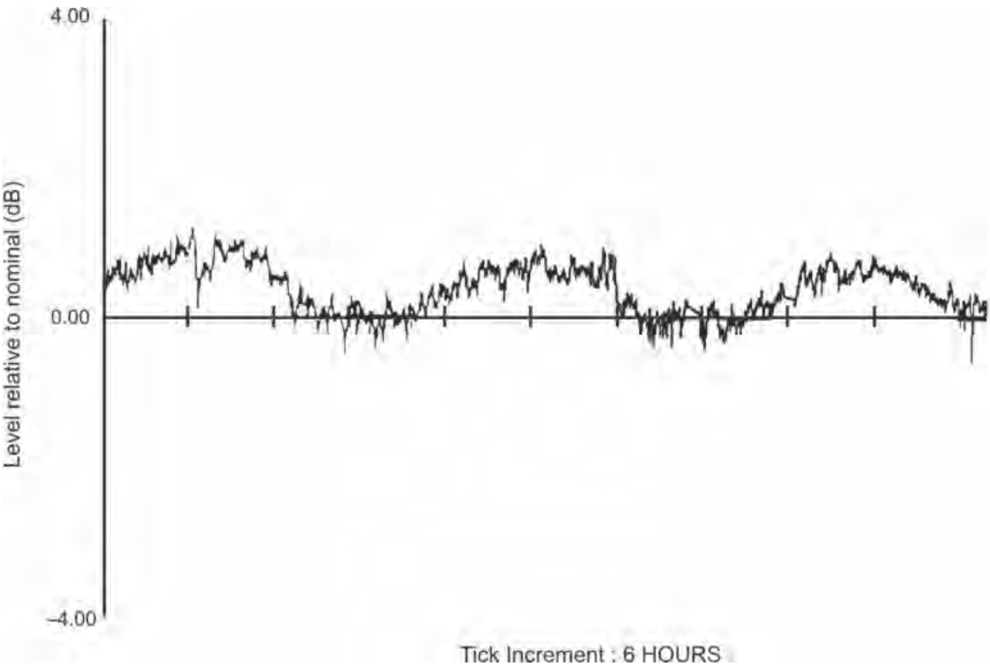
A satellite is said to be EIRP limited when it is unable to provide the desired power for the given traffic load. On the other extreme the system can become spectrum limited when the available spectrum is insufficient to comply the service needs. The latter situation may occur despite the availability of satellite bandwidth when the allocated spectrum to an operator falls short of real demands.

9.5 Radio Frequency Monitoring

To keep a vigil on transmission levels and frequency, intra and intersystem interference, unauthorized transmissions, and to identify faulty mobiles or gateways, satellite transmissions must be monitored continuously. RF emissions are readily monitored on spectrum analysers at suitably sited monitoring stations; a computer-controlled spectrum-analyser measurement system offers operators the capability to measure accurately numerous parameters. The remote stations are typically controlled centrally from the network control centre using dedicated links. To maintain accuracy calibration is performed regularly often several times a day. A typical sample of measurements (including alarms in case of an anomaly) includes:

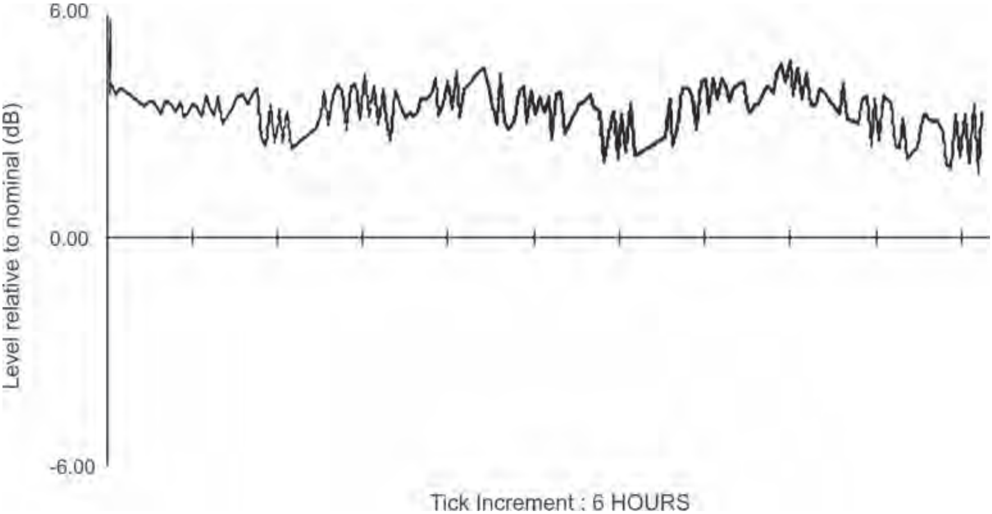
- EIRP of a single modulated or unmodulated RF carrier, a group of carriers or a full satellite transponder;
- Carrier centre frequency;
- C/No and Eb/No;
- Periodic automatic monitoring of authorized carries;
- Detection of unauthorized transmissions;
- Detection of interfering carrier.

Figure 9.19(a–c) represents typical measurements of a satellite spectrum monitoring system (Inmarsat, 1994) illustrating monitoring of various parameters. Figure 9.19(a) shows a plot of a well-behaved carrier for a period over about 2.625 days sampled every 20 min. The cyclic variation is caused by variation in the gain of the transponder due to traffic loading.

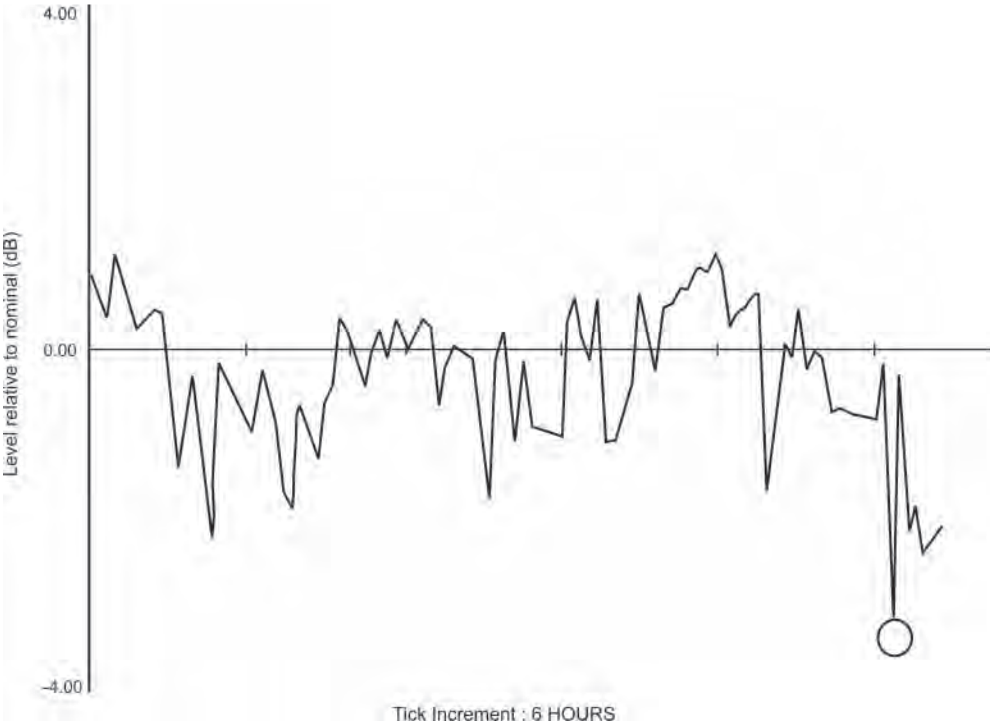


(a)

Figure 9.19 (a) Variation in the EIRP of a carrier over 2.625 days



(b)



(c)

Figure 9.19 (b) Variation of an overdriven carrier over a period of 2.625 days. (c) Behaviour of an unstable carrier over a period of 1.625 days. (All parts source: Inmarsat, 1994. Reproduced with permission of Inmarsat.)

Figure 9.19(b) shows a carrier operating above its nominal level and hence requiring attention by the transmitting station. An alarm would be raised on the operator's console in such a situation and the NOC would instruct the offending LES to correct the EIRP. Figure 9.19(c) illustrates an incorrectly behaving carrier with a fluctuating EIRP and a spike at the circled point. The NOC would investigate the problem associated with the carrier.

A number of other tasks constitute an essential part of operation; these include interference management, network anomaly investigations, invoking back-up procedures in case of a transponder fault and satellite failure, calibration of reference carriers, assistance to earth stations and mobile earth station at commissioning by antenna radiation pattern measurement, transmission stability measurement and end-end signal quality and protocol tests, etc. Off-line tasks include frequency planning, network usage analysis and coordination with participating entities.

One of the problems in managing the RF integrity of an MSS network is the difficulty in monitoring forward link emissions of all the spot beams, because of their spatial separation – exceeding hundreds of km with the current technology. The scope of the problem depends on the number of spot beams and whether a geostationary or a non-geostationary system is being considered. Emissions in the return link are received centrally at a few large earth stations in one or a few spot beams and are thus relatively straightforward to monitor and synthesize on the ground. As already mentioned, EIRP monitoring and control are critical in the forward direction.

Consider first the monitoring of the forward service link of a geostationary satellite system. When the number of spot beams is small, say < 10, measurement stations can be deployed in each spot beam (or at a junction of spot beams to minimize their number) and retrieve the measurements over a land or satellite link at a central site where signals are reconstructed and synthesized. In such cases, the number of monitoring sites can be reduced by locating each site to maximize the number of visible spot beams from each monitoring station, as illustrated in Figure 9.20. In practice, such ideal locations may not always be feasible; it would be extremely difficult to install monitoring stations within spot beams that illuminate oceans or inhospitable terrains.

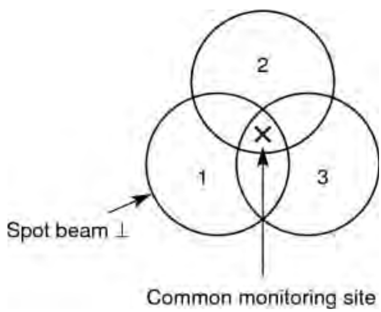


Figure 9.20 The concept of minimizing the number of monitoring sites by maximizing the number of simultaneously visibility of spot beams. In this example, three spot beams can be monitored from the same site

Such schemes are impractical in geostationary systems deploying hundreds of beams. We suggest a scheme here wherein monitoring information is transmitted to a central site by cooperative mobile earth stations or a number of low cost ruggedized transreceiver installed in various spot beams. Such terminals would periodically transmit measured parameters such as RF level, Eb/No and centre frequency of key RF carriers for collation at a central site.

Furthermore, we suggest below a few alternative monitoring schemes that can be used to augment the approaches outlined earlier, or deployed on their own.

- Transmissions can be monitored at each fixed station through a directional coupler at the output of the transmitter and measurements transferred to the central site through a land or satellite link, where signals are processed and synthesized to obtain the RF profile of participating stations and the network. A simple mathematical transformation together with spacecraft transponder gain data can provide a reasonably accurate estimate of satellite EIRP. The technique can provide pseudo-monitoring of satellite transmissions in the forward link and is adequate for policing transmissions from the network's participating ground stations but does not monitor real satellite emissions, thus missing satellite-generated or extraneous spurious transmissions and interference.
- Statistical monitoring of spot beams in the forward and return direction can be achieved in systems that use power control and position reporting by extracting from each active mobile's signalling message a measure of signal quality in the forward direction and an estimate of the return link satellite EIRP. By compiling this information statistically, it is then possible to build a model of EIRP distribution over each spot beam as well as boundaries. Any subsequent deviation from the model would indicate a potential problem. Failure of a spot beam identifier or signalling channels would be identified through analysis of signalling messages. For instance, an absence of return link signalling messages would indicate a problem in the forward signalling or failure of the spot beam identifier channel; an increase in the level of spot beam identifier channel EIRP would manifest through mobiles signalling from outside of spot beam limits. Such a monitoring scheme would be non-real time. Note that interference would not be detectable in such a scheme. The scheme will require to filter out anomalous samples such as those emanating from faulty mobiles or those in deep fade.
- The spacecraft architecture could include a sub-system to measure the signal strength of carriers or groups of carriers and transmit the data to the ground. Simple and coarse monitoring schemes could be based on monitoring current flow through spacecraft RF amplifiers.
- Off-line analysis of usage pattern of each RF carrier can provide vital information about integrity of each channel.
- Optimal combination of monitoring information from various sources such as identified above using schemes such as Kalman filter.

Measurement of total satellite EIRP over 24 h requires periodic sampling of the spacecraft EIRP in each beam and an appropriate summation when antenna gains differ between beams. In such cases, EIRP is referenced to a suitable point, such as the input of the spacecraft antenna:

$$P_n = EIRP_n - G_n$$
$$P_n(\text{Watts}) = 10\{(P_n/10)$$
$$P_t = SP_n(\text{Watts}) \tag{9.4}$$

where P_n = EIRP (dBW) of beam n at the reference point, $EIRP_n$ = total edge of coverage EIRP (dBW) of carriers in spot beam n and G_n = edge of coverage gain (dB) of spot beam n. P_t = total satellite EIRP at the reference point, S = total number of beams.

Due to the presence of voice-activated demand assigned carriers, EIRP measurements must be averaged.

Continuous monitoring of the emissions of non-geostationary satellite systems would require a number of monitoring stations dispersed throughout the coverage. Monitoring schemes such as proposed for the super-geostationary satellite could be adapted. Intermittent monitoring in such systems is feasible with one or more monitoring stations placed in each orbital plane. Monitoring at gateways, statistical monitoring through information gathered from signalling and off-line monitoring scheme as discussed previously, offer other applicable alternatives.

9.5.1 Radio Frequency Interference

MSS systems are susceptible to radio frequency interference (RFI) for the following reasons:

Mobiles can move freely within the coverage area and hence are susceptible to local radio interference. MSS allocations are used by other satellite and terrestrial services such as fixed satellite service (FSS) and radio relays in some regions as secondary allocations. In practice, it is difficult to enforce ceasing of transmissions from such offending sources.

Harmonics from local transmissions can cause interference or saturation of the mobile receiver front end. Saturation occurs because the front end of a mobile is sensitive and often wideband, allowing strong unwanted radio signals to pass through.

Due to the large number of mobiles and possibly gateways, as well as the number of carriers transmitted from each gateway, the probability of interference from equipment malfunction is high.

It has been observed that in practice the probability of intra-system interference is more likely than inter-system interference because of the large number of intra-system frequency reuses typical of MSS networks and the number of operational transmitters. For example, if a spot beam size increases due to a higher level of spot beam identifier, or a specific carrier EIRP exceeds specifications, then co-channel interference levels are likely to increase due to insufficient spatial isolation of the affected spot beam.

9.5.2 Radio Frequency Interference Management

Recall that in order to maximize spectrum utilization by mutual sharing, a small percentage of interference is budgeted in radio link design. Here we refer to interference to be at a level such that it is harmful, that is it causes unwarranted degradation to the users' received signal quality.

In the preceding section we observed that interference can be caused by equipment malfunction or can be inter-system. Equipment malfunction can occur at a mobile or a fixed earth station. Humans can cause interference intentionally with malicious intent or inadvertently such as by transmitting at the incorrect frequency or polarization. Inter-system spillover can be caused by terrestrial systems or by inadvertent transmissions at incorrect frequency from another network. Transponder noise can spill into another operator's band because of an insufficient transmit-filter skirt. Intra-system interference, one of the main sources of interference, is caused by numerous sources. Retransmissions of the received signal is quite common where the received band (or a part of it) gets coupled to the transmit chain resulting in retransmissions of the band at a low level. These signals appear as echo of the main signal causing annoyance.

According to an industry organization called the Satellite Interference Reduction Group (sIRG) who promote a number of initiatives to reduce interference in satellite networks (sIRG, 2012):

> The satellite industry loses millions of dollars per year down to cases of interference and a great deal of manpower has to be given over to discovering its causes. RF interference is caused by human error, bad installation, lack of training, poor equipment or system design and a lack of adherence to industry standards and guidelines. Occasionally, interference may be malicious, but this is rare and the main issues of interference lie largely within the heart of the satellite industry itself. The orbital spacing of satellites is being reduced and the fill rate is getting higher. It is getting crowded up there, leading to increased interference.

RFI management involves detection and elimination of harmful interference. RFI causes an increase in the noise floor of the affected channels, causing degradation to the BER (or E_b/N_0). If interference emanates from channels of similar transmission format, it can result in an intelligible cross-talk. If the interference is near the noise floor it tends to appear as noise – often spectrally shaped – to the detection circuit. Interference signals sometimes remain buried below the wanted signal. In practice interference can take a variety of characteristics – sweeping carrier, bursts, hopping carrier with regular or intermittent occurrence.

RFI can affect a single carrier or a group of carriers and can occur locally or over a wide area. Detection of RFI to demand assigned carriers becomes problematic as the occurrence is confined to single calls, which typically last 2 and 3 min for circuit-mode transmissions and may occur only in a part of the service area. In practice on obtaining a noisy channel, chances are that the chances are that user terminates the call and attempts to make another call. Thus, the impact of RFI on one or a few channels can remain undetected over long periods – particularly if confined within small areas. RFI in such cases manifests itself in call statistics, as a reduction in holding time, an increase in call failures on the interfered RF channel. An operator can obtain such information by user feedback maintaining a database of previous interference reports. In many cases, such incidents can be traced to faulty user equipment and thus careful cataloguing of such information is vital. Interference occurring globally, for example due to spurious transmissions through a satellite, can be detected by regular monitoring. Unauthorized transmissions are detected on the basis that they are absent from the authorized frequency list.

Characterizing interfering signal assists in isolating the interference source. One technique is to maintain spectral characteristics of the carriers used in the network and correlate the interference with each. By identifying the carrier type narrows the possible offending source, if not isolate it (Figure 9.21). Adjacent channel interference or cross-polarization interference are often the source of interference. Figure 9.21(a) shows example spectral signatures of an unfiltered binary phase shift keying (BPSK) channel at 1.2 kSps in the L band extending about 10 kHz. Figure 9.21(b) shows the spectral signature of a filtered BPSK 6.0 kSps carrier. Figure 9.21(c) illustrates the spectrum of a filtered offset-quadrature phase shift keying (O-QPSK) at 4 kSps occupying about 10 kHz.

It is relatively simpler to detect interference to signalling channels as it affects all or a significant number of users. Furthermore, the performance of signalling carriers in a system is closely monitored and any degradation in their performance can be detected rapidly through the associated system malfunction. For example, interference to a broadcast carrier

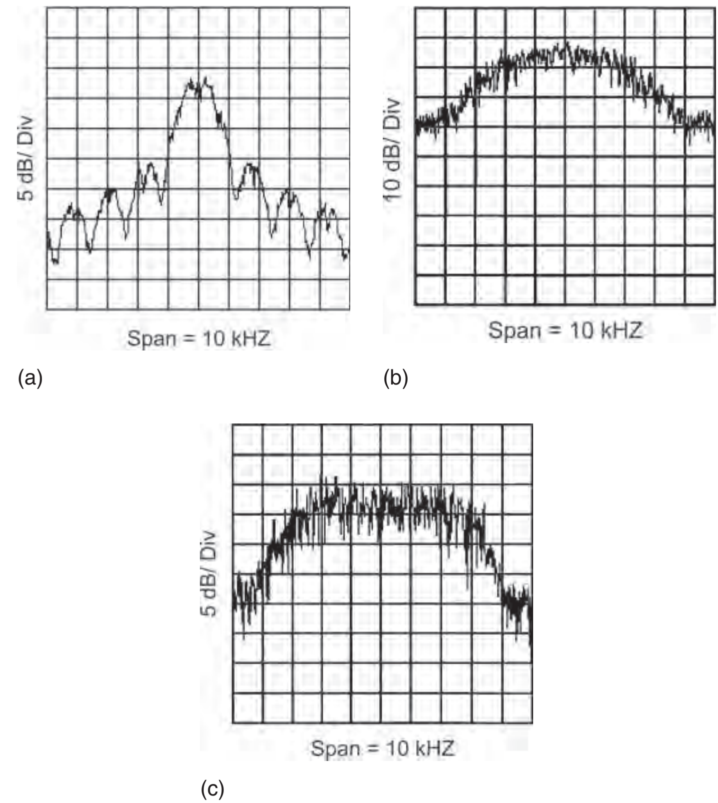


Figure 9.21 (a) Unfiltered BPSK spectrum analysis in the L band; spectrum analyser setting: span = 10 kHz, resolution bandwidth = 300 Hz and video bandwidth = 100 Hz, sweep = 839 ms. (b) Filtered BPSK spectral signature for a carrier transmitting at rate of 6.0 kS/s at the L band; spectrum analyser setting: span = 10 kHz, resolution bandwidth = 300 Hz, video bandwidth = 300 Hz, sweep = 670 ms. (c) Filtered O-QPSK spectrum analysis in the L band. spectrum analyser setting: span = 10 kHz, resolution bandwidth = 1 kHz and video bandwidth = 1 kHz, sweep = 200 ms. (All parts source: Inmarsat, 1994. Reproduced with permission of Inmarsat.)

will manifest in a distortion to the network traffic trend. Intra-system interferences can be eliminated readily through internal automated and manual procedures.

Interference elimination can be an arduous task and in severe conditions when vital system channels are involved, interference can seriously affect the functioning of an MSS system. Countermeasures in such extreme conditions include transferring operations to back-up frequencies, which are known to the fixed network and 'burnt' in mobile terminals. Techniques used in practice for interference elimination include:

- Correlation of spectral signature with a known set;
- Time correlation with possible sources (e.g. television transmitters, radars, etc.);
- Manual procedures to identify the offending fixed station (e.g. instructions to check transmitted spectrum in case of interference by retransmissions; depending on the type of interference, specific instructions to participating earth stations);
- In case of an over-deviating or incorrectly assigned channel, switching it off by removing it from the assignment list;
- In case of inter-system interference, coordinating with the identified operator;
- Locating the offending source.

Interferer source location techniques for geostationary systems are based on triangulation, Doppler or spectral signature detection techniques. In one geolocation system the time difference of arrival (TDOA) and frequency difference of arrival (FDOA) are measured for the interfering signal through two satellites. Using the ephemeris of each satellite the locus of TDOA and FDOA are estimated on the surface of the Earth. The interferer lies at the intersection of these two curves (Downey, Constantino and Chu, 2009). In a MSS system, the Doppler and time signatures can get distorted due to motion of mobiles and even with fixed terminals estimate of TDOA is relatively more difficult. Thus if the system must rely on TDOA there is need for a third satellite to get another curve. In an MSS network, even after locating the interfering source the operator may not have a communication channel to instruct the operator of the remote offending source. A tractable solution in such a situation is to introduce features in signalling to force interfering mobiles to switch off transmissions on command and include features in mobiles to minimize inadvertent transmissions due to equipment malfunction.

Since a majority of interference events in commercial satellite systems is unintentional, sIRG has been promoting the use of an industry-wide carrier identification system wherein an identifier is tagged to each transmission as an embedded code that would provide details to trace and contact the source. This would significantly reduce the effort in eliminating large numbers of events. Several interference location systems claiming accuracies of less than 1 km are available commercially and have been deployed in practice by major fixed satellite operators to counteract an increase in instances of interference and consequent loss in revenue.

9.6 Quality of Service

In a general terminology, QoS is a user perception of the telecommunication service. Network managers use quantifiable measures in order to provide the desired quality.

1. Grade of service: The user perceives the GOS as congestion, which is the net effect of the terrestrial and space segments. Congestion of the space segment can be obtained at the NCS where radio resources are assigned. Gateways may not have adequate capacity, especially if the traffic has built up rapidly in a short period, or when unforeseen traffic surges occur and congestion arises in such situations. Economics do not warrant earth station operators installing extra capacity that remains idle for considerable periods of time. We have already discussed issues related to space segment capacity and mentioned that it is governed by spectrum availability and the efficiency with which the spectrum is utlized by an operator.
2. Regular BER and E_b/N_o measurements, ARQ repeat assessment and test transmissions on randomly chosen traffic channels can provide useful indicators of signal quality. Such measurements can be incorporated in to RF and baseband monitoring equipment.
3. Customer feedback is a vital source of information, as it is extremely difficult to monitor each location covered by MSS systems.
4. Statistical monitoring of call records provides another measure of off-line quality assessment.
5. Packet data transmission systems define QoS in a number of ways, such as average throughput, delay and jitter which can be monitored through measurements and test calls.
6. General user perception can be assessed by regular test calls.

9.7 Licensing Issues

A vital requirement of a prospective operator is to obtain a license from the regulatory authorities of the countries to which service is targeted. In most cases, the MSS services extend to a number of countries and therefore operating license has to be obtained from various authorities, which makes this process complex and time-consuming. Operators such as Inmarsat, Globalstar and Iridium have spent and continue to spend consider time and effort on this issue as the process is fraught with political and regulatory hurdles. Countries are often worried about security and potential loss of revenue to local telecommunication operators.

Licences may often be awarded to more than one operator to encourage competition; operators may belong to the country of jurisdiction or to another country. In such cases, the country may favour local operators, or license each operator to encourage a competitive market environment.

There are wide variations in the approach adopted for granting MSS spectrum license, although traditionally the approach has been to award licenses on first-come, first-served basis following established procedures recommended by the ITU (see Chapter 3). With increasing commercial competition, regulatory authorities, such as the FCC in the USA, are using alternative methods to encourage efficient use of spectrum. Sometimes a charge is levied to recover administrative costs. Costs may sometimes be increased to encourage a more efficient use of spectrum with the rationale that the operator will then try to increase the returns by improving spectrum efficiency. Moreover, when spectrum shortage is acute, the licensing costs are increased to reduce the number of competitors. Authorities may invite potential operators to participate in competitive bidding and select the licensee on the basis

Table 9.2 Possible approaches for spectrum licensing with their merits and demerits

Type of licensing	Advantage	Disadvantage
First come first served	Well proven Simple to implement	Favours early entrants Does not encourage competition Accommodating late entrants becomes increasingly difficult Not well suited when there is acute spectrum shortage Spectrum usage not well controlled
Bidding	Possible to select the best applicants	Favours established operators
	Encourages efficient use of spectrum	Favours those with committed financial resources
	Encourages innovative technology	May increase the cost of the space segment thus pricing the operator out of business
Auction	Offers best price when spectrum is scarce	High initial cost may lead to investment uncertainty
	Spectrum awarded to the operator willing to maximize spectrum usage	Negative impact on long term growth as the operator may concentrate on recovering costs on profitable segments Favours existing operators, as new entrants need to establish an infrastructure in addition to funding license Increases cost to users Service may be offered only to profitable service areas thus users requiring full coverage may be at a disadvantage Care necessary to keep spectrum fragmentation within acceptable bounds or operators may not be able to profit
Conditional (license subject to conditions such as proof of concept, financial resources, etc.)	Safe approach, as spectrum is given after techniques/finance has been demonstrated	Operator may require up-front investments
	Spectrum not tied up	Risky if there is delay in project, e.g. launch failure, financiers backing out

(Adapted from ITU, 2005.)

of criteria such as technical excellence, funding arrangement, level of risk in the proposal, etc. A license may only be granted on a conditional basis, such as proof of technical concept or funding. More recently, spectrum auctions have been used for licensing terrestrial mobile systems. A major concern when developing a charging policy for MSS licenses is that if each jurisdiction charges an operator a licensing fee, the charges may become significant and prohibitive, as satellite operators usually provide services to a large number of countries. Some of the possible approaches with their respective merits and demerits are summarized in Table 9.2 (ITU, 2005).

A study conducted by the UMTS forum, using an economic business model for terrestrial systems, demonstrated that increasing the price beyond the administration cost is likely to have a negative impact on the development of future UMTS services (UMTS, 1998). The analysis for terrestrial UMTS indicated that in such a scenario profitability and pay-back periods will increase and be detrimental beyond a threshold.

Revision

1. What are the entities of an MSS network management system, and their specific goals?
2. What are the parameters that must be monitored to ensure a satisfactory quality of service in an MSS network?
3. Explain the process of spectrum management in an operational MSS, including the process of generating frequency plans.
4. What are the similarities and differences in the frequency planning process of GEO and NGEO satellite systems?
5. Explain the methods used in short-term and medium-term spectrum planning of an operational mobile satellite system.
6. Outline the methods used to monitor satellite transmissions and interference management in an MSS network.
7. Various approaches are used in granting spectrum licenses. List the advantages and disadvantages of approaches applicable to MSS.

References

Bjelajac, B. (1996) CIR based dynamic channel allocation schemes and hand-over prioritisation for mobile satellite systems. Vehicular Technology Conference, Atlanta, GA, 28 April–1 May, pp. 313–318.

Castelino, D.J., Hurley, S. and Stephens, N.M. (1996) A Tabu search algorithm for frequency assignment. *Annals of Operations Research*, **63**, 223–250.

CISCO (2013) Traffic Analysis for Voice over IP, Available at: www.cisco.com/swa/i/icon_pdf.gif (accessed 5 April, 2013).

Cozzens, M.B. and Wang, D.I. (1984) The general channel assignment problem. *Congressus Numerantium*, **41**, 115–129.

Davis, L. (ed) (1991) *A Handbook of Genetic Algorithms*. Van Nostrand, New York.

Del Re, E., Fantacci, R. and Giambene, G. (1994) Performance analysis of a dynamic channel allocation technique for satellite mobile cellular networks. *International Journal of Satellite Communications*, **12**, 25–32.

Downey, M.L., Constantino, P and Chu, J.C. (2009) Interference detection+geolocation Tools. Sat Magazine (Feb.), www.satmagazine.com/cgi-bin/display_article.cgi?number=2035208231#top (accessed 27 August 2013).

Finean, R. (1996) The role of satellites in UMTS. PhD thesis. School of Electronic Engineering, Information Technology and Mathematics, University of Surrey, Surrey.

Goldberg, D.E. (1989) *Genetic Algorithms in Search, Optimisation, and Machine Learning*, Addison-Wesley Publishing Company Inc., New York.

Hale, W.K. (1980) Frequency assignment: theory and applications, *Proceedings of IEEE*, **68**, 1497–1514.

Holland, J.H. (1975) *Adaption in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, MI.

Holland, J.H. (1992) Genetic algorithms, *Scientific American*, **267**, 66–72.

ILOG (2006) www.ilog.com/corporate/profile/index.cfm.

Inmarsat (1992) Project 21: The Development of Personal Mobile Satellite Communication, June.

Inmarsat (1994) An overview of Inmarsat SSMS (Satellite Spectrum Monitoring System), Issue 1, compiled by H. Shaw.

ITU (1987) *Handbook on Telephonometry*, ITU, Geneva.

ITU (1989) Transmission Impairments, Delay. ITU-T G114.

ITU (2005) *Handbook on National Spectrum Management*. ITU, Geneva.

ITU (2010) Detailed specifications of the radio interfaces for the satellite component of International Mobile Telecommunications-2000 (IMT-2000), Recommendation ITU-R M.1850, (01/2010).

Jahn, A. (1999) Stochastic optimisation of satellite frequency assignment. IMSC '99, Sixth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre and the Jet Propulsion Laboratory, pp. 172–179.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983) Optimisation by simulated annealing, *Science*, **220**, 671–80.

Kohler, E., Handley, M. and Floyd, S. (2006) Datagram Congestion Control Protocol (DCCP), IETF, RFC 4340, Mar.

Morgan, W.L. and Gordon, G.D. (1989) *Communications Satellite Handbook*, John Wiley & Sons, Inc., New York.

Oliver, I.M., Smith, D.J. and Holland, J.R.C. (1987) A study of permutation operators on the travelling salesman problem, in Proc. 2nd Int. Conf. on Genetic Algorithms, pp. 224–230.

Pujante, A. and Haro, L. (1997) Optimisation of satellite frequency plans with balanced carriers by simulated annealing, *Electronic Letters*, **33**, 934–5.

Richharia, M. (1992) Satellite Operations: Management of Satellite Resources for Mobile Satellite Services. MSc seminar, University of Surrey, Surrey.

Richharia, M. and Trachtman, E. (2006) Capacity enhancement techniques in mobile satellite systems, 3rd Advanced Satellite Mobile Systems Conference (ASMS), Herrsching am Ammersee Germany, 29–31 May, 8 pages.

Richharia, M., Trachtman, E., Westbrook, L. *et al*. (2006) Frequency planning approaches in mobile satellite systems, 3rd Advance Satellite Mobile System Conference (ASMS 2006), Munich, Germany, 29 May–31 May, 2006, 9 pages.

Sammut, A.J.R. (1999) Radio resource management for satellite personal communication networks. PhD thesis. School of Electronic Engineering, Information Technology and Mathematics, University of Surrey, Surrey.

Sarwar, G., Boreli, R. and Lochin, E. (In Press: 2013) On the quality of VoIP with DCCP for satellite communications. *International Journal of Satellite Communications and Networking*, **30** (4) 163–180. ISSN 1542-0973.

UMTS Forum (1998) Cost Impacts. Report no. 3, UMTS.

10

Commercial Issues

10.1 Introduction

A commercial satellite communications venture is based on the same business principles as other high technology ventures and in this context, Figure 10.1 attempts to capture various network and business entities and value chain of a commercial mobile satellite service (MSS). This chapter discusses the issues of interest from a technological perspective.

In the following discussions we will refer to the space segment provider as the system operator. Other components of an MSS comprise a number of specialist vendors for provision of space and ground sub-systems as portrayed in the figure. The land earth stations and gateways are owned by the system operator or, depending on the business strategy, the service provider. The service provider offers services to the end users directly or through local retailers. Mobile terminal manufacturers, after obtaining type approval from the operator, supply terminals to the users through their distribution chain. Embedded within the value chain are several ancillary businesses for provision of specialist services such as applications development, hardware and software maintenance.

The space segment including the supporting ground infrastructure constitutes the most significant up-front cost, as a glance at Table 10.1 should testify – investments range from \$0.738 billion to \$2.9 billion. When an MSS has achieved a successful operating service, introduction of new or enhanced products over the deployed infrastructure requires a lower investment, unless, of course, the product requires its total refurbishment. But in any event, such investments are quite substantial. Therefore the significance of a realistic business plan cannot be understated.

Suffice to say that the market research conducted by operators during the 1990s concluded that the personal MSS market would support and sustain profitably. Figure 10.2 shows the anticipated number of big-LEO (low earth orbit) subscribers perceived by market projections of the late 1990s projected up to 2014. Analysts' views varied regarding the number of systems that could be sustained, ranging between two and four. Iridium, ICO and Globalstar generally figured in the final reckoning. The actual number of subscribers at the end of 2011 was of the order of 2.2 million (although rising steadily in recent years) compared to the projection of over 30 million. It is hardly surprising that several start-ups went bankrupt due to the error in the market predictions, thus emphasizing the firm need to understand

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

Main assets
- Satellites
- System infrastructure
- Gateways

Business entities
- Space segment provider
- Gateway operator
- Manufacturers
- Vendors
- Service provider

Product
a) End user
  - User Terminal
  - Telecommunication services

b) Business to business
  - Satellites
  - Earth stations
  - Software
  - Hardware
  - Operation support
  - Satellite time

Market (example)
- Individuals
- Fleet owners
- Ship owners
- Airlines
- Government
- Journalists
- Broadcasters

Figure 10.1 Entities of a commercial MSS venture

Table 10.1 Cost estimates of various mobile satellite systems, illustrating the massive initial investments needed several years prior to service availability

System	Orbit/ coverage capability	Initial estimate of cost by operator (B$)
Inmarsat (global express)	GEO/global (except polar regions)	1.2
Thuraya	GEO/regional	>1.0
Globalstar (second generation)	LEO/global	0.738
Iridium (next generation)	LEO/global	2.9

(Source: Inmarsat, 2012; Thuraya, 2012; Globalstar, 2009; Iridium, 2012.)

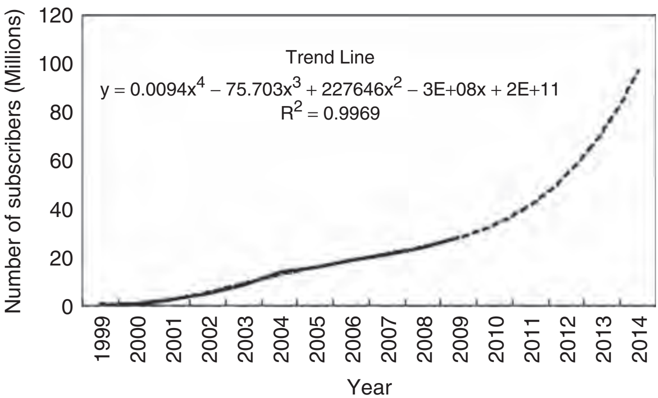


Figure 10.2 Anticipated number of big LEO subscribers perceived by market surveys of late 1990s and trended estimates up to 2014. (Data source: Corbley, 1996.)

the MSS market including the impact of external influences, vastly underestimated in the projections of 1990s. The stupendous growth of terrestrial mobile systems, the advent of roaming (which virtually eliminated the anticipated market of the travelling businessman), the high operational and user terminal and service costs, large size, weight and low aesthetic appeal of the user terminal in comparison to cell phones – were all unforeseen.

Similarly, another market study conducted at that time estimated the subscriber base for little-LEO satellite systems as about 168 million units, comprising 42 million tracking units, 36 million monitoring units and 90 million messaging units in due course (Crossman, 1999). Little-LEO systems support messaging and low bit-rate applications such as asset tracking, machine-to-machine communication, weather monitoring and supervisory control and data acquisition (SCADA) through inexpensive pager-like receivers. Again, these projections vastly over-estimated the market. The billable subscriber communicators of ORBCOMM (the main Little-LEO operator) towards the end of 2012 totalled about 689 000 (ORBCOMM, 2012) against the predictions of about 168 million.

As is now well known, both the incumbent big-LEO operators Iridium and Globalstar faltered in their stride at the outset, filing for bankruptcy. Thankfully, each emerged successfully from the financial trauma, while other ventures failed. Similarly, the little-LEO operator ORBCOMM, after a hesitant financial beginning, managed to retrieve its financial shortcomings after nearly a decade.

The intention here is not to analyse any specific system but to extract factors and issues that influence growth in this particular MSS segment. Predominant influences appear to be changes to the market conditions because of the rapid evolution of competing technologies during the long gestation period of the projects lasting 8–12 years. During the early 1990s, when many big-LEO systems were conceived, cellular telephone sizes were much bigger, large areas of the world were not covered by cellular systems, and international roaming arrangements were in their infancy. By the time the systems were introduced cellular phones were smaller, multi-functional, lighter, attractive and fashionable; the uncovered terrestrial coverage had shrunk enormously, while data rates increased, as the third generation systems rolled out. A user could travel to large parts of the populated world and communicate on the same phone, thanks to inter-operator global roaming arrangements and standardization in the cellular industry. Why would a user not expect from satellite systems a similar quality, type of service, costs and hardware as from a cellular system?

This highlights the need to address the correct market segment and the need of a realistic marketing strategy. People located in areas devoid of a telecommunication infrastructure, are likely to be receptive to the idea of a satellite service despite its limitations. Such a user being aware of the limitations is cooperative when using the satellite service, for example by ensuring that the satellite view remains unobstructed. The products must be made available to such users, which brings in the need for a realistic marketing strategy. Lastly, investors are anxious to receive returns within a stipulated time beyond which funds begin to dry up – so, how can operations be sustained and ageing satellites replaced? A programme of market research, financial analysis and a programme of product and service enhancements to users' needs in face of competition are all a part of a well-conceived business model. Such a model must consider:

- funding (*costs*: capital, operation, maintenance; *source*: own, stock market, bank loans, etc.);
- revenue and returns;
- risks (delays, increased cost, regulatory hurdles, political difficulties, fund scarcity, launch

Consider some basic issues applicable to a mobile satellite personal communications service. These world-wide or regional systems offer voice, data, facsimile, e-mail, messaging, and so on, on a personal telephone or a pluggable subscriber identity module (SIM) card.

Satellite personal systems are generally best targeted for regions where existing services are unavailable or unreliable. Examples are large areas of under-developed parts of Africa or Asia, expanses of developing countries such as India and remote areas of developed countries such as Canada or the USA.

Consider the growth model of cellular systems. The world-wide cellular system penetration pattern clearly illustrates the growth spearheaded by developed countries, followed by uptake of the service by the developing countries when costs and technology mature and become affordable. Can this model be applied to MSSs? Satellite terminal costs and call charges (plus subscription charges, if applicable) are relatively high, which makes the service too expensive for normal usage. Individuals from richer countries and specialists such as journalists, explorers, and so on, are likely to be the key consumers – but this begs the question of whether there are enough of them. While at the outset the cellular systems address densely populated regions with much lower investment and a rich client base, the satellite market addresses the thin-route regions where the number of users is difficult to quantify and the population may not be as affluent and sophisticated. This is one reason why invariably all the MSS operators offer interoperability with terrestrial systems, with dual-mode handsets and commercial agreements with terrestrial operators. It leads to economies of scale through a wider user base. Recent advancement towards tighter integration with terrestrial systems at a network level has, consequently, bolstered the MSS market. Nevertheless, MSS has remained a specialist market and is likely to remain so in the foreseeable future.

10.2 System Planning

To be financially viable, satellite system technology must be market led rather than technology driven. In such an environment, commercial systems benefit greatly in terms of schedule and cost, leading to healthier returns, by using a systematic planning approach. Figure 10.3 shows the main entities and their interaction in an exercise of this nature (adapted from Lazear *et al*., 1997).

A new satellite system or a product within an existing venture is initiated through marketing studies and based on telecommunication needs, preferred size and type of hardware, quality of service expectations, geographical areas, consumers' purchasing power, preferred cost of service, market size and anticipated growth, etc. User requirements have to be converted to revenue and returns through a business model that benefits by incorporating the entire lifecycle of the product (more later). A number of iterations may be necessary between system designers, marketing and business strategists before a realistic set of requirements, technology and schedule materializes. During the design process, there is also a need to incorporate regulatory issues, such as spectrum availability, licensing policy in the target service areas, time lag in completion of each requirement, possibility of obtaining the preferred orbital location and frequency bands, and others. The initial model – which may be based on the operator's own historic data, published reports, documents in the public domain, government business initiatives, empirical models extracted from the literature, and etc. – has to be refined by interaction with manufacturers and vendors. Furthermore, the system design

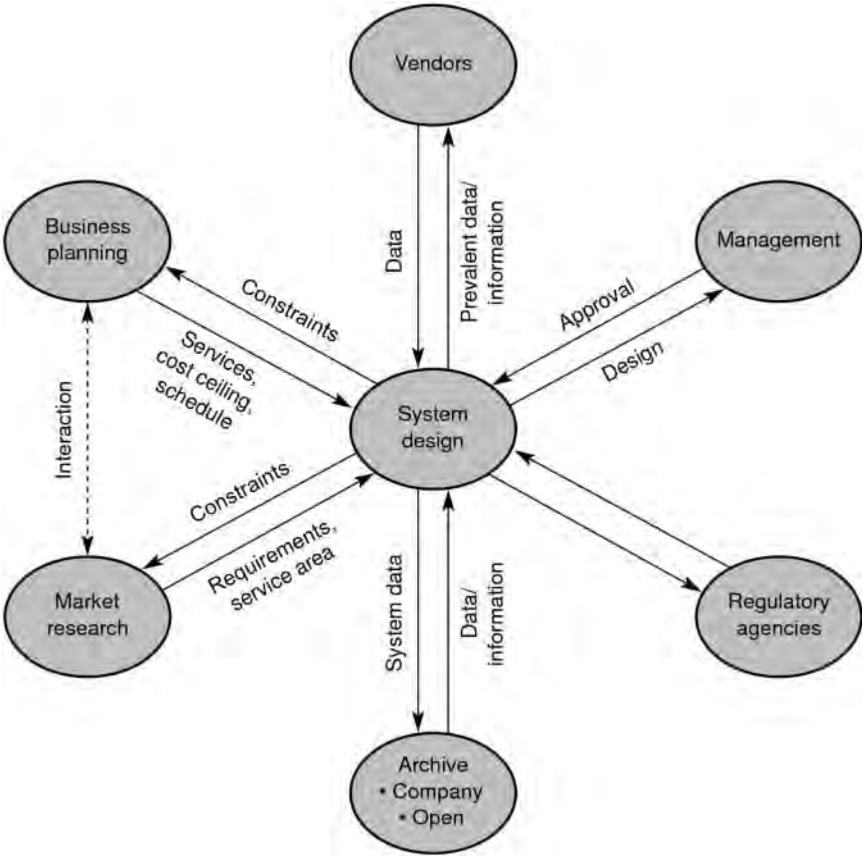


Figure 10.3 Main entities and their interaction in system planning

concepts require industrial validation by way of technology, risks and schedule. At the end of such an interactive design process, a business plan encapsulating system specifications, operational philosophy, a detailed programme schedule, capital and revenue, is ready.

Commercial satellite systems are planned with the goal of achieving the maximum return. Using a system approach, which includes the whole life of the venture, can identify areas of potential savings and risks. This type of analysis also provides a clearer view of the commercial venture for investors and shareholders (especially when billions of dollars are at stake!). Such an approach has found favour with organizations such as ESA and NASA for conducting various types of feasibility analysis, as is evident from their studies and initiatives in this area (Sultan and Groepper, 1999).

A preliminary cost analysis of the mission based on available information such as to provide a rough order of magnitude (ROM) costs, is a useful start to assess viability and readjust the requirements. There are several generic cost estimations models for space segment that may be used for the purpose. In one approach, the space segment cost is partitioned into recurring and non-recurring components and a regression fit is applied to the existing cost

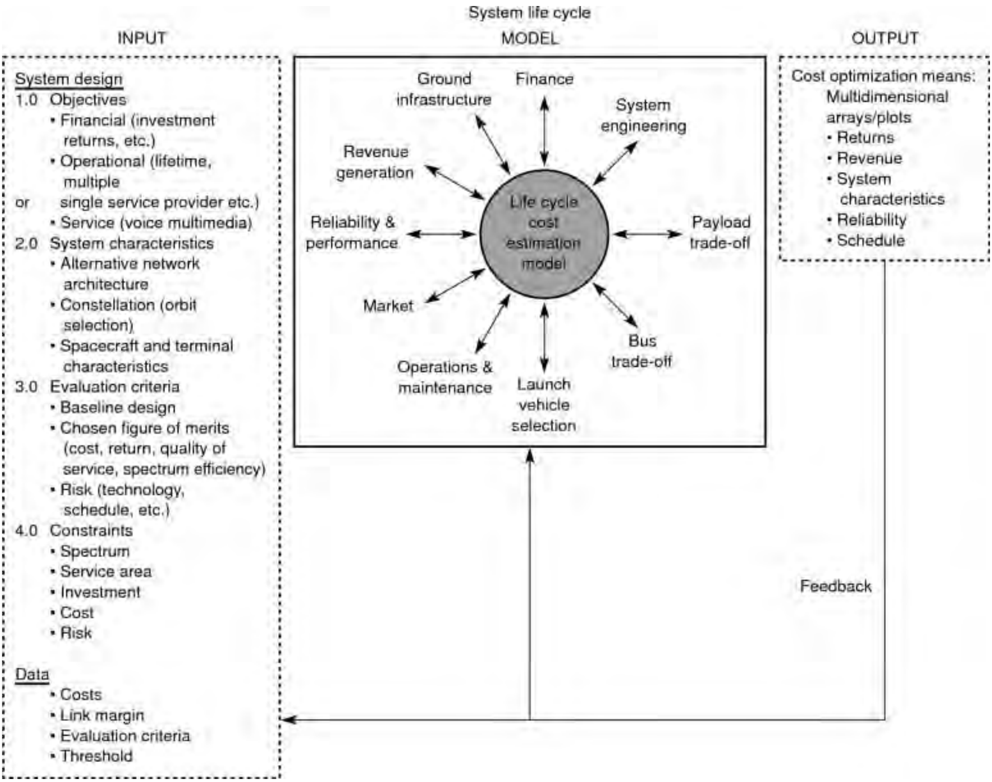


Figure 10.4 A top level model of a lifecycle cost-benefit analysis. (Adapted from Sultan and Groepper, 1999.)

database to provide an empirical model for each component. Lifecycle programme cost estimation, however, requires in addition, ground segment capital costs, operational and maintenance cost for replenishing the space and ground segments, etc. When such a model is developed, returns can be estimated in terms of profitability. Sensitivity analysis of system parameters may be necessary in terms of profit or another figure of merit. An example would be the profit/unit spacecraft effective isotropic radiated power (EIRP) ($/W) or profit/unit bandwidth ($/kHz) or profit in terms of weighted combination of spectrum and satellite EIRP. Figure 10.4 shows a top-level model of such a lifecycle cost-benefit analysis. The inputs to the model may comprise:

- *mission objectives*: financial objectives, lifetime of the mission based on operational considerations, service environment (e.g. land, maritime, aeronautical), service type (e.g. throughput – low, medium or broadband), product range and their lifetime, together with system constraints;
- *system characteristics*: alternative network architectures and concepts, orbital and constellation characteristics, spacecraft and terminal characteristics, system drivers (e.g. user expectations in terms of terminal size, throughput);
- *evaluation criteria*: Baseline designs, figure of merit (cost, financial returns quality of service, spectrum efficiency), risk (technology, schedule, etc.), receiver complexity, system

capacity and cost; penetration rate and revenue; service cost and penetration rate, cost trade-off between owned versus leased space segment;
- *knowledge*: cost models, spacecraft power/mass estimation models, user terminal cost models, risk strategy, growth trend of similar services, inflation trend, traffic distribution;
- *system data*: link margin, frequency of operation, available spectrum, available finance, expected return, historic data of similar products such as unit cost, call charge, and so on;
- other considerations and inputs as listed in Figure 10.4.

The output expected from the model should provide information that allows system optimization. Performance evaluation criteria include: profit and revenue as a function of time; required satellite EIRP and bandwidth for profitability within acceptable time limits; service penetration rate to achieve profitability within the target timeframe, etc.

The feedback loops offer the planner the opportunity to vary inputs and adjust the performance to achieve the desired goal.

The ESA study identified and grouped the main cost drivers of a space mission as illustrated in Figure 10.5 (Sultan and Groepper, 1999).

Engineering trade-off analysis is used by planners at the outset to cut costs with minimal risk and expenditure. The ESA study estimated a cost reduction by a factor of 2 for production of 15 spacecraft when using economies of scale. A cost reduction of up to 35% is possible in medium and large space missions with good management and implementation methods, although quantifying measures for personnel efficiency, motivation and management is difficult. A higher cost reduction can be achieved where better control over teams, resources and facilities is possible. Research and development requires initial investments, but the effort can result in up to 35% reduction in cost over periods of about 10 years when research costs are excluded. Obviously, the magnitudes of savings mentioned here would depend on the assumptions used in the study, but should be indicative of the sensitivity of these cost-saving measures.

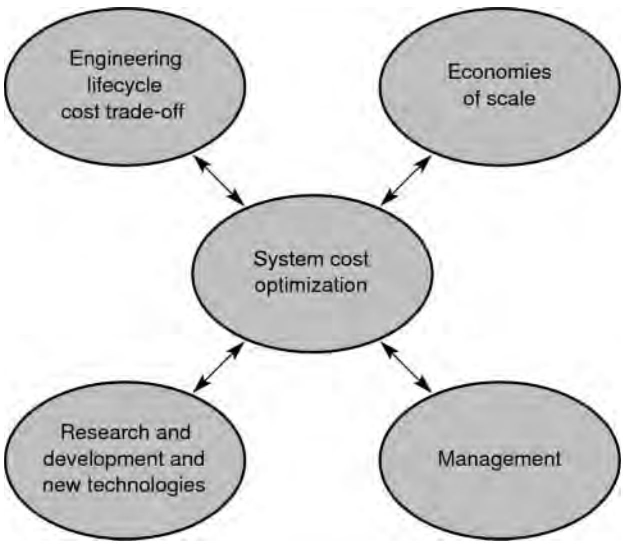


Figure 10.5 Main cost drivers of a space mission. (Adapted from Sultan and Groepper, 1999.)

Initial engineering studies benefit greatly through iterations as a better appreciation of requirements, conceptual design and cost sensitivities is gained. For example, the perception that the lowest spacecraft cost is the best option may not necessarily hold when the venture is viewed in its entirety. This study phase involves close interaction between the business and engineering teams and is vital to the eventual success of a venture.

The study applied the methodology for Inmarsat-3 lifetime optimization using data available in the literature and from previous studies. The results presented here are illustrative of the approach, without claiming any resemblance to implementation. Geostationary systems using 6, 19, 18 and 52 spot beam systems were investigated to serve a range of services to various types of user terminals for maritime, land and aeronautical services. A wide range of parameters was used in the trade-off analysis. Figure 10.6(a) compares four spacecraft

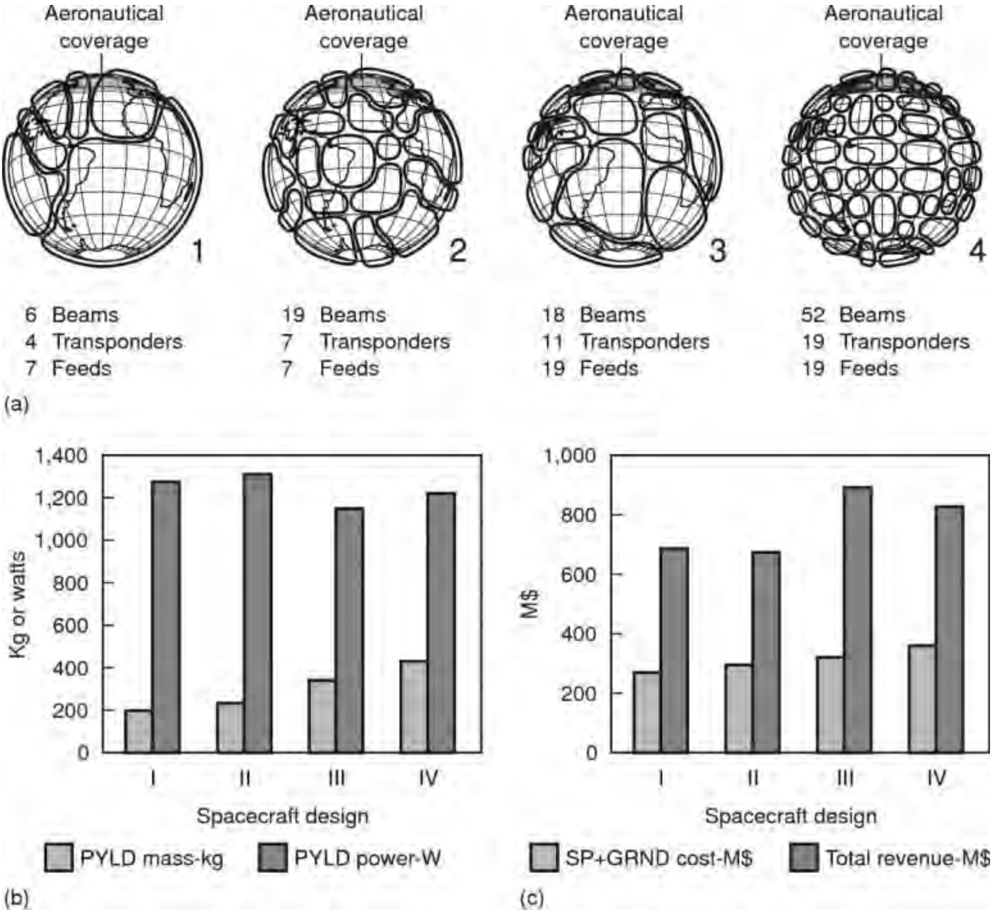


Figure 10.6 (a) Four configurations considered for global coverage by three or four identical geostationary satellites. (b) Spacecraft design versus payload mass (kg) and power (W). (c) Total cost (i.e. space plus ground segment) and revenue over a 10-year period for each design. (All parts source: Sultan and Groepper, 1999, IMSC '99, The Sixth International Mobile Satellite Conference, Ottawa, 1999, co-sponsored by Communications Research Centre and the Jet Propulsion Laboratory.)

Table 10.2 System breakdown cost of an FSS service

Segment	Percentage of total cost
Satellite	18
Launch	13
Spacecraft control	5
Network control	5
Ground terminals	35
Operation and maintenance	24

(Adapted from Lazear *et al.*, 1997. Fifth International Mobile Satellite Conference, Pasadena, California, June 16–18, 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97-11, Jet Propulsion Laboratory, Pasadena, California, June 16, 1997.)

designs for global coverage by three or four identical geostationary satellites, Figure 10.6(b) shows spacecraft designs versus payload and mass for these designs and Figure 10.6(c) demonstrates the cost and revenue of the system over a 10-year period. Design three shows the best configuration from a total revenue point of view and offers the lowest mass, even though it is not the simplest or lowest cost design. This result reinforces the need for a parametric system modelling trade-off analysis.

Table 10.2 shows an estimate of the system breakdown cost of a fixed satellite service (FSS) as an illustration (Lazear *et al.*, 1997).

Satellite communication systems tend to be commercially risky in several ways, and in particular, risks increase when introducing a new technology or addressing a new market. Technical risks include: system engineering risks in terms of architecture, complexity, etc.; space segment risks in terms of software and hardware development, speed of production, launch, operations, and so on; risk in the ground segment in terms of hardware and software development, external interfaces, operational issues such as interference, faulty transmissions, amongst others; regulatory risks in terms of obtaining spectrum, operating licences in target service regions etc. In terms of programme, the risks involve cost, schedule, financing, programme management, risk management, marketing in terms of terminal costs and availability, competition, call cost, and so on. Gaffney et al. (1995) performed independent risk taxonomy of a number of non-geostationary MSS satellite systems during the planning and development phase of these systems. Risk parameters in the analysis included:

1. *system engineering*: interference control, forward error correction and speech coding technology;
2. *space segment*: satellite manufacturing, spacecraft antennas, on-board processing, inter-satellite links, effects of radiation belts;
3. *ground segment*: earth station technology and hand-held terminal antennas;
4. *non-technical issues*: cost, schedule, regulatory, financing and market.

The study included the evolution of a number of proposed systems – Odyssey, Iridium, Globalstar, ELLIPSO and Constellation – with regard to satellite launch, system cost and

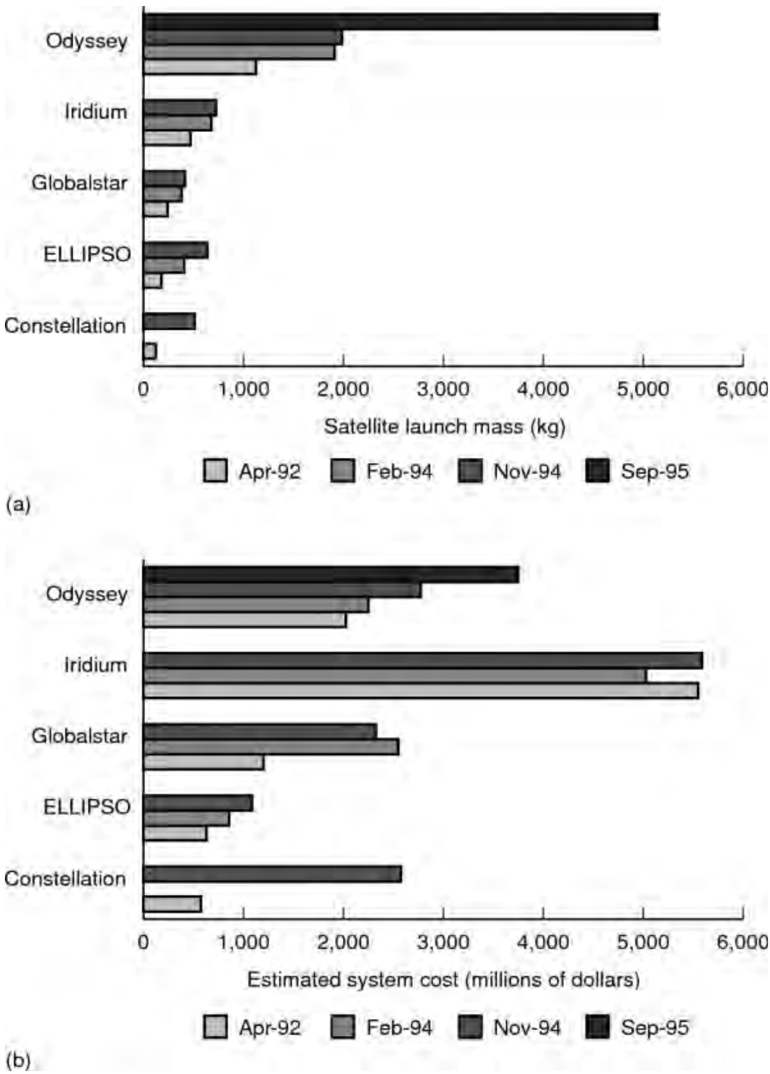


Figure 10.7 (a) Launch mass evolution of non-geostationary MSS satellite systems (Gaffney *et al*. 1995) (b) System cost escalation of non-geostationary MSS satellite systems. (Both parts source: Gaffney *et al*. 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

schedule. It was noted that each system had undergone significant revisions since its initial filing. Figure 10.7 parts (a) and (b) illustrate, as examples respectively, the escalation in satellite launch mass and system cost using data in open literature. Note the escalation in launch mass in all the systems, which was due to a revision in system availability, that necessitated an increase in spot beam gains. This caused a corresponding increase in the system cost and a knock-on effect on the schedule as well. System costs covered all the major cost drivers, that is, design, development, production, launch, some ground segment costs and operation costs, but a meaningful comparison across systems was not possible because of the differences in the way each proposer estimated costs. The study concluded that the highest risks were in the areas of cost, schedule, financing, interference control and complexity escalation. Quite evidently, risks were well founded, as we know that several of these proposals failed, namely, Odyssey, ELLIPSO™ and Constellation, and the remaining ones experienced considerable financial hardship.

A business plan has to be approved by financiers, investors and creditors to ensure the viability of the project. It also lays out firm guidelines for development and informs the employees and stakeholders of the stated goals and plans. The next phase of the project involves design, development, launch and deployment of satellites, product development, product marketing and launch and finally, operations. Careful monitoring of product growth, pricing, etc. is essential to keep the product competitive and viable. The entire programme requires careful management at all levels. The design flow of an MSS product design using such a methodology is illustrated in Figure 10.8 (adapted from Lazear *et al.*, 1997).

Figure 10.9 portrays some of the main elements and interaction of the product marketing plan with the operational concepts and the financial strategy (Lazear *et al.*, 1997). The

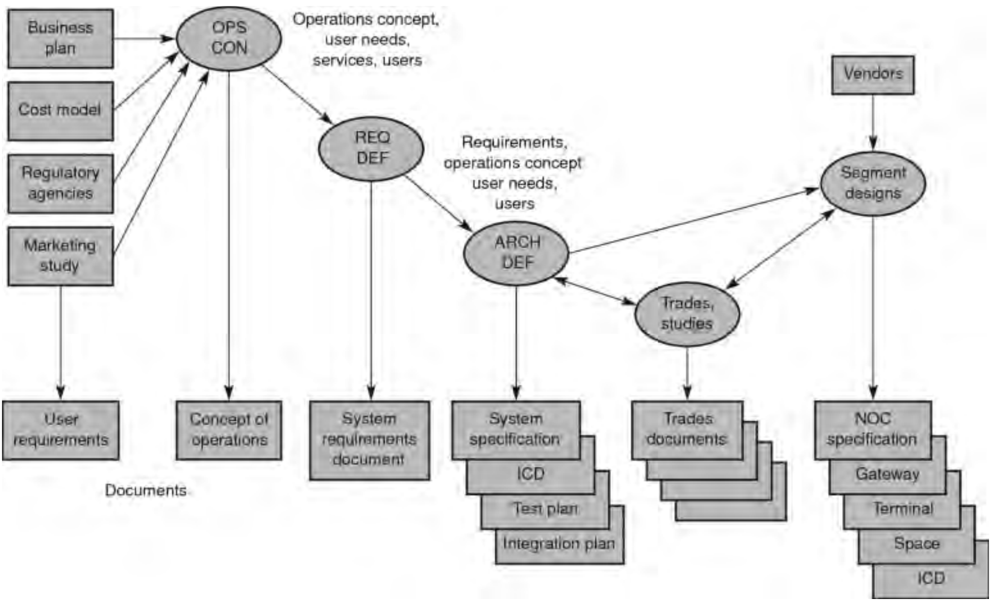


Figure 10.8 Design flow of a MSS product. OPS CON = Operational concept, REQ DEF = Requirements definition, ARCH DEF = Architecture definition, NOC = Network operations centre, ICD = Interface control document. (Source: Lazear *et al.*, 1997 (adapted). Fifth International Mobile Satellite Conference, Pasadena, California, June 16–18, 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97–11, Jet Propulsion Laboratory, Pasadena, California, June 16, 1997.)

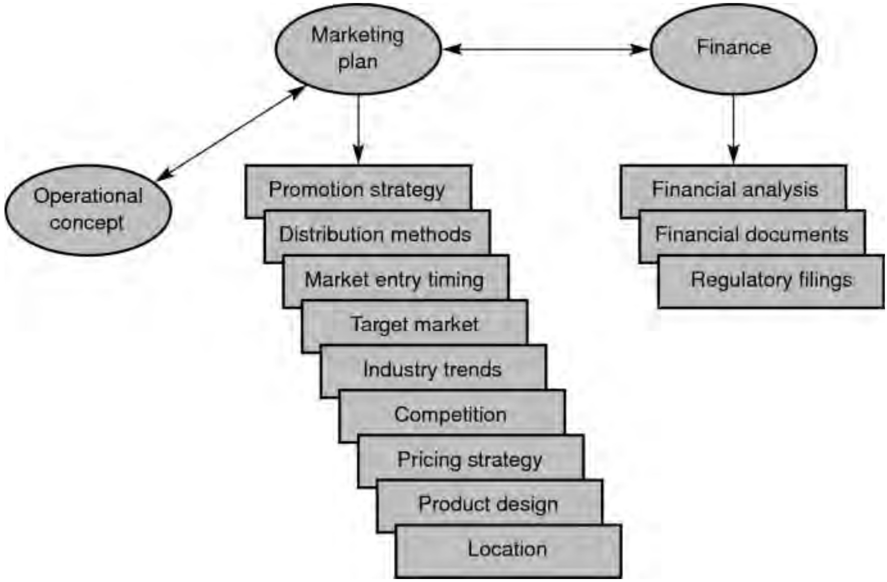


Figure 10.9 Marketing plan and influences. (Source: Lazear *et al.*, 1997. Reproduced with permission of the Fifth International Mobile Satellite Conference, Pasadena, California, June 16–18, 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97–11, Jet Propulsion Laboratory, Pasadena, California, June 16, 1997.)

marketing plan is closely tied with certain aspects of operations such as market entry timing; similarly, certain aspects of finance such as the pricing strategy influence the marketing plan.

MSS services cover vast geographical expanses: countries, continents and for international service, the world. The success of an MSS business is strongly influenced by user equipment cost and service cost in relation to affordability, local infrastructure in conjunction with other factors. Equipment cost and complexity can be reduced by powerful satellites; in addition, economies of scale are achievable after the introduction of service and the technology matures. The service cost usually comprises two components: subscription charge and call charges. The subscription charge is included as a measure for the operator to cover some of its maintenance costs in keeping the network operational for each user. The main element of call cost is satellite resource usage. The trade-off lies between terminal size, bandwidth and the service link satellite EIRP (primarily). In systems supporting small, low-cost user terminals, such as hand-held units, satellite EIRP requirement is high; whereas in comparison, EIRP to provide an identical service to large (and hence expensive) user terminals, the EIRP requirement is low. Initial trade-off analysis, discussed previously, takes such technical factors and non-technical factors like competition and trend into consideration to arrive at an optimum call/service cost.

Figure 10.10 shows the large number of interacting factors likely to affect usage and hence revenue to the operator. Consider *revenue* as the main driver. It depends on *system usage*, which in turn is influenced by various factors – user terminal cost, extent of users' satisfaction with the quality of service; pilferage of revenue by competition, effectiveness

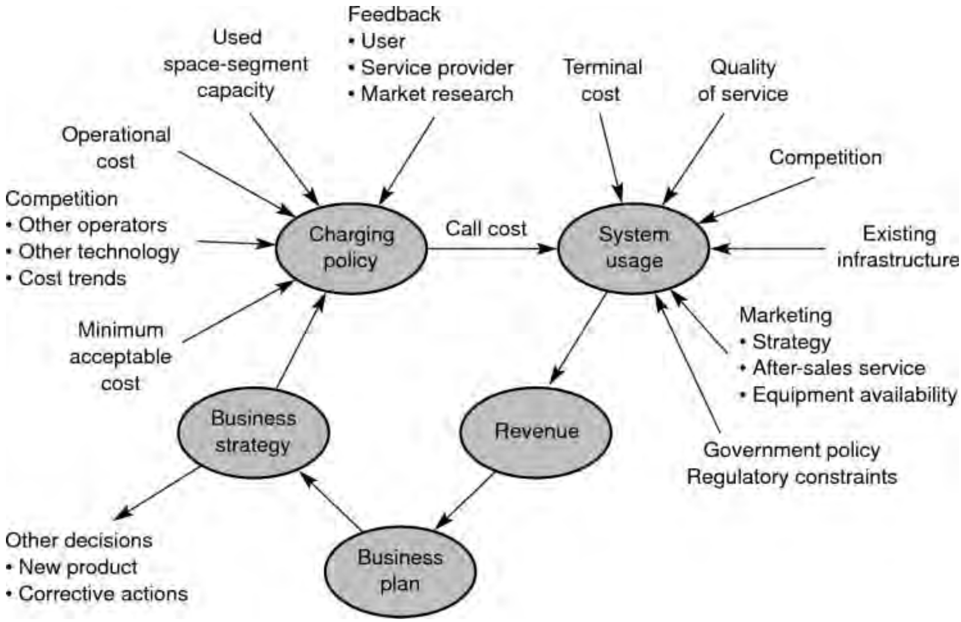


Figure 10.10 Factors likely to affect the usage and the revenue of an operator

of the infrastructure; marketing issues such as equipment availability, promotion, after-sales service; and of course, the call and service cost. Call cost depends on the *charging policy*. Charging policy is governed by a number of independent considerations, which may be a minimum threshold for business viability, geographical dependence (i.e. may be based on user affordability), telecommunications trend (reducing call charges, migration to IP data services), competition from other operators or services, consumer feedback, investment return, operational and fixed costs, and so on. The *business strategy* deals with the overall business direction that considers revenue goals, market conditions, shareholders interest, and others. The revenue details are fed into the company's *business plan* to assess the performance of each product to assist in making the strategic decisions. The revenue is monitored by strategists who take corrective action to meet the sales target in compliance with expected investment returns. For instance, if the usage is not up to the desired level, strategists may apply a corrective action depending on the circumstances, such as ensuring adequate availability of user equipment, altering call charges to a more acceptable level, etc.

Figure 10.11 demonstrates the (notional) sensitivity of space segment call cost on the revenue per year for a spectrum-limited regional geostationary satellite system for various fill and frequency reuse factors to carry circuit-mode traffic. The space segment capacity assumptions are: Spacecraft EIRP = 70 dBW; Available bandwidth = 5 MHz; Bandwidth per channel = 5 kHz; Frequency utilization efficiency = 90%; Average EIRP/circuit-mode channel = 25 dBW.

It was assumed that the system is bandwidth limited, that is, the available bandwidth exhausts before the satellite goes into EIRP saturation. The reuse factor F_{ru} relates the total

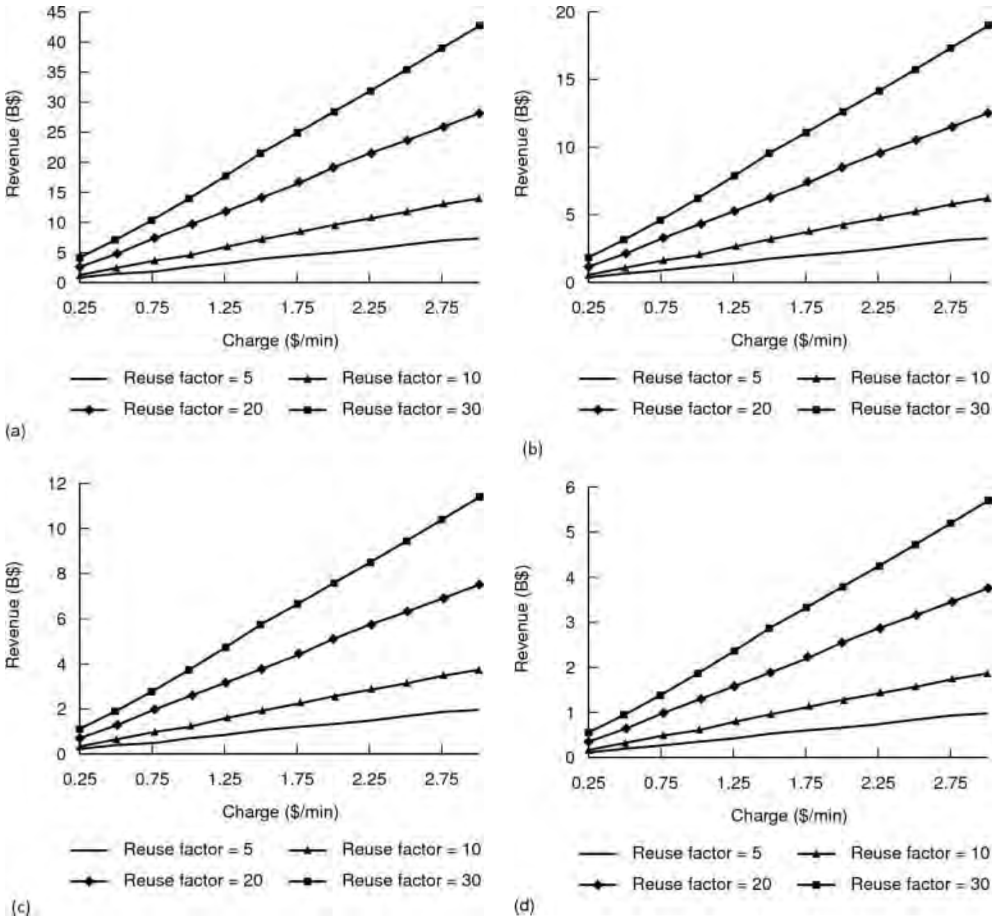


Figure 10.11 Sensitivity of space segment call cost of a circuit-mode system on revenue for a regional geostationary satellite system at various fill and reuse factors. The usage is given as a percentage of available capacity. (a) Theoretical maximum – full utilization (b) weekday: 50%; weekend: 25% (c) weekday: 30%; and weekend: 15% (d) weekday: 15%; weekend: 5%. (All parts source: Graphics AR.)

number of spot beams (N_s) and cluster size (C_s) as follows:

$$F_{ru} = N_s/C_s \tag{10.1}$$

For example, reuse factors of 5 and 30 correspond respectively to 35 and 210 spot beams for a cluster size of seven spot beams.

As expected, revenue increases in direct proportion to call charge and better spectrum reuse. Table 10.3 demonstrates the sensitivity of yearly revenue corresponding to various reuse and fill factors when the nominal call charge is set to 50 cents per minute (assuming a circuit-mode connection). Examining any row shows the improvement in revenue with higher usage. In interpreting the impact of improving the reuse matrix in the table (any column) it has to be understood that the density of the uniformly distributed traffic per spot

Table 10.3 Revenue per year (B$) for various reuse and satellite load percentage at a call charge of 50 cents/min; W_w = weekday; W_e = weekend

Reuse	Satellite load % – W_w :W_e			
	100 : 100	50 : 25	30 : 15	15 : 5
5	1.18	0.51	0.30	0.14
10	2.37	1.01	.60	0.29
20	4.73	2.03	1.22	0.57
30	7.10	3.04	1.82	0.86

beam was kept constant and thus in effect the net traffic in the network increases directly in proportion with an increase in reuse. If the net traffic (Erlangs) was unchanged then improving the reuse matrix from say 5 to 10 would have no impact on the revenue unless the system was interference limited at a reuse of 5. There is little advantage in revenue terms on tightening the reuse matrix for the same traffic intensity; however, the spectrum requirement reduces due to an improvement in spatial reuse and that should enable the operator to lower the call cost.

10.3 Service Distribution Model

Charges paid by the user consist of costs derived from the service distribution system, which in turn is conditioned by the system architecture and the operational philosophy. The cost components could consist of:

- space segment charge;
- home gateway charge;
- gateway charge of another service provider, when applicable;
- terrestrial network charge (PSTN/PLMN/PDN/ISP (public switched telephone network/public land mobile network), etc.);
- service provider's charge, when applicable;
- distributor's charge if the service provider re-sells.

To illustrate various types of distribution systems, let us consider the schemes applied in practice. We will consider representative geostationary earth orbit (GEO), big-LEO, medium earth orbit (MEO) and the little-LEO systems – Inmarsat, Iridium, Globalstar, ICO and Orbcom (Ingley, 1999).

- Inmarsat's 'legacy' services refer to those services that are available from their architecture prior to the introduction of the broadband global area network (BGAN). These services are distributed through gateway operators as shown in Figure 10.12. Inmarsat is the owner of the space segment and the network infrastructure to run these services. The services are offered to the end-users through land earth stations (acting as gateways) owned by service providers whose call charges comprises fees paid to Inmarsat, the terrestrial operator, the gateway's operational and overhead costs and a profit. Since there is a degree of competition between service providers who are dispersed around the world and a wide variation in local economics, there is variability in the end-user charge. Users

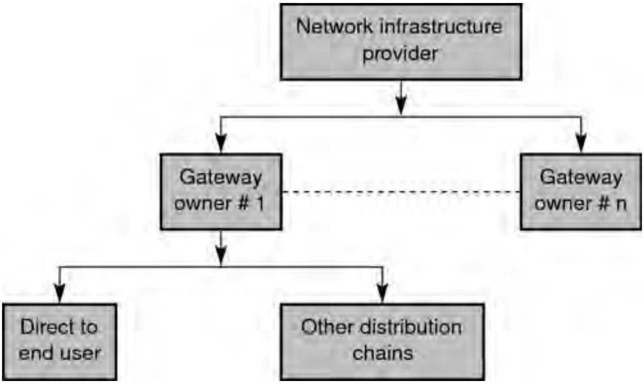


Figure 10.12 A distribution system where services are offered by gateway operators

select a gateway on the basis of cost effectiveness or affiliation. Call charges depend on the service. The charge levied by Inmarsat is driven by commercial considerations. It may sell satellite time at a standard rate or in wholesale at a reduced price to make it attractive to the operators who are able to sell the purchased satellite time to users at a discounted rate; other incentives include a reduced off-peak call charge.

Inmarsat owns the gateways in the BGAN network and therefore sells the BGAN services to service providers as a point-of-presence.

We will review the distribution system originally proposed by Iridium. Iridium bills the operating company (service distributors) for the end-to-end space segment call charges, consisting of space usage charges (operating cost and profit), non-resident gateway charges and terrestrial charges. The operating company then adds the regional gateway charges and profit to set its charge to the service provider; finally, the service provider adds its own costs and profits before billing the user. In the original proposal the service providers comprised PTTs (Post Telephone Telegraph companies), cellular operators and, in some instances, the operating company was also the service provider. The Iridium system minimizes the cost of the terrestrial network by routing calls via intersatellite links.

Dual-mode handsets communicate with various cellular standards by snapping in a cassette that supports a local cellular standard (refers to the original proposal). This service, called global cellular, was expected to provide revenues to Iridium as the customer bill included an Iridium component added to the payment levied by the cellular operator.

- Globalstar's original distribution system has been designed to support a decentralized distribution system allowing a country or a group of countries to exercise full control of local distribution; initially, up to 200 gateways were planned but this was later scaled to 50, which were considered adequate for global coverage. Service providers own and manage gateways and arrange for terrestrial routing. They purchase satellite time in bulk from Globalstar; either the service provider or the local distributor, which may be a cellular or personal communication system (PCS) operator or a PTT, bills the customer. Thus, the bill consists of a satellite component, an overseas gateway component when applicable, service provider's operating cost and profit and a terrestrial charge. Globalstar bills the service provider for the satellite time and overseas gateway usage.

- To minimize dependencies on terrestrial systems, ICO's global communication architecture consists of a backbone terrestrial optical fibre network interconnecting all satellite access earth stations. The satellite access nodes (SANs) and the terrestrial network (ICONET) are owned by ICO. There were about 60 service partners, each of whom could have one or more service provision arrangements. The service partners and service providers manage the customer relations, set call charges, distribute terminals through various channels and bill the users. The user charge comprises satellite and terrestrial usage and the service partner/service provider's operation cost plus profit. ICO charges the service partners/providers for satellite and terrestrial usage.
- The ORBCOM system provides low bit rate data-only service and hence has a different user base to the big-LEO systems. ORBCOM distributes its services directly to users, through value added sellers and through international licenses. The cost components are similar to the ones in big-LEO systems but the cost is based on used data bytes.
- Universal Mobile Telecommunication System (UMTS) business model: UMTS was proposed as the next generation mobile system for Europe in the 1990s, which would include a satellite component. The UMTS business model was envisaged to comprise the space segment owners, network operators, service providers, the subscriber and user. It was envisaged that the satellite component would be an extension to the PLMN comprising a system that served both cellular and satellite cells or a system serving satellite cells only. In the case of the former, two possible business models are illustrated in Figures 10.13 and 10.14 (adapted from SMG5 TD 224/93rev1). Figure 10.13 illustrates a configuration where the satellite UMTS segment is an extension to terrestrial UMTS (thus owned by the same entity) – the space platform is shared by a number of independent UMTS networks. Figure 10.14 illustrates a configuration where the satellite UMTS is owned by

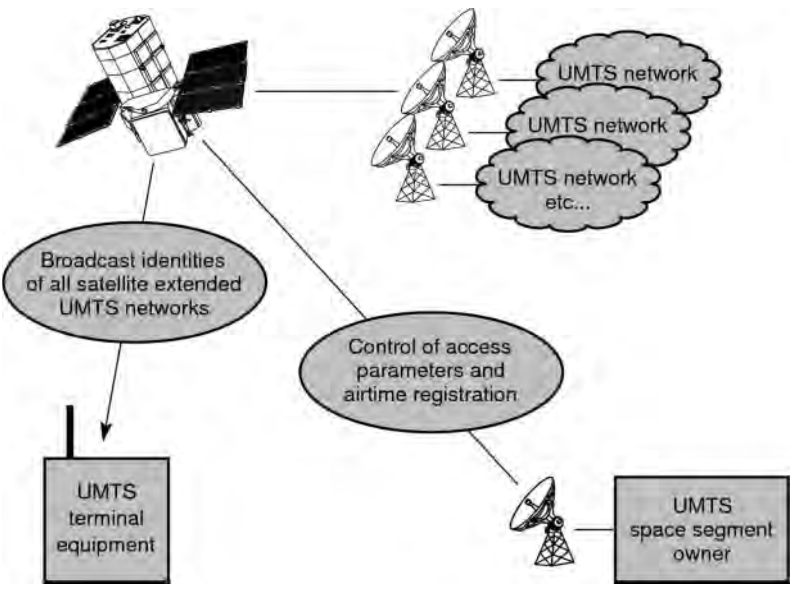


Figure 10.13 Satellite component integrated with multiple UMTS PLMN. (Source: Dondl, 1995. © 1995 IEEE. Reproduced with permission.)

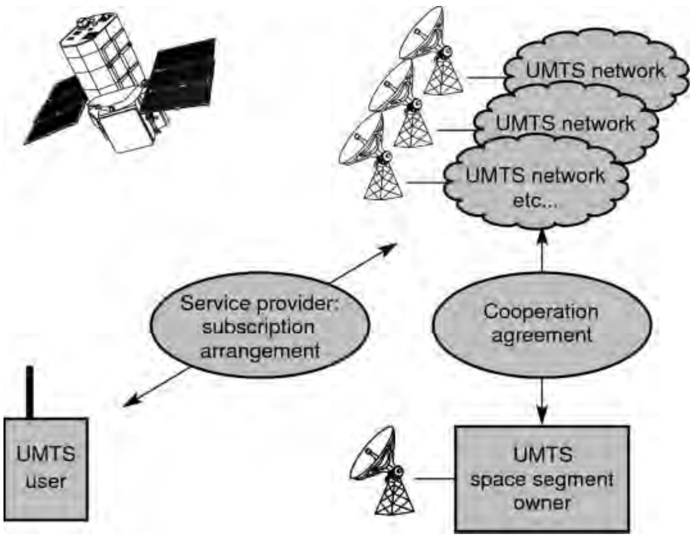


Figure 10.14 Satellite component shared by several UMTS networks. (Source: Dondl, 1995. © 1995 IEEE. Reproduced with permission.)

an independent party who provides satellite extensions to a number of terrestrial UMTS by a subscription arrangement. In another approach, the satellite component would reside outside the terrestrial UMTS business arrangement. In such a case, the model reduces to that of the present types of arrangement, such as used by Inmarsat legacy systems.

10.4 Billing Issues

Customer care and billing are important and complex aspects of commercial MSS. The task of the billing system becomes compounded for international systems that must have gateways in a number of countries and whose service providers serve customers in a variety of currencies and bill them in local languages and currencies. The billing method depends on the business architecture of the system (see Section 10.3). In the following the space segment infrastructure provider is referred as system operator.

- In a *segmented billing scheme*, the system operator bills service providers, leaving the task of managing the customer billing to the service provider. The satellite usage of each service provider is monitored by the system operator, who bills the service provider for space segment usage. The service provider has its own billing system for managing the terrestrial component and individual billing. In a variant of this scheme, the space infrastructure provider may utilize the call monitoring system of service providers, to reduce the complexity and cost of billing. Each service provider sends bills to customers in a format and language understood locally. Thus, the system operator's billing system is simplified as the operator needs to bill a relatively lower number of customers who can all work in a local language/currency.

- In a *semi-centralized billing system*, end-to-end call/data usage monitoring is done by the system operator. The operator then sends the usage records with space segment charges and non-resident gateway charges to the caller's home gateway operator, who cascades them down to the service provider after adding its overheads; and finally, the service provider bills the customer with a mark-up consisting of its cost and profit.
- In a *centralized billing system*, the end-to-end usage monitoring is performed by the system operator. The operator owns the entire system, selling the service through local service providers. The call records, together with space segment, gateway and terrestrial charges, are passed on to the service provider's business centre, which marks up the cost to account for its operating cost and profit before passing it over to the customer.

10.4.1 Investment Routes

Commercial satellite projects arrange finances through various financial instruments: loans from investment banks, public equity markets, private investors, high yield debt notes, bonds, vendor money, strategic equity and other financial means. Generally, a loan requires some security/guarantee due to the high risks perceived by financiers, and its availability depends on the state of the creditor's market. A short-term loan is arranged through incremental banking arrangements. Capital also flows through mergers, acquisitions and spin-offs. Sometimes, funding takes on complex dimensions, involving financiers who may themselves take loans from several banks, such as used by Arianespace Finance for financing the launch customers of Arianespace (Foley, 1997). Far fewer analysts cover the satellite industry for purchase recommendations, etc. compared with other telecommunications technologies. Financial market resources are limited; the finance is divided between various sectors, such as FSS, MSS, DAR, satellite imaging, and so on, according to the confidence the investors have in the proposals and the trend. Obviously, such strategic forecasts are revised regularly, affecting stock prices and investments; when financing dries up, historically the satellite industry tends to be among the first to lose funding. In the global economic downturn in the early years of 2010 decade the mobile satellite systems outperformed the general trend showing a positive growth throughout.

Financiers favour companies that reduce regulatory, development and business risks by having strong sponsors and a good management team, obtains as many licenses as possible and show evidence of competitive advantage. Financiers find the satellite projects rather challenging, because of perceived risks such as launch failures and long gestation periods of satellite communication systems. Earlier in the chapter, we explained that the cost of MSS projects may run into billions and tends to escalate due to unavoidable delays in projects of such complexity. Figure 10.15 illustrates the sensitivity of a business to rate of penetration for various hypothetical scenarios. The x-axis represents time (years) and the y-axis the corresponding expenditure or revenue as necessary. When the revenue exceeds the expenditure the system reaches profitability. For example, the scenario illustrated in case (a) demonstrates the financial performance of a system that reaches profitability after suffering loss during the gestation period of a few years, closely matching the expectations. Case (b) represents performance of a system that never reaches profitability and case (c) represents the scenario when the system picks up revenue after a slow start and enters the profitability late in its lifecycle.

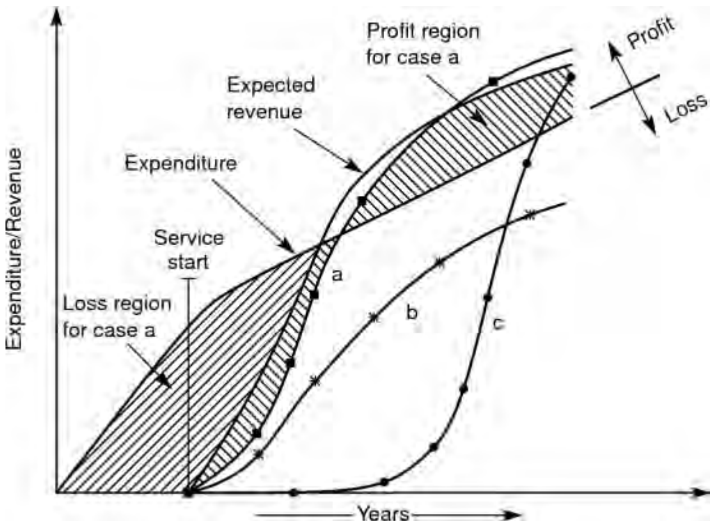


Figure 10.15 Sensitivity of business to rate of penetration. Case (a) closely matches expectations; case (b) never makes a profit and case (c) represents late profit.

Setbacks, such as Iridium's bankruptcy and the 12-satellite launch failure of Globalstar, dent the confidence of the financial community. Investors tend to calm down if companies display confidence in the product – take, for example the instance when the CEO of a satellite company volunteered to take the burden of all debts if the financial community did not invest; yet another company spent considerable effort in proving the technology by launching satellites in orbit to win investor confidence. In this respect, well-established companies with deployed networks and earnings are preferred. Aligning a company with major players, finding partners and corporate backing from large companies are other ways in which companies gain confidence.

10.5 Regulatory Issues

Regulatory issues address compliance with national, regional and international Radio Regulations that include issues such as intersystem coordination as well as licensing, which may vary between countries. These procedures are very lengthy, circuitous and therefore not conducive to a commercial environment. A necessity to streamline regulatory procedures is generally felt by commercial operators.

The regulatory process for the licensing of MSSs is rather complex and therefore takes several years for completion. Furthermore, even within the International Telecommunications Unions (ITUs) global framework, regulations can differ by country, depending on priority and regional differences. Typically, the licensing process involves spectrum allocation, national licensing and coordination of the allocated frequency with external systems. Spectrum and frequency allocations are likely to incorporate international involvement. Frequency coordination is required to guarantee that the operators do not cause interference to each other. The coordination guidelines are laid out in ITU articles 11 and 13 of the Radio Regulations.

An operating license is necessary for an operator to market services in a country. Some countries which may not license certain operators are concerned that terminals belonging to such unlicensed operators may be used within their country. MSS operators address such concerns by building technical features capable of preventing this type of fraudulent use. For instance, the system may prohibit communication if a call is requested from a communicator located within a 'barred' country. Strict discipline, cooperation and agreement between parties with infringement clauses are necessary.

Other matters, in addition to frequency assignments, are related to type approval, global roaming and non-discriminatory access to market. Local partners are recognized as a necessity for best marketing access through local support and marketing skills, and at the same time, facilitates license procurement. Partners could benefit from subsidies and participation in the operator's system design/operations. For these reasons, some companies such as Globalstar have built systems that can support gateways in individual countries or in a region. However, other operators prefer to own all the gateways to facilitate interoperability and maintain better overall control.

While the service markets may be segmented geographically, there are obvious advantages when the space segment and operations are maintained by a single entity – therefore MSS operators tend to keep operations within their jurisdiction. Benefits include network-wide performance evaluation, anomaly investigation, and data collection for operational, market and strategic planning.

Operators often find it difficult to obtain an operating license in some countries. The concerns of the telecommunication authorities of such countries may relate to their share of revenue, security concerns, political considerations, and government policy, etc. The revenue concerns stem from the fact that users may bypass the country's network using a foreign gateway, thus depriving the country of the revenue. The problem could be exacerbated if the service is offered from a regional gateway located in another (unfriendly) country. Roaming to other countries may also result in loss of revenue to the home service provider. From the operator's viewpoint, the difficulty lies in the fact that the telecommunication costs in certain countries are so large as to make the service unattractive, and they are faced with the burden of billing a visiting terminal. Such matters are managed in the cellular systems such as Global System for Mobiles (GSM) by mutual agreements.

Many issues addressed above are under discussions at the ITU and other regulatory regimes. Such debates play an important role in harmonizing the growth of the technology and in the protection of sovereign rights of nations to regulate telecommunications by laying down international standards, promoting international cooperation, etc.

10.6 Traffic Forecast

A crucial requirement of an MSS business planning is communication traffic forecasting to enable planners to make timely and informed operational and strategic decisions. In Chapter 3, we introduced a generic method of long-term spectrum forecast. Chapter 9 addressed methods for short-term forecasts using measured traffic data flow through a network suited for the short-term business and operational needs of prevalent operators. Long-term market forecasting, on the other hand, is necessary by regulatory authorities for spectrum planning; standardization bodies require forecasts for dimensioning satellite components of long-term initiatives such as UMTS; an MSS operator uses forecasts for long-term strategic planning; mobile terminal manufacturers may use them to plan their strategic manufacturing initiatives.

Here we will consider various approaches to long-term traffic/spectrum forecast and introduce methods used by the ITU, developed in European RACE II programme, the UMTS Forum and by Eurocontrol and Federal Aviation Administration. The forecasts are significantly dependent on the assumptions, and hence a wide variation in the forecasts is possible. Therefore, forecasters typically provide forecasts for low, moderate and high growth of a service.

Depending on the methodology, a variety of parameters is included (directly or indirectly) as an input to a MSS forecast model, a representative set is listed next:

- general technology growth trends likely to influence the MSS market in the forecast period;
- state of applicable technology in the forecast period;
- influence of competing technology such as terrestrial mobile systems, other MSS operators and personal/mobile communication technologies such as mobile FSS systems;
- an evolution path to the next generation;
- market penetration of similar service(s) and rate of penetration
- QoS;
- number of potential operators and traffic segmentation;
- tariffs and terminal cost;
- geographical areas of interest, their gross domestic product (GDP) and related demographic data;
- for personal communications: age, occupation, GDP of the target group, expenditure per head in similar technologies and acceptability of such services;
- for enterprises/businesses/specialists: annual budget for similar service;
- sociological and technology trends in region(s) of interest;
- influence of regulatory constraints such as spectrum segmentation;
- risks in obtaining licensing of services, especially for international systems.

Assumptions include acceptable levels of tariffs, impact of competition, acceptable terminal costs and feasibility in terms of manufacturing costs, component costs, production volume, terminal complexity, service take-up, usage per terminal and market penetration rate derived from existing trends.

A notable technological development has been the growing convergence of telecommunications, information exchange, computing, entertainment systems, business needs and the universal acceptance of the Internet as a telecommunication transport workhorse. The so-called information society of this decade therefore requires technologies that offer these services to individuals and business transparently. This premise forms one of the bases of forecasts for mobile communication systems.

10.6.1 UMTS Methodology

Consider, as an example, a methodology applied by the UMTS forum in forecasting the third generation mobile communication market in Europe and the world (UMTS Forum, 1999).

The methodology is generic enough to be applicable for MSS forecasts. Since the primary purpose here is to introduce the concepts, the forecasted results serve as an illustration only. The purpose of the study was:

1. to determine the level and nature of third generation mobile services in the world, with emphasis on the European Union;
2. to determine the key issues influencing the evolution and their impact on the market analysis;
3. to develop a robust analysis to forecast the world-wide mobile services market up to 2015;
4. to forecast emerging requirements and intermediate steps towards the introduction of UMTS/IMT-2000 (International Mobile Telecommunication-2000).

The salient features of the methodology are summarized in Figure 10.16. Societal trends are combined with technology evolution trends to develop a number of scenarios for market development and a forecast of each produced. The scenarios and forecasts were validated and industrial opinions regarding future market requirements and development were gathered through a series of workshops for industrialists. Finally, all research findings were consolidated, identifying the key drivers, enablers, barriers and uncertainties that would influence the future market. Clearly, the scope of the study was wide, with far-reaching implications for the future of mobile telecommunications. The study undertook forecasts for both terrestrial and satellite markets.

It was anticipated that the third generation mobile systems would offer broadband services, such as graphics and video, to mobile users transparently, taking advantage of the convergence of the fixed and the mobile networks. Based on these trends, it was assumed that broadband services would be introduced as enhancements to existing technologies, such as DECT, before the commercial launch of third generation systems.

Behaviour and attitudes in society that would likely affect acceptance of third generation systems were identified and their rate of development was assessed. Three key trends were identified:

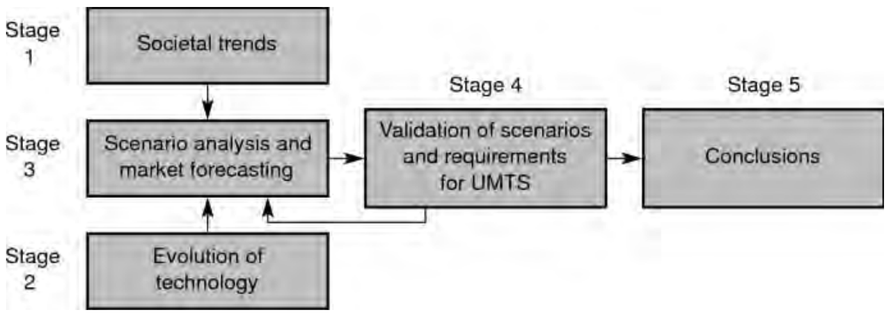


Figure 10.16 Forecasting methodology used by UMTS forum. (Source: Adapted from UMTS, 1999. Reproduced by permission of UMTS.)

- The market for fixed networked multimedia, such as pay-TV, video/audio-on-demand, interactive entertainment, educational and information services and communication services such as video-telephone and large file transfer, was generally growing at rates of over 60%;
- Computer information retrieval and communications technology was being assimilated rapidly, evidenced by phenomenal Internet uptake;
- There was a growing trend for accessing information and entertainment services by mobile users.

An underlying assumption was that the demand for multimedia services in the mobile environment would be influenced significantly by the growth in the fixed sector, and hence would follow it. The services considered in this category included: passive audio-visual services, for example direct to home satellite television, video-on-demand; passive audio services, for example, audio-on-demand as an alternative to CD.

Over 30 applicable technologies were studied; of these, critical technologies that could influence the functionality, attractiveness and cost of mobile multimedia terminals were identified as semiconductor, display and interface technologies. The relevant technologies for the mobile multimedia services were identified as delivery and management technologies such as the Internet, Java, database, spectrum-enhancing and service creation technologies.

It was assumed that factors influencing service take-up would be service affordability, governed by the price of the service and terminal, and service attractiveness. It was further assumed that the price of the service would be determined by competitiveness in service delivery and the cost of network access; the price of terminals would be influenced by component cost and production volumes; and the service attractiveness would be affected by the offered variety, usability and utility.

Four scenarios were modelled, comprising the permutation of two key factors – the take-up of fixed multimedia service in mass markets and the primary location of intelligence whether in the network or in the terminal – which were likely to affect multimedia service evolution up to 2005.

In the *slow evolution* scenario, mobile multimedia growth would be adversely impacted by various factors such as unsuccessful liberalization or limited attractiveness of the service. In the *business-centric* evolution scenario growth would be driven by the business sector with a slow uptake in the consumer market further impeded by limited liberalization, high prices and limited IT literacy; In the *evolved mass market* model, the mobile multi-media is accepted in all the sectors after a slow start aided by various positive influences such global standardization and reduced service/terminal prices; In the *commoditized mass market* scenario the growth is rapid aided by a variety of positive impetus such as liberalization, reduced service/terminal price, stable regulatory regime, and so on.

Each scenario depicted the interaction between elements of a conceptual market dynamics model under different assumptions in social trends, technology evolution, regulatory/political environment and industry structure/competition. The scenarios were then characterized in terms of market size, its drivers, enablers and barriers. Some of the main characteristics are summarized in Table 10.4. The scenarios were useful in making a sensitivity analysis thereby arriving at the specific issues and industrial development likely to influence the market forecast, which could then be used for forward planning by the

Table 10.4 Factors that could influence the evolution of the proposed scenarios

Factor	List
Drivers	Growth of related services in the fixed network due to increase in Internet use and reduction in usage costs
	Demand for rapid remote access to information for business and personal use
Enablers	Encouragement to competition and cheap access to services
	Development of IMT-2000 specification by ITU
	Possibility to exploit general packet radio service (GPRS) in the multi-media delivery evolution path
	Improvements in technologies such as interface design, display, spectrum efficiency enhancement and semiconductor device
Barriers	High cost and limited spectrum
	User concerns on security and failure to resolve them
	Slow IT literacy penetration in mass market
Uncertainties	Growth rate in mobile multi-media market, network capacity and spectrum availability.
	Network model – network intelligence versus device intelligence

(Adapted from UMTS, 1999.)

industry, market, etc. The scenarios could also reveal the effects of uncertainties, such as slower than expected growth, on the forecast. The possibility of marked deviation from trend was emphasized. Such departures could occur through unforeseen developments or breakthroughs. Table 10.4 lists the factors considered as possible candidates to influence the evolution of these scenarios.

Figure 10.17 portrays the influences in the evolved mass market model leading to encouragement in service take-up and entry of service providers. Successful liberalization, development of global standards, low chip costs, high levels of IT literacy, use of artificial intelligence and intelligent agents all have a positive influence, encouraging entry of service providers and increased service take up.

Table 10.5 lists the forecast produced for the satellite market by the UMTS forum (1999) for world-wide UMTS/IMT-2000 compatible users for the years 2005 and 2010. The forecast for non-multimedia subscribers did not include the hand-held MSS sector. The forecast did not include installations of the FSS that provide a personal multimedia service as this sector fell outside the UMTS/IMT-2000 system definition. The forecast was derived from Inmarsat's work in this area as explained later in this section. It is worth noting that the forecast grossly overestimated the market as the total MSS subscribers are estimated to be of the order of 2.2 million (including the hand-held MSS sector) in the year 2011 (NSR, 2012), far short of 18.475 million subscribers predicted by the UMTS forecast to the end of 2010.

One problem perpetually experienced by planners is lack of market-related data when a service is new. In such cases, indirect methods have to be used. Regional population and addressable market potential in conjunction with data such as existence (or not) of terrestrial coverage, affordability of the target population, handset cost and service charge, anticipated penetration rate are indicative of the likely demand. A traffic dimensioning methodology, was developed within the Satellite Integration in the Future Mobile Network (SAINT) project of the European RACE II programme and addresses such a scenario (Hu and Sheriff,

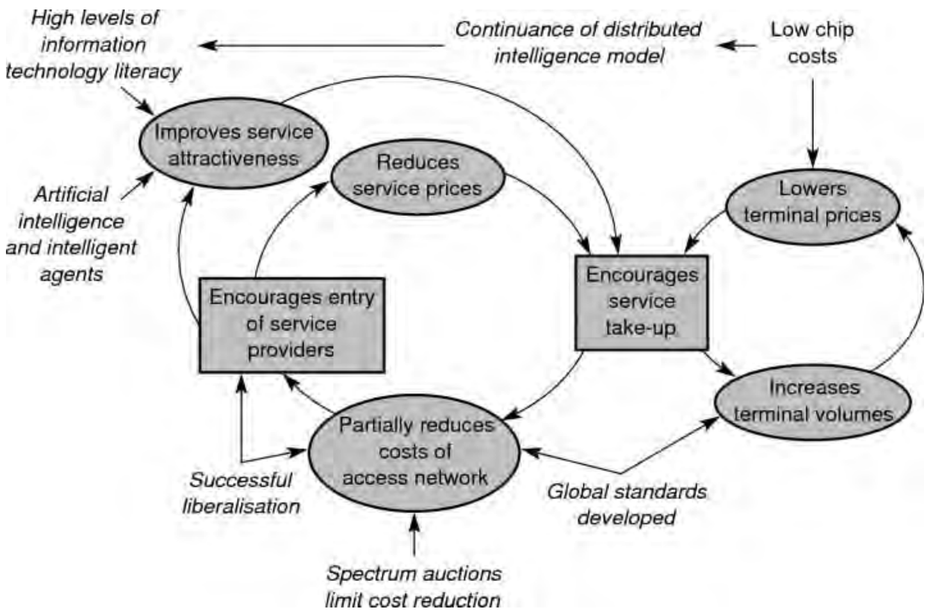


Figure 10.17 Forces that influence uptake of service in the evolved mass market. (Source: Adapted from UMTS, 1999. Reproduced by permission of UMTS.)

1995, Hu and Sheriff, 1997). The algorithm, developed for dimensioning the satellite component of Universal Mobile Telecommunication System (S-UMTS), uses logical assumptions and demographic data to predict satellite voice traffic in Erlangs for over 210 countries, presenting them on 36×72 traffic grids on world maps. An earlier study conducted for the European Commission was used as the basis for developing the model (KPMG, 1994).

The model presented here is as reported in the more recent of the two publications. Based on S-UMTS terminal types (five in all) and the role of S-UMTS, two categories of users were identified – mobile users and fixed users. Then traffic volumes for each user group were predicted as follows:

- assess S-UMTS feasibility in a region based on existence of terrestrial coverage;
- determine gross potential market (GPM);
- estimate S-UMTS service penetration, factoring in the take-up rate T_p.

The penetration P is defined as the percentage of people subscribing for S-UMTS. Parameter P reflects the ability of the population to subscribe for the service, which depends on the ratio of per capita income to the tariff of the service. Relative ability to pay versus penetration percentage graphs used were based on terrestrial cellular service data. Penetration rate was assumed to follow the well-known product cycle comprising a slow introductory phase, followed by rapid growth, saturation and decline. The penetration rate at time t was modelled by curve-fitting historic data as follows:

$$T_p(t) = M/(1 + ae^{bt}) \tag{10.2}$$

Table 10.5 World-wide satellite market

Year	2005	2010
MSS subscribers (000s)		
Non-multimedia	4 875	7 500
Multimedia	6 585	10 975
Total	11 460	18 475
Average usage per subscriber (kbytes/month)		
Non-multimedia		
Voice	8 709	8 491
Low speed data	6 208	5 587
Multimedia		
Voice	1 194	1 561
Low speed data	2 584	3 380
Asymmetric	26 154	34 247
Interactive	1 781	2 334
Total annual traffic (million MBytes)		
Non-multimedia		
Voice	509	764
Low speed data	491	736
Multimedia		
Voice	94	206
Low speed data	204	445
Asymmetric	2 067	4 510
Interactive	141	307
Total	3 506	6 968

(Adapted from UMTS, 1999.)

where $T_p(t)$ = penetration rate (i.e. take-up) at time t, parameters a and b are obtained by regression analysis and parameter M describes the saturation level.

The parameters M, a, b differ between countries and are therefore calculated for each country on the basis of cellular growth rate. Penetration in year t, $P_r(t)$, is obtained by factoring P and $T_p(t)$.

The number of subscribers N_t in the year of interest for a GPM, is then given as

$$N_t = (\text{GPM})\, P_r(t) \tag{10.3}$$

GPM is the number of potential customers. An S-UMTS service becomes feasible only when terrestrial services become uneconomic, that is where the population density is below a threshold D_t. The threshold depends on the per capita income of the region. Hence, the first step is to determine those regions where S-UMTS could be introduced. D_t was determined on the basis of per capita income for the years of interest (1998–2010). For 2010, D_t for high (>20 000 ECU per capita), medium (6000–20 000 ECU per capita) and low GDP (<6000 ECU per capita) were estimated as 3 people/km^2, 30 people/km^2 and cities of 1 million respectively.

Note 1

The European Currency Unit (ECU), based on a number of European currencies, was an accounting unit of the European community prior to the introduction of the Euro.

Mobile users were categorized as cellular fill-in users (see next) and international travellers, that is, people who rely on satellite communication when they travel to areas where there is no terrestrial coverage.

The GPM of the cellular fill-in, G_f, is defined as the number of people not served by the terrestrial UMTS. The number of cellular fill-in customers N_c in year t is then given as

$$N_c = N_{pop}(1 - p/100)P_r(t) \tag{10.4}$$

where N_{pop} = population of the region under consideration and p = percentage of population covered by the terrestrial UMTS.

The GPM for international travellers was estimated as the number of air travellers, N_a, leaving a country. The call charge per minute was set at a higher level; service take-up rate was set the same as cellular fill-in but penetration was lower due to the higher tariff. The number of subscribers N_i in year t, for this category, is then given as

$$N_i = N_a P_a(t) \tag{10.5}$$

where $P_a(t)$ = penetration rate in year t.

The GPM for fixed S-UMTS users, N_f is estimated as follows.

$$N_{fg} = N_h - N_T \tag{10.6}$$

where N_h is the number of houses and N_T is the number of telephone mainlines in residential areas. Factoring the penetration of telephone lines per household in a manner similar to mobile users and factoring it by the take-up rate, $P_f(t)$, the number of subscribers N_f is given as

$$N_f = N_{fg}.P_f(t) \tag{10.7}$$

Data of up to 225 countries, obtained from UN and ITU publications, were used. The number of S-UMTS subscribers depended strongly on handset subscription rate and call cost. Figure 10.18(a) and (b) shows the subscriber forecasts for low and high rate tariffs. The respective handset price in the low and high cost entries would be 500 ECU falling by a factor of 0.7 each year; the corresponding subscription rate would be 60 and 240 ECU respectively reducing at the same rate as the handset price; and the call rate would be 0.1 and 0.5 ECU respectively reducing by 0.5 annually. The number of mobile subscribers for the lower cost entry is about 90 million against only 35 million for the high cost entry.

Figure 10.19(a) and (b) shows the split in S-UMTS users by region for 2010 when S-UMTS was targeted to have the maximum penetration. Figure 10.19(a) shows that majority of mobile users would be in North America, the former USSR, Europe and Asia; and Figure 10.19(b) shows that fixed users would be mainly from Asia, South America and Europe. Further, it was observed that 52% of the market would be for fixed users and the remainder for mobile users.

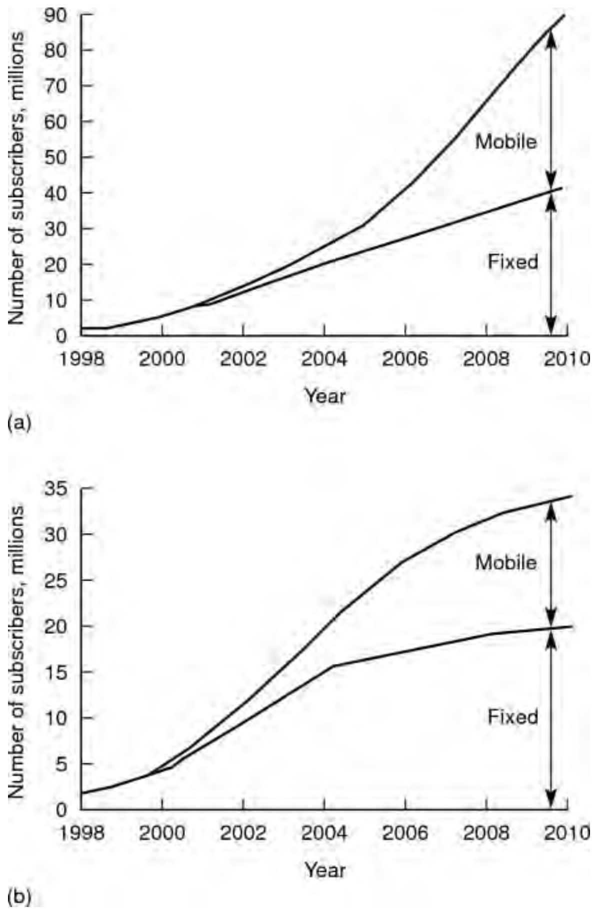


Figure 10.18 Subscriber forecast for (a) low tariff (b) high tariff. (Both parts source: Hu and Sheriff, 1997. Reproduced by permission of IET.)

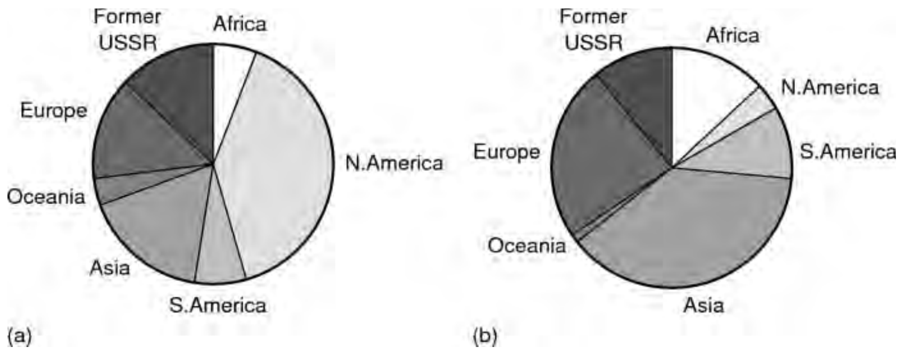


Figure 10.19 Regional distributions of S-UMTS (a) mobile subscribers (b) fixed (non-permanent) users. (Source: Hu and Sheriff, 1997. Reproduced by permission of IET.)

The reader will recall that the intention of this chapter is to introduce forecast techniques rather than to provide market forecasts. Nevertheless, it noteworthy that the numbers of subscribers fell far short of the projections of the year 2010 – compare the total number of subscribers attributed to MSS in 2011 of 2.2 million to the projections of 35 million for the high-cost entry. This example as the previous (i.e. UMTS forum forecast) highlights the remarkable significance of the assumptions used and the difficulty in predicting consumer pattern of a new service.

10.6.2 ITU Traffic and Spectrum Forecast Methodology

Satellite component of IMT-2000 and systems beyond IMT-2000 refers to concepts and standards to provide ubiquitous, high-data rate, content rich mobile telecommunication services throughout world. The ITU has facilitated the development of methodologies to forecast traffic and spectrum requirements of MSS services including those of the satellite component of beyond IMT-2000. ITU-R report M.2023 includes methodologies to forecast traffic and spectrum up to the year 2010 (ITU-1, 2000); using similar techniques, ITU-R report M.2077 extends these forecasts up to the year 2020 (ITU-2, 2006). The methodology to calculate satellite spectrum requirements used in these reports is detailed in ITU-R recommendation M.1391-1 (ITU-3, 2006). The forecasts were used in preparation of world radio conferences (WRC) held respectively in the years 2000 (WRC-2000) and 2007 (WRC-07).

Here we introduce these methodologies with a few sample results. These results are presented only as illustrations to demonstrate the application of the methodologies and not for practical use. Such forecasts are revised regularly based on available information and the ever-changing market dynamics of this sector. For example, ITU-R report ITU-R M.2218 1 (ITU-4, 2011) supplements the results of the report M.2077 by including higher data rates as well as extending the frequency range from 1–3 to 6–16 GHz. Further, it includes the provision to support backhaul communication to support terrestrial mobile systems that provide mobile service (e.g. mobile phone and WiFi) to passengers in captive environments such as planes, ships railways and in disaster-struck areas.

The UMTS Forum MSS and IMT-2000 compliant market forecasts compiled in 1997 for 2005 and 2010 mentioned earlier in this section were largely based on Inmarsat's work. Since then the methodology has been refined. This section outlines this updated approach.

10.6.2.1 Services and Data Rates

It was anticipated that IMT-2000 satellite services would be an extension of terrestrial 3G services with modifications necessary to suit MSS users who are likely to use handsets or larger variant of terminals such as handset-laptop combinations. Four categories of service were anticipated ranging from 4–384 kbps, including voice (4–16 kbps), messaging and low-speed data (9.6–16 kbps), asymmetric multimedia (up to 144 kbps); and interactive multimedia services (up to 384 kbps).

It was anticipated that IMT-2000 satellite services would serve the following environment categories:

- rural – vehicular, pedestrian and semi-fixed (non-permanent);
- remote – vehicular, pedestrian and fixed;
- maritime and aeronautical;
- mobile base stations (e.g. on trains or buses);
- locations where terrestrial IMT-2000 services may not have become available yet.

10.6.2.2 Market

The market projections were based on forecasts from two major global satellite organizations – a USA company and Inmarsat, an international MSS operator. Both forecasts indicated an increase in demand. Although the net result for the MSS spectrum was very similar, their assessment varied geographically. The variation is attributed to differences in their market perspective and methodologies; it further highlights the sensitivity of long-term forecasts on the assumptions. Both the studies considered *frequent users* – those who rely mainly on MSS communication and *infrequent* users – those who rely on terrestrial services relying on MSS only when roaming outside the terrestrial coverage. The US forecast of frequent users was very conservative compared to that of Inmarsat. Although both the market studies used primary research for the projections, their research areas differed. Inmarsat focused on usage of multimedia service for frequent users whereas the American study focused on voice and data usage of global subscribers. The forecast of the US company included primary market analysis that screened more than 200 000 people and interviewed over 23 300 individuals from 42 countries and 3000 corporations that use remote operations. Inmarsat's forecasts of multimedia subscribers was based on user and potential user survey in 15 countries covering over 1000 companies and individuals in conjunction with forecast and user profile models developed over its operational experience of about 20 years. Their forecasts were segmented into non-multimedia subscribers, only a portion of who were expected to be IMT-2000 compliant and IMT-2000 compliant multimedia subscribers. Forecasts for non-multimedia subscribers were based largely on previous primary and secondary research and a number of industry quoted forecasts of such services.

10.6.2.3 Usage

Spectrum requirements are considerably sensitive to the subscriber usage. The American study used a combination of primary market research and secondary market analysis for deriving usage characteristics. MSS usage levels were provided in megabits per second per month and year for each of four service categories based on assumptions regarding the time subscribers spend outside of terrestrial wireless coverage areas, service categories and user type (e.g. a journalist). Based on historic data and prevailing user characteristics it was assumed that usage per subscriber stays static for voice and messaging over time whereas usage per subscriber of medium multimedia and high-speed data were assumed to increase at an annual rate of 5% between 2005 and 2010.

Inmarsat forecasts for non-multimedia satellite-based services appear to have been based largely on previous primary and secondary research along with industrial forecasts. The multimedia usage forecasts were based on substantial primary and secondary research.

10.6.2.4 Forecast Methodology

Figure 10.20 illustrates the flow chart of the forecast process as provided by the American organization. The primary market research was based on the number of cellular users who expressed interest in IMT-2000 MSS services. These service descriptions were based on secondary research and analysis of the service type suited to this market. The usage forecast was derived through a combination of primary research and secondary analysis. The usage levels were based on the assumption of time spent outside of terrestrial coverage, depending on the user category and type of service. For example, voice and messaging usage occurs in areas without cellular coverage. A journalist would spend a larger time outside terrestrial coverage than a businessman. Assumptions were made on the usage per subscriber as mentioned in the preceding section. Usage levels were estimated for each of four service categories and specified as megabits per second per month. The traffic volumes were translated in to busiest hour traffic in the busiest cluster using various practical considerations. The busy hour usage was converted to spectrum usage with a spatial and spectral efficiency factor achievable within the timeframe.

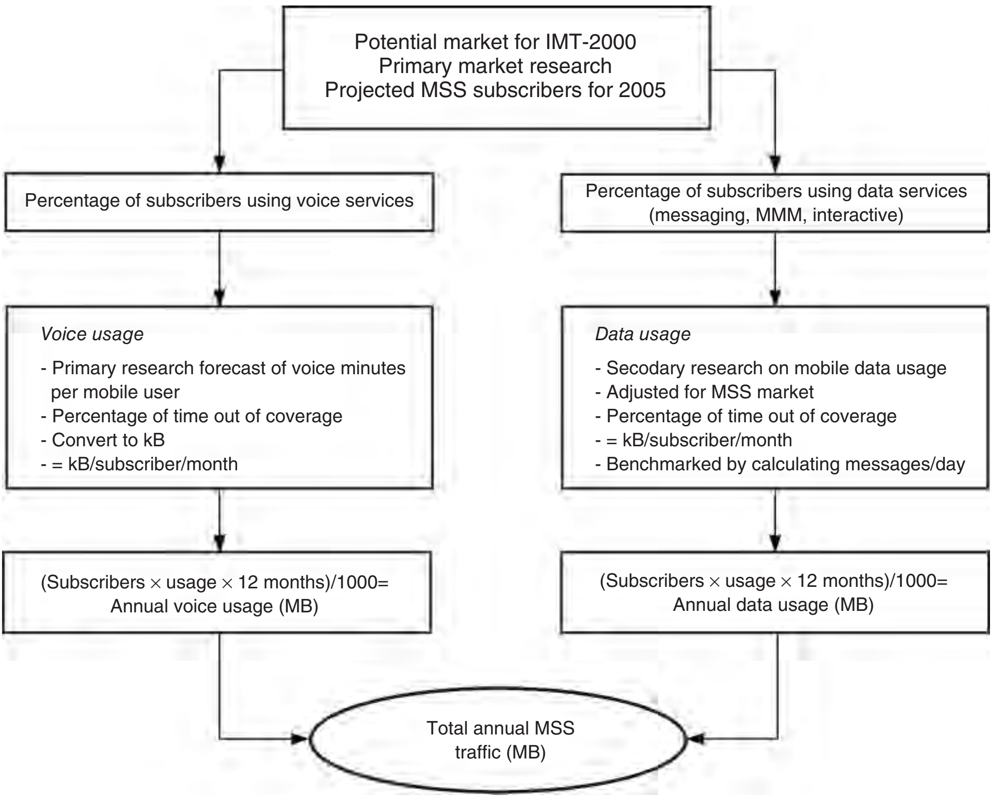


Figure 10.20 Flow chart to forecast spectrum usage by the US global satellite organization. (Source: ITU-1, 2000. Reproduced by permission of ITU.)

The ITU-R report M.2023 was updated to forecast spectrum over the period 2010–2020 taking into considerations data available since the publication of the original report ITU-R report M.2077 (ITU-2, 2006). It built upon subscriber forecasts, traffic models and application of the latest spectrum calculation methodology of ITU-R recommendation M.1391. The forecast was derived taking as the starting point known number of subscribers in 2006 amounting to 1.4 million, the traffic growth (since the first report), growth rate and penetration of cellular systems and nomadic technologies such as IEEE 801.16 and 802.20, GDP capita with growth rate and population distribution. The variables depended on the region (Asia, North America, South America, Europe, Africa and the Arab States), the user segment (e.g. professional) and usage environment (urban, rural).

10.6.2.5 Results

Number of Subscribers

Both the studies forecast for the geographical area where usage is the highest. Inmarsat's forecast of the total MSS spectrum estimated as 2×145 MHz (both directions, distributed equally) for the year 2010 was recommended for planning purposes.

Table 10.6 and Figure 10.21 show the number of subscribers for the years 2010–2020. The adjusted market review factors in an estimate on the basis of a 29% growth in mobile subscribers during the period 2002 and 2005. According to an ITU survey, the number of MSS subscribers increased from 643 000 by end of 2002 to 1 402 000 by the end of 2005. Two scenarios corresponding to pessimistic (9%) and optimistic (14%) were included to set the bounds. The proportion of subscribers in land, maritime and aeronautical environment indicates only a minor (%) variation of 99, 0.8, 0.2 respectively in 2010 to 99.2, 0.7 and 0.1 in land, maritime and aeronautical sectors respectively as summarized in Table 10.7.

According to a recent industrial survey, there were about 2.2 million in-service user units in 2011 and the annual forecast growth is anticipated to increase at a compound annual growth of 7.9% until 2021 resulting in total number of subscribers to 4.8 million in 2021 (NSR, 2012). These numbers are closer to the pessimistic bound in the ITU estimate, which amounts to 2.15 and 4.68 for the years 2011 and 2020 respectively. Note that these numbers

Table 10.6 Subscriber forecast up to 2020 using low, adjusted and high traffic scenario

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Subscribers for low traffic scenario (million)	1.976	2.154	2.348	2.559	2.790	3.041	3.314	3.613	3.938	4.292	4.678
Subscribers for adjusted market review (million)	2.170	2.425	2.766	3.195	3.574	3.918	4.249	4.720	5.097	5.479	5.881
Subscribers for high traffic scenario (million)	2.365	2.696	3.074	3.504	3.994	4.554	5.191	5.918	6.746	7.691	8.768

(Adapted from ITU-2, 2006.)

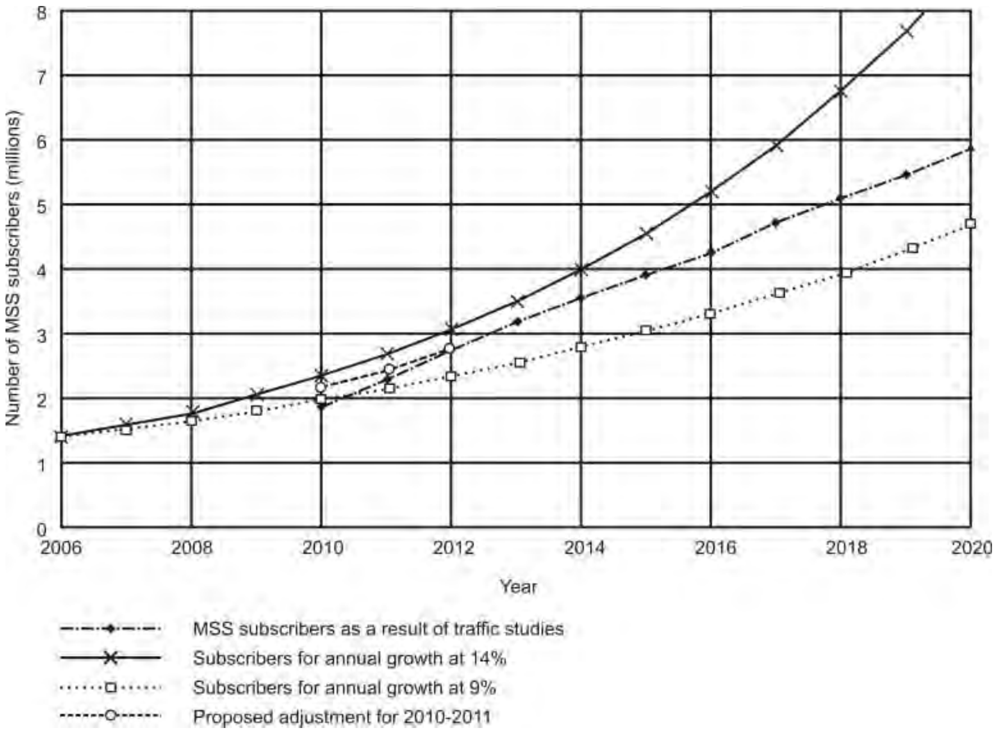


Figure 10.21 Subscriber forecasts for the period 2010–2020. (Source: ITU-2, 2006. Reproduced by permission of ITU.)

Table 10.7 Forecast breakup in the three sectors – aeronautical, land and maritime

Subscriber proportions (%)	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Land	99	99.2	99.2	99.2	99.2	99.2	99.2	99.2	99.2	99.2	99.2
Maritime	0.8	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Aeronautical	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
Total	100	100	100	100	100	100	100	100	100	100	100

(Adapted from ITU-2, 2006.)

do not project the revenue capability, which depends on the usage and service cost which is specified as average revenue per user (ARPU). For example, the NSR study observes that although the maritime market will continue to grow robustly in the next 10 years the ARPU from narrow band maritime units will continue to fall from the present \$180 per month. Whereas although there will only be a 10% increase in broad band subscriber unit of the maritime segment, these subscribers are anticipated to provide an ARPU of \$1200

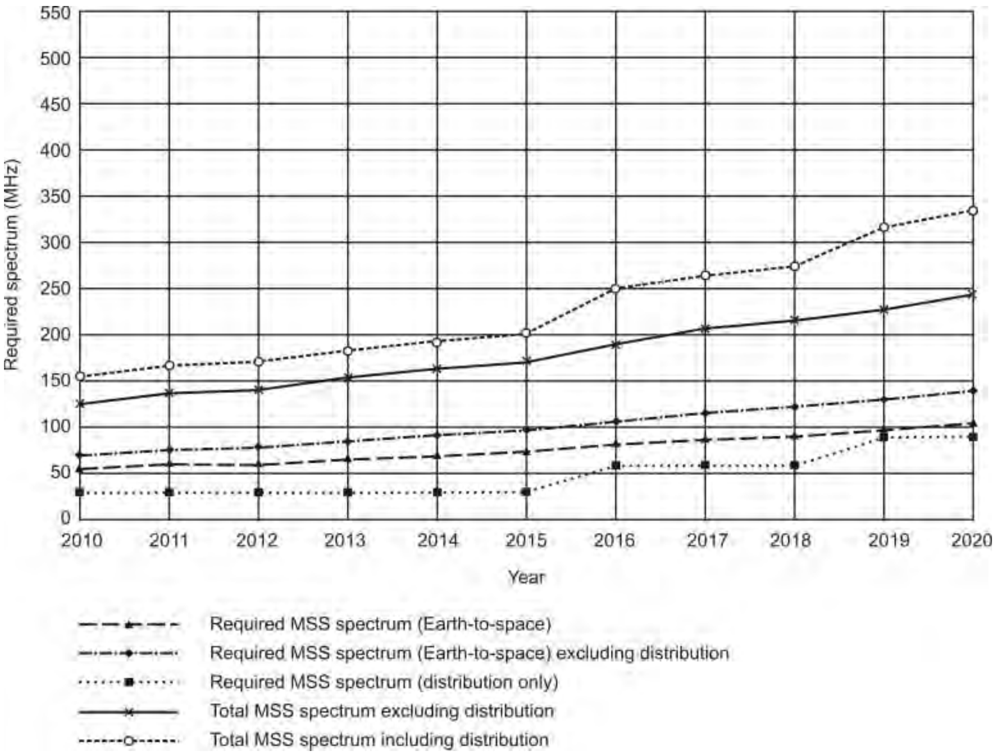


Figure 10.22 Spectrum estimate for the low traffic scenario for the years 2010–2020. (Source: ITU-2, 2006. Reproduced by permission of ITU.)

per month to boost the overall revenue of the sector to about \$2.5 billion in 2021. Similarly, despite the relatively small number of aeronautical units the aeronautical sector revenue is anticipated to increase from \$870 million in 2011 to \$2.3 billion in 2021.

10.6.2.6 Spectrum Requirements

Various assumptions relying on historic and anticipated developments were used in estimating the spectrum in the ITU studies. These include: number of systems sharing the (total) traffic, portion of daily traffic during busy hour, month-to-day conversion ratio, busy hour offset factor, number of beams in a frequency reuse cluster, coding rate, carrier data rate, grade of service, carrier bandwidth, efficiency factor, average effective data rate and signalling overhead factor. Due to the large uncertainty in predicting the traffic usage trend, bounding growth rates of low and high traffic scenarios were considered. Figure 10.22 illustrates the spectrum estimate of the low-growth scenario. It was observed that:

- spectrum requirement from non-multimedia services remain approximately constant;
- there is an increasing spectrum demand for multimedia services;

- requirements from asymmetric multimedia services (i.e. e-mail, Internet, intranet, file download type applications) and multimedia distribution (high bandwidth, content rich services) dominate;
- contributions from the remaining multimedia and non-multimedia services are rather constant and relatively small.

10.6.2.7 Spectrum Estimation Methodology

The conversion of the traffic intensity to spectrum was based on the Recommendation ITU-R M.1391 as follows (ITU-3, 2006). For multimedia traffic the spectrum requirement S (MHz) assuming a uniform traffic distribution is given as:

$$\mathrm{S} = \mathrm{N_b B\ N_c} \tag{10.8}$$

$$S = \mathrm{Nb} \cdot B \cdot \left\lceil \frac{T_{BH} \cdot 8\,000}{3\,600 \cdot \mathit{eff} \cdot R} \right\rceil \tag{10.9}$$

where

$\mathrm{N_b}$ = Number of beams in a frequency reuse cluster
$\mathrm{N_c}$ = Number of carriers (must be an integer)
The symbols $\lceil \cdot \rceil$ rounds the calculated number to the next largest integer
T_{BH} = Busy hour traffic in one beam (Mbytes)
$\mathrm{T_{BH}}$ (kbits) = $\mathrm{T_{BH}} \times 8\,000$; division by 3600 converts hourly traffic into unit of seconds
B = Carrier bandwidth (MHz)
eff = Efficiency factor to take into account the average loading of each carrier
R = Average effective data rate of a carrier (kbps)

The traffic in the busy hour, T_{BH}, is calculated by adding the traffic requirement of all categories of traffic (land, maritime, aeronautical, etc.). When the traffic forecast is given in megabytes per month (for data traffic) or min/month (for example voice traffic) a conversion to Mbytes in the busy hour is done as follows:

$$T_{BH} = \sum_i T_i \tag{10.10a}$$

$$T_i = \frac{T_{Mi} \cdot p_{BHi} \cdot p_{HSi} \cdot H_i}{MD_i \cdot N_b} \tag{10.10b}$$

$$T_i = \frac{T_{Mi} \cdot 60 \cdot R_{VC} \cdot p_{BHi} \cdot p_{HSi} \cdot H_i}{8000 \cdot MD_i \cdot N_b} \tag{10.10c}$$

where

T_{Mi} = Forecast global traffic per month for traffic category i; if this is given in megabytes, Equation 10.10(b) is used; if it is given in minutes Equation 10.10(c) is used
R_{VC} = Coding rate (kbps)
p_{BHi} = Portion of the diurnal traffic that occurs in the busy hour for traffic category i
p_{HSi} = Portion of the global traffic that occurs in a hot spot cluster for traffic category i
H_i = Busy hour offset factor (between 0 and 1) for traffic category i
MD_i = Month to day conversion ratio for traffic category i
N_b = Number of beams in a frequency reuse cluster

These equations assume that the traffic is uniformly distributed between the beams resident in the hot spot cluster. This simplification can underestimate the spectrum requirement.

Non-Multimedia Traffic

Three types of non-multimedia traffic are considered – low-speed data, messaging and voice telephony, each assumed to be carried on a separate carrier type and is designated by the subscript i in the following equations. Assuming that all non-multimedia traffic is circuit-switched, busy-hour traffic is specified in Erlangs. The Erlang-B formula is used to convert the traffic to the required number of circuits. The spectrum required, S_i, for the *ith* service is given as the spectrum occupied in the busiest cluster.

$$S_i = N_b \cdot ErlangB(T_{Erl,i}, GoS_i) \cdot B_i \tag{10.11}$$

where:

N_b = number of beams in a frequency-reuse cluster
$T_{Erl,i}$ = busy-hour traffic in one beam (Erlang) for traffic type i
GoS_i = grade of service (blocking probability) for traffic type i
B_i = carrier bandwidth (MHz) for traffic type i.

$$T_{Erl,i} = \frac{T_{M,i} \cdot H_i \cdot p_{HS,i} \cdot p_{BH,i} \cdot 8000}{N_{beams} \cdot MD_i \cdot R_i \cdot 60 \cdot 60} \tag{10.12a}$$

$$T_{Erl,i} = \frac{T_{M,i} \cdot H_i \cdot p_{HS,i} \cdot p_{BH,i}}{N_{beams} \cdot MD_i \cdot 60} \tag{10.12b}$$

where

$T_{M,i}$ = forecast global traffic per month for traffic type i; if this is given in Mbytes, Equation 10.12(a) is used, if it is given in minutes, Equation 10.12(b) is used

H_i = busy hour offset factor (between 0 and 1) for traffic type i (converts the busy-hour traffic of traffic type i to the traffic in the overall busy hour.)
$p_{BH,i}$ = portion of the diurnal traffic that occurs in the busy hour for traffic type i
$p_{HS,i}$ = portion of the global traffic that occurs in a busiest (hot-spot) cluster for traffic type i
MD_i = month-to-day conversion ratio for traffic type i
N_{beams} = number of beams in a cluster
R_i = carrier data rate for traffic type i

The total spectrum requirement S (MHz) for the non-multimedia traffic is the summation of the requirement of the three different traffic types, that is:

$$S = \sum S_i \quad \text{MHz} \tag{10.13}$$

It is assumed that busiest cluster is the same for all the three services.

Broadcast and Multicast Service

Broadcast or multicast service is considered as a particular case of point to point service with the following attributes:

- Traffic is independent of number of subscribers over the service area; it is assumed that broadcast/multicast traffic forecast would have considered a minimum subscriber threshold to ensure system viability.
- It is assumed that the traffic forecast provides a data volume (e.g. Mbyte) over a period (e.g. a month); therefore month to day conversion factor would typically be 30 assuming continuous transmissions.
- The concept of busy hour does not generally apply as the channels remain active uniformly throughout; therefore the busy hour traffic can be set to 1/24 of the daily traffic volume; and busy hour offset factor (H) will be unity; exception to these assumptions arise when radio channels are switched on or off periodically in parts of the coverage (i.e. some beams).
- The carriers are fully loaded and thus the fill factor equals 1; therefore the carrier loading efficiency (eff) equals 1.
- The concept of busiest (hot spot) cluster may not apply due to the broadcast nature of the service, except when a larger number of channels are transmitted over those beams that illuminate populated areas.

With these assumptions, Equation 10.9 reduces to:

$$S = N_b \cdot B \cdot \left\lceil \frac{T \cdot 8000}{3600 \cdot R} \right\rceil \tag{10.14}$$

Similarly, Equation 10.10(b) reduces to:

$$T = \frac{T_M \cdot p_{BH}}{MD \cdot N_b} \tag{10.15}$$

10.6.3 Eurocontrol/FAA Approach to Traffic Model Methodology

A document known as Communications Operating Concept and Requirements (COCR) to define the Future Radio Systems (FRSs) prepared by Eurocontrol and Federal Aviation Administration (Eurocontrol/FAA, 2007) identifies concepts, requirements and trends that will form the basis for the FRS (Eurocontrol/FAA, 2007). The FRS concept is intended to comply future safety and regularity of flight communications requirements, that is Air Traffic Services (ATS) and safety related Aeronautical Operational Control (AOC) communications dealing with ground to air and air to air communication.

COCR proposes an approach to estimate the capacity of FRS over a period extending as far out as 2030. FRS loading is obtained by estimating message sizes, message frequencies, performance requirements and airspace aircraft densities projected into the future. The loading is fed in to a queuing model to estimate the required capacity of the FRS. The queuing model has been outlined in Chapter 7.

The service instances (i.e. typical number of times a service is used within an operational area), service volume flight durations (service volume refers to an airspace domain, e.g. Airport, enroute, etc.) and service volume peak instantaneous aircraft count (PIAC) are used to derive message arrival rates. The COCR identifies the addressed services. Having obtained the loading, each service is assigned to a queue in accordance with the communication class of service (COS). Each COS is assigned a priority relative to the others. The individual service arrival rates and message sizes are used to calculate the queue message arrival rates and sizes. Using this approach it is possible to estimate the capacity required in various phases of the flight. Satellites are ideally suited for the enroute communications (i.e. when the flight has settled after take-off) with capability to support communication in other phases.

The COCR document uses PIAC projections based on two separate computer-based air traffic models supplemented with estimates from subject matter experts for new airspace types and in sectors not accounted for in the referenced computer models. These air traffic models are:

- EUROCONTROL'S System for Traffic Assignment and Analysis at a Macroscopic Level (SAAM) tool, which simulate air traffic and provides data about the traffic through specified airspace volumes.
- The MITRE Corporation's Center for Advanced Aviation System Development Mid-Level Model uses Terminal Area Forecast (TAF), which is a compilation of scheduled airline flights growth. TAF is further refined through the observations of the Enhanced Traffic Management System to determine unscheduled traffic. The model provides PIAC counts for 2004, 2013 and 2020. The PIAC counts for 2030 are extrapolated for the estimates used in the COCR.

10.7 End-User Perspective – A Case Study

10.7.1 A Maritime Perspective

The mobile population aspires to a high-quality, low-cost, low maintenance service throughout the region of interest. In the dynamic domain of mobile communications the user is left with a bewildering choice – this, for example is the situation for the cellular systems that support a plethora of service providers and vendors. A similar situation has arisen in the maritime broadband satellite communication segment due to a spurt in demand for broadband and the consequent entry of numerous operators that provides a wide choice.

Here we address the alternatives available to the maritime MSS community, beginning with the considerations as to whether a satellite communication system with the associated costs is indeed necessary for a given scenario.

It is argued that that the advantages obtained by using satellite communications on ships can offset the lifetime cost of the satellite system. This premise needs to be established on a case-by-case basis through a trade-off analysis weighing the cost versus the benefits. A maritime satellite communication system can provide the following advantages (KVH-2, 2012):

- Increases a vessel's operational efficiency through enhancements in navigation, voyage planning, on-board logistics and collaboration with shore-based support;
- Optimizes engineering and IT functions by enabling remote systems monitoring/diagnostics, software updates and electronic access to e-documentation and minimizes travel costs and labour charges for repairs;
- Facilitates safety and security through vessel tracking to avoid piracy attacks, documentation of on-board events, liaison with maritime security organizations and assistance in meeting regulatory requirements;
- Enhances crew management efficiency and boosts crew morale through access to e-learning programmes, email, phone, Internet, and by enabling communications with family and friends;
- Improves customer service through real-time status reports.

To obtain the best available solution it becomes necessary for the user to analyse the requirements in terms of the required throughput and usage, service area, life-time cost of the system that includes costs associated with user terminal, installation, maintenance, after sales support and airtime, necessity of Global Maritime Distress and Safety System (GMDSS) support (see Chapter 13) and reliability.

The capacity (typically, megabytes/per month) is determined by the anticipated usage of communication applications, that is email, web-browsing, reception and dispatch of images, video streaming, voice communication, and so on. A typical business use on a ship is said to be about 10–15 MB per day (KVH-1, 2012). Forward planning caters for future expansion as new innovative applications continue to emerge.

Ship's operational routes establish the desired service area. Global coverage is generally available in L and C bands, although K_u band coverage at present tends to favour land. Figure 13.9(a) shows the shipping lanes across the world as an illustration. Note that a majority of routes are covered by geostationary satellite systems. For routes beyond about $\pm 76^\circ$ a non-geostationary satellite system would be necessary.

The life time costs include airtime, hardware (including contractual obligations, support, etc.), installation including configuration of on-board connectivity, logistics (e.g. schedule, management of installation, procurement, etc.), on-board connectivity aspects (LAN, VoIP,

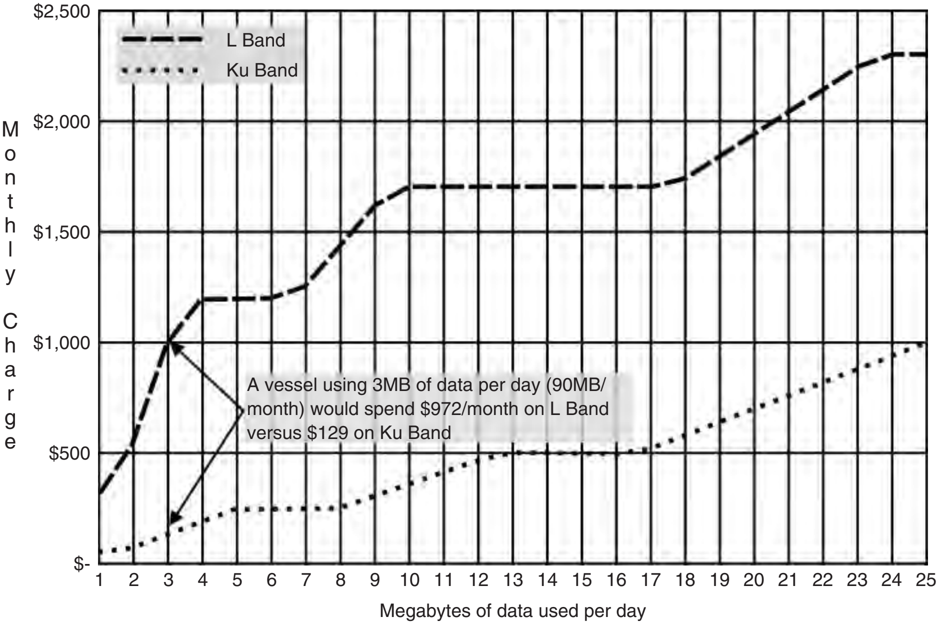


Figure 10.23 A comparison of monthly cost at various daily usage for an upper end L band user terminal and state-of-the-art maritime MVSAT. (Source: KVH, 2012. Reproduced with permission of KVH.)

bandwidth management system, security, end-end application set-up, etc.) and after-sales maintenance and support.

Figure 10.23 (KVH-2, 2012) is an illustrative monthly air-time cost comparison of Ku and L band systems with various levels of daily traffic. The example demonstrates an advantage of nearly 87% ($972 versus $129 per month) for the K_u band system when the required capacity is 3 MB per day.

10.7.1.1 Available Alternatives

With the availability of affordable, high capacity mobile MSS and FSS services covering all the major shipping routes, the maritime MSS community has several alternatives to choose from.

Systems are available to operate in various MSS and FSS frequency bands – L, S, C, K_u and K_a.

L band systems operate in the primary MSS band. The L band generally provides a more reliable radio link compared to other bands with a connectivity to the GMDSS system (see Chapter 13). L band systems are mature, and offer turnkey solutions at reasonable terminal cost capable of easy installation; however, due to the scarcity of the L band spectrum the available throughput is limited. Airtime costs tend to be high because the systems serve low directivity small user terminals and hence require high – and hence expensive – EIRP transmissions.

The legacy C and K_u band maritime very small aperture technology (VSAT) terminals operate in the FSS bands where much wider bandwidth is available. In addition these terminals are directive so that satellite EIRP is lower in comparison to the L band. This combination allows higher bit rate at a lower airtime cost compared to L band MSS systems. The first generation (legacy) mobile very small aperture terminal (MVSAT) terminals are assembled by procuring subsystems of different manufacturers, and therefore, are complex and expensive often requiring more than a day for installation. The coverage of legacy C and K_u band systems is often limited to shipping routes. Note that the C band transmissions can be restrictive near coasts due to radio regulation, although this band offers a more robust radio link than K_u or K_a band.

Optimized versions of K_u and K_a band MVSATs can provide seamless connectivity across the world. Since the terminals are designed to support mobility these modern versions are compact, low cost and spectrally efficient resulting in an overall improvement in performance and cost.

Hybrid systems such as K_u-C band expand the K_u band coverage through global beam operating in C-band. The user terminal switches seamlessly between these bands to provide continuous coverage. Similarly, hybrid K_u-L band systems expand the K_u band coverage globally through the L band GEO or LEO system. Figure 10.24(a) shows as an example the coverage of a hybrid K_u-L band system using a K_u VSAT system overlaid with an L-band system to provide a near global coverage and Figure 10.24(b) shows the coverage of a K_u-C band MVSAT coverage.

Table 10.8 lists the characteristics of some representative systems available in recent years (KVH-2, 2012). Notice the higher user terminal cost, larger throughput and lower air-time cost of mobile VSAT in comparison to L band MSS system. However L band terminals are optimized for mobile applications tailored to provide turn-key solutions, and additionally, provide access to the GMDSS system. L band systems offer a highly reliable solution for low/medium bit rate services. The space segment architecture of L band MSS systems are robust such as to provide a rapid back up in case of a satellite failure, which may not be the case for the mobile VSAT system.

In the final reckoning the user selects the most suitable overall option, depending on the requirements set out initially.

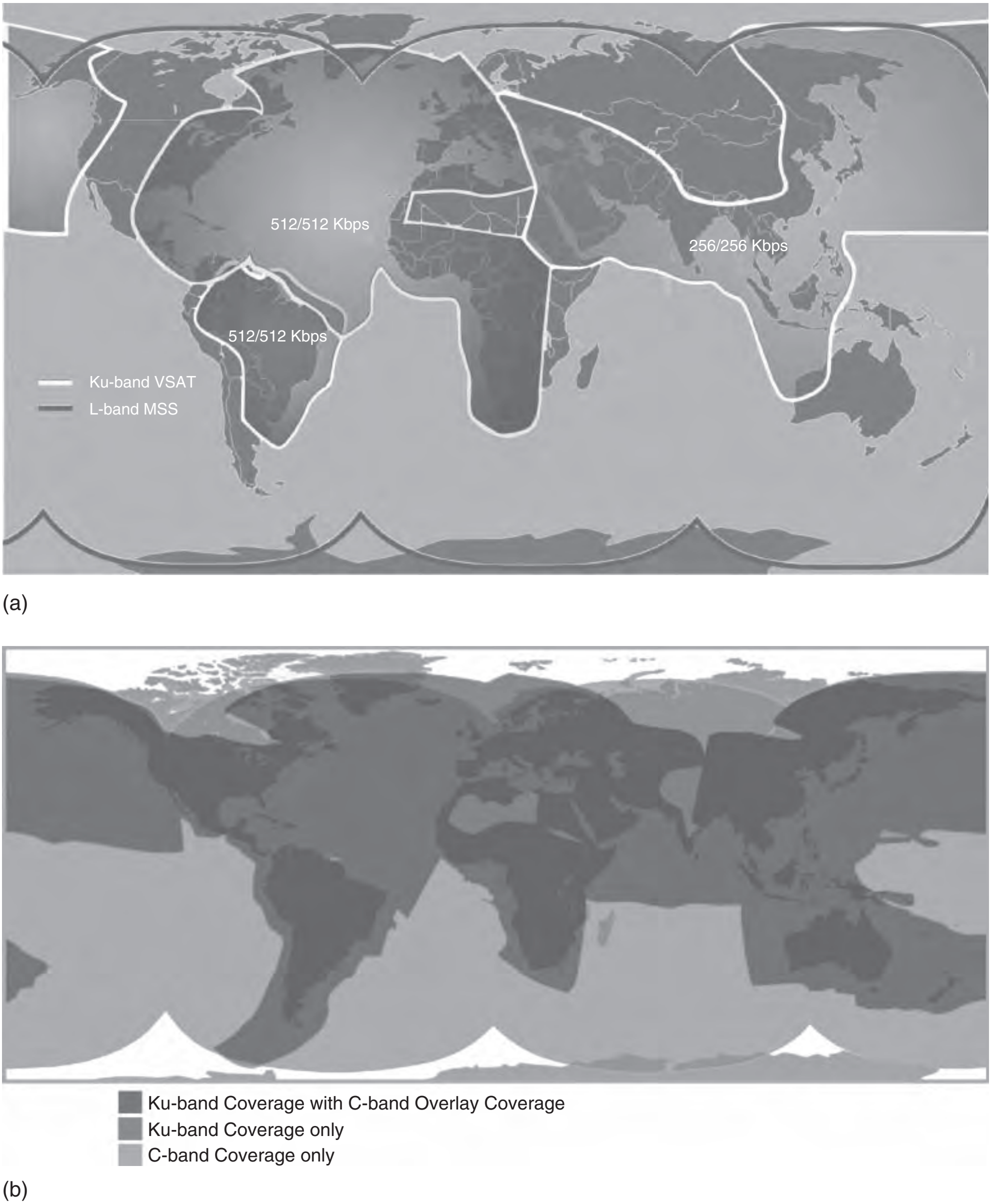


Figure 10.24 (a) Seamless coverage provided by a hybrid K_u-L band MSS-MVSAT system. (b) Seamless MVSAT coverage provided by a K_u-C band coverage. (Both parts source: KVH-2, 2012. Reproduced by permission of KVH.)

Table 10.8 A comparison of various types of maritime satellite communications

Service	Frequency band	Orbit	Coverage			User terminal antenna diameter[a] (cm)	Throughput[a] uplink/downlink (kbps)	Typical cost of user terminal (US $, 2012 base)	Typical airtime charge per month (US$) – metered usage		
			Global-except polar region	Global-including polar region	Part global – covering major shipping lanes				30 Mbyte/month	90 Mbyte/month	300 Mbyte/month
MSS	L	GEO	Yes	No	N/A	30–60	150–432/150–432	7 500–19 995 (depending on capability)	300	972	1700
MSS	L	LEO[b]	Not applicable	Yes	N/A	57	134/134[b]	4 995	–	–	–
FSS	C (legacy)	GEO	No	No	Yes	365	512/1000	14 600	–	–	–
FSS	K_u (legacy)	GEO	No	No	Yes	127	512/1 000	59 995	–	–	–
FSS	K_u (optimized)	GEO	Yes (as an option)	No	Yes	39–130	128–1000/2000–4000	16 995–35 995	49	129	349
MSS	K_a[c]	GEO	Yes	–	N/A		(Anticipated up to 50 Mbps in downlink)	Not available	–	–	–
Hybrid FSS/MSS	C/L	GEO/LEO	Yes	Yes	N/A	Not available	Not available	Not available	–	–	–
	K_u/L	GEO/LEO	No	Yes	N/A	Not available	Not available	Not available	–	–	–
	K_u/C (optimized)	GEO/GEO	Yes	No	N/A	130	1 000/4 000	74 995	–	–	–

[a]Terminal dependent.
[b]Next generation systems will improve throughput (refers to Iridium system).
[c]System yet to be operational (Anticipated service entry – 2013: refers to Inmarsat system).
N/A = Not applicable.
(Adapted from KVH-2, 2012.)

Revision

1. **State** the reasons for the financial failure of various NGSO mobile satellite system start-ups in the first decade of the millennium. What lessons can be learnt from the downturn?
2. With the help of a diagram discuss the role of the system entities and their interaction at the planning stage of (i) a new MSS venture and (ii) a new product of an incumbent MSS operator.
3. What are the advantages of a systematic planning approach for the development of a commercial mobile satellite system?
4. Discuss the concept and advantages of a lifecycle cost–benefit analysis with examples of inputs and outputs into the process.
5. Identify the factors that contribute towards revenue improvement of an MSS operator. Explain their individual and collective role.
6. Outline the similarities and differences between the service-distribution schemes used in MSS practice.
7. Space segment usage and call charge impact the revenue of an MSS operator directly. The operator attempts to maximize the usage of the available space segment radio resource namely spectrum and EIRP. Estimate the revenue generating capability of a single-satellite regional system specified below operating a circuit-mode voice service at call charge ranging from \$0.25–2.00 per min. State the assumptions and suggest whether the system is spectrum or power limited.

   Spacecraft EIRP = 73 dBW; available bandwidth = 6 MHz; bandwidth per channel = 5 kHz; spectrum utilization efficiency = 90%; average EIRP/channel = 25 dBW; cluster size = 7; spectrum re-use capability (reuse factor) = 5.
8. Outline the main features of the forecast methodology used by:
   i. The UMTS forum during the planning stage of the third generation mobile systems
   ii. The ITU to forecast spectrum requirements up to the year 2020.

   Compare the salient features of these two methodologies.

References

Eurocontrol/ FAA (2007) Communications Operating Concept and Requirements for the Future Radio System, COCR Version 2, Future Communications Study Operational Concepts and Requirements Team, May.

Corbley, K.P. (1996) Accessing Satellite and Cellular Systems – Dual Mode Handsets Provide the Option. *Via Satellite*, 76–90.

Crossman, M. (1999) The little LEO scorecard. *Satellite Communications*, 70.

Dondl, P. (1995) Standardisation of the satellite component of the UMTS, *IEEE Personal Communications*, **2**, 68–74.

Foley, T. (1997) Satellite Financing – which Direction Will the Market Take. *Via Satellite*, Vol. **14**, 22–30.

Gaffney, L.M., Hulkower, N.D. and Klein, L. (1995) Non-GEO mobile satellite systems: a risk assessment. IMSC'95, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. A23–A27.

Globalstar (2009) Globalstar Completes \$738 Million Financing, Globalstar press release, July 1, 2009, http://eu.globalstar.com/en/index.php?cid=7010&pressId=549 (accessed 27 August, 2013).

Hu, Y.F. and Sheriff, R.E. (1995) Traffic model for the satellite component of UMTS, IMSC 1995. The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 508–511.

Hu, Y.F. and Sheriff, R.E. (1997) The potential demand for the satellite component of the universal mobile telecommunication system. *Electronic and Communication Engineering Journal*, **9**(2), 59–67.

Ingley, C. (1999) Global vision: making the right connections. *Satellite Communications*, **23**, 38–45.

Inmarsat (2012) web site, available at: www.igx.com/ (accessed 27 August, 2013).

Iridium (2012) web site, available at: www.iridium.com/About/IridiumNEXT.aspx (accessed 27 August, 2013).

ITU-1 (2000) Spectrum Requirements for International Mobile Telecommunications-2000 (IMT-2000), Report ITU-R M.2023, ITU.

ITU-2 (2006) Traffic Forecasts and Estimated Spectrum Requirements for the Satellite Component of IMT-2000 and Systems beyond IMT-2000 for the Period 2010 to 2020. Report ITU-R M.2077, ITU.

ITU-3 (2006), Methodology for the Calculation of IMT-2000 Satellite Spectrum Requirements. Recommendation ITU-R M.1391-1, ITU.

ITU-4 (2011) Traffic Forecasts and Estimated Spectrum Requirements for Future Development of the Mobile-Satellite Service in the Range 4-16 GHz, ITU.

KPMG (1994) Satellite communications and their consequences for European telecommunications trade and industry. Report to European Commission, March, ITU.

KVH-1 (2012) Connectivity at Sea: A Guide to Advances in Maritime VSAT, KVH, www.kvh.com (accessed 27 August, 2013).

KVH-2 (2012) Comparing mini-VSAT Broadband to Legacy Solutions, KVH, www.kvh.com (accessed 27 August, 2013).

Lazear, Y.M., Capots, L.H., Langlais, R.D. and Price, K.M. (1997) System planning methodology for satellite communication services. IMSC 1997, Fifth International Mobile Satellite Conference.

NSR (2012) *Mobile Satellite Service, NSR Research – Report Briefing*, 8th edn, Northern Sky Research LLC, Cambridge, MA www.nsr.com/ (accessed 27 August, 2013).

ORBCOMM (2012) www.orbcomm.com/Collateral/Documents/English-US/2012Q1Earnings.pdf (accessed 27 August, 2012).

Sultan, N. and Groepper, P.H. (1999) Mobile satellite life cycle cost reduction: a new quantifiable system approach. IMSC 1999, Sixth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre and the Jet Propulsion Laboratory, pp. 246–251.

Thuraya (2012) www.thuraya.com/who-we-are (accessed 27 August, 2012).

UMTS Forum (1999) The Future Mobile Market. Report No. 8, London: UMTS Forum, March.

11

Representative MSS Systems

11.1 Introduction

In this chapter, we illustrate practical applications of concepts discussed in previous chapters by presenting examples of a range of MSS (mobile satellite service) systems representing different approaches to system design. The basis in the selection these systems is technical diversity rather than commercial viability or performance. Some of these were not implemented but their concepts are innovative and representative of this book's theme. This comprehensive survey will give the reader an in-depth exposure to the technical aspects and merits of mobile satellite systems. The operational and commercial status is not relevant in this context.

We will consider examples of GEO (geostationary orbit), MEO (medium earth orbit), LEO (low earth orbit) and hybrid orbit satellite systems. Inmarsat and ACeS (Asia Cellular Satellite) are representative of classic global and regional geostationary satellite systems, and EUTELTRACS system provides an innovative regional low bit rate solution operating within secondary allocated K_u band spectrum; Intermediate Circular Orbit (ICO) system (not implemented), although dated in some respects, includes various innovative aspects of non-geostationary system; the Iridium and Globalstar systems – both operational systems – are examples of big-LEO satellite systems using diverse configurations and technology; ORBCOMM is an example of a low-cost innovative little-LEO system that operates in the Little-LEO VHF (very high frequency) band deploying small satellite to offer low-cost, low bit rate two-way packet-switched communication services in large parts of the world: the Ellipsat system (not implemented) represents a system configured over an innovative hybrid constellation design. LightSquared's ATC (Auxiliary Terrestrial Component) enabled proposal represents an emerging class of satellite-terrestrial hybrid architecture that draws strengths from each system while reusing MSS frequencies terrestrially. In each example, the baseline design has been described rather than the evolved system; regular adjustments to system capabilities driven by dynamic business strategies are common. Table 11.1 lists the distinct features of each of these systems.

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

Table 11.1 Distinctive features of systems described in the chapter

System	Specific feature
Inmarsat	An operational geostationary global system that offers voice, broadband data and IP services on fixed/nomadic, portable and mobile user terminals
Asia cellular satellite system (ACeS)-Inmarsat	A regional geostationary satellite system offering circuit-mode voice and data service on handheld units in partnership with Inmarsat
EutelTRACS	A pan-European, low bit rate geostationary satellite system
Lightsquared ATC proposal	An ATC enabled hybrid geostationary MSS network proposal in USA to provide 4G-LTE services. The proposal was shelved due to concerns about harmful interference caused by ATC transmitters to GPS receivers
ICO Global Communication system (original proposal)	A MEO satellite system proposal for voice service on handheld units; uses diversity and transparent transponder; its evolved version supports broad-band communication
Globalstar	A big-LEO satellite system capable of global hand-held voice service; uses transparent transponder, inclined orbits, CDMA, diversity and soft-handover
Iridium	A big-LEO satellite system to provide global hand-held service; uses regenerative transponder, intersatellite link and polar orbit
ORBCOMM	A little-LEO low bit rate satellite communication system using small low-cost satellites operating in VHF band
ELLIPSAT proposal	A big-LEO satellite system proposal based on an innovative hybrid constellation for regional/global hand-held service; uses transparent transponder and allows gradual evolution of coverage; favours coverage of populated regions

11.2 Geostationary Satellite Systems

11.2.1 Inmarsat System

The International Mobile Satellite Organization (Inmarsat) is an international MSS provider, offering seamless communication services through a constellation of geostationary satellites to maritime, land and aeronautical sectors throughout the world except those areas that lie outside of geostationary satellite footprints.

The services are provided by a constellation of 10 geostationary satellites (April 2013) with plans to introduce an additional fourth generation satellite and a fleet of three fifth generation K_a band satellites by the year 2014. The fourth generation L-band satellites provide IP (Internet Protocol) services at download speeds of up to 492 kbps to a variety of fixed and mobile user terminals. Work is in hand to introduce a K_a band system with peak rates of up to 50 Mbps to mobile and fixed user terminal from 2014 on its fifth generation satellites.

While continuing to evolve its present service capabilities, the company continues to support its previous generation service portfolio, referred below as legacy service. The legacy services are based on classical MSS architecture that supports a large number of independent service providers sharing a global space segment infrastructure. The shared technical and management structure of the company in conjunction with international cooperation and non-commercial operational philosophy in the formative stage of the industry provided a solid base for the development of the MSS. Although the distributed (but inefficient in terms of resource utilization) architecture of the legacy system was superseded in the new generation by a centralized architecture, we have retained the legacy systems for its technical interest.

11.2.1.1 Features

Inmarsat operates its legacy services on a ground segment comprising a number of fixed Earth stations, a wide variety of mobile earth stations (MESs), and a space segment consisting of a constellation of bent-pipe geostationary satellites. Legacy services are services introduced progressively since 1990s on its second and third generation satellites and continue to be supported together with the new services.

The fourth generation satellites, in addition to supporting the legacy services, provide a portfolio of broadband services at data rates up to 492 kbps over a different class of ground segment architecture that support IP services based on the third generation partnership project (3GPP) core network (CN) and a standardized air interface.

The satellite constellation provides a seamless coverage of the Earth in the regions within the footprint of geostationary satellites. Depending on the service, the world is partitioned in to three or four regions to provide a seamless coverage. Each operational satellite is backed up by an in-orbit spare satellite ready to take over operation in case of failure of an operational satellite. Capacity on spare satellites may be leased to third parties on a pre-emptive basis such that the satellite can be pressed into operations in the eventuality of a failure. The safety communication services are compliant to Global Maritime Distress and Safety System (GMDSS) recommendations (see Chapter 13).

11.2.1.2 Legacy Systems

Gateways – known variously as coast earth stations (CESs), land earth stations (LESs) and ground earth stations (GES) respectively by the maritime, land and the aeronautical user community – serve as interfaces to the terrestrial public networks. We refer to them here generically as LESs.

To maintain the integrity of the network, each LES has to comply with a specific set of RF (radio frequency) requirements. Table 11.2 represents a sample of LES parameters; these specifications are revised by Inmarsat when necessary and hence the table should not be used as such for practical purposes without verification. Feeder links operate in the C band whereas service links operate in the L band. Each LES requires a number of verification tests before being authorized to operate within the network. There were about 40 LESs owned by a number of service providers distributed around the world at the peak of third-generation operation (early 2010s).

MESs are categorized according to the environment in which they operate – aeronautical, land or maritime. Within each category, the mobiles are further subdivided into ‘standards’, according to the specific range of service provided.

Table 11.2 Representative technical parameters of fixed earth stations of the legacy network

Parameter	C band	L band
Transmit pass band (MHz)	6417.5–6454.0	1626.5–1660.5
Receive pass band (MHz)	3599.0–3629.0 and 4192.5–4200 (this band supported first generation satellites)	1525.0–1559.0
Polarization:		
transmit	Right hand circular (RHC) and left hand circular (LHC)	RHC
receive	LHC and RHC	LHC
Typical antenna gain:		
Transmit (dBi)	54.0	29.5
Receive (dBi)	50.5	29.0
Ground station EIRP/per carrier (various types of carriers are transmitted) (dB W)	45–70	12–36
Receive G/T (dB/K)	30.7 (for operation with second and third generation satellites)	2.0
EIRP stability (dB)	± 0.7	+ 1 to −2
Frequency tolerance (Hz)	± 100	± 200

Compiled by Author (Courtesy, Inmarsat.)

Each MES manufacturer obtains a one-off type approval from Inmarsat before manufacturing and marketing a specific type of terminal. In order to introduce a new terminal into the network, it is formally commissioned by Inmarsat.

Salient features of the major legacy services supported on the third generation satellites are summarized in Table 11.3. However, as the products continue to evolve, the interested reader should visit the Inmarsat web site for the current portfolio.

Inmarsat's primary role is to provide space segment infrastructure; it earns its revenue by levying charges to each LES operator for space segment use. LESs are generally operated by the service providers. Service provider's charge to the end user includes the space segment charge, operational costs and a profit component.

Inmarsat's third generation satellites include a navigation transponder that is capable of enhancing the accuracy, availability and integrity of GPS (global positioning system) and GLONASS (Global Navigation Satellite System) navigation signals. The signals transmitted in the navigation band are able to provide:

- a system integrity channel that would carry, amongst other useful navigation signals, the status of GPS and GLONASS satellites to ensure that data from faulty satellites are not used at the receiver;
- additional ranging signals to improve GPS availability;
- wide-area differential correction signals that can be used by receivers to improve accuracy.

The following sections summarize features of legacy services.

Table 11.3 Characteristics of mobile earth stations supported in the Inmarsat legacy network

Service	Approximate year of introduction	Communication capability	Receiver pass band (MHz) (Note 1)	MES G/T (dB/K)	MES EIRP (dBW)	Antenna type	Applications
Fleet 77	2003	Up to 64 × 2 kbps	1525–1559	−4	33	Parabolic steerable, 0.8–1 m	Voice, data (ISDN – 64 and 128 kbps); mobile packet data service, fax 9.6 kbps (Group 3), 64 kbps (Group 4), distress and safety
C	1990	Low bit rate store and forwarded communication at 600 bps with interfaces to X.400 and X.25 PAD networks, group call	1530–1545	−23	12	Omni-directional	International access to electronic mail, database access, global telex, X.400 and X.25 fax delivery
Aero-C	1992	As Inmarsat-C	1530–1545	−23	11	Omni-directional	Messages up to 32 000 characters, non-safety related communications on corporate, GA aircraft and helicopters
Aero-L	1990	Real time duplex communications at 600 bps	1530–1559	−26	13.5	Omni-directional	Real-time flight and passenger related communication, for example engine monitoring

(*continued overleaf*)

Table 11.3 (*continued*)

Service	Approximate year of introduction	Communication capability	Receiver pass band (MHz) (Note 1)	MES G/T (dB/K)	MES EIRP (dBW)	Antenna type	Applications
Aero-H and H+	1990	Circuit and packet mode communication at 9.6 kbps	1530–1559	−13	10.5–25.5	Steerable mechanical or phased arrays (12 dB gain)	Real-time medium bit rate communication, for example voice, fax; H+ offers all features of H and voice is offered at a lower cost by the use of a more efficient voice codec at 4.8 kbps
Aero-I	1998	Cockpit/passenger voice telephony and data in spot beams; packet mode ISO 8208 data 600 bps to 4.8 kbps in global beam	1530–1559	−19	19.5 (max)	Steerable mechanical or phased arrays (6 dB gain)	Passenger voice telephone, facsimile, PC data within Inmrsat-3 spot beam footprint, packet mode data, cockpit voice and data, air traffic control, secure voice access to major air traffic control centres

Note: Add 101.5 MHz to obtain transmission frequency band.
Compiled by Author (Courtesy, Inmarsat.)

Services

The Inmarsat-C system supports two-way, store and forward communication at a data rate of 600 bps for land and maritime users on terminals weighing only a few kilograms. Inmarsat-C terminals use small non-tracking antennas, to simplify installation and maintenance. The service is suited for non-real-time short data messages such as position reporting or half-page of text (~ 250 characters). About 90% of messages are delivered in less than 10 min. It is also possible to send telex, e-mail and facsimile. The system can provide specialized communications – broadcasts, data polling and reporting. Broadcasts support group message delivery either within a specific geographical area or a specific group. User terminals can be programmed to report information, either at fixed intervals or in response to a poll request. Data reporting allows short reports to be sent from remote terminals with lower delay and at a lower cost. The system was introduced to complement the more expensive service existing at the time and provide services to smaller ships and yachts. It is not necessary to be logged on to receive a message, as the message is saved in an electronic mailbox, from which the message is downloaded at log-on.

Fleet broadband (known as Fleet 77) is a maritime service with capability to support voice and ISDN (Integrated Services Digital Network) and packet switched (PS) data. The terminals meet the distress and safety requirements specified by the International Maritime Organization (IMO) in resolution A.888 for voice pre-emption and prioritization within the GMDSS. The maritime user terminal comprises above-deck equipment containing a parabolic antenna and electronics connected to below-deck equipment containing electronic units, power supplies and interface connections to telephone, computers, ISDN Terminal Adapters, and so on.

A call from a Fleet 77 user terminal is routed via an Inmarsat satellite to an LES and then directed into the national and international phone, telex and data networks.

A number of legacy services are available for the aeronautical sector covering safety and regular communications – they include Aero-I, Aero-H, Aero-H+.

Aero-L provides two-way 600 bps real-time data communication using terminals with omni-directional antennas. The service complies with proposed International Civil Aviation Organization (ICAO) requirements to support future air navigation systems. The service is used mainly for air traffic control, operational and administrative purposes.

The Aero-H and H+ services provide real-time, 10.5 kbps packet data communications for ACARS (Aircraft Communications Addressing and Reporting System), FANS (Future Air Navigation System) and ATN (Aeronautical Telecommunication Network) communication to assist flight-deck and airline safety operations using medium gain (10–12 dB) antenna terminals within Inmarsat-3 global beam footprint around the world. They also provide data rates up to 9.6 kbps per channel for multi-channel voice, fax and data links for passenger, operational and administrative communication. The Aero H+ service is an evolution of Aero-H that can additionally operate within Inmarsat-3 spot beams to provide a more efficient and low cost voice service voice using 4.8 kbps voice codec within the spot beam coverage of Inmarsat-3. The Aero H+ service can also operate within the full footprints of Inmarsat-4 satellites.

The Aero-I service offers medium bit rate voice and 4.8 kbps circuit-switched data service through small, medium gain, low-cost terminals within spot beam coverage of Inmarsat-3 satellites. Packet data services are available world-wide via the global beam including Inmarsat-4 footprint.

Aero-C provides 600 bps store and forward data communications to aircraft and is an extension of the Inmarsat-C system. The service, excluding flight safety communications, is suitable for regional aircraft operators, in particular those operating in remote regions and aircraft operators in need of reliable service, but who do not envisage operating in the ICAO's proposed Air Traffic Management System.

Space Segment

A network operations centre (NOC), located at Inmarsat headquarters in London, maintains the integrity of the legacy network through continuous monitoring of satellite transmissions and traffic flow through the network. TT&C (tracking, telemetry and control) centres, placed around the world, can be managed centrally from its satellite control centre.

Satellites

Inmarsat's first generation satellites were phased out during 1991–1992, and replaced by second generation satellites to meet the increased traffic demand. Beginning in 1995, the second generation satellites were replaced by third generation (Inmarsat-3) satellites, which deploy up to seven spot beams. Tables 11.4 and 11.5 summarize the main characteristics of these satellite generations. Inmarsat-2 satellites provided around four times the L band EIRP (effective isotropic radiated power) and additional 10.5 MHz bandwidth compared to first generation satellites.

Third generation satellites deliver about 7.5 times the power of Inmarsat-2 satellites and provide 16 MHz wider bandwidth in the forward direction and 11 MHz in the return direction, and in addition deploy spot beams. The enhanced G/T of spot beams supports smaller/cheaper mobile terminals. Satellites can allocate the EIRP and spectrum to spot beams flexibly (in non-real time), depending on traffic requirements, to allow effective

Table 11.4 Inmarsat satellites: the main parameters of the mobile communications transponders as specified

Generation	Identification (generic)	Frequency band	Frequency (MHz) and polarization (RHCP or LHCP) transmit	Frequency (MHz) and polarization (RHCP or LHCP) receive	Specified G/T (dB/K)	Specified EIRP (dBW)
II	Inmarsat-2	C	3600.0–3623.0 (LHCP)	6425.0–6443.0 (RHCP)	−14	24.0
		L	1530.0–1548.0 (RHCP)	1626.5–1649.5 (RHCP)	−12.5	38.75
III	Inmarsat-3 (Note 1)	C	3600.0–3629.0 (LHCP and RHCP)	6425.0–6454.0 (RHCP and LHCP)	−13	27
		L	1525.0–1559.0 (RHCP)	1626.5–1660.5 (RHCP)	−11.5 (global) −5.5 (spot)	39 (global) 47.4 (spot total) (exchangeable)

Note: C × C and L × L transponder specifications are not included.
Compiled by Author (Courtesy, Inmarsat.)

Table 11.5 Inmarsat satellites: general parameters

Parameter	II Generation	III Generation
Owners	Inmarsat/North Sea Marine Leasing Company	Inmarsat
First launch	November 1990	December 1995
Numbers deployed (mid 2000)	4	5
Design lifetime (yr)	10	13
Launch weight (kg)	1230–1320	1900
Solar power (kW)	1.4	~2
Type of stabilization	Three-axis	Three-axis
Type of coverage	Global	Global and spot
Number of spot beams	None	Up to seven

Compiled by Author (Courtesy, Inmarsat.)

use of satellite resources throughout their lifetime. The satellites also include a navigation payload to provide an international complement to the global position system and a C-to-C transponder to carry point-to-point traffic. Figure 11.1 (Richharia, 1994) illustrates the EIRP and bandwidth range of each Inmarsat satellite generation used in the legacy network. Inmarsat third generation satellites include two sets of identical C-band frequencies in opposite circular polarization to achieve the desired isolation (X = right-hand circular polarization (RHCP) up, left-hand circular polarization (LHCP) down; Y = LHCP up, RHCP down). However, L band frequencies for each C band polarization are RHCP and can therefore only be reused between spot beams that exhibit sufficient isolation. The spectrum is notionally divided into L (land/maritime), M (maritime) and A (aeronautical) bands, for historic reasons (the bands are now generic). The G band corresponds to the global beam frequency band. Each band is further segmented into sub-bands, the smallest band being 450 kHz. Each sub-band can independently be switched to any beam, except the G1 band, which is fixed. The gain of each sub-band can be controlled independently. There are three 'secondary' transponders. The C × C transponder provides a direct C band to C band link using Y polarization. It can be used for inter-station communication. The L × L transponder provides direct L band to L band links and was provided to link mobiles directly, but was not used because of operational considerations. The navigation transponder provides navigation signals in both L and C bands, as described earlier.

The monitoring and control of all spacecraft are performed at the Satellite Control Centre (SCC) located at Inmarsat headquarters in London. The telemetry and tele-command stations, located at Fucino (Italy), Beijing (China) and Southbury/Santa Paula in the USA are linked with the SCC to provide real-time data for monitoring and control.

Coverage

The Earth has been segmented into four regions designated as Atlantic Ocean Regions – East and West (AOR-E and AOR-W), Indian Ocean Region (IOR) and Pacific Ocean Region (POR). Each region is served by a single satellite or a satellite cluster. Initially, the AOR

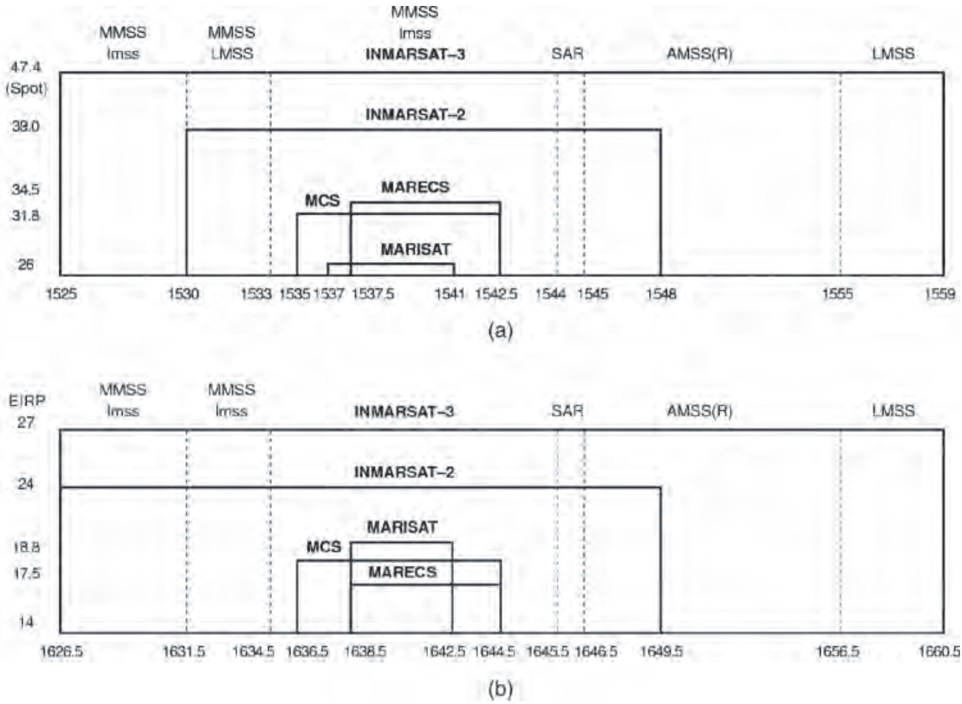


Figure 11.1 Bandwidth and EIRP of each satellite generation up to III generation (Richharia, 1994). (a) L band transmission channelization. The Y-axis shows L band EIRP of satellites. (b) L band receive channelization. The Y-axis shows C band EIRP of satellites.

consisted of only one part – today's AOR-E, which left a small coverage gap in the Pacific Ocean. The Atlantic West region was added to fill the coverage gap. Associated advantages included addition of space segment capacity to share the heavy traffic in this part of the world, and coverage redundancy.

The service area in each ocean region is limited to areas within which a satellite appears above about 5° elevation from mobiles, corresponding to about ±76° latitude. Interestingly, aeronautical services via omni-directional antennas are known to provide coverage up to almost 86° latitude because of the altitude of aircraft. Successful communications below 5° elevation from mobiles of other services are reported regularly. Figure 11.2(a) (Courtesy, Inmarsat) shows 5° elevation coverage from Inmarsat's operational satellite and Figure 11.2(b) (Courtesy, Inmarsat) shows spot beam coverage of each Inmarsat-3 satellite.

Demand Assigned Operation

Depending on the requirement, Inmarsat uses various types of demand assigned (DA) satellite accessing schemes for communications. These schemes provide efficient utilization of satellite power and bandwidth. For example, Fleet 77 uses a centrally controlled DA scheme. A frequency (or time slot) pool is maintained at an NCS (network control station) specific to each service and assigned to users on demand. To minimize network failure caused by NCS outage, each NCS is backed up by a stand-by NCS. Inmarsat-Aeronautical service uses a

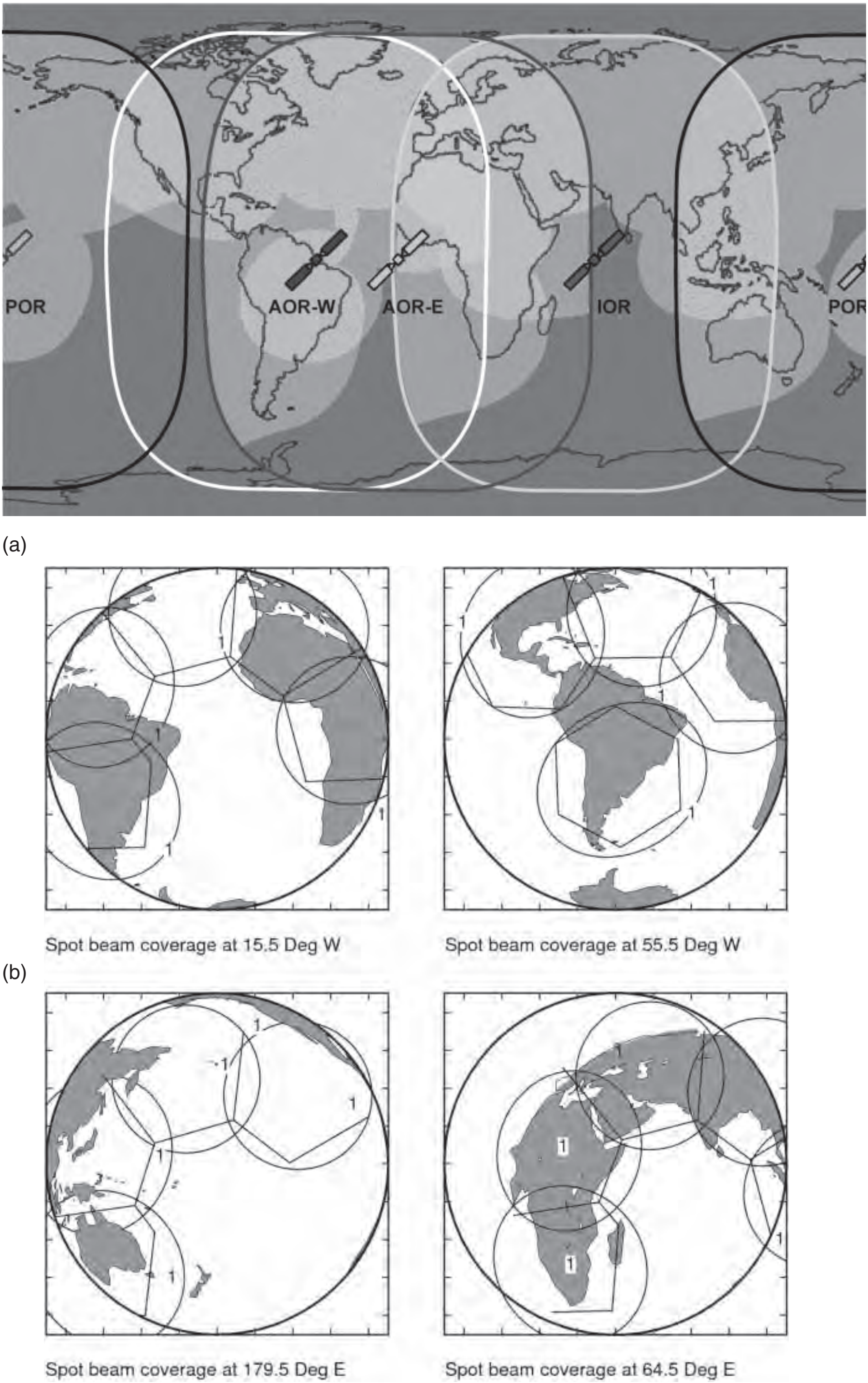


Figure 11.2 (a) Five degree elevation coverage of Inmarsat satellites. (b) Spot beam coverage of Inmarsat-3 satellites. (Inmarsat. Reproduced with permission of Inmarsat.)

distributed resource management architecture, where each GES manages its individual frequency pool independent of other GES. This type of distributed approach is more resilient to a single point (i.e. NCS) failure, but less efficient in terms of spectrum utilization.

In addition to the DA traffic channels, a number of fixed assigned channels are used for signalling between the NCS, LES and MES components of each service. Such channels are essential for network operation.

In a single channel per carrier system, it is necessary to transmit a reference pilot signal via each satellite to enable each LES to correct for frequency uncertainties caused by satellite, LESs local oscillator and the Doppler effect due to satellite motion. Uncorrected transmissions/reception from/at LESs can cause interference to/from adjacent channels because carriers in the network are closely spaced, the occupied bandwidth ranging from 2.5–100 kHz. Because of the location dependence of the Doppler Effect, pilots are transmitted for the Northern Hemisphere, equatorial region and Southern Hemisphere to provide the most accurate reference for each region. To improve the reliability of the network, there are provisions for stand-by AFC (automatic frequency control) transmissions, should AFC transmissions from the designated LES fail.

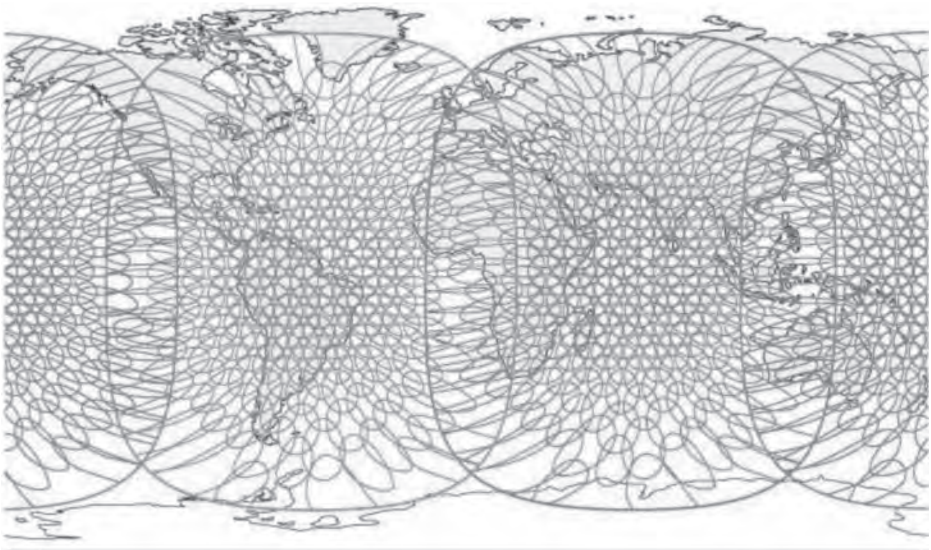
There are numerous other types of carriers for specific applications, such as special services, order wire, tests, and others.

Satellite Transmission Monitoring – Legacy Services

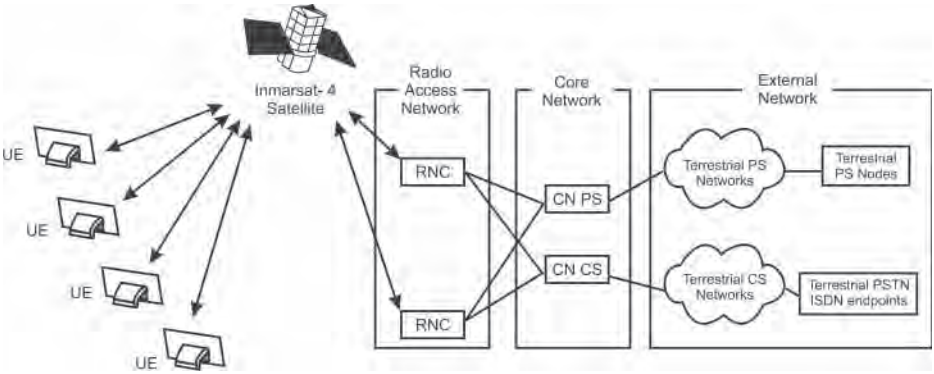
A computer-controlled spectrum analyser monitoring system known as the Satellite Spectrum Monitoring System (SSMS) is used for maintaining network RF integrity. The SSMS consists of a number of remote Earth stations linked to the NOC via land or satellite links. To monitor traffic flow in real time and maintain a record of traffic flow through the network, data contained in the signalling channels of each service are extracted and processed by another monitoring system. In addition to assistance in real-time fault detection and diagnosis, the processed data from the off-air monitoring system are used for billing and to provide useful statistics such as call holding time, satellite usage each day, type of calls, grade of service, and others. The statistics have a number of operational uses, such as congestion management, monitoring usage of satellite EIRP, planning of space segment resources, traffic forecast, and so on.

11.2.1.3 Mainstream Services

The Inmarsat Broadband Global Area Network (BGAN) system is the transport platform used by Inmarsat for its existing mainstream services spanning enterprise, maritime, aviation and government sectors. The BGAN system permits interoperability with a terrestrial 3G CN by re-use of 3G Non Access Stratum (NAS) layers and is recognized by the ITU (International Telecommunications Union) as an IMT-2000 (International Mobile Telecommunication-2000) compatible standard constituting a satellite extension of the terrestrial 3G network. Since the terrestrial UMTS Terrestrial Radio Access Network (UTRAN) air interface is unsuitable for satellite channel, the BGAN system design is based on an air interface, which provides a complete Access Stratum Protocol Stack optimized for the geostationary satellite environment. BGAN services were lunched on three classes of land portable user terminals, later extended to include mobile terminals operating in air, on land and sea (Richharia and Trachtman, 2005).



(a)



(b)

Figure 11.3 (a) Representative coverage of Inmarsat-4 satellites and (b) BGAN architecture. (Both parts source: Inmarsat. Reproduced with permission of Inmarsat.)

The system architecture is presented in Figure 11.3(b). The user equipment (UE) communicates with the Radio Network Controller (RNC) nodes via Inmarsat-4 satellites (I4). The satellite provides transparent amplification and frequency translation of the received signals. Representative payload characteristics of I4 are listed in Table 11.6 and coverage is illustrated in Figure 11.3(a).

Transmission between the RNC and satellite is via C band feeder links operating in dual circular polarization. The feeder link beam provides a global coverage. The L band service link transmissions are in RHCP. The L band frequency is re-used by dividing the coverage by up to 228 narrow spot beams and 19 wide spot. The global beam overlay assists in tasks such as signalling and provision of GMDSS compliant services.

Table 11.6 Inmarsat-4 communication payload characteristics

Number of transponders	4
C-L transponder frequency range (MHz)	6424.0–6575.0/1525.0–1559.0
L-C transponder frequency range (MHz)	1626.5.0–1660.5/3550.0–3700.0
L × L	Capability to interconnect any two user terminals
C × C	Capability to interconnect any two satellite access stations to provide 5 MHz
User beams	Narrow, wide and global
Polarization	C band: dual polarized; L band: RHCP
Frequency reuse	Feeder link: 2 (dual polarization) Narrow spot beams: 20 times (uniform traffic distribution) Wide spot beam: 4 times (uniform traffic distribution)
L band G/T	Narrow spot beams (minimum): > 10 dB/K Wide spot beams (minimum): 0 dB/K Global beam: −10 dB/K
L band EIRP	Aggregate uniformly loaded narrow spot beam edge-of-coverage EIRP: 67 dBW with 28 dBW edge-of-coverage EIRP in global beam; Aggregate uniformly loaded wide spot beam edge-of-coverage EIRP: 56 dB W with 39 dBW edge-of-coverage EIRP in global beam Global beam: 41 dBW
L band transponder channel width	200 kHz; tuneable to any segment of transponder and switchable to any spot beam
Number of transmission channels	630 at beginning of life
Channel to beam switching time	< 2 s
C band EIRP (dB W)	31
C band G/T (dB/K)	−10.5

Compiled by Author (Courtesy, Inmarsat.)

The BGAN radio access network (RAN) consists of a number of radio network subsystems (RNSs). Each RNS consists of a BGAN RNC and associated BGAN RF Subsystem. The RAN handles all radio related aspects of the BGAN ground system network infrastructure. The RNC interfaces to the CN for switching and routing calls and data connections to and from the external networks.

The BGAN CN is a suite of UMTS (Universal mobile telecommunication system) network nodes. The BGAN CN is aligned to a 3GPP release 4 architecture having separate PS and circuit switched (CS) domains. Media Gateway/MSC (mobile switching centre) Server

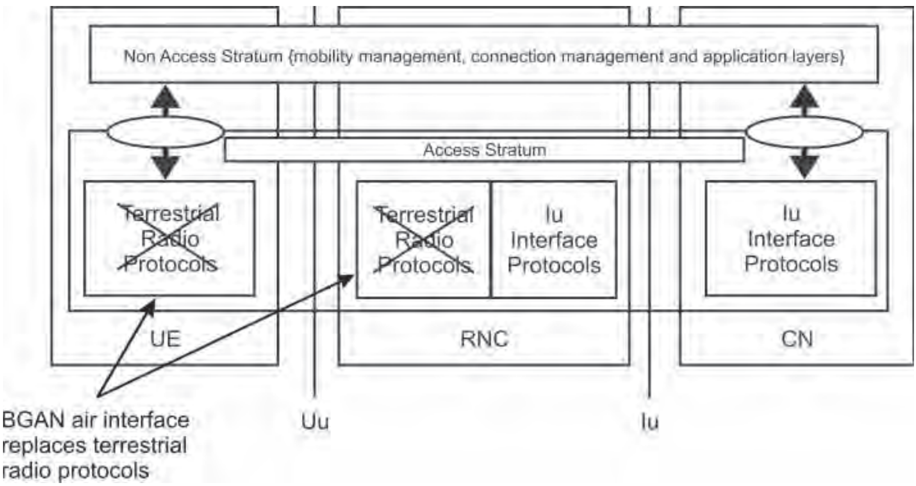


Figure 11.4 BGAN protocol architecture. (Source: Richharia and Trachtman, 2005. Reproduced with permission of Inmarsat.)

nodes (for user and control plane transmission respectively) are provided for circuit switched communication (PSTN (public switched telephone network) telephony and ISDN). Serving GPRS Support Node (SGSN) and Gateway GPRS Support Node (GGSN) provide IP PS communications.

Figure 11.4 shows the BGAN protocol architecture. Since UMTS allows services transport over alternative radio access technologies, the terrestrial radio access technology is replaced by Inmarsat's air interface retaining the NAS of UMTS. The 'Iu' interface between the CN and RNC as well as CN architecture remains unchanged thereby allowing an industry standard UMTS CN to be used.

The system provides a suite of packet and circuit switched communication services to native BGAN users and to roaming customers who possess a roaming agreement between their home network and Inmarsat.

BGAN supports:

- Telephony and ISDN calls;
- An 'always-on' Internet/Intranet connection with fixed or dynamically assigned IP addresses;
- Short Messaging Service;
- Advanced messaging services such as 'multimedia messaging services (MMSs)' as applications supported over the PS GPRS (general packet radio service)/IP service;
- UMTS location based services such as local travel information.

There are two classes of IP services: standard and streaming. The standard service provides back-ground service on shared IP bearer at rates up to 492 kbps for applications such as web access, file transfer, email, and so on. The streaming class, available on-demand, provides guaranteed send/receive data rates of 32, 64, 128, 256 kbps for streaming applications.

The air interface protocols take account of satellite link attributes like high delay and variable error rate. The protocols include adaptation, bearer connection, bearer control and physical layers.

Each layer communicates with its peer, the layer above and the layer below using a set of protocol unit definitions. The BGAN protocol interfaces at the edge with the upper layers of UEs at one end and the CN at the other.

The Adaptation Layer is responsible for:

- Registration Management (e.g. spot beam selection, system information handling);
- GPRS Mobility Management (GMM) handling;
- mobility management handling within the satellite network;
- radio bearer control (e.g. handling signalling related to setup, modification and release of radio bearers).

The Bearer Connection Layer is responsible for:

- buffering and flow control of information to the layer above;
- QoS (Quality of Service) policing (shaping of information streams where required);
- segmentation and reassembly of information;
- Automatic Repeat Request (ARQ) (when required); and ciphering.

The Bearer Control Layer is responsible for:

- low-level aspects of UE behaviour;
- ensuring fairness between connections and managing the access to the physical layer;
- provision of transport for the BGAN system information from the RNC to the UEs;
- ensuring availability of an appropriate level of resource within the RNC.

The physical layer consists of a TDM (time division multiplex) transmission in the Forward direction and frequency division multiplexed TDMA (time division multiple accessing) in the return direction. To achieve an efficient bearer packing density, sharing of the bearers is done at a packet level by assigning the origination and destination addresses to each packet.

The air interface supports multiple bearer types with mechanisms to support different resource management and signalling schemes on each bearer. The BGAN traffic typically passes through a single Satellite Access Station (SAS), where the radio resource is managed by the resident RNC. The RNC selects a bearer such that the transmission efficiency is maximized for the prevalent channel conditions. The bearers collectively provide data rates in the range of 4.5–512 kbps.

Owing to the wide range of UE classes, it is not possible to provide a single solution that optimizes the transmission rate for the entire UE portfolio. Instead, the system utilizes a judicious combination of modulation and code rates depending on the transmission needs and direction, channel conditions and the UE class. Modulation schemes include QPSK (quadrature phase shift keying), 16-QAM (quadrature amplitude modulation) and $\pi/4$-QPSK at symbol rates of 8.4–151.2 kbps with up to 15 turbo-code steps. Bearer bandwidths range between 21 and 189 kHz.

The forward transmit channel units can apply different code rates on an FEC (Forward error correction) block basis as directed by the Bearer Control Layer in order to adapt transmissions to different UT types on the same forward bearer. The return bearer comprises burst transmissions in slots of 5 or 20 ms duration, which are assigned by the RNC in a schedule that is transmitted on a forward bearer. These return schedules also define the symbol rate and modulation in each slot. The system allows code rate adaptation to contend the prevailing channel condition. The desired code rate is achieved by puncturing the turbo-code generated parity streams using one of a number of pre-defined puncturing matrices. The matrices are designed such that coding gain is variable in 1 dB step. This approach requires signalling provision to control the coding rate.

Each UT includes a transceiver, in-built GPS receiver unit, UMTS aligned Universal Integrated Circuit Card (UICC)/Universal Subscriber Identity module (USIM), connections/adaptors for external terminal equipment (TE) attachment, battery management (for both mains and battery operation) and a man machine interface (MMI) (via TE).

The baseline system developed for fixed-portable point-to-point communication was later enhanced to include mobile user terminals in air, land and sea (Rivera, Trachtman and Richharia, 2005; Richharia, Kaluvala, *et al.*, 2005). The propagation model used for the development of the baseline system was unsuited for mobile systems and hence propagation models applicable for the mobile environment and relatively large BGAN bandwidth of 200 kHz were developed from established databases and theory; it was observed that while maritime and land propagation channel were coherent, the aeronautical channel exhibited dispersion when traversing routes over calm sea, necessitating the use of equalizers at the aeronautical receivers to provide robustness under all flying conditions (Richharia, Kaluvala *et al.*, 2005, Gambaruto, Richharia and Trachtman, 2008). The air interface was modified such that it would be backward compatible with minimal changes to the existing system to ensure commercial viability. Changes applied included enhanced features such as elevation angle dependent propagation margin adjustments, a Doppler compensation scheme for aircrafts, robust synchronization to promote rapid recovery following a fade and handover support for aircrafts, definition of new user terminal classes, development and management of new bearer types, amongst others. Table 11.7 illustrates the representative features of the BGAN user terminal classes emerging from this exercise (adapted from Rivera, Trachtman and Richharia, 2005). The schedule for integration of these classes in to the existing system would depend on their respective commercial viability that was investigated separately. Some of these classes (e.g. omni-directional variants) were not commercially viable at that time and hence were dropped from the next release.

It was recognized that a cross-section of users in addition to the core point-to-point BGAN service aspire for a multicast service as this would be more efficient than a point-to-point service when the same information has to be transferred to several subscribers. In a multicast service, the information is transmitted on a bearer to which the multicast group is tuned. Such a service could not be supported by the BGAN CN, as such. 3GPP specifies a multimedia broadcast and multicast service (MBMS) for multicast applications in a terrestrial environment (TS23.246 3GPP-MBMS); however, it became clear that the specifications as such could not be applied to BGAN in order to meet Inmarsat's requirements. Therefore, Inmarsat embarked on a project to define a multicast network within the BGAN system framework such that BGAN user terminals could access the service efficiently with minimal change to protocols, RAN and UT software without changing the existing CN while retaining the

Table 11.7 Representative features of (a) directional and (b) the proposed omni-directional classes of BGAN user terminal

Class	Sub-class	Minimum G/T at 5° elevation (dB/K)	EIRP (dBW)
(a)			
Land portable	A3	−10.5	20
	A4	−13.5	15
	A5	−18.5	10
Aeronautical	High gain	−13	20
	Intermediate gain	−19	15
Maritime	High gain	−7	22
	Low gain	−15.5	14
Land	High gain	−10	20
	Low gain	−15	15
(b)			
Aeronautical	Omni-directional	−24.5	11
Maritime	Omni-directional	−24.5	7

(Adapted from Rivera *et al.*, 2005.)

IETF (Internet engineering task force) defined MBMS interfaces at the edges (Febvre *et al.*, 2007). Figure 11.5 shows a configuration of the proposed BGAN multicast network.

The multicast service provider transmits contents to subscribers via the broadcast/multicast (BM) domain of the CN consisting of Broadcast and Multicast Service Centre (BMSC) and Broadcast and Multicast Service Node (BMSN). A UT attaches to the multicast domain only after attaching itself to the PS domain. The service provider sends the multicast content to the satellite operator who transmits the content to the desired spot beams after UT group attachment. The subscriber interacts with the multicast provider on a standard BGAN point to point link.

Inmarsat also added low-cost circuit-switched voice communications on hand-held and fixed UTs to its portfolio through a collaborative agreement with ACeS on the latter's operational system in the Asian region. The service was extended globally through enhanced handsets and Inmarsat-4 satellites.

Further enhancement to the BGAN system capacity is expected through a more advanced version of Inmarsat-4 called Alphasat I-XL based on Alphabus platform developed by Astrium and Thales Alenia Space in Europe (Astrium, 2012), launched in 2013. The satellite utilizes a more advanced digital signal processor used in payload, a 11 m diameter aperture in addition to ESA's (European Space Agency) technology demonstration payloads. The satellite adds 7 MHz of the WRC-03 allocated lower part of the MSS band, an additional 120 channels, four-colour frequency reuse rather than seven, higher G/T and EIRP. The satellite will be integrated with the I4 constellation to enhance the BGAN coverage and

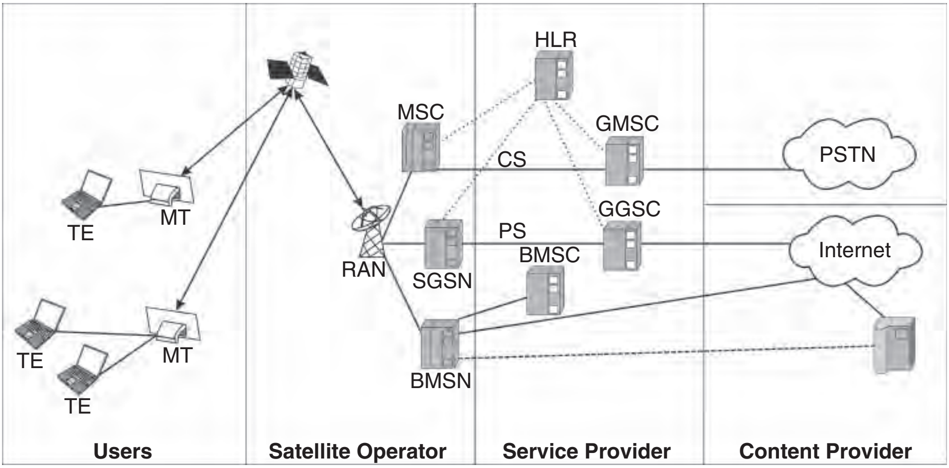


Figure 11.5 The proposed BGAN multicast network configuration. (Source: Richharia and Trachtman, 2005. Reproduced with permission of Inmarsat.)

Table 11.8 Alphasat I-XL payload parameters

Frequency band (MHz): C-L transponder	6425.0–6575.0/ 1518.0–1559.0
Frequency band (MHz): L-C transponder	1626.5–1660.5 and 1668.0–1675.0 MHz/ 3550.0–3700.0
EIRP (dB W): global/wide/narrow	48/60/70
G/T (dB/K): global/wide/narrow	−8, 1, 13
Feeder Link EIRP (dB W)	35
G/T (dB/K)	−10

(Courtesy, Inmarsat.)

capacity; enable development of smaller hand-held terminals; allow a tighter reuse pattern. A broadcast service based on OFDM (orthogonal frequency division multiplexing), 16QAM and turbo-coding or a standard such as S-DMB (satellite digital multimedia broadcasting) may be included. Table 11.8 lists a few salient payload characteristics (Trachtman, 2006).

The satellite also includes four hosted ‘piggy-backed’ payloads (ESA, 2012).

11.2.1.4 Inmarsat Next Generation

A constellation of three Inmarsat-5 geostationary satellites, expected to be launched during 2013 and 2014, will provide global mobile broadband communication in K_a band for

Table 11.9 Main characteristics of Inmarsat-5 satellites

Manufacturer	Boeing
Spacecraft platform	702 HP
Launch vehicle	Proton
Life (yr)	15; Launch: 2013 into 2014
DC power (kW): beginning-of-life power/end-of-life	15/13.8
Solar panels	Two wings each with five ultra triple-Junction GaAs solar cells, length = 33.8 m
Propulsion system	Liquid apogee engine 445 N
Launch	Xenon-ion propulsion system
In-orbit	4×22 N (axial)
Control thrusters	4×10 N (radial)
Antenna diameter (m)	8.08
Number of service beams	89 (6 steerable high capacity beams)
Transponder type	Transparent
Frequency band	
Feeder link	C
Service link	L

(Data source: Boeing, 2012.)

deep-sea vessels, in-flight connectivity, streaming high resolution video, voice and data services throughout the world. The mobile VSAT (very small aperture technology) system will provide data rates of up to 50 Mbps in the down-link and 5 Mbps in the up-link with a potential to integrate seamlessly with Inmarsat's L band services to provide a hybrid solution with GMDSS capability. The system known as Global Express is expected to provide global coverage by late 2014 (Inmarsat-Global Express, 2012). Table 11.9 provides the salient features of Inmarsat-5 satellites and Figure 11.6 illustrates an artist's impression of the satellite (Boeing, 2012). Figure 11.7 illustrates a hypothetical coverage map from the satellite.

11.2.2 Asia Cellular Satellite (ACeS) System

The ACeS System is a regional system, providing digital voice, facsimile, data and paging services to hand-held, mobile and fixed terminals in regions encompassing China and Japan to the north, Indonesia to the south, the Philippines to the east and India and Pakistan to the west (Nguyen, Buhion and Adiwoso, 1997) as depicted in Figure 11.8(a). In 2006, the company signed a collaborative agreement with Inmarsat to provide these services under the Inmarsat brand. The network is designed on a GMR-2 (Geo Mobile Radio) interface, which is based on the GSM (Global System for Mobile) CN to ensure integration with the terrestrial standard and to maximize reuse of proven terrestrial technology (see Chapter 8). Integration with the cellular networks allows seamless roaming with terrestrial GSM systems. Figure 11.8(b) shows ACeS system architecture, which comprises satellites (called Garuda - Figure 11.9) controlled by the satellite control facility (SCF) located in

Figure 11.6 An artist's impression of the Inmarsat-5 satellite. (Source: Boeing. Reproduced by permission of Boeing.)

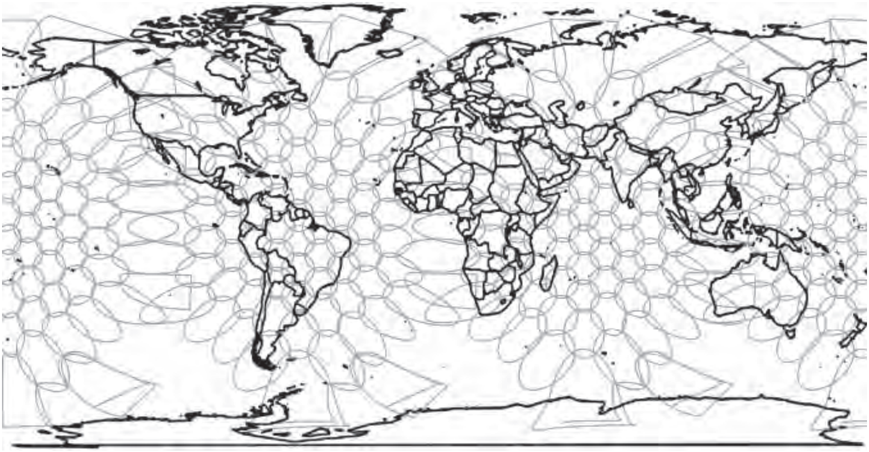


Figure 11.7 A hypothetical footprint of the Inmarsat-5 satellite

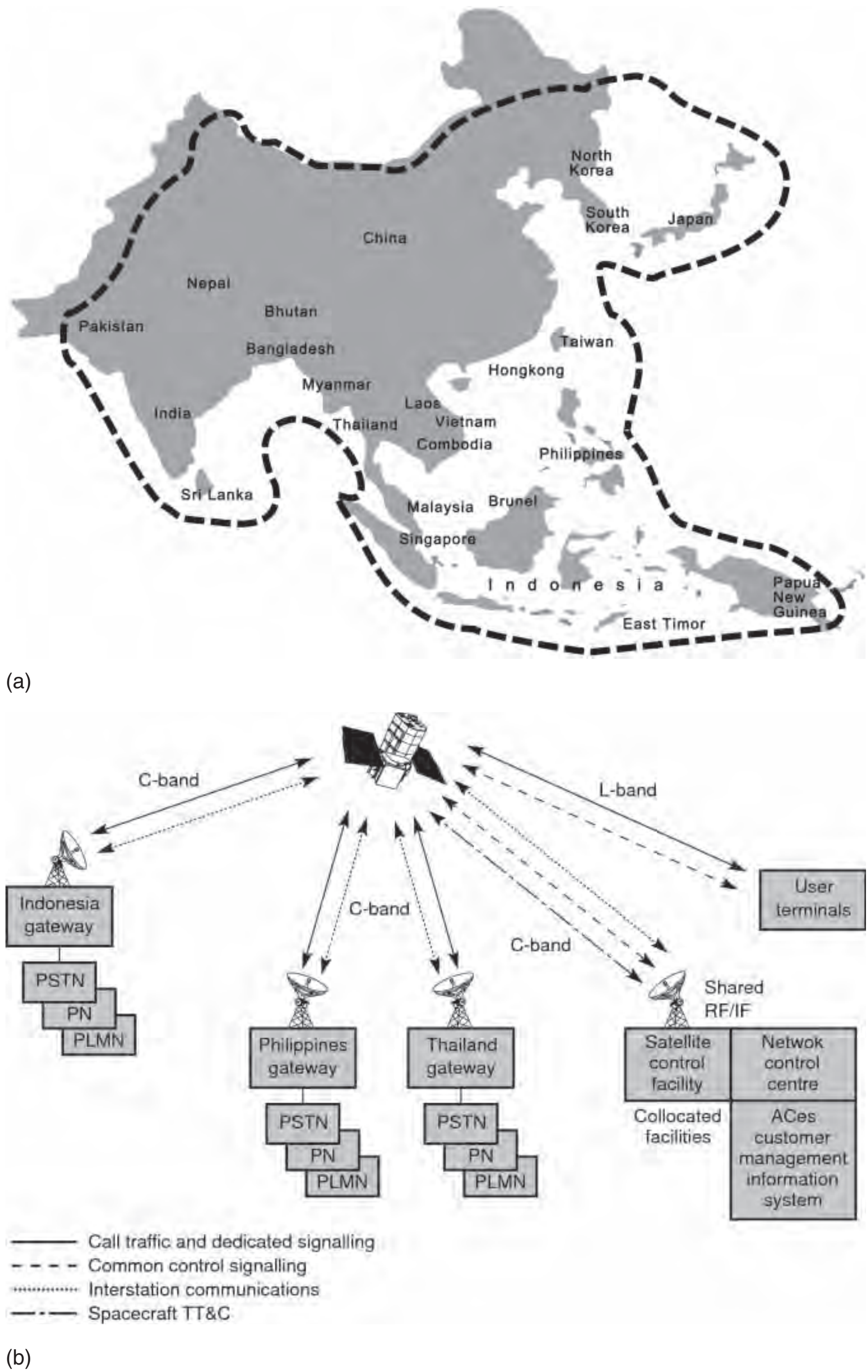


Figure 11.8 (a) ACeS coverage diagram and (b) ACeS system architecture. (Both parts source: Nguyen *et al*., 1997, Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, California, June 16-18, 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97-11, Jet Propulsion Laboratory, Pasadena, California, June 16, 1997.)

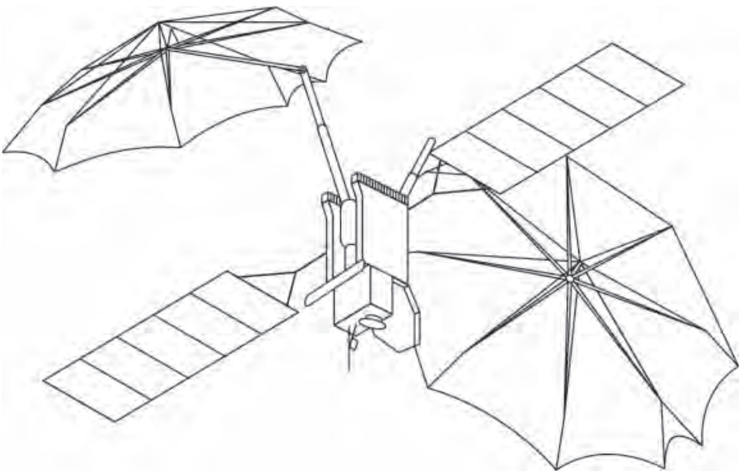


Figure 11.9 Garuda-1 satellite view. (Source: Nguyen *et al.*, 1997, Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, California, June 16–18, 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97–11, Jet Propulsion Laboratory, Pasadena, California, June 16, 1997.)

Batam Island, Indonesia, the network control centre (NCC), gateways, customer management information system (GCMIS) and the user population using a variety of user terminals.

The salient features of the satellite are a dynamic digital channelizer, a beam-forming network, and multi-port power amplifiers with the capability to switch power and bandwidth to spot beams according to resource needs. The digital channelizer routes frequency sub-bands between C and L band beams. The routing table is fully configurable through ground commands from the NCC via the SCF. Figure 11.9 shows an artist's impression of a Garuda satellite. The satellite operates in the extended C-band frequency range of 6425–6725/3400–3700 MHz and 1626.5–1660.5/1525–1559 MHz, deploying 140 spot beams in the service link formed by a separated transmit/receive 12 m deployable reflector antenna system – separated to avoid the passive inter-modulation problem – and based on flight-proven deployable mesh reflector technology. Using a seven-cell pattern, a frequency reuse factor of 20 is achievable. With a G/T of 15 dB/K and an L-band EIRP of 73 dB W, the satellite can provide up to 11 000 simultaneous voice links to hand-held terminals with a link margin of 10 dB, and up to 28 000 voice links at a lower margin.

The ground segment comprises the NCC, co-located with SCF, which controls and manages the ACeS system; the customer management information system that manages customer billing, and other customer-related activities; gateways located at a number of places interfacing with the PSTN, public land mobile network (PLMN) and private network (PN), and including a GCMIS. The main functions of the NCC are monitoring/configuring the channelizer routing table, control of the frequency reuse scheme, network resource allocation to gateways, monitoring the network load, system broadcast, call set-up and performance monitoring. The ACeS customer management information system gathers resource usage data, and performs customer and gateway accounting functions.

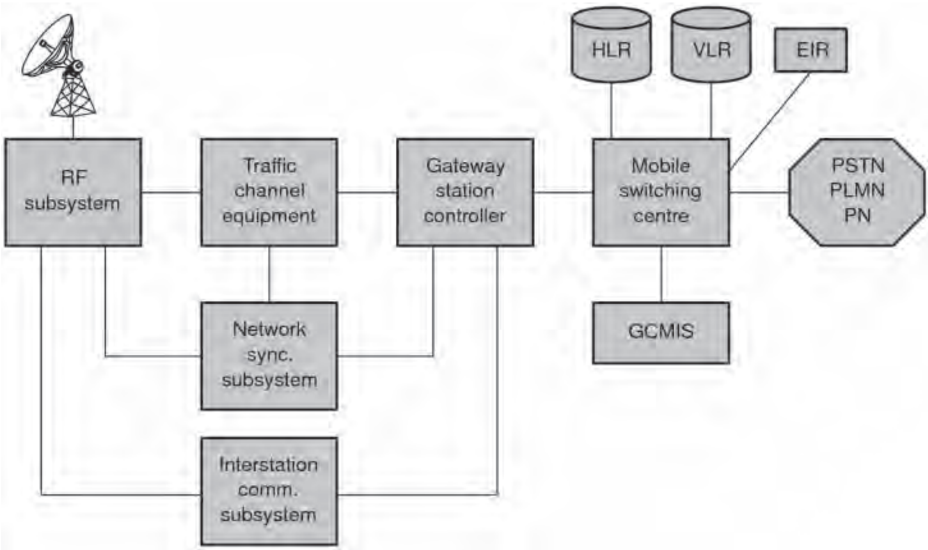


Figure 11.10 ACeS gateway functional diagram. (Source: Nguyen et al., 1997, Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, California, June 16–18, 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97–11, Jet Propulsion Laboratory, Pasadena, California, June 16, 1997.)

Figure 11.10 represents the functional block schematic of a gateway. In addition to connecting the ACeS network to the terrestrial network, the gateway manages frequency sub-bands and TDMA time slots allocated to it by the NCC; performs call set-up in conjunction with the NCC; controls frequency and timing offset, mobility management and support functions such as user authentication. The RF subsystem is used for radio link connectivity with the satellite and users through a 13 m tracking antenna. The traffic channel equipment provides modulation, demodulation, baseband processing and RF power control. The gateway station controller performs call set-up and release in conjunction with the NCC. The MSC interfaces the ACeS network to the terrestrial network, handling the SS7 (Signalling System 7) signalling of the fixed network. The home location register (HLR), visitor location register (VLR) and equipment identity register (EIR) are used for mobility management and authentication respectively. The network synchronization subsystem (NSS) is slaved to the master NSS resident in the NCC, which comprises ranging equipment, a timing pulse generator and GPS equipment. The inter-station subsystem consists of equipment for communications between the NCC and the gateways through the C × C link. Finally, the GCMIS is a customer management information system for maintaining the subscriber and terminal database and subscribers' accounts.

The system primarily supports three types of user terminals. Cellular-sized hand-held terminals are satellite/terrestrial dual mode with automatic or manual switching capability, while mobile and fixed terminals use larger, directive antennas. The system also supports GSM features such as call transfer, call waiting, call holding, conference calls, etc. A high-penetration alerting facility informs the user that a call is not able to connect due to blockage, so that the user may move to a more favourable location. Dual-mode terminals can roam the terrestrial network on the same number. Optional services include voice mail, store and forward facsimile, and paging via high power transmissions to improve service availability.

The radio link, operating in the C band over the feeder link and the L band for the service link, uses a propriety TDMA/FDMA (frequency division multiple accessing) accessing scheme, based on GSM signalling. Mobile-to-mobile connectivity can be achieved by L × L cross-strapping in the satellite's digital channelizer.

11.2.3 EUTELTRACS

EUTELTRACS is a European commercial MSS operated by the European Telecommunications Satellite Organization (EUTELSAT) in collaboration with Qualcomm, to provide an integrated message exchange and position reporting service throughout Europe as shown on Figure 11.11 (Colcy and Steinhäuser, 1993). In 2012, the service was available through EUTELSAT 36B satellite that serves Europe, North Africa the Middle East and central Asian republics. Features of the service include (EUTELSAT, 2012):

- Real-time data communication.
- Automatic vehicle tracking: locates a vehicle or vessel to an accuracy of 100 m (GPS or EUTELSAT proprietary positioning).
- Seamless, pan European coverage – using EUTELSAT 36B.
- Roaming-free communication across Europe.
- Data collection and transmission from vehicle or vessel.
- Transmission of alarm and distress messages.
- Value-added optional features: for example, mobile terminal access facility to external databases for weather or traffic conditions, etc.
- High value cargo security and protection.

The commercial arrangement consists of national or regional service providers; equipment of each provider interfaces with the hub station located just outside Paris that is

Figure 11.11 A generic footprint of EUTELSAT satellites. (Source: Colcy and Steinhäuser, 1993. The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

operated by EUTELSAT who have the overall responsibility for the service. The responsibility includes provision of space segment capacity. All aspects related to system hardware, software, mobile terminals and marketing are managed by Alcatel Qualcomm. Regional service providers operate their individual service network management centres and deal with the end users, invoicing them for transactions, equipment, software, training, and so on. The system offers a fully integrated message exchange and position reporting system; it has a low-entry-cost architecture capable of exploiting capacity on non-dedicated satellites using well-proven technology; it is a secure system with a single point of contact for all services and billing. Applications include mobile data broadcasting, supervisory control and data acquisition (SCADA) and monitoring fishing activities, as well as low bit rate regional aeronautical communications The system architecture is based on the OmniTracs system, which provides two-way satellite communications and geolocation trailer tracking technology for the over-the-road transport market being operated by Qualcomm in the USA since 1988 (Jacobs, 1989). The main components of EUTELSAT system, shown in Figure 11.12, comprise:

- Customer's terminal and dispatch centre.
- Service provider's network management centre (SNMC) used for sending/receiving messages to/from customers and to access position information for the customer's fleet.
- Hub station that utilizes two antennas to operate with two satellites, associated RF equipment and a hub terminal facility whose function is to process, control and monitor traffic flow and control satellite access.
- EUTELSAT satellites, comprising a 'data' satellite, for the main communications and a 'ranging' satellite to assist localization functions. A ranging signal is transmitted by the secondary hub antenna solely for localization and timing information.
- A mobile communication terminal, mounted on a vehicle, provides communication and reports to the hub, the signal arrival time-difference between data and ranging satellites sent by the user used at the hub to derive the terminal's position for the position reporting service.

The forward link comprises a high-power TDM stream spread with a chirp signal to minimize the effects of multipath and interference, resulting in a 2 MHz bandwidth signal. The data rate is 4.96 kbps, BPSK (binary phase shift keying) rate half-Golay coded (called, 1X data rate) or 14.88 kbps, QPSK rate three-quarter-coded (called, 3X data rate). The result is always 9.92 ksps PSK (phase shift keying) in 9.92 kHz bandwidth. The data rate for each terminal is determined dynamically depending on the terminal's reception environment.

The return link uses the full 36 MHz of the transponder to support a maximum of 45 000 users transmitting at low levels. The low bit rate link uses a one-third rate convolution code of constraint length 9 with Viterbi decoding in conjunction with a powerful interleaving scheme to mitigate likely interference in its operational band. A 32-ary FSK (frequency shift keying) scheme is combined with direct spreading sequence at 1 MHz rate in an MSK (minimum shift keying) modulator. The spread signal randomly frequency-hops over the full return band to reduce the effects of interference. The hopped sequence is synchronized at the user terminal and the hub for data demodulation. During return link transmissions, the transmitter is enabled for 50% of the time at 15.12 ms, during which either one 32-ary FSK symbol at 55 bps or three symbols at 165 bps are transmitted. During the period

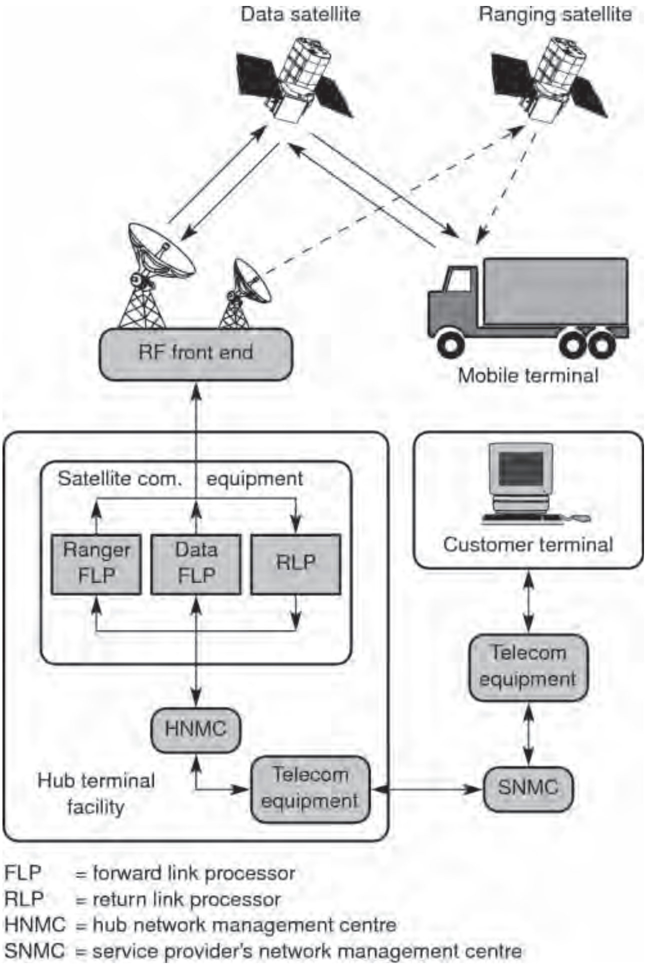


Figure 11.12 Main components of EutelTRACS. (Source: Colcy and Steinhäuser, 1993. The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

when transmission is disabled, the tracking system tracks the satellite using the down-link transmissions, while the receiver performs frequency and time tracking. The selection of the mobile data rate is dynamic, depending on its transmit environment, and is performed under the control of the hub network management computer. Before setting up a transmission, the mobile terminal must acquire the forward link broadcast and request for resource allocation. Calls are set up through the forward signalling channel using special system packets.

Data integrity is obtained by an acknowledged store and forward protocol. Each packet is acknowledged after an error-free reconstruction. If no acknowledgement is received from the mobile, the packet is retransmitted by the hub up to 12 times in an hour before declaring it unacknowledged. No new messages are transmitted to a mobile until the current message has been delivered successfully. In the return direction, the hub network management centre

Table 11.10 Delay distribution statistics in the EutelTRACS system

Number of tries	Delay time range	% of messages delivered successfully in each environment		
		Clear	Nominal	Marginal
1	20–30 s	99.2	85.6	73.3
2 or 3	2–3 min	0.6	9.9	14.9
Up to 8	12–20 min	0.2	4.5	8.1
Up to 12	<1 h	0	0	2.9
Never delivered	–	0	0	0.8

(Source: Colcy and Dutronc, 1990. Reproduced with permission of John Wiley & Sons, Ltd.)

(HNMC) acknowledges error-free receipt. The mobile retransmits a message up to 50 times before discarding the packet(s).

The performance of a store and forward system can be measured as delay distribution of the message delivery, which includes queuing, acknowledgement and other processing delays. EutelTRACS quantifies the performance in four ranges of number-of-attempts for successful delivery. Delay distributions were obtained for three types of mobile environment:

1. Clear line of sight; example: fixed sites.
2. Nominal land mobile; example: land mobile travelling in flat, hilly country, city suburbs.
3. Marginal; example: low elevation angle, mountainous area, large city.

Table 11.10 summarizes the distribution along with the results of a measurement campaign (Colcy and Dutronc, 1990; Colcy, Dutronc and Ames, 1990). Note that in the clear and nominal conditions, all the messages are delivered within 20 minutes. In the worst case considered, that is marginal, 97.1% of the messages are delivered within an hour and 99.2% of the messages are delivered successfully (0.8% messages remain undelivered).

The EutelTRACS system also provides position reporting. The technique involves the use of two separate geostationary-orbit satellites for timing measurement at the mobile. Measurements include a precise estimate of round trip delay and the time difference between the two waveforms transmitted by the hub. The terminal tracks the data and ranging satellites alternately to derive the arrival time difference between the satellites enabled by the primary and secondary hub transmitting an identical triangular spreading signal of long enough duration that allows alternate tracking by user terminal of both the satellites without ambiguity. A multi-lateration technique that provides consistent, reliable, economic and accurate results was chosen for obtaining the position estimates. Round trip delay is measured for all message packets as a part of normal demodulation. The secondary hub up-links a low power signal to the ranging satellite, which is structurally identical to the forward message, but without carrying the message packets. The mobile acquires the ranging signal transmitted by the ranging satellite, derives the timing with respect to the primary signal and returns to the primary satellite. The timing information is transmitted to the hub with the next return message or with acknowledgement packets.

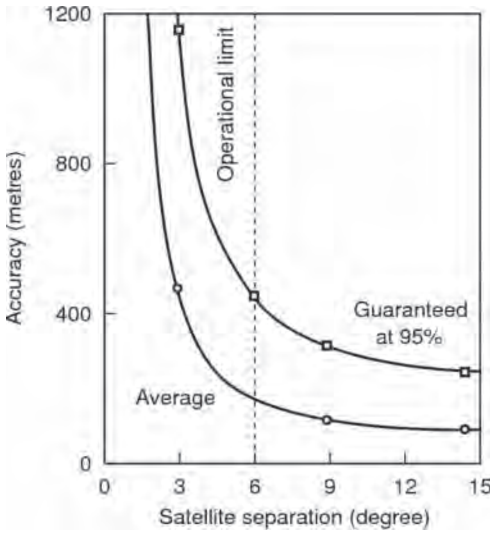


Figure 11.13 Random accuracy versus satellite separation, a maximum of 15° on the graph shows the current maximum separation possible from EUTELSAT satellites. (Source: Colcy and Steinhäuser, 1993. The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

The system acquires satellite position independently of the satellite control centre by multi-latering from a number of geographically spread fixed Earth stations. The mobile altitude necessary for accurate position estimation is obtained from the United States Geological Survey (USGS) world database at the hub terminal facility, given as the distance from the centre of the Earth based on World Geodetic System 84 (WGS 84) ellipsoid. The altitudes above this reference are taken from the USGS world database, which is regularly updated with satellite survey data. The grid has a precision of 10 arc minutes and height precision is 30.48 m. A linear interpolation is used within this grid, giving a worst-case error of half the peak-to-peak variations in rough terrain.

The error in the position reporting system has been analysed in terms of two components – a bias error, mainly caused by altitude model inaccuracy manifesting as north/south bias in the position solution, and a random error caused by errors in timing measurements, satellite position and influenced by satellite separation. The error is estimated as the root sum square of both components. Figure 11.13 shows the random accuracy versus satellite separation; the maximum separation possible from EUTELSAT satellites available at that time was 14.5° (Colcy and Steinhäuser, 1993). A separation of 6° was considered as the operational limit.

Table 11.11 lists bias errors, while Table 11.12 gives a summary of random accuracy versus satellite angular separation, measured in a trial. The position error is given as the root sum square of the two components. As an example, the average error in a hilly terrain can be estimate as 90 m for 14.5° separation.

Table 11.11 Accuracy due to bias errors versus environment

	Flat terrain	Hilly terrain	Mountainous
Average accuracy (m)	10	40	200
Guaranteed accuracy (m) (95% cases)	20	120	420

(Source: Colcy and Steinehäuser, 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

Table 11.12 Accuracy due to random errors versus satellite separation

Satellite separation (°)	3	6	9	14.5
Average accuracy (m)	460	170	120	80
Guaranteed accuracy (m) (95% cases)	1 150	440	320	240

(Source: Colcy and Steinehäuser, 1993. IMSC '93, The Third International Mobile Satellite Conference, Pasadena, California, 1993, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

11.3 LightSquared MSS-ATC Proposal

11.3.1 Background

The previous edition of the book described the L band MSS services provided by American Mobile Satellite Corporation (AMSC) in the American region (Johanson *et al*., 1993). Various commercial changes have occurred since – AMSC was rebranded as Motient Corporation; later a new entity called Mobile Satellite Ventures (MSV) was formed by merging operations of Motient corporation with TMI Communications a Company of Canada; MSV changed its name to SkyTerra in 2008; In 2010 Harbinger Capital Partners acquired SkyTerra, which then became part of LightSquared consortium (LightSquared, 2012).

LightSquared continues MSS operations integrating SkyTerra-1 satellite to their space segment to provide voice, push-to-talk, data and GPS tracking services primarily for public safety, security, fleet management and asset tracking; the ageing MSAT satellites are being replaced by powerful SkyTerra series of satellites. These satellites will constitute the satellite component of 4G-LTE open wireless broadband network being developed for the US by LightSquared.

LightSquared acquired a license through its forming parent company MSV to develop an integrated satellite-terrestrial network using an Auxiliary Terrestrial Component its allocated in conjunction with MSS frequency band. The license was later relaxed to provide a terrestrial-only service by removing the need of a mandatory satellite component but the license was later withdrawn due to interference likely to be caused to GPS receivers in the service area.

Keeping to the theme of the book, we outline the technical approach adapted by LightSquared/MSV's hybrid network solution due to its innovative appeal (Parsons and Singh, 2006) recognizing that the system could not be launched.

11.3.2 MSS-ATC Hybrid Network

The proposed hybrid MSS/ATC network consists of a spectrally-efficient multi-spot beam satellite system, which can provide mobile services on mobile phone sized, aesthetically pleasing user terminals integrated seamlessly with a network of ATC transmitters dispersed throughout the service area to service those regions where satellite services are unreliable and uneconomic such as in urban areas. The satellite network would service areas outside the coverage of the terrestrial component and thus the two networks would complement each other to provide a ubiquitous coverage. The space segment comprises satellites with high EIRP and sensitive receivers in the service link, which can provide the desired services to mobile phone sized handsets over a shared CN and an air interface that is either identical to or closely resembles the complementary mainstream terrestrial air interface to maximize commonality in the handset. An underlying premise regarding the air interface was that the terrestrial component would reuse L band MSS frequencies. Such a configuration makes the satellite transmissions appear as transmissions from a terrestrial base station. Sharing the chipset, antenna and handset peripherals with the terrestrial component and ubiquitous network connectivity would provide substantial economies of scale for the handset as well as the service costs due to a phenomenal increase in the addressable market from a few hundreds of thousand for a pure MSS to millions due to the attractiveness of such a service (Parsons and Singh, 2006). Handset cost was estimated to reduce 5–10 times compared to MSS-only sets through economies of scale. The scalability of the network allows the terrestrial component to be deployed incrementally. The system achieves high spectral efficiency by nesting terrestrial transmitters inside spot beams such that the MSS frequencies used in terrestrial system are not used in the overlaid spot beam to avoid interference. These terrestrial frequencies are used in other spots that are sufficiently isolated spatially. A mobile assisted handover adapted from prevailing terrestrial handover technique enables movement between the networks. The overall network capacity is increased by deploying two satellites.

An all-IP open network architecture was envisaged. The user would use the same network at home and when away when using the hybrid system. Such an IP network would support a bundle of telecommunication applications, and since it was shared with the fixed system, would benefit the mobile users as well. Alternatively, the system would provide only a mobile broadband service including voice.

The target markets included public safety and homeland security, consumer telematics, fleet management, direct broadcast satellite (to provide a return-link for interactive television), maritime and aeronautical rural markets.

Figure 11.14 illustrates the key elements of the proposed network architecture.

The space segment comprises two satellites, located at 101 and 107.3°W. Signals from mobiles are received by both satellites and retransmitted to ground to provide a diversity gain at the satellite gateway. The terrestrial ATC would be based on a mass-market terrestrial air interface standard supported on an all-IP open architecture including the satellite component to serve a variety of modern devices. Optimal use of spectrum utilization is made with a provision to partition spectrum geographically and between satellite and terrestrial networks on a needs basis. Some of the (patented) novel techniques used in the network include ground-based beam forming of satellite cells (allowing for optimum beam shaping to maximize received signal strength), space and polarization diversity reception by the satellites and ATC induced interference suppression technique.

The space segment characteristics are summarized in Table 11.13. Figure 11.15 illustrates the coverage map of the satellites for North and Central America with the G/T. The air

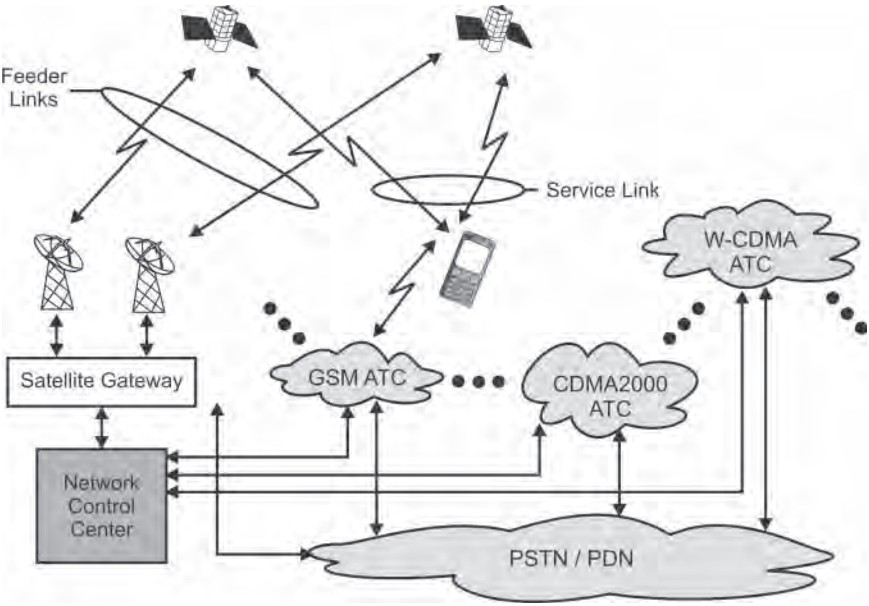


Figure 11.14 The key elements of the MSV's hybrid network proposal. (Source: Parsons and Singh, 2006. Reproduced with permission of Mobile Satellite Ventures.)

Table 11.13 Space segment characteristics of the hybrid system

Satellite orbit locations	101 and 107.3°W
Service links	1525–1559 MHz (forward) 1626.5–1660.5 MHz (return)
Feeder links	12.75–13.25 GHz (up-link) 10.75–10.95 GHz and 11.20–11.45 GHz (down-link)
Aggregate EIRP (dB W)	79
G/T (dB/K)	21 over primary coverage area
Number of spot beams and size	Variable number and size −500 and 0.4° typical
Supported protocols	Wideband (3G or 4G)
Processing	Digital channelizer on board, digital adaptive ground-based beam forming with interference suppression capability, space and polarization diversity reception
Prime contractor	Boeing
Launch vehicle	Sea Launch, Proton
Design life	15 y inclined

(Adapted from Parsons and Singh, 2006.)

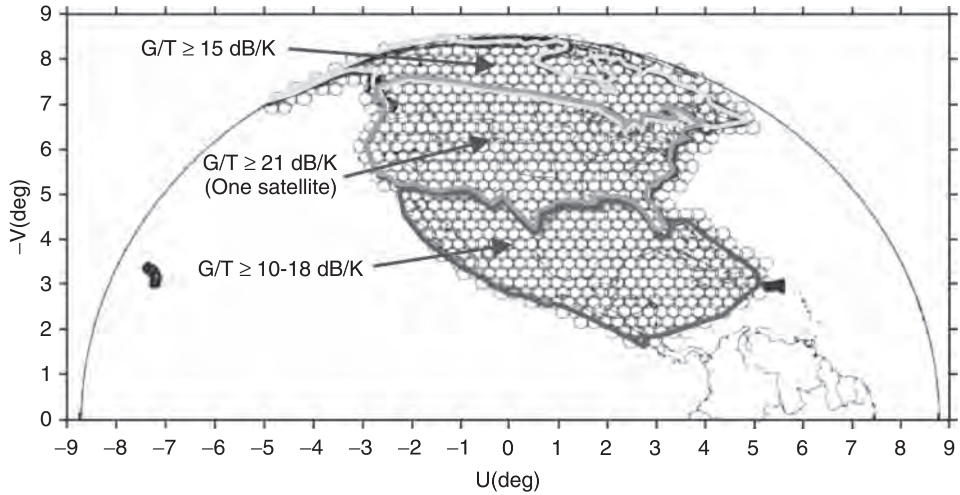


Figure 11.15 Space segment coverage of North and Central America. (Source: Parsons and Singh, 2006. Reproduced with permission of Mobile Satellite Ventures.)

interface standard is expected to be a satellite adaptation of a widely used cellular air interface such as WiMAX (Worldwide interoperability for Microwave Access) or LTE (Long term Evolution).

11.4 Big-LEO Systems

11.4.1 Iridium

11.4.1.1 Overview

The announcement of the formation of a global big-LEO system called Iridium satellite communications system (Iridium) in late 1989 by Motorola of the US to provide voice communication service on handheld user terminals heralded one of the most innovative phases of the satellite communications industry. Maintaining its lead, Iridium became the first big-LEO operational system in November 1998.

The system is named after the element Iridium, which has 77 electrons in its atom (i.e. atomic number = 77) corresponding to the 77-satellite constellation originally proposed by Motorola engineers. The size of the constellation was later scaled down to 66, but the original name (rather than Dysprosium, which has an atomic number of 66) was retained for its popular appeal.

An overview of the Iridium system components is portrayed in Figure 11.16 (Hutcheson and Laurin, 1995).

The space segment consists of a meshed 66-satellite LEO constellation comprising regenerative satellites interconnected via intersatellite links (ISLs).

The ground network consists of a commercial gateway in Arizona; a satellite network operations centre in Virginia; a technical support centre in Arizona; and five interconnected

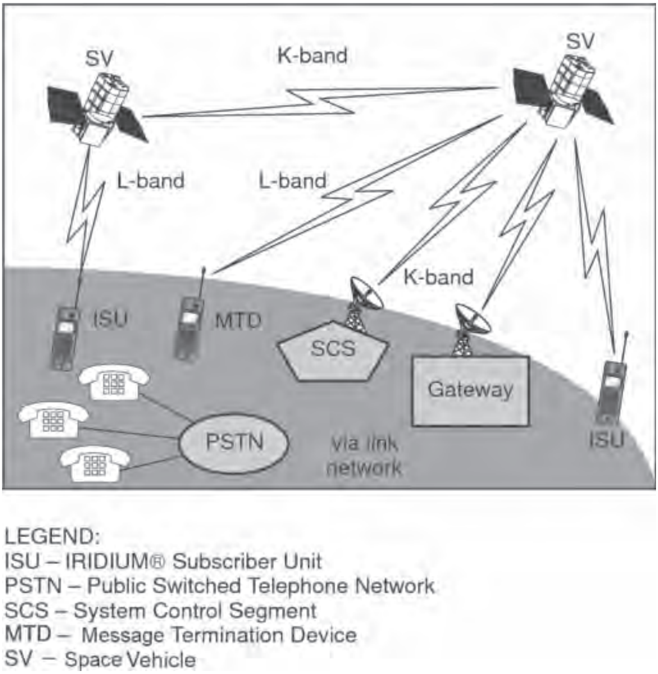


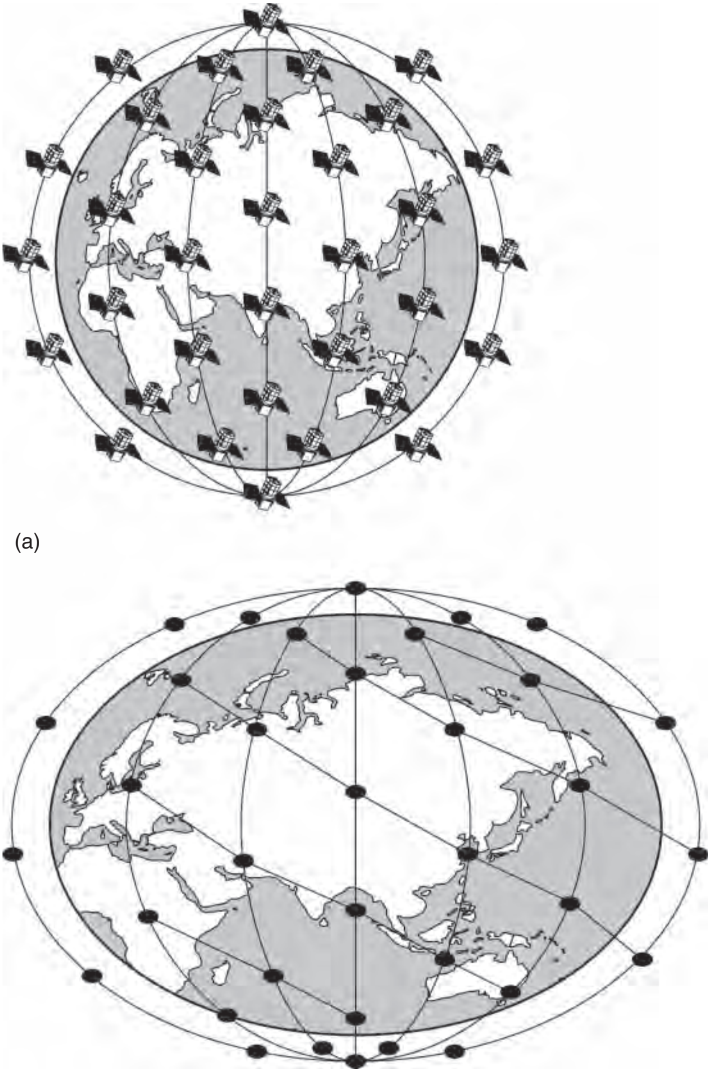
Figure 11.16 An overview of the Iridium system components. (Source: Hutcheson and Laurin, 1995. Reproduced with permission of Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

TT&C stations located in Canada, Alaska, Norway and Arizona. The ground infrastructure is made resilient by rerouting arrangements similar to that in the space segment. The US Department of Defense owns a gateway in Hawaii to support US government traffic (Iridium, 2012).

The system offers a number of low-medium bit rate services in land, maritime and aeronautical sectors. In the land sector it offers voice communication to pocket-size telephones, facsimile, two-way messaging, duplex 2.4 kbps data bearers, position determination, paging over a variety of maritime and aeronautical communication platforms. Handsets can select a local terrestrial cellular system, such as GSM, CDMA (code division multiple access), PDC (Personal Digital Cellular), D-AMPS (Advanced Mobile Phone System), when a user is within coverage of a participating terrestrial cellular system, and communicating through the space segment when outside the cellular coverage. The system offers a world-wide roaming service across a number of wireless protocols on the same subscriber number and a single bill. Pocket-sized pagers provide reception of alphanumeric messages in a number of international character sets and use an off-the-shelf disposable battery. The terminal cost and call cost target a wide clientele, including individuals. The maritime and aeronautical services include voice, facsimile and data services to ships and commercial, business and general-aviation aircrafts.

11.4.1.2 Space Segment

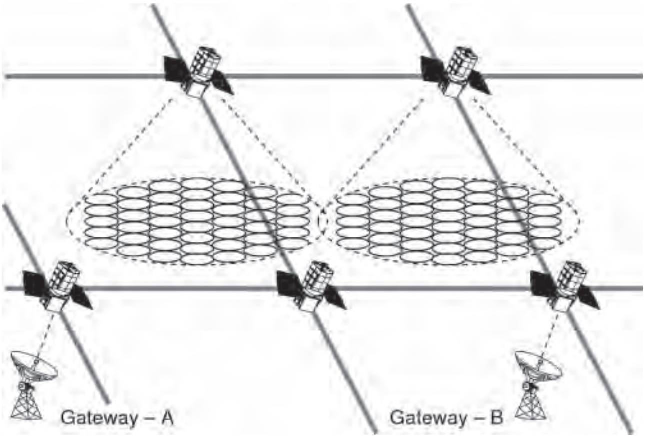
The constellation consists of satellites placed in six, 86° inclined orbital planes at an altitude of 780 km, with 11 equi-spaced satellites in each plane, to provide a seamless world-wide coverage at a minimum elevation angle of 8°. Figure 11.17 parts (a) and (b) illustrate a pictorial view of the constellation and its ISL connectivity, respectively.



(a)

(b)

Figure 11.17 (a) Pictorial representation of Iridium constellation, (b) intersatellite connectivity in Iridium network



(c)

Figure 11.17 (c) representative satellite footprint near the equator. (All parts source: Hutcheson and Laurin, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

To reduce risk in launch delays, Iridium satellites were launched through three launch providers – Boeing, China Great Wall and the Khrunichev State Research and Production Centre of the Russian Federation. These providers support launch services respectively through Delta 2 with six satellites per launch; Long March 2C/SD with two satellites per launch; and Proton 2K with seven satellites per launch. The first satellite was launched on 5 May 1997.

Each satellite deploys 48 circular service link spot beams of ~4800 km diameter. These spot beams provide a frequency reuse scheme similar to the terrestrial cellular system. Figure 11.17(c) shows the antenna pattern of two adjacent satellites near the equator. As the spot beams move with the satellites, the reuse scheme takes into consideration the dynamics of the constellation. In all, there are 2150 active beams on the surface of the Earth, which give about 180 times frequency reuse; within the USA alone, the system achieves a reuse of five times. Three phased array antennas, located on the side panels of the satellite, form 16 cellular beams each. Each beam supports an average of 236 channels. A spot beam can be used by more than one gateway. Iridium constellation visibility analysis and its comparison with other constellations are reviewed in Section 2.3.8.

Due to the limited visibility of satellites from ground in a LEO constellation, a large number of gateways are required to establish long-distance call connectivity if a terrestrial routing scheme is used. In the Iridium system, the routing arrangement was simplified by routing each call in space to the gateway nearest to the destination via ISLs to eliminate the need of intermediate terrestrial routing.

Each satellite is linked via a fixed waveguide slot antenna to the forward and rear satellite of the same plane and via a mechanically steered waveguide slot array to each adjacent orbit. Antennas provide a 5° elevation beamwidth so that steering is required only in the azimuth plane. The gain of each array is 36 dBi. Each ISL uses a QPSK modulation scheme and operates at 25 Mbps, thus the channels are spaced at 25 MHz intervals and eight 25 MHz

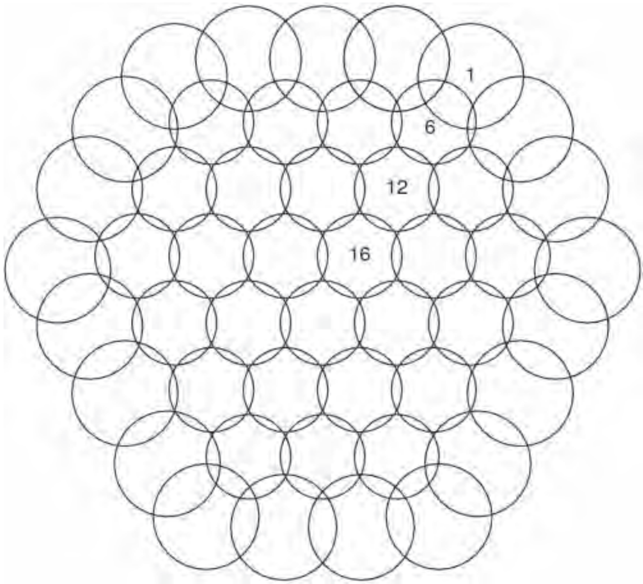


Figure 11.18 Iridium's spot beam coverage pattern near the equator. (Source: Freeman, 1996. Reproduced with permission of John Wiley & Sons, Ltd.)

bands can support all intersatellite communication of the network. RF signals are demodulated, decoded, switched and recoded/remodulated for transmission. The link provides a BER (bit error rate) of 1×10^{-7} with ½ rate FEC coding. Under the most severe loading, up to 1100 Iridium users can access a satellite and up to 5000 calls can transition through the satellite; the fully loaded throughput of the on-board processors is estimated as 100 Mbps. Link budgets for the service links are given in Section 3.4.

Service link spot beams are arranged in a regular pattern, as seen in Figure 11.18 and are satellite-fixed, that is move with the satellite. Being a polar constellation, coverage tends to concentrate at high latitudes resulting in large overlaps and therefore spot beams are switched off progressively as satellites approach higher latitudes to conserve satellite power and avoid unwarranted coverage redundancy and interference. The system does not use satellite diversity and instead a high link margin of ~ 16 dB is built into the RF link budget.

11.4.1.3 Radio Link and Network

Table 11.14 lists the frequency bands used in the feeder and service links. The system uses the band 1616–1626.5 MHz in both the up and down service link using the time division duplex (TDD) accessing scheme in a ping-pong fashion (described later). The feeder link uses 27.5–30.0 GHz band in the up-link (operational range: 29.1–29.3 GHz) and 18.8–20.2 GHz band in the down-link, (operational range: 19.4–19.6 GHz) and as stated, the ISLs operate in the 22.55–23.55 GHz band (operational range: 23.18–23.38 GHz). The system employs a combination of FDMA/TDMA-TDD signal multiplexing to make efficient use of limited spectrum.

Table 11.14 Iridium frequency bands

Link	Frequency (GHz)
Service up- and down-link	1.616–1.6265
Feeder link (up-link/down-link)	27.5–30.0/18.8–20.2
ISL	22.55–23.55

Figures 11.19 and 11.20 respectively represent the service link frequency plan and the TDD multiplexing format; each RF channel could support four full-duplex voice channels.

Each transmission is Doppler corrected such as to arrive at the receiver at the correct frequency. Mobiles use up to four 8.28 ms TDD bursts in 90 ms frames, the peak transmit power being 3.7 W at an average of 0.34 W, the remaining four slots are used for reception. Transmissions are QPSK modulated at 50 kbps, with carrier spacing of 41.67 kHz. Iridium user terminal use a propriety voice compression technique called Advanced Multi-Band Excitation (AMBE), a trademark of Digital Voice Systems Inc., providing voice at 2.4 kbps that is convolutionally coded at a one-third rate. The BER for voice transmission for the worst case is specified as 2×10^{-2} – typical BER range 1×10^{-3} to 1×10^{-4}. For

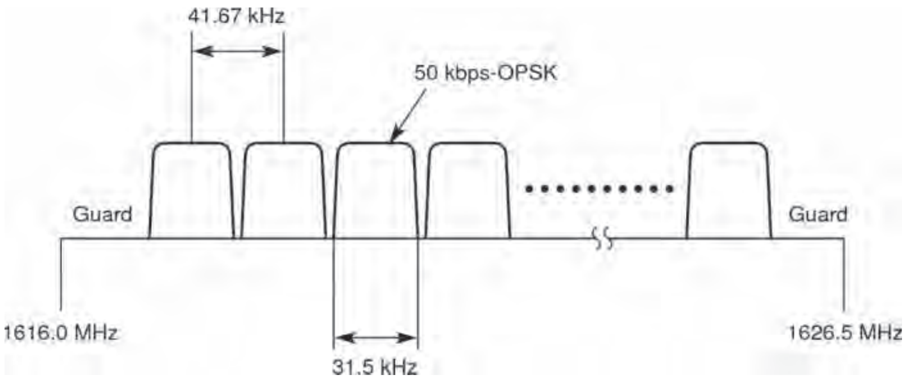


Figure 11.19 Service link frequency plan. (Source: Freeman, 1996. Reproduced with permission of John Wiley & Sons, Ltd.)

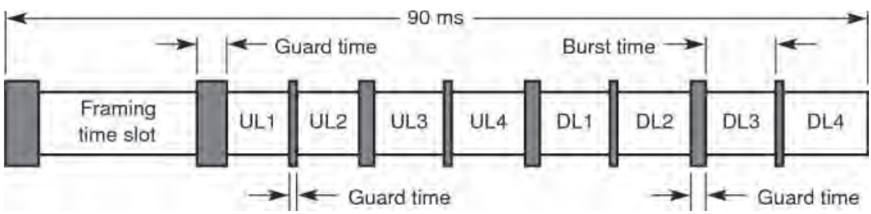


Figure 11.20 Time division duplex frame format. (Source: Freeman, 1996. Reproduced with permission of John Wiley & Sons, Ltd.)

data transmission, lower BER of the order of 1×10^{-5} are obtained by using more powerful coding and protocols, installed in the subscriber's unit. The subscriber units are similar in size and shape to a mobile phone using a quadrifilar helix antenna of 1 dBi gain, 8.5 dBW peak EIRP and −23 dB/K G/T.

Iridium gateways operate in the K_a band in the feeder link, using ½ rate FEC and transmission rate of 12.5 Mbps. Six channels are assigned at a spacing of 15 MHz. The supported BER is better than 1×10^{-7}. Two gateway antennas placed about 63 km (34 nautical miles) apart achieve spatial diversity gains, mitigating the effects of rain and remove outage due to solar interference; the availability on these links is of the order of 99.8%. Connectivity arrangements with regard to business management are discussed in Chapter 10.

The constellation is managed by the system control segment, comprising a ground station that performs TT&C functions in addition to computing routing data and frequency plans and loading them on satellites.

Gateways comprise K-band tracking earth stations that connect the space segment to the public networks via an international switching centre. The configuration uses (at the outset) SS7 multi-frequency compelled response signalling.

11.4.1.4 System Architecture

The architecture of the system is modelled after the well-established terrestrial standard, GSM (Hutcheson and Laurin, 1995) to allow easy integration with terrestrial systems and permit the Iridium system to benefit from advances in the terrestrial standard. Thus, each gateway incorporates GSM functions for call processing, and additionally, incorporates Iridium-specific functionality to manage communications with the satellite constellation, as depicted in Figure 11.21.

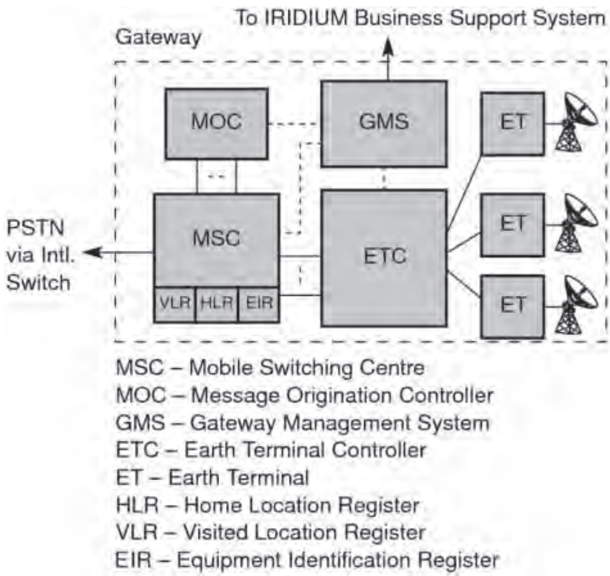


Figure 11.21 Main elements of an Iridium gateway. (Source: Hutcheson and Laurin, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

A MSC constitutes the main switching element connecting the gateway's earth terminal controller (ETC) to the public network via an international switching centre. The ETC, akin to the GSM base station system, is connected to the MSC by an 'A' interface. (Note: 'A' is the nomenclature of the interface between MSC and base station system). On the space segment end, it controls three K band Earth station terminals, one, carrying traffic, the second providing redundancy and rain diversity where separation is adequate, and the third ready to acquire a new satellite to take over connection from the current satellite. The VLR and HLR maintain location information about subscribers in much the same way as in GSM. The EIR is a database for keeping the identity of physical subscriber equipment. The message origination controller (MOC) supports Iridium's paging services. Operations, administration and maintenance support is provided by a gateway management system (GMS), which is also connected to Iridium's business support system (BSS). BSS has the function of managing usage and making settlement statements between Iridium Inc. and gateway operators, for which it receives call detail records from each gateway's GMS over an Ethernet link.

Each Iridium subscriber has a home gateway that maintains a record of the subscriber in an HLR. Depending on the situation, an Iridium subscriber can be identified by one of the several types of numbers. Mobile subscriber Integrated Services Digital Network (MSISDN) number is used by a land calling party; a temporary mobile subscriber identification (TMSI) is sent over the radio link while establishing connection, which is changed periodically to protect user identity; and an Iridium network subscriber identity (INSI), a number stored in the user's phone, sent on the radio path when a valid TMSI is unavailable. Table 11.15 shows the structure of a user's MSISDN number (Hutcheson and Laurin, 1995).

The 'geo-political entity' and 'service provider' fields in conjunction identify a subscriber's home gateway. When a call is made to or initiated from a subscriber, the system locates the subscriber and sends the location information to the home gateway using SS7 capability. The gateway evaluates the information and grants the subscriber access if the call is permissible at the location ensuring compliance with restrictions in a territory where access to the system is not permitted.

The home gateway also assigns call management authority to a visited gateway when subscriber is away from the home territory. After such an assignment, a visited gateway controls the call management. The home gateway uses the location information to determine the subscriber's location area code (LAC), which is used to determine the visiting gateway via an LAC map. The map defines the service area of each gateway on a world map in terms of LAC. Each gateway has the LAC association of every gateway; and up to 2047 LACs are available to each gateway.

Table 11.15 Number structure of Iridium's MSISDN

Iridium country code	Iridium significant number		
	Geo-political entity	Service provider	Subscriber number
Three digit (assumed)	Three digits 200–799	Two digits 0–99	Seven digits

(Source: Hutcheson, J. and Laurin, M., 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

The visited gateway receives subscriber's details from the home gateway via SS7 signalling and keeps a copy of the records in its VLR until the subscriber moves away to another gateway's jurisdiction. All call management is performed by the visited gateway as long as the subscriber is within the LAC served by it. This includes setup, maintenance, tear down and management of supplementary services.

When a call is made from the fixed network through a gateway where the subscriber is not registered, the serving gateway can determine the location of the subscriber through the home gateway and set up the call via ISLs. Similarly, a mobile subscriber can call a PSTN number outside the present jurisdiction of the home or a visited gateway. The location of the associated connecting gateway in this situation can be determined through the dialled PSTN number and the call can then be established using a transit connection between the mobile's serving gateway, ISLs and connecting gateway (i.e. where the called PSTN number is resident). Figure 11.22(a) and (b) shows the concept of such a transit connection.

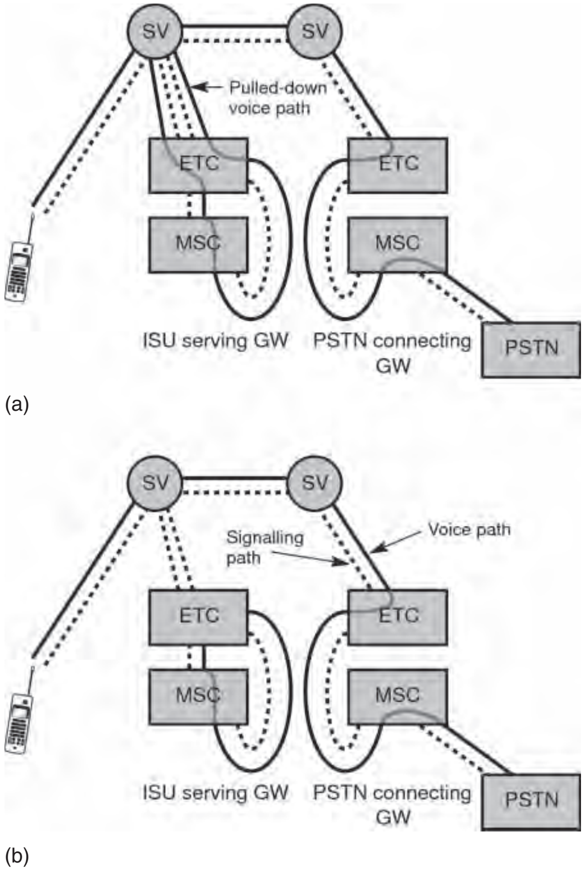


Figure 11.22 (a) Mobile originated voice call is routed through the serving gateway, intersatellite link and the destination gateway; SV = satellite vehicle. (b) Mobile-originated voice call is established directly to the destination gateway through an ISL. The voice connection through the serving gateway MSC is maintained. (Both parts source: Hutcheson and Laurin, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1995, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

In Figure 11.22(a), when a mobile call is to be routed through a PSTN served by a different gateway, the voice call is looped back through the serving gateway for call completion. In Figure 11.22(b), the voice call is established directly to the destination gateway through an ISL; however, the signalling path through the serving gateway is maintained, giving it full control of the call. Thus, if necessary, the call can be terminated. Moreover, the voice connection through the serving gateway's MSC is maintained, which allows the progressing call to be pulled back in when necessary, for example if a call-waiting tone has to be introduced. This feature of the system, known as 'cut-through', can also be used for establishing a mobile–mobile call to minimize delay. Cut-through provides a more efficient delivery, a lower path delay and reduced loading on the K band feeder link.

ISLs, GSM call processing architecture and geographically controlled system access feature permits flexibility in adding gateways when necessary for business or provide redundancy arrangements between gateways. Because the home gateway is determined by MSISDN and the visited gateway by the geographical map, it is possible to assign service areas to gateways as desired. For example, an operator with multiple gateways can partition subscribers in terms of best business, subscriber mailing address, service centre, and so on. If a new gateway is added, new or existing subscriber blocks can be added to the gateway's home register and its LAC map would be distributed to all gateways. Geographic regions can even be shared between visited gateways by providing LAC maps with common territory.

11.4.1.5 Iridium Next Generation

Iridium has embarked on replenishment of its space segment with the latest technical advancements to comply mobile telecommunication growth trend. (Orbital, 2012). The project known as Iridium NEXT, expected to be introduced in 2015–2017 timeframe, aims to provide:

- seamless backward compatibility with the existing system;
- substantially greater bandwidth to support up to 3 million users compared to 2 million users of the present generation;
- improved data speeds up to 1.5 Mbps compared to a maximum of up to 128 kbps;
- K_a band services at up to 8 Mbps;
- enhanced voice quality;
- more flexible bandwidth allocation;
- enhanced services including support of private gateways;
- support of powerful new devices and innovative applications;
- facility of secondary (hosted) payload (such as for climate change surveillance, low-resolution Earth imaging and space weather) to support third party needs.

The space segment will consists of 66 operational satellites with six in-orbit spare and nine spares on ground. The system will use the same frequency bands with satellites generating 48-beams within a footprint of about 4700 km in L band. The K_a band will be used in the feeder link and ISLs. The hosted payload may have a mass of up to 50 kg, average power of 50 W and data rate of up to 1 Mbps.

11.4.2 Globalstar

The Globalstar system was propounded by Loral Aerospace Corporation with Qualcomm Inc., the former being responsible for the space segment and the latter taking responsibility for most of the ground and user segments. The baseline system is based on a combination of cellular CDMA technology (US EIA/71A IS-95 standard) and LEO satellite system technology. The CDMA technology is developed from the Qualcomm's experience with the OmniTracs radio determination and messaging service and the development of the US's CDMA terrestrial standard IS95 (Wiedemanand and Viterbi, 1993).

The goal was to keep the system simple, affordable and low risk by employing well-proven technologies that also offered high spectral efficiency. Despite a launch failure, resulting in the loss of 11 satellites, the system began operation in 1999. Enabling technologies used in the Globalstar system include (Schindall, 1995):

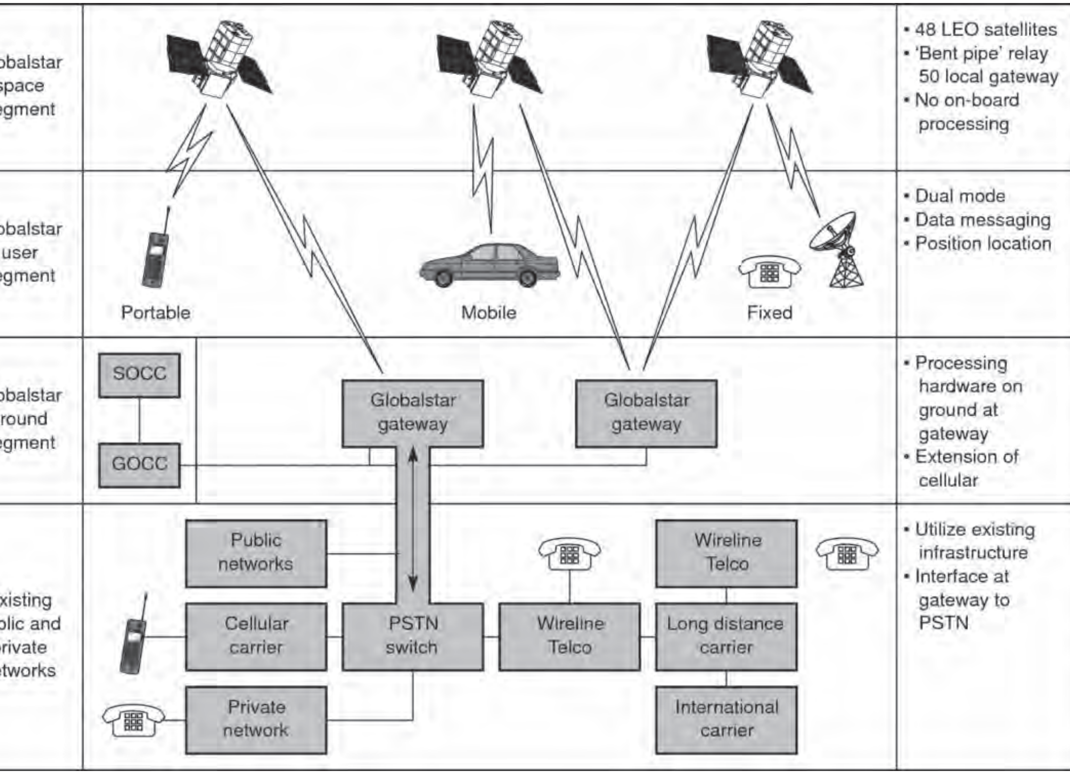
- spread spectrum/CDMA with efficient power control;
- high efficiency vocoder with voice activation;
- diversity, including soft handover;
- spot beams with weighted antenna gains.

The baseline system offers voice and a 7.4 kbps data bearer service. The service is targeted at rural and thin route requirements of the public, government and private networks – attempting to meet telecommunication needs where terrestrial services are expensive or non-existent. The system covers about 80% of the Earth comprising primarily the mid-latitude regions. The network is configured so that each service provider can offer services independently through gateways under their jurisdiction and control – thus service can be tailored for a specific region without involvement of gateways belonging to other service providers. The main system components, comprising the space segment, user segment, ground segment connected to public and private networks, are shown in Figure 11.23.

11.4.2.1 Space Segment

The Globalstar space segment comprises a 48-satellite Walker constellation (48/8/1 in shorthand, i.e. eight satellite planes and six satellites per plane) as illustrated pictorially in Figure 11.24 (Globalstar, 2012). The 390 kg (launch mass of first generation) satellites are placed in eight 52° inclined circular orbits at an altitude of about 1414 km, which has an orbital period of 114 minutes. The altitude was chosen such that it:

- Lies below the Van Allen belt but above the belt likely to be cluttered with debris and be affected adversely by atmospheric drag.
- Provides relatively low free space loss and low propagation delay so that the maximum end-to-end delay is less than 100 ms, of which only 18 ms are incurred in the satellite links.
- Covers large-enough areas per satellite with minimum number of satellites.



11.23 Main components of Globalstar's MSS. (Adapted from Dietrich, 1997. Fifth International Mobile Satellite Conference, Pasad
ia, June 16-18 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97–11, Jet Propulsion Laboratory, Pasadena, Califo
1997.)

Figure 11.24 A schematic of the Globalstar constellation. (Source: Globalstar. Reproduced with permission of Globalstar.)

Seamless coverage with dual-satellite diversity at $\sim 10°$ elevation is available up to $\pm 70°$ latitude and intermittent coverage is available up to about $\pm 75°$. The constellation visibility is optimized for the $\pm$ 30–60° latitude region, where the highest market was perceived. Figure 2.29 represents satellite visibility statistics as a function of latitude. The diameter of each satellite's footprint on the Earth is about 5760 km. Low-risk spacecraft technologies of the first generation satellites include a Sun tracking solar cell array and nickel-hydrogen battery cells. On-board GPS receivers are used for position and attitude control to simplify orbital maintenance and provide an on-board timing and frequency reference. Moreover, the orbital control requirements are relatively relaxed due to the large overlap between adjacent satellites. The system does not use ISLs. The simplified approach permits the use of modest-sized 390 kg satellites measuring $1.78 \times 0.96 \times 0.58$ m capable of being launched in groups of six to twelve by commercial launchers. The design life of the first generation satellite is 7.5 years.

Each satellite is three-axis stabilized and consists of transparent transponders and 16 isoflux-shaped spot beams in the service link comprising a central beam with two rings of outer beams – six beams in the inner ring and nine rings in the outer. Isoflux antenna systems exhibit a shaped pattern to compensate for the differences in flux density on the ground due to range variations across the coverage. The beams are generated with phased array technology using 91 elements that have individual high power amplifiers (and low noise amplifiers in the receive pattern). The C-band antenna systems provide an isoflux pattern to compensate for the path loss variations. The first generation spacecraft power system consists of solar arrays that have a beginning-of-life power of 1.9 kW and uses a large nickel-hydrazine battery as a back-up, which gets recharged (primarily) over oceans where low traffic intensity is expected, as the system mainly targets land-based applications.

11.4.2.2 Air Interface

Globalstar originators' experience with the terrestrial CDMA system enabled them to harness the terrestrial technology to advantage for the satellite environment. A CDMA scheme can provide frequency reuse with spot beam isolation as low as 2–3 dB, whereas in FDMA or TDMA isolation necessary for frequency reuse is 15–20 dB. The CDMA system permits graceful degradation of channel quality, unlike TDMA and FDMA that exhibit a hard capacity limit. At the limit of capacity for a given quality target, the operator may either permit degradation of channel quality or redistribute traffic between satellites in view – thus exceed the capacity limit under peak loading conditions. Spectrum management is relatively easy, as frequencies may be reused in all the spot beams.

The system utilizes frequency division multiplexed spread spectrum CDMA (FD/SS/CDMA) at S band (2483.5–2500 MHz) in the forward link (space to earth) and at L band (1610–1.626.5 MHz) in the return link. CDMA is a direct sequence spread spectrum scheme spread with QPSK modulation. The transmission band is spread over 13 1.25 MHz bands. The forward link CDMA channels are 1.23 MHz wide Walsh coded pseudo-random sequences. The return channels use CDMA with very long pseudo random spreading sequences and a quadrature spreading code of length of 2^{15}. The network distinguishes each user by a time offset determined by the user's address, thus providing a large address space.

Power control is used both in the forward and return links to promote an equitable sharing of the CDMA channel. Open and closed loop control is exercised. In open loop control, the power is increased when the mobile undergoes shadowing without any feedback as is necessary in the closed loop control.

The CDMA scheme allows path diversity by combining signals from multiple satellites (and possibly multipath) to mitigate fading effects. The diversity technique also simplifies handover between satellites. The majority of the handovers occur due to rapid movement of satellites with typical beam-beam handover occurring every 2–4 min. As soon as a gateway begins to receive signals of an on-going call from an adjacent beam, it begins to transmit a time-shifted version of the signal via the other beam. The RAKE receiver begins to track the new signal and finally shifts to the new signal when it is able to give adequate quality thus completing the handover. An identical procedure is used for satellite–satellite handover. This autonomous handover reduces protocol overheads.

The system provides around 11 dB fade margin in the service link with an additional 1 dB margin for Ricean fading. The power control scheme allows increments of power by up to 10 dB in steps of 0.5 dB. When the 11 dB margin has been exhausted, the system can route the signal through another satellite by utilizing path diversity. Typical values of Eb/No for voice communication in the forward link (i.e. at user terminal) is around 3.9 dB with two circuits in operation and 5.7 dB in the return link with satellite diversity.

Transmitted signals are picked up by all satellites in view of the user. The gateway passing the call manages power control as well as path diversity. It receives signals from all satellites in view and combines them. Each path is monitored for level and stability and the user is diverted to the best path at a pre-set threshold. A variable rate voice codec operating between 1.2 and 9.6 kbps, averaging 2.4 kbps, is used with one-third FEC in the up-link and half in the down-link. The vocoder rate is changed every 20 ms depending on channel conditions. In the absence of voice, the bit rate drops to 1.2 kbps, which reduces the transmitted power and hence reduces self-interference thereby increasing system capacity (see Sections 4.2.1 and 4.4.2). The voice-activated carriers operate at mobile powers of up to 2 W.

11.4.2.3 User Terminals

The system supports three types of user terminals – *fixed* for residential or rural locations, *portable* for mobile access to satellite components and *portable dual-mode* units for accessing satellite and terrestrial systems. The terminals use a rake receiver that allows reception of up to three signal paths simultaneously, irrespective of their being direct and combines them coherently to give a high signal level to the decoder. Fixed terminals use a small antenna and a 3 W power amplifier and, in addition to the rake receiver, incorporate software to manage calls and other functions. Portable units are slightly bigger than conventional mobile telephones using a specially designed antenna to provide near omni-directional coverage and 1–1.5 dB gain. The antenna protrudes slightly above the head to minimize blockage from the user's head. The output power averages to around −10 dBW, with a peak of about −4 dBW. Globalstar-specific functions include call management, local earth model computation for position location and others. The car-mounted version of the phone uses an outdoor antenna coupled to a power amplifier. The dual-mode terminal is an integration of the Globalstar portable unit with a mobile phone, reusing components wherever possible to preserve the features of cellular systems as much as possible. The phone design includes a subscriber identification module (SIM), allowing the user to change handset at will.

11.4.2.4 Gateways

The gateways are operated by local service operators and each gateway has an independent system control, giving the operators full control for call routing. Each gateway uses up to four 5.5–6 m tracking antennas, which ensures that all satellites in view – normally three – are tracked, while the fourth awaits the arrival of a new satellite. The feeder link operates in the 5 (up-link) and 7 GHz (down-link) band. If a lower number of satellites are visible, as at high latitudes, gateway antennas are reduced accordingly. Figure 11.25 shows a block schematic of the Globalstar gateway. Further details of network connectivity of the Globalstar system in relation to commercial activity are discussed in Section 10.3.

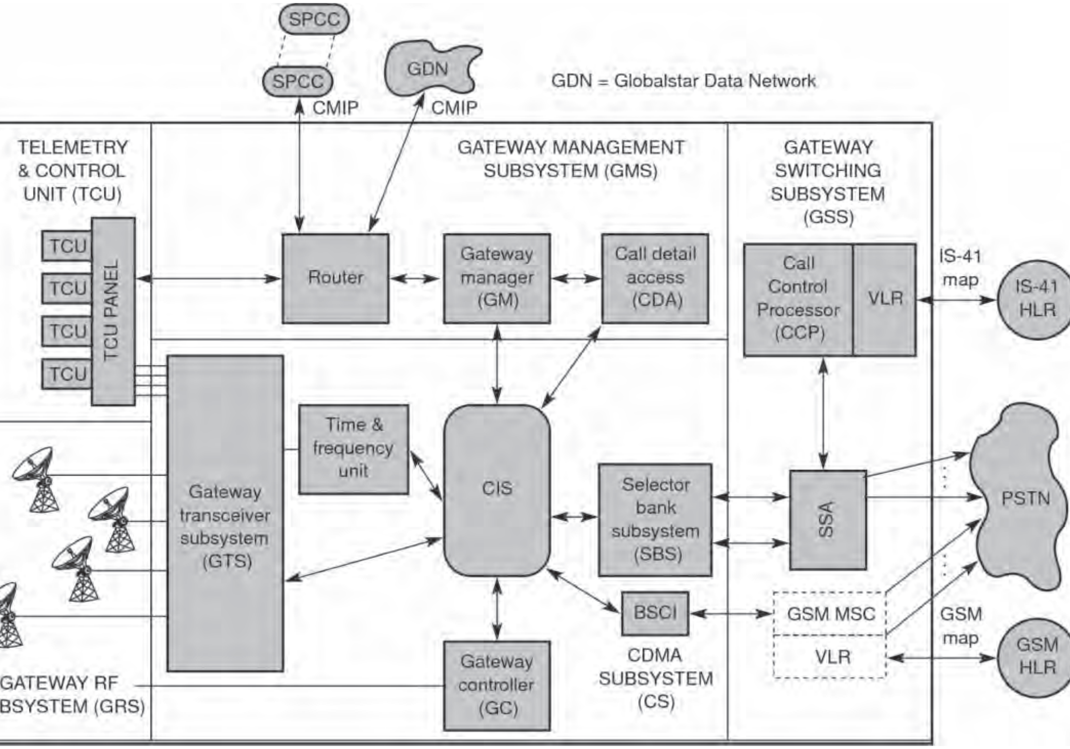
Telemetry and Control Unit (TCU) and Gateway RF sub-system GRS

The TCU acts as a telemetry and control interface between the constellation and the satellite operations control centre (SOCC). The TCU interfaces with the SOCC via the router in the gateway management system (GMS). The TCU interfaces with individual satellites via the gateway transceiver subsystem (GTS) and the GRS.

The GRS interfaces the gateway to Globalstar users via the Globalstar satellite constellation.

Gateway Management Sub-system (GMS)

The GMS performs non-real-time configuration and management of the gateway via the gateway manager (GM) and interfaces with external management entities including service



1.25 Block diagram of a gateway in Globalstar system. (Adapted from Dietrich, 1997. The Fifth International Mobile Satellite Confere
a, California, June 16–18 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97–11, Jet Propulsion Laboratory, Pasad
ia, June 16, 1997.)

provider's control centre (SPCC), SOCC and Globalstar Operations Control Center (GOCC) using the common management information protocol (CMIP). The SOCC and GOCC interconnections utilize the Globalstar's global data network (GDN) that links various sites and gateways. [Note: The GOCC allocates capacity to gateways and collects operational and billing data; The SOCC is responsible for managing the spacecraft and the constellation]. Call detail access (CDA) is a separate, fault-tolerant workstation within the GMS, with stricter reliability requirements than the rest of the GMS. The CDA uses a confirmed-transfer protocol to retrieve accounting from CDMA interconnect subsystem (CIS).

CDMA Sub-system (CS)

The CS performs real-time operation of individual calls, maintaining the integrity of each physical link and performing physical layer format conversion between the CDMA waveform on the GRS side and terrestrial signals on the gateway switching subsystem (GSS) end. Gateway controller (GC) is responsible for operation and supervision of the CIS and the GRS. GTS is responsible for the physical layer implementation of the Globalstar air interface. RAKE receivers are used. Under the control of the GC, control elements in the GTS set up and operate overhead and traffic channels as required. CIS provides packet-level and timing reference connectivity between all subsystems in the gateway. The selector bank subsystem (SBS) provides an interface between the GSS and the CIS, and performs layer-two operation and radio link management of individual traffic channel circuits. The SBS also performs service option-specific processing of traffic channel data. Service options include voice, data and short message services. Base station controller interface (BSCI) provides an interface between the CDMA subsystem (CS) and the GSM MSCs. The BSCI implements the BSC side of the A1 Interface, providing the SS7 transport, the protocol discrimination function and passing messaging between the GSM MSC and the CIS. The BSCI can be configured to terminate multiple A1 Interface links between multiple GSM MSCs. The configuration and set-up of the BSCI is controlled through the GMS interface (by way of the CIS). Time and frequency unit (TFU) provides a highly reliable and stable source of timing and frequency references to the CIS and to the GRS. The TFU output is synchronized to the GPS.

Gateway Switching Sub-system (GSS)

The GSS interfaces the gateway to the PSTN and controls the state of each call.

11.4.2.5 Globalstar Second Generation

In order to improve the reliability of its space segment, Globalstar began to deploy second generation satellites manufactured by Thales Alenia Space since October 2010 and Arianespace launched them in batches of six satellites by on Soyuz launch vehicle from the Baikonur Cosmodrome in Kazakhstan. The 32 second generation satellites integrate with first generation satellites to provide reliable coverage in the target service areas. Network characteristics are anticipated to remain unchanged.

11.5 Little-LEO System

11.5.1 ORBCOMM

ORBCOMM, a subsidiary company of Orbital Sciences Corporation (OSC), in cooperation with OSC, has developed and deployed a LEO mobile satellite system for world-wide low-cost, low bit rate two-way packet-switched communication services. OSC is a developer of small satellite technology and a low-cost air launch system, Pegasus. The ORBCOMM system can provide the service in the VHF band via a constellation of 36 LEO satellites. The VHF band permits low cost technology so that unit costs are lower than the terminals of big-LEO systems. The system is suited for applications requiring 100–200 byte two-way short message transmissions from/to remote areas to monitor and control remote fixed or mobile assets, email, etc. (Schoen and Locke, 1995; ORBCOMM, 2012).

The system also supports a ship broadcast system known as the automatic identification system (AIS), which is used by ships to transmit the vessel's identification and critical data essential for navigation and maritime safety. In preparation for deployment of the AIS system two satellites – one in the polar and the other in the equatorial orbit have been launched and there are plans to introduce 18 additional AIS equipped satellites.

ORBCOMM complements its satellite services by terrestrial cellular extensions to provide low-cost solutions when the subscriber unit lies within a cellular footprint. Management of subscription, devices and data is supported via a web-based portal.

The system provides a means for derivation of geolocation at each user terminal (to an accuracy of about 500 m) by suitably combining Doppler frequency shift measured by the user terminal with the satellite's ephemeris data broadcasted by each satellite. Satellites derive ephemeris (i.e. position, velocity and time) from an on-board GPS receiver.

Examples of applications include tracking of assets during shipment, remote interrogation of electronic tags, collection of environmental data, monitoring of remote sensors from field command posts, military communication, monitoring off-shore oil reserves and remote mines; SCADA, etc. In a typical application, a user terminal interfaces with a monitoring device, such as an electronic tag, through an RS-232 link. An application programme is used to activate the monitoring at the appropriate time and data is sent via the ORBCOMM space segment to the monitoring end.

Key features of the system are low transmission delay, low-cost, proven VHF electronics, simple low-cost/low-mass spacecraft resulting in a low-cost space segment and network features providing satellite-ground earth station (GES) handovers, GES and satellite based store and forward support and interfaces with terrestrial public networks. ORBCOMM's commercial goals rely in its choice of low-risk, low-cost technology and a niche market.

The key elements of the system shown on Figure 11.26 consist of gateways, the satellite constellation and mobile users. Gateways interface with terrestrial networks to provide remote users access to public terrestrial networks. A NCC controls the functioning of the network and satellite control centre controls and monitors performance of the satellites and gateways.

Functionalities built into gateway and user terminal ensure that the radio connection is made to an appropriate satellite when several satellites are in view. The user terminal remains unaware of a GES – satellite handover. In the case that the satellite being used by a user terminal sets during an on-going communication, the terminal searches for a new satellite

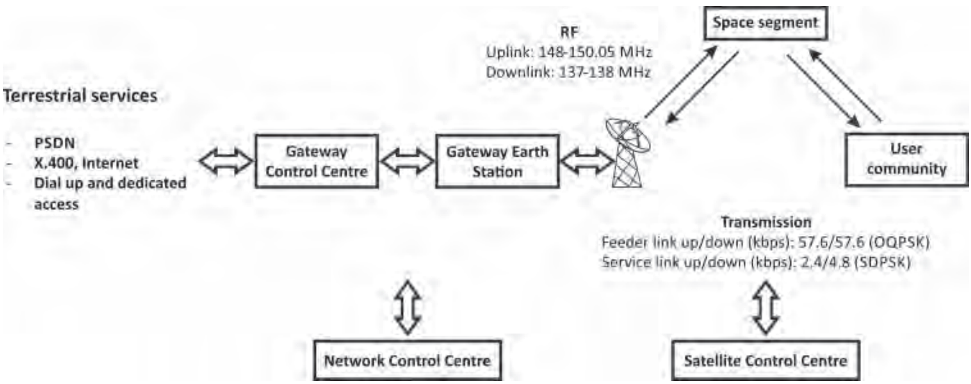


Figure 11.26 The main elements of the ORBCOMM network.

connected to the same gateway and when successful, continues communication via the new satellite. The broadcast channel provides the list of satellites in current use.

11.5.1.1 Space Segment

The complete constellation comprises 36 satellites, optimized such that users in temperate zones are in view of a satellite at 5° elevation for 98% of the time. ORBCOMM is permitted to launch up to 47 satellites in four 45° inclined planes of 8 satellites each, two highly inclined planes of 4 satellites each, and one equatorial plane of 7 satellites (ORBCOMM LLC, 2012). The target for 98% outage caused by satellites not being visible is <2 min and for 99% outage the target is <5 min. The orbital altitude of the constellation is 775 km giving a field-of-view of about 5100 km on the surface of the earth. [Note: Deployment status of the operational constellation at any time can differ on the basis of practical considerations.]

The low constellation altitude in conjunction with small satellites enables lower launch costs, of the order of $15 million (cost base - mid 1990s) for each Pegasus launch that orbits eight satellites per launch. In the case of an operational satellite's failure, the remaining satellites are repositioned within their plane to minimize the impact of the lost satellite. The system can absorb a loss of up to four satellites without significant impact on the system capacity.

Satellites are three-axis gravity gradient stabilized, transmit average power of 70 W and weigh about 43 kg (95 lb). Each satellite measures about 1.04 m (41″) in diameter and 16.5 cm (6.5″) in depth prior to deployment. These dimensions and weight permit up to eight satellites to be launched at a time. The transmit and receive signals are RHCP but use different antennas. ORBCOMM has launched several more powerful second generation satellites (called OG2) to enhance performance in terms of reduction in latency, allowing smaller antennas and lower power consumption at the user terminals with increase data rates and message payload size.

Satellites comprise message routing and queuing computers accessible by VHF radio links through eight receivers and three transmitters. Satellites include GPS receivers for attitude determination and provide orbital location and velocity information to ground receivers that support position determination capability. Satellite down-links operate in the 137–138 MHz little-LEO band. The band was chosen because it is shared with the space operations service, the use of which is diminishing in this band as the band is relatively narrow for modern spacecraft operations. The 148.0–149.9 MHz band used in the up-link was selected because of its proximity to the down-link, allowing the use of the same antenna for transmit and receive, and it would be easier to coordinate the frequencies. Satellites include an ultra-high frequency highly stable beacon at 400.1 MHz to improve accuracy in position determination due to a larger associated Doppler shift at this frequency for receivers capable of receiving the beacon. The operational band is shared with a number of terrestrial mobile systems. To minimize interference, the satellites use a channel management technique known as the dynamic channel activity assignment system (DCAAS). Each satellite monitors power levels on 2.5 kHz segments over the entire up-link band and channels are ranked 1 to N, from best to worst according to noise power; additionally, the spacecraft also keeps a record of packet error in each active channel. The spacecraft transmits this list to receive terminals on a broadcast channel for establishing a session. Table 11.16 summarizes the main features of the satellites (Schoen and Locke, 1995).

Table 11.16 Main characteristics of ORBCOMM first generation satellites

Mass	~95 lbs
Solar array power (orbital average)[a]	160 W
Transmitters	VHF (user link): 2 VHF (feeder link): 1 UHF (beacon): 1
Receivers	VHF (user links): 7 VHF (feeder links): 2
Attitude control	Autonomous/GPS
Design life	Four years
Cost (mid 1990s)	< $2 million

[a] Assumed as beginning of life power.
UHF, ultra high frequency.
(Source: Schoen and Locke, 1995. IMSC 1995, The Fourth International Mobile Satellite Conference, Ottawa, Canada, co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA.)

11.5.1.2 Gateway Earth Stations

A gateway system consists of a gateway earth station (GES) and a gateway control centre (GCC). A GES typically comprises two radome-enclosed 14–17 dBi tracking antennas with associated RF and baseband equipment. All equipments, including the antenna, have redundancy and are designed for unattended operation. GESs up-link in the 148.0–150.05 MHz band and receive in the 137.0–138 MHz band. The 56.7 kbps OQPSK (offset-quadrature phase shift keying) modulated waveform is transmitted at about 33 dB W in a TDMA mode. Each GES has a connection area of about 3100 km. The primary function of a gateway is to interface the users with public and private network and the Internet. Other functions of a GES are:

- acquisition and tracking of satellites using orbital parameters supplied by the NCC;
- to establish two-way communication to transfer messages and telemetry between satellites and GCC;
- to monitor status of its own hardware/software;
- to monitor system-level performance of the satellites to which it is connected.

The GCC interface the gateway to the terrestrial networks and can control more than one GES. GCC monitors the performance of each GES under its control and the status of ORBCOMM message switch system. Gateways have been installed at locations throughout the world.

11.5.1.3 User Terminals

The compact user terminals (~ 850 g – first generation), transmit at about 7.5 dBW and have a G/T of −28 dB/K with ½ wavelength (~1 m) whip antenna adequate to receive the ~ 12 dBW transmissions from the satellite. The Symmetric Differential Phase Shift Keying (SDPSK) raised cosine filtered modulated signals are transmitted in the 148–150.05 MHz band at a rate of 2400 baud. The received signal is DPSK modulated at the 4800 band in the 137–138 MHz band. Random access is used for passing control information or for short communication and reserved channels are used for message transfer. User terminals receive network information from satellites' broadcast channels.

A typical user terminal includes keypads, RS-232 interface port, integrated GPS receivers, encryption capability using Digital Encryption Standard chips and an LCD screen. The terminals are type-certified by ORBCOMM to ensure that the design complies with their operating licence. Each terminal has a unique ID given by the manufacturer and a unique X.400 address at the time of its activation, each of which is verified by the NCC for authorization.

Manufacturers produce terminals under a licensing agreement with ORBCOMM. Under this agreement, ORBCOMM provides the communications software, which allows terminals to communicate with the space segment.

11.5.1.4 Network Control Centre

The NCC manages the entire ORBCOMM system. The network is responsible for managing the operation of the satellite constellation, its processes and analysis of telemetry data. It monitors the message traffic of the network through a wide area dedicated network known as ORBnet.

11.6 MEO System

11.6.1 ICO System

A MEO system called the Intermediate Circular Orbit system was proposed by the (former) ICO Global Communications Ltd. created in 1995 to implement the system developed by Inmarsat project-21 programme (Inmarsat, 1992; *New York Times*, 1991). We will not attempt to follow the evolution or the status of this company, as this is outside the scope of the work but instead describe the ICO system concepts as conceived at the outset, as a representative of this class of MSS system. The goal was to provide world-wide circuit-mode voice telephony at 4.8 kbps, data services at up to 9.6 kbps, and a suite of messaging and value-added services to single- or dual-mode hand-held terminals using a MEO constellation depicted in Figure 11.27(a). The system was later modified to increase data rates. Here we will highlight the technical attributes of the baseline system (Bains, 1999).

The system architecture is shown in Figure 11.27(b). It is partitioned in to user segment, space segment and the ground segment consisting of ICO-Net (Intermediate Circular Orbit Network) comprising 12 interconnected SASs and gateways connected to public networks.

The service links operate in the personal communications band at 1.98–2.01 GHz in the up-link and 2.17–2.2 GHz in the down-link with QPSK modulated TDMA used for satellite access. Each Satellite supports at least 4500 TDMA telephone channels.

The feeder link operates in 5 and 7 GHz bands, which are the WRC-95 allocation for feeder links of non-geostationary satellite systems.

A feature of the system is a high-penetration notification with acknowledgement to alert a user to an incoming call when a normal communication link cannot be established due to excessive signal shadowing.

11.6.1.1 Ground Network

The *ground network* comprises 12 ground stations known as satellite access nodes (SANs) covering the world, interconnected with a terrestrial network known as ICONET, as shown in Figure 11.27(c).

A SAN interfaces the space segment and the terrestrial network. Each SAN comprises five antennas and the associated RF; a switch for routing traffic within ICONET and to the public networks; and databases to support mobility management, that is the HLR and VLR. Each SAN tracks satellites within its view and routes traffic to a user through the most robust link. Mobility management follows the GSM standards for which the information between SANs are exchanged over the ICONET. Gateways bridge SANs to various fixed and mobile terrestrial networks.

11.6.1.2 Space Segment

The *space segment* comprises a MEO constellation chosen as it could offer 'the best overall service quality for the desired market' by providing the following benefits:

- high elevation angle coverage;
- good satellite path diversity;
- slow satellite movement, of the order of 1°/min, as viewed from the Earth;
- A reasonable implementation-goal and schedule compromise.

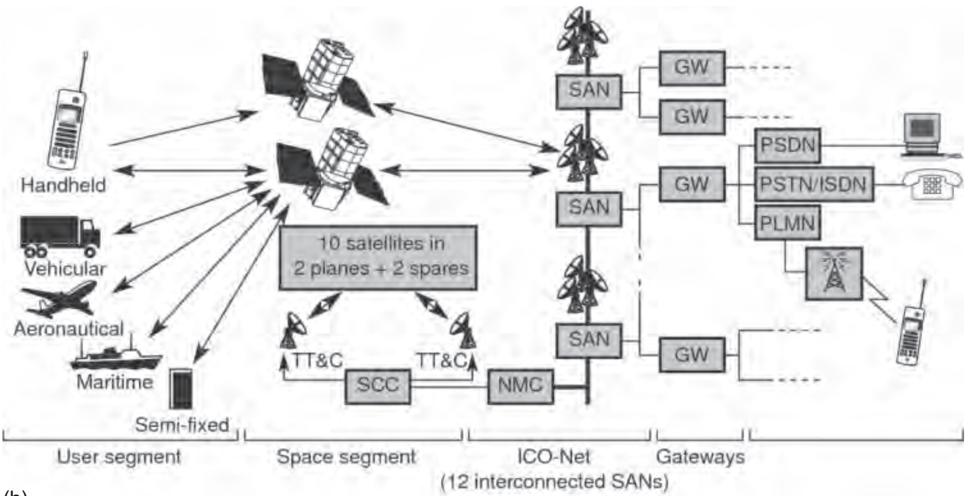
The size of a LEO constellation coupled with the short lifetime of satellites in such an orbit would impose logistical/manufacturing difficulties in constellation maintenance; satellites in a GEO constellation would become too complex due to the need for a large number of spot beams, and transmission delays of satellites would be relatively long.

The ICO constellation consists of 10 satellites in two 45° inclined orbits arranged such that there are five satellites and a spare in each plane at an orbital altitude of 10 355 km. Figure 11.27(a) represents a pictorial view of the constellation. The constellation provides world-wide coverage at an elevation angle exceeding 10° over throughout the entire earth and up to 40° at mid-latitudes. Satellite footprints from such an altitude cover about 30% of the Earth's surface at any instant as illustrated in Figure 11.28. Figure 2.29 shows visibility statistics of satellites as a function of latitude (see Section 2.3.8). Two to four satellites are generally visible at any instant from a SAN and the user community.

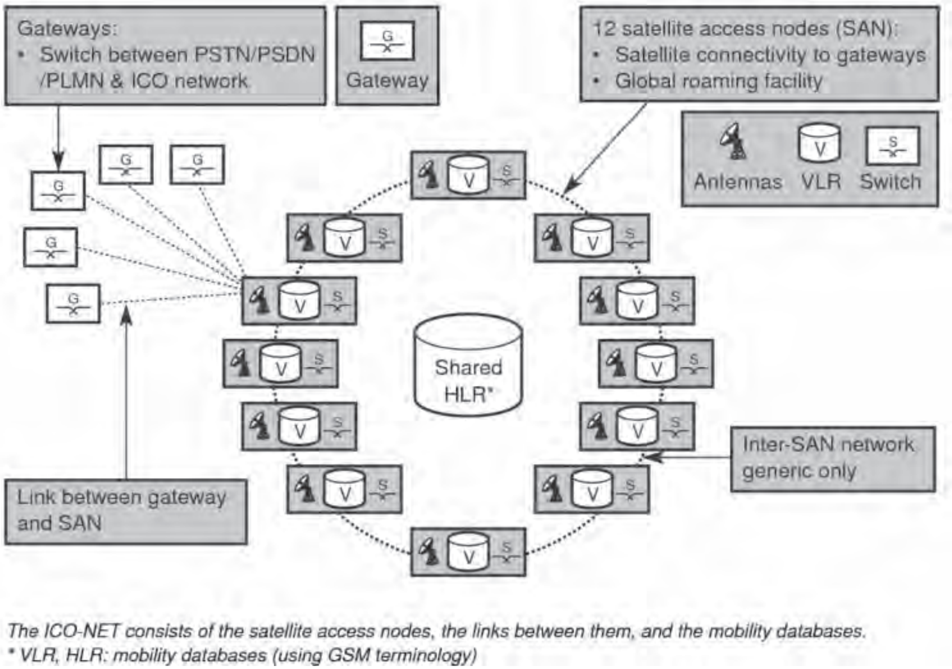
Satellites deploy a transparent transponder to permit flexibility in transmission format. Maximum use of digital technology provides flexibility in satellite reconfiguration, with

(a)

Figure 11.27 (a) ICO constellation pictorial representation



(b)



(c)

Figure 11.27 (b) ICO system overview and (c) ICO-NET architecture. (All parts source: Bains, 1999. IMSC '99, The Sixth International Mobile Satellite Conference, Ottawa, 1999, co-sponsored by the Communications Research Centre and the Jet Propulsion Laboratory.)

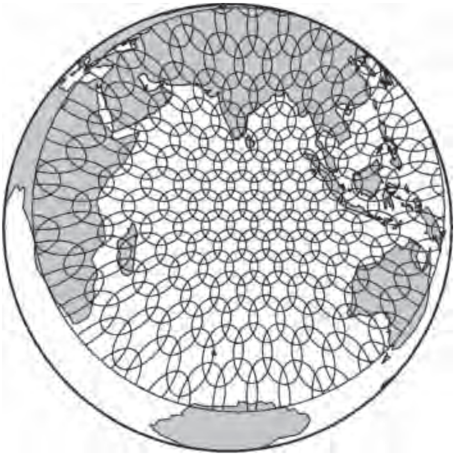


Figure 11.28 Spot beam coverage from a single satellite. (Source: Bains, 1999. Reproduced with permission of IMSC '99, The Sixth International Mobile Satellite Conference, Ottawa, 1999, co-sponsored by the Communications Research Centre and the Jet Propulsion Laboratory.)

advantages in manufacturing and mass production. Functions where digital subsystems replaced analogue solutions include spectrum channelization and beam generation. Transmit and receive antennas used in the service link are separate to achieve a better inter-modulation protection than a composite arrangement and to facilitate manufacturing. Each satellite provides 163 transmit/receive spot beams in the service link using antennas exceeding 2 m diameter. Transmit EIRP can provide an 8 dB link margin at the user terminals in typical operational conditions.

Gallium Arsenide Solar cells generate 8700 W of DC power. Each satellite, weighing about 2600 kg, can be injected directly into the MEO, thus dispensing the need for an apogee-kick motor (used in GSO (geostationary satellite orbit) launch) and permitting multiple-satellite launch. The design lifetime of each satellite was estimated at about 12 years.

The function of the satellite control centre (SCC) is to manage the constellation by regularly adjusting the orbit of each satellite in normal operation and in contingency when a satellite fails. The SCC monitors satellite health and supports the launch and deployment phase. An important operational task of the SCC is to configure feeder-to-service link connectivity to distribute RF channels between spot beams according to traffic distribution.

11.6.1.3 User Segment

The *user segment* consists of competitively priced single and dual-mode hand-held sets comparable to mobile phones including features like external data ports/internal buffer memory to provide data communications, messaging and use of SIM. To comply with radiation safety standards, the average power would be about 0.25 W, similar to those of mobile phones (ranging ~ 0.25–0.6 W). Dual-mode terminals have the capability to select the ICONET or a participating terrestrial system. Other types of user terminals include vehicular, aeronautical, maritime and fixed/semi-fixed installations.

11.7 Hybrid Orbit Systems

11.7.1 ELLIPSO™ System

11.7.1.1 Services

The ELLIPSO™ system was one of the five big-LEO systems granted licences in the USA for commercial MSS operation (Castiel and Draim, 1995; Draim and Davidson, 1999). The system did not materialize but we include it here for to its exemplary constellation design. The system was conceived to provide public mobile communication services in Canada and most of the populated world. ELLIPSO™ intended to provide services to users whose communication needs were not served by existing mobile or fixed telephone systems. The target market would therefore be residents of remote areas, remote installations such as offshore mining and exploration sites, tourists, mountaineers, emergency services, security services, cross-country transportation fleets, aviation, etc.

The service would be provided through mobile phone-sized sets, fixed and semi-fixed installations. The hand-held phones would resemble the terrestrial phone in appearance as well as radiated power, weight, size and operation. Other phone models included vehicle-mounted units and ruggedized platforms for maritime and military applications. Semi-fixed/fixed installations included coin/card-operated solar payphones and private residential phones consisting of rooftop antennas wired to in-building phone jacks to plug in regular phones and a ruggedized transportable version. Other terminal designs included data-only terminals, paging and polling terminals.

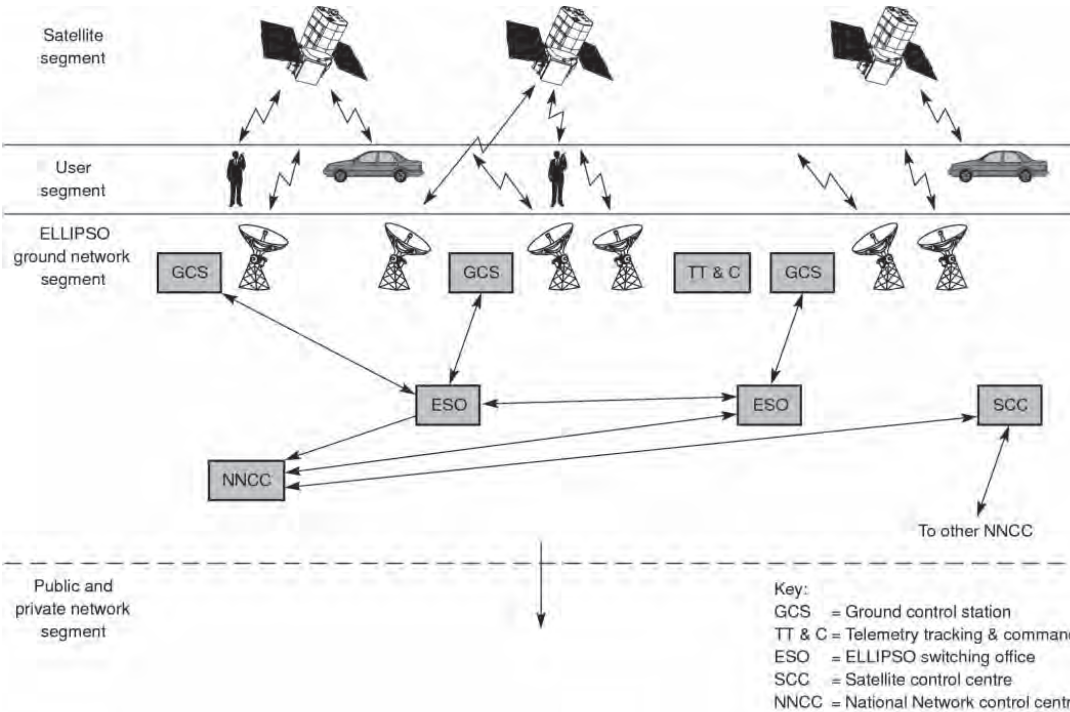
There were plans to sell satellite time wholesale. Services included telephony, data, facsimile, paging, voicemail, messaging and geo-positioning services. Figure 11.29 shows the main constituents of the system.

Due to the rather tentative nature of the market, the system was designed to profit with a conservatively sized market. There are elements in ELLIPSO™'s system design that allow capacity to be tailored to suit geographical variations in demand. An initial market of fewer than 1 million subscribers would obtain profitability, of which, 400 000 users would be targeted in the USA alone, and the remainder in the rest of the world.

11.7.1.2 Space Segment

The system would deploy a unique patented constellation design, which combines highly elliptical and circular orbits judiciously to favour coverage over desired regions of the Earth during daytime hours when higher capacity is required. This approach allows an efficient matching of demand and capacity, with a lower number of satellites than in an unbiased constellation.

The Earth's distribution of land masses and population forms the basis of the constellation design. The northern hemisphere above 40°N latitude contains several times more land mass



re 11.29 Main components of ELLIPSO™ satellite system. (Source: Castiel and Draim, 1995. IMSC '95, The Fourth International ite Conference, Ottawa, 1999, co-sponsored by the Communications Research Centre and the Jet Propulsion Laboratory.)

than the southern hemisphere south of 40°S – the former includes Europe, about half of the United States and Canada, the CIS and parts of Japan. In the southern hemisphere, most of the land lies within about 52°S and most of the world's population is confined to areas above the 40°S latitude.

Inclined elliptical orbits are used to provide coverage to high latitudes in the northern hemisphere and a circular equatorial orbit provides coverage in the remaining portions. The orbital design comprises two 116.5° inclination, elliptical Sun-synchronous orbits, named BOREALIS™, each having a perigee altitude of 520 km and an apogee of 7846 km over high northern longitudes, such that they provide high elevation angle coverage over Canada and North America. The reason for the choice of Sun-synchronous orbits is that such orbits exhibit the same relationship with respect to the Sun and hence it is possible to have an eclipse-free sun light and maintain the same time-of-day relationship throughout the year (see Chapter 2). The orbital parameters are chosen such that the satellites remain in the northern hemisphere during daylight hours for an average of two-thirds the orbital period. In practice, this property can be achieved by selecting the orientation of the plane with respect to the earth and tilting the perigee–apogee line and epoch so that the apogee is over the desired northern region at the desired time of day. Launching satellites in an elliptical orbit is less expensive than in a circular orbit with a similar coverage. The satellites remain quiescent below the equator, which reduces power consumption while the batteries are charged. ELLIPSO™ system designers estimate that four satellites in a Sun-synchronous orbit provide service equivalent to six circular orbit satellites. There are five satellites in each orbital plane, with a period of approximately 3 h.

The lower northern latitudes, tropical regions and populated regions of the southern hemisphere are served by satellites in an equatorial circular orbit called CONCORDIA™. The plan is to deploy six satellites at an altitude of 8040 km (4.8-hour period) so that continuous coverage is available up to the 55°S latitude, which contains the most populous regions. The Tropic of Cancer approximately divides the service area of the two constellations. There is a wide band of coverage overlap. Constellation deployment can be adjusted to meet market demands. The constellation provides dual coverage to achieve diversity gains above the 40°S latitude and gives single coverage up to latitude of 55°S.

Figure 11.30 shows a representation of the ELLIPSO™ constellation. Figure 2.29(d) presents coverage statistics of both the orbits taken over a two-week period.

All the satellites are of identical design to minimize risk and ease manufacture. Each satellite, weighing 1350 kg, is three-axis stabilized and deploys two large solar arrays of about 40 m^2 area. The transponders are transparent, using well-proven designs. The antenna provides 61 spot beams in the service link, arranged in a circular symmetry. The accessing scheme is CDMA, allowing the same 12 MHz band to be used in each spot beam. The service link operates at 1610–1621 MHz in the up-link and 2480–2500 MHz in the down-link. A single beam is used in the feeder links that uses K_u band in the up-link and C band in the down-link. Satellites use reaction wheels and thrusters for attitude control and station keeping, with enough battery reserve to allow full eclipse operation. The mean lifetime of each satellite is five years. Up to six satellites can be launched at a time.

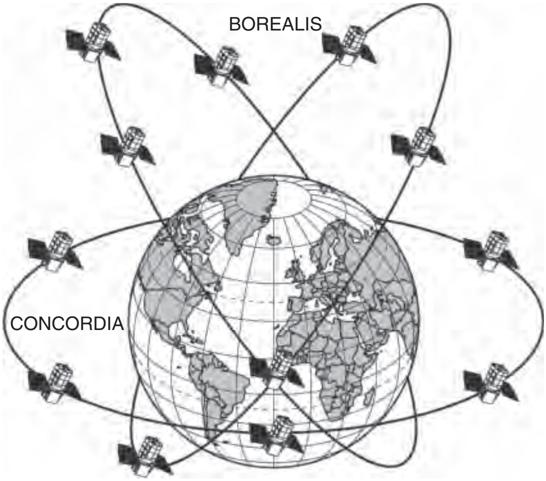


Figure 11.30 Representation of ELLIPSO^{TM} constellation. (Source: Castiel and Draim, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1999, co-sponsored by the Communications Research Centre and the Jet Propulsion Laboratory.)

11.7.1.3 Air Interface

The ELLIPSO^{TM} voice codec uses a technique known as the code excited linear predictive (CELP) algorithm that can provide high quality voice at 4.15 kbps. In addition, a number of digital services in the range 300–9600 bps are supported. These include Hayes modem data, facsimile, message forwarding, paging and geolocation.

Signals are convolutionally coded and interleaved, and the pseudo-random code spreads the signal by a factor of 1000. Signals may be transmitted over a 3 MHz wide band or a 7.5 MHz with the wideband transmissions generally assigned for mobile and hand-held applications.

User terminals in idle state remain locked to a broadcast channel. The broadcast channel announces system status and incoming calls and is also used by terminals for maintaining synchronization. Calls are established through exchange of a sequence of packets over the broadcast channel in the forward direction, and response and requests originating from a mobile in the return direction. After a call has been successfully established, the ground control station (GCS) assigns a spreading code key to the user, which is used during the call. In-band signalling, as, for example, necessary for power control, is multiplexed with data signals. Some short messages or functions, such as paging and position location, are carried over the signalling channels. The system is capable of handling advanced signalling and call services offered by the PSTN.

ELLIPSO^{TM}, like Globalstar, chose CDMA as the multiple access scheme, for similar reasons. In particular, CDMA offers frequency reuse in each beam and satellite, the capability to combine signals when using path diversity, the capability to provide soft handover,

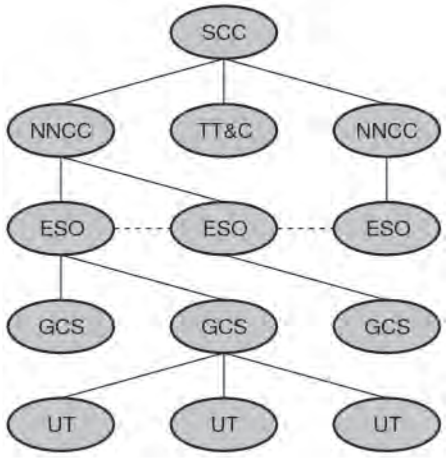


Figure 11.31 ELLIPSO^{TM}'s ground network architecture. (Source: Castiel and Draim, 1995. IMSC '95, The Fourth International Mobile Satellite Conference, Ottawa, 1999, co-sponsored by the Communications Research Centre and the Jet Propulsion Laboratory.)

less difficulty in spectrum coordination, capacity enhancement through voice activation, resistance to fading, the potential to make precise time measurements for geolocation and the potential for privacy.

11.7.1.4 Ground Network

ELLIPSO^{TM}'s ground network comprises the GCSs, national network control centres (NNCCs), ELLIPSO^{TM} switching offices (ESOs), system coordination centre (SCC) and TT&C centres. Their primary functions are:

- call connection;
- power control for optimum utilization of capacity;
- selection and maintenance of optimum transmission path;
- interface to public network;
- subscriber record keeping, validation and transaction accounting;
- dissemination of network information to all entities, including user location information maintenance, dissemination and geolocation;
- system-related functions – planning, system health monitoring/maintenance, resource allocation, creation of optimum network configuration.

The ground network arrangement, shown in Figure 11.31, has been modelled on the GSM architecture, equipment and standards to facilitate integration with terrestrial networks. The network utilizes modern switching and trunking facilities, and CCITT (International Telegraph and Telephone Consultative Committee) and ANSI (American National Standards Institute) signalling systems. Features include CCITT SS7, attributes such as call forwarding and call identification, domestic and international subscriber roaming, and so on.

A GCS uses three antennas; two for communication while the third is kept ready to acquire a new satellite. The GCS interfaces with the system's switching office for communications and signalling and with the area's NNCC for system status and control information. The GCS also determines the user's location for establishing connectivity. Functionally, a GCS is similar to a terrestrial base site controller and emulates it to benefit from terrestrial technology and facilitate integration.

The ESO interfaces with the PSTN and includes X.25 and SS7 network interfaces. The ESO is also connected to NNCCs for control and status purposes, transfer of call records and subscriber information. The ESO contains a fault-tolerant switch, data processing system and subscriber and network databases for managing subscriber profiles and call connectivity. It can be located at any convenient position – at a GCS, NNCC or a terrestrial MSC – and may serve one or more GCSs.

The NNCC controls the ELLIPSO™ market in a region and is the central facility for network planning, management and accounting. Its functions are handling of subscriber affiliation, record keeping, accounting, roamer call routing, GCS management, satellite and network resource management system. The NNCC interfaces with the ESO, GCS and SCC to carry out these functions. The interface with the SCC is used for transactions reporting, global roaming management and system resource management.

The SCC manages the overall ELLIPSO™ system functions, interfacing with each NNCC of the network and with the TT&C. Its functions, in conjunction with these interfaces, include world-wide call transactions, maintenance of the subscriber database together with their current status/location, coordination of routing for international roaming and management of network and spacecraft health and status. Global system planning and resource allocation is also a part of the SCC's function. It was envisaged that separate SCCs may be established for BOREALIS™ and CONCORDIA™ constellations.

ELLIPSO™'s TT&C would be situated in two locations so that together they can see each satellite in every orbital revolution. The function of the TT&C is to monitor, control the function and integrity of the spacecraft, maintain orbit, determine ephemeris and assist the NNCC in managing satellite allocation and handoffs.

The location of each network component would be decided on the basis of market needs, which were believed by ELLIPSO™ to be influenced by population distribution, demand for ELLIPSO™ services, the status of the prevailing telecommunications infrastructure, amongst others. Technically, it would be possible to provide a world-wide service with a minimum of 14 GCSs strategically located. However, factors such as trunking costs and territorial interests would influence their number and location. For example, it is envisaged that markets such as the USA, spanning large geographical areas, would require more than one GCS to provide acceptable subscriber connectivity statistics, recalling that common satellite visibility between the two is essential for making a connection. GCSs would likely to be located at the extremes of sub-satellite travel over the market area; and intermediate stations would be added if better services could be offered through trunking economy.

References

Andrews, E.L. (1991) Satellite group plans a pocket-phone system. *New York Times* (Sep. 15), available at http://webcache.googleusercontent.com/search?q=cache:http://www.nytimes.com/1991/09/15/world/satellite-group-plans-a-pocket-phone-system.html (accessed 28 August, 2013).

Astrium (2012) Alphasat www.astrium.eads.net/en/programme2/alphasat.html (accessed 28 August, 2013).
Bains, N. (1999) The ICO system for personal communications by satellite. Sixth International Mobile Satellite Conference, Ottawa, Canada, June, pp. 88–93.
Boeing (2012) Inmarsat-5. www.boeing.com/defense-space/space/bss/factsheets/702/Inmarsat-5/Inmarsat-5.html#specs (accessed 28 August, 2013)
Castiel, D. and Draim, J.E. (1995) The ELLIPSO™ mobile satellite system. IMSC 1995, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 409–418.
Colcy, J.N. and Dutronc, J. (1990) Land mobile communications in Ku band: result of a test campaign on EUTELSAT IF1. *International Journal of Satellite Communications*, **8**, 43–63.
Colcy, J.N., Dutronc, J. and Ames, W.G. (1990) The EUTELTRACS position reporting system characteristics and performance. NAV 90 – The 1990 International Conference of the Royal Institutes of Navigation, Warwick, UK.
Colcy, J.N. and Steinhäuser, R. (1993) EUTELTRACS: The European experience on mobile satellite service. IMSC 1993, Pasadena, CA, 16–18 June, JPL Publication 93-009, 261–266.
Dietrich, F.J. (1997) The Globalstar satellite communication system design and status. International Mobile Satellite Conference, 1997, Fifth International Mobile Satellite Conference, Pasadena, California, June 16–18 1997, co-sponsored by NASA/JPL and DOC/CRC; JPL Publication 97-11, Jet Propulsion Laboratory, Psadena, California, June 16, 1997.
Draim, J.E. and Davidson, C.S. (1999) The ELLIPSO™ system – An optimal solution for the Canadian mobile satellite communications market. IMSC’99, Sixth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre and the Jet Propulsion Laboratory June, pp. 76–80.
ESA (2012) Telecommunications and integrated applications www.esa.int/esaTE/SEMWXV0T1PG_index_0.html (accessed 28 August, 2013).
Eutelsat (2012) web site www.eutelsat.com/products/mobile-fleet-management.html (accessed 28 August, 2013).
Febvre, P., Bouthors X., Maalouf, S. and Bath, B. (2007) Efficient IP-multicast *via* Inmarsat BGAN, a 3GPP satellite network, *International Journal of Satellite Communications and Networking*, Special issue on ASMS 2006, **25**(5), 459–480.
Freeman, R.L. (1996) *Reference Manual for Telecommunications Engineering*, 2nd edn, John Wiley & Sons, Inc., New York.
Gambaruto, E., Richharia, M. and Trachtman, E. (2008) Land vehicular propagation channel characterization of BGAN system. *International Journal of Satellite Communications and Networking*, **26**(2), 85–117.
Globalstar (2012) How Globalstar Works, About Globalstar/www.globalstar.com (accessed 26 July 2012).
Hutcheson, J. and Laurin, M. (1995) Network flexibility of the Iridium® global mobile satellite system. IMSC 1995, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 503–507.
Inmarsat (1992) web site, www.wtec.org/loyola/satcom/ab_inmar.htm (accessed 28 August, 2013).
Inmarsat (n.d.) web site, www.inmarsat.com (accessed 28 August, 2013).
Inmarsat-Globalexpress (2012) web site www.igx.com/ (accessed 28 August, 2013).
Iridium (2012) Iridium: The Global, Reliable Network, www.iridium.com/ (accessed 28 August, 2013).
Iridium-Next (2012) Iridium NEXT Satellite Generation, www.iridium.com/ (accessed 28 August, 2013).
Jacobs, I.M. (1989) An Overview of OmniTRACS: The First Operational Two Way Mobile Ku Band Satellite Communication System. *Space Communications*, (Dec), **7** (1) , 25–35.
Johanson, G.A., Davies, N.G. and Tisdale, W.R.H. (1993) The American mobile satellite system: implementation of a system to provide mobile satellite services in North America. *Space Communications*, **11**, 121–128.
LightSquared (2012) web site www.lightsquared.com/ (accessed 26 July 2012).
Nguyen, N.P.; Buhion, P.A. and Adiwoso, A.R. (1997) The Asia cellular satellite system. IMSC 1997, Proceedings of the Fifth International Mobile Satellite Conference, Pasadena, CA, 16–18 June, Co-sponsored by NASA/JPL and DOC/CRC, JPL Publication 97–11, Jet Propulsion Laboratory, Pasadena, CA, pp. 145–152.
ORBCOMM (2012) ORBCOMM system overview, www.m2mconnectivity.com.au/sites/default/files/more-information/System_Overview_Rev_G.pdf (accessed 28 August, 2013).
Orbital (2012) Iridium Next’ Fact Sheet, Orbital Sciences Corporation, www.orbital.com/NewsInfo/Publications/IridiumNEXT_Hosted_Fact.pdf (accessed 28 August, 2013).
Parsons, G.M. and Singh, R. (2006) An ATC Primer: The Future of Communications, Mobile Satellite Ventures.
Richharia, M. (1994) Inmarsat space segment. Infocom 94, Bombay, India, December.

Richharia, M. and Trachtman, E. (2005) Inmarsat's broadband mobile communication system, GLOBECOM Workshop on Advances in Satellite Communications: New Services and Systems, St Louis, MO, 2 December, p. 6.

Richharia, M., Trachtman, E. and Fines, P. (2005a) Broadband global area network air interface evolution. 11th K_a and Broadband Communications Conference and 23rd AIAA International Communications Satellite Systems Conference (ICSSC) 2005, Rome, Italy, September 25–28, p. 12.

Richharia, M., Kaluvala, N., Fines, P., *et al*. (2005b) Aeronautical channel model for broadband L Band satellite communications. ICSSC 05, Rome, Italy, September 26–28, 2005, p. 12.

Rivera, J.J., Trachtman, E. and Richharia, M. (2005) The BGAN extension programme. *ESA Bulletin*, **124**, 62–68.

Schindall, J. (1995) Concept and implementation of the Globalstar mobile satellite system. IMSC'95, Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. A-11–A-16.

Schoen, D.C. and Locke, P.A. (1995) ORBCOMM – initial operations. IMSC 1995, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 397–400.

Trachtman, E. (2006) BGAN, Its extension and evolution. International Workshop for B3G/4G Satellite Communications, Seoul, South Korea, November 23, 2006, http://kosst.or.kr/data/Eyal.pdf (accessed 25 November 2012).

Wiedeman, R.A. and Viterbi, A.J. (1993) The Globalstar mobile satellite system for world-wide personal communications. IMSC 1993, Pasadena, CA, 16–18 June, PL Publication 93-009, pp. 291–296.

12

Mobile Satellite Broadcast Systems

12.1 Introduction

There are a number of similarities between mobile satellite broadcast and mobile satellite service (MSS) technologies, despite a difference in their respective radio regulatory regime and service features. Hence, it was felt that an exposition of the mobile broadcast systems should be a useful addition to the book.

Mobile satellite broadcast receivers and MSS receivers are similar in size and weight, operate in a similar propagation environment and provide service on small terminals to mobile, portable and fixed users. The frequencies used for these services lie in L and S bands where the propagation mechanisms are similar. MSS multicast application bears similarities to regular broadcast service in that the multicast service is used for broadcasts of shared material to targeted user groups (e.g. those with requisite subscription). However, due to a marked difference in the nature of service, mobile satellite broadcast systems belong to the broadcast satellite service (BSS).

The Digital Video Broadcast-Return Channel Satellite + Mobility (DVB-RCS + M) platform supports broadband interactive services on mobile and nomadic terminals. However, the standard targets mobile very small aperture technology (VSAT) applications primarily, although it can be used for interactive broadcasts. We will not consider this system further in this chapter as it has been covered in Chapter 8 and, instead, address L and S band satellite systems for radio, television and multimedia broadcasts to small hand-held, vehicular and nomadic user terminals.

Early in the development of commercial satellite mobile broadcasts dating back to 1990s, the focus was on radio broadcast as evidenced by the first generation mobile satellite broadcasts system – for instance, Sirius-XM Radio Inc. The scope of satellite radio broadcast has since been subsumed in a generic architecture encompassing radio, data, multimedia and television. Similarly, broadcast standards and associated International Telecommunication Union's Radio Telecommunication Sector (ITU-R) recommendations have evolved to keep abreast of user needs and advancements in ground/space technologies. Despite a mixed commercial success of the technology, there is a keen interest in promoting the technology commercially in Europe and elsewhere (Evans and Thompson, 2007).

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

The term *satellite digital radio* (SDR) or *satellite radio* is often used to describe satellite audio broadcasts. In Europe a generic description of such systems is called, satellite digital multimedia broadcast (SDMB), a term which emerged from the European work in Satellite-Universal Mobile Telecommunications System (S-UMTS) (Narenthiran, Karaliopoulos and Evans, 2004).

We also introduce the salient characteristics of multimedia broadcast and multicast service (MBMS) since it can provide an interactive broadcast service within the prerogative of MSS, thus, enabling an MSS to leverage the wide-area advantage of satellite systems.

Beginning with the mobile satellite broadcast technology evolution, the chapter introduces service requirements, system configuration, space segment and the associated transmission technology. This leads to the applicability of Open system Interconnection (OSI) architecture in a broadcast context with examples of prevalent transmission systems and receiver architecture. Digital video broadcast-satellite services to handheld (DVB-SH) architecture is discussed next, as a representative of a modern standard. The chapter concludes with a brief description of MBMS.

12.2 Evolution

Pioneering work in Europe, the USA and under auspice of the ITU in 1980s created sufficient interest in the BSSs (sound), leading to spectrum allocation for the services in WARC-92. Frequency range in between 500 MHz and 3 GHz was considered the most appropriate for the service as it provided a relatively benign propagation environment along with the mature technology status of these bands. Below 500 MHz, spacecraft antenna size becomes prohibitive, and above ~3 GHz the satellite effective isotropic radiated power (EIRP) requirement per channel would be prohibitively high with the available technology. The WARC-92 conference awarded the spectrum in L and S bands to DBS (sound) (see Table 12.1). Some nations applied certain restrictions in the use of these bands until 2007.

Developments in digital systems, in general, and telecommunications/computing/multimedia, in particular, opened interesting opportunities in the broadcast field. For example, it became possible to include features such as multiplexing downloadable data with audio programmes and real-time interactivity through terrestrial or mobile satellite system. The ITU consequently recommended a number of generic requirements for digital audio broadcasting (DAB) services (ITU-R, 1995). A DAB system would be expected to:

- provide a range of receiver signal qualities up to CD quality; and stereophonic, two- or multi-channel sound;
- use state-of-the-art source/channel coding, modulation and digital signal processing techniques;
- be more spectrally and power-efficient than conventional (analogue) schemes in the presence of multipath and shadowing;
- operate with terrestrial systems synergistically through hybrid/mixed satellite-terrestrial systems and maximize commonality in dual-mode receivers used in such systems to benefit from economies of scale;
- provide programme-related facilities such as service identification, programme labelling, programme delivery control, copyright control, etc.;

Table 12.1 Spectrum awarded to DBS (sound) WARC-92

Frequency range	Bandwidth (MHz)	Countries
1452–1492 MHz	40	All countries except those listed below
2310–2360 MHz	50	India and USA
2535–2655	120	Bangladesh, Belarus, China, Japan, Pakistan, Republic of Korea, Russian Federation, Singapore, Singapore, Sri Lanka, Ukraine

- provide value-added services such as business data, paging, graphics, and so on, using a variety of data rates;
- provide coverage for national, regional or international services.

Early in the development of the technology, two systems were evaluated for recommendation by the ITU. These are the European Eureka 147 system, called Digital System A in the ITU recommendation and another system called Digital System B sponsored by the Voice of America and the Jet Propulsion Laboratory of the United States of America. The European approach was based on COFDM (Coherent Orthogonal Frequency Division modulation) over a 1.5 MHz band on the hypothesis that frequency diversity of this order is necessary to combat frequency selective fading and multipath observed in land mobile systems; whereas the US developers used adaptive equalization that allowed transmissions within 50 kHz (mono-phonic frequency modulation (FM) quality) – 200 kHz (CD quality).

Subsequently, a number of companies, notably in the USA, proposed satellite direct sound broadcast systems offering CD quality for both national and international markets. Of these, the Sirius Satellite Radio Inc. and XM Satellite Radio Corporation system after surmounting numerous financial, regulatory and commercial hurdles emerged as viable alternatives but with competition from other media and between each other merged as Sirius-XM Radio, which together continue to provide services in the USA. A venture called 1worldspace folded due to commercial difficulties. XM Satellite Radio system and 1world space system are based on geostationary earth orbit (GEO) space platform whereas Sirius system is a highly elliptical orbit system. Table 12.2 summarizes the technical attributes of these systems (compiled in Richharia and Westbrook, 2010).

Technology advancements resulted in the ITU recognizing at least five systems, discussed later in this chapter. Notably, the broadcast emphasis has shifted from sound-only to multimedia broadcast, including television.

Examples of such modern systems in use include the S-DMB (Satellite – Digital Multimedia Broadcast) system in South Korea (Hirakawa, Sato and Kikuchi, 2006) and a system in China based on a Chinese standard known as CMMB (China Mobile Multimedia Broadcast) (Burger *et al*., 2007). Various system proposals are under consideration in Europe – these include Worldspace (enhanced version of digital radio system with repeaters), EUROPA-MAX (three-satellite, HEO SDMB system), Ondas (three-satellite, highly elliptical orbit (HEO) digital radio system), NEMO (GEO mobile multimedia system with terrestrial repeaters), SOLARIS (GEO SDMB system), Unlimited Mobile TV (SDMB system) (Evans and Thompson, 2007).

Table 12.2 A comparative assessment of a few commercially deployed satellite radio systems

	1worldspace	XM Satellite Radio Inc.	Sirius Satellite Radio Inc.
Orbit/number of satellites	Geostationary/2	Geostationary/2	24 H, highly elliptical (Tundra)/3
Frequency band (MHz) up-link/down-link	7025–7075/1452–1492	7050–7075/2320–2324 and 2328.5–2332.5	7060–7072.5/2332.5–2336.5 and 2341–2345
Target service areas	Africa, Asia, Middle-East, Far-East and Europe (planned)	CONUS and Canada	CONUS and Canada
Location	21°E and 105°E	85°W and 115°W	Time variant, moving around 100° W
Service	Radio and data broadcasts to fixed and portable receivers. Not optimized for mobile reception.	Radio and datacast to fixed, portable and primarily mobile receiver	Radio and datacast to fixed, portable and primarily mobile receivers
Modulation/Access: Satellite [Terrestrial]	QPSK/TDM [No repeaters]	QPSK/TDM [COFDM]	QPSK/TDM [COFDM]
Channel coding	Reed–Solomon (RS) – outer and rate ½ convolutional – inner	Reed–Solomon –outer and rate ½ convolutional –inner	Reed–Solomon –outer and rate ½ convolutional –inner
Source coding	MPEG Layer III	Propriety	Propriety
Transmission rate Mbps – before channel coding	1.536 per TDM; 2 TDM per spot beam (3 spot beams per satellite)	4.0	4.4
Number of CD quality (64 kbps) channels	48 per spot beam	50	50
Receiver	Portable and fixed	Plug and play, vehicular, desktop, personal portable and cockpit mounted	Plug and play, integrated with vehicle, desktop, personal portable (add-on: WiFi connectivity for streaming, storage)
Approximate number of terrestrial repeaters	Not used (planned)	1500 in 60 markets	150 in 45 markets
Number of RF channels	6	6	3

CONUS: Contiguous United States.
(Source: Richharia and Westbrook, 2010. Reproduced with permission from John Wiley & Sons, Ltd.)

12.3 Mobile Broadcast System Requirements

12.3.1 Service Requirements

The requirements of a mobile satellite broadcast service are governed and constrained by the operating environment (i.e. fixed, nomadic, urban, rural, etc.), receiver cost and form, service charge, service type and the anticipated usage of such a service. As such, these requirements differ from the conventional broadcasts. For example, the size of the viewing screen (of a mobile television service) is considerably smaller, which influences the transmission format and radio bandwidth and the user may be interested in a different type of content than the home user. Transmissions for display on small screens require a considerably lower bandwidth (e.g. 200 kHz) than the conventional satellite broadcasts (several megahertz).

Ideally, the broadcast system should be capable of the following service provisions (ITU-R, 2012):

- high-quality content delivery;
- flexible configuration of each service, that is, audio, video, ancillary and auxiliary data;
- access to content and services, possibly controlled through conditional access, protocols or other content protection mechanisms;
- seamless access throughout the broadcast network;
- fast identification and selection of content and services;
- mechanisms to minimize power consumption and physical size of the receivers;
- stable and reliable service coverage for receivers in target reception environments;
- interactivity, (e.g. interactive content);
- efficient and reliable delivery mechanisms;
- technical aspects that enable interoperability between broadcast and telecommunication networks (e.g. content format);
- service quality comparable to fixed reception;
- countermeasures for multipath, Doppler and vehicular motion.

In practice, the requirements would be conditioned by commercial and technical factors. A basic system would comprise an Electronic Service Guide (ESG) to announce the portfolio of available contents with its associated location (e.g. local, wide-area) and a hand-over mechanism to ensure seamless coverage in situations when the user moves from a satellite-only coverage to terrestrial retransmissions zone. A more expansive set would attempt to emulate features of an advanced terrestrial broadcast system like digital video broadcast-handheld (DVB-H) (ITU-R, 2011). Such a system has provision to support:

- Mobile TV programme broadcasts, possibly with associated auxiliary data (e.g. links to the service provider's web pages, video clips, sound tracks, games, etc.);
- Enhanced mobile TV to provide interactivity to include features such as online TV shopping, chat, gaming and quiz plus voting, and others, where the interaction could be achieved through a mobile communication system;

- Scheduled download facility (announced in the ESG) to provide audio-visual content or executable software;
- Provision for (online) service purchase, service access and content protection, including pay-per-view purchase;
- Roaming facility that allows contents access when outside the home network;
- High quality of service to ensure compatibility with conventional terrestrial systems and provide error-free downloads of data files in presence of Doppler, multipath, shadowing and interference;
- System and receiver architecture that facilitate low power consumption at the receiver;

12.3.2 Receiver Types

The features, characteristics and requirements of mobile broadcast receiver depend on their operating environment, the addressed clientele and the application support, that is sound, television and/or multimedia.

Handheld receivers are constrained by the size, weight, power drainage and processing limitations, and if used for television, then the screen size and resolution. Figure 12.1(a) and (b) illustrates a basic and an enhanced version of a hand-held receiver.

Portable receivers are less constrained than handheld units in battery and processing power. Similarly, vehicular receivers have lower physical and power constraints but must include countermeasures to compensate the relatively higher Doppler and multipath fluctuations with an external tracking or non-tracking antenna. Receiver architectural issues, including an example, are discussed later in this chapter.

(a) (b)

Figure 12.1 (a) An example of a basic hand-held user terminal. (b) An example of an enhanced hand-held user terminal. (Both parts source: ITU-R, 2011. Reproduced by permission of ITU.)

12.4 System Configuration

Mobile satellite broadcast systems, when operating standalone, are suitable for remote area coverage with little terrestrial broadcast infrastructure and/or to serve a niche market (e.g. hi-fidelity, advertisement-free dedicated music channels for motorists or motels). Alternatively, they can be overlaid on an existing terrestrial broadcast system to fill coverage gaps – as in the Eureka 147 system (addressed later). Conversely, a terrestrial system can augment satellite service by retransmission in areas of high user population density where satellite signals suffer extensive blockage as, for example in a dense city centre like the Sirius-XM systems. Where feasible, satellite and terrestrial signals are combined at a receiver to improve signal quality whereas in satellite coverage gaps, the terrestrial gap-filler provides the service.

Figure 12.2 illustrates the concept of such a *hybrid architecture* as proposed in Digital system D_H, one of the recommended system of the ITU-R (ITU-R, 2001). This system, explained later, provides a powerful countermeasure against shadowing and multipath by utilizing satellite diversity and a hybrid architecture where the coverage gaps are filled by terrestrial transmission using a multi-carrier modulation (MCM). There are three distinct

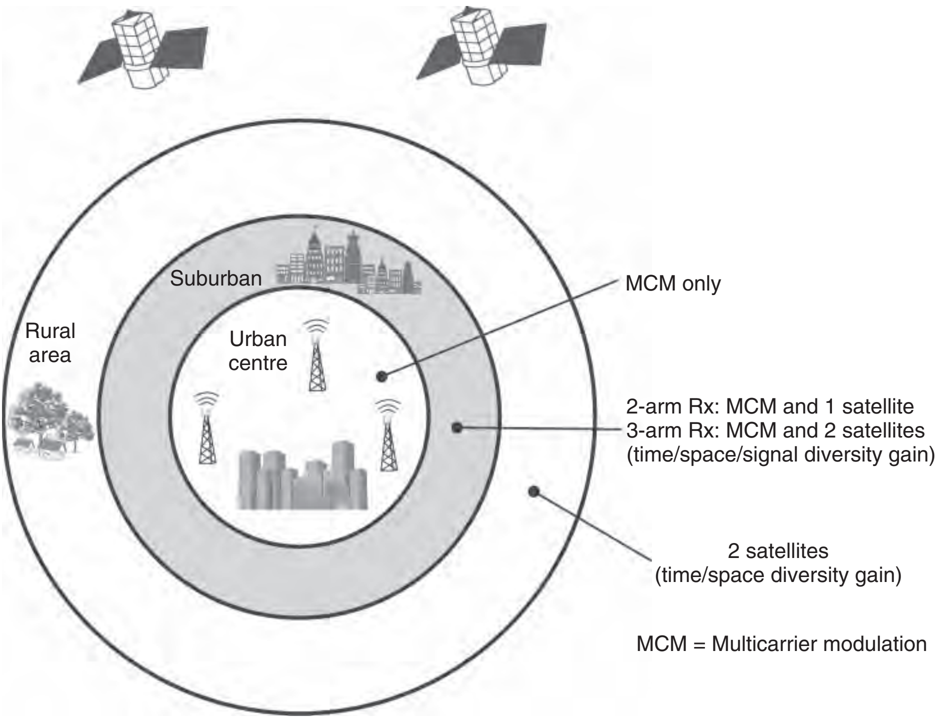


Figure 12.2 A hybrid architecture of ITU-R recommendation Digital System, D_H, illustrating various reception zones and two-satellite diversity. (Source: ITU-R, 2001. Reproduced by permission of ITU.)

reception zones – satellite-only in remote areas, one or two satellite signals together with terrestrial MCM signals in suburban regions and MCM-only signals in dense urban regions.

In a *mixed mode*, the transmissions of satellite and terrestrial signals are in the same format at different frequencies in the same frequency band to simplify dual-mode receiver design. Here, when both terrestrial and satellite transmissions are received, the user either selects the more appropriate (e.g. the one which offers a better quality, or, in response to a service announcement), alternatively the intelligence can be built into the receiver for a seamless handover.

A majority of proposed and operational mobile broadcast systems utilize the hybrid approach. Figure 12.3 shows an example of a hybrid system configuration as proposed in ITU recommendation Digital System E (outlined later in the section). The system comprises a content provider's facility linked to an earth station that processes the signals and up-links them to a satellite at 14 GHz. The signal is received by the satellite, down-converted to K_u band (11 GHz) and S band (2.6 GHz) and broadcasted to a widely dispersed user community for direct reception. Three classes of user terminals are supported – hand-held, vehicular and fixed. To extend services where satellite signals are blocked extensively, various types of terrestrial re-transmitters (gap-fillers) are proposed. *Frequency-conversion gap-fillers* receive 11 GHz signals from the satellite and retransmit the signals at high power after conversion to S band to cover places such as urban centres. These high power transmitters require a different frequency for reception and transmission in order to avoid a regenerative feedback loop, which tends to occur when signals are received and retransmitted at high power on the same frequencies. The *direct-amplifying gap-fillers* are low power transmitters for relatively small areas; and *spotlight gap-fillers* are low power direct amplifying transmitters to service city canyons, streets and other narrow confined areas.

12.5 Space Segment

One of the prime considerations in selection of the space segment for a satellite mobile broadcast service is the Earth-Space geometry that has a profound impact on the link reliability. In Chapter 3, we learnt that fading loss is inversely proportion to the elevation angle. In this respect, geostationary satellite systems are better suited for the regions where elevation angles exceed around 20° and hence broadcast services at mid and high latitude regions become unreliable (see Figure 2.7a). A technique to mitigate the problem is to use a highly elliptical satellite orbit, for example Molniya or Tundra orbits (see Section 2.2) that provide a quasi-stationary high elevation coverage at mid-high latitudes. This concept has been studied in the European Space Agency's Archimedes project, which concluded that the Tundra orbit is a promising candidate for a European DAB, as satellites in this class of orbit appear at elevation angles of ~55–90° within Europe resulting in reduced link margin and allowing simple upward pointing antennas on vehicles (Galligan, 1989). An added advantage of the configuration is that the satellite eclipse occurs near the perigee when satellites are not in service; however, at least two satellites phased 180° apart would be necessary to provide continuous coverage (at $> 52°$ elevation). Since the configuration requires handover between satellites, the system architecture gets more complex than an architecture using a geostationary orbit. The Molniya orbit was discarded in this particular study as radiation doses to satellites in such orbits are higher causing accelerated degradation to satellite equipment and altitude excursion during operations is large causing large variations to the mean received signal; moreover launch from the European launcher (Ariane) is difficult and

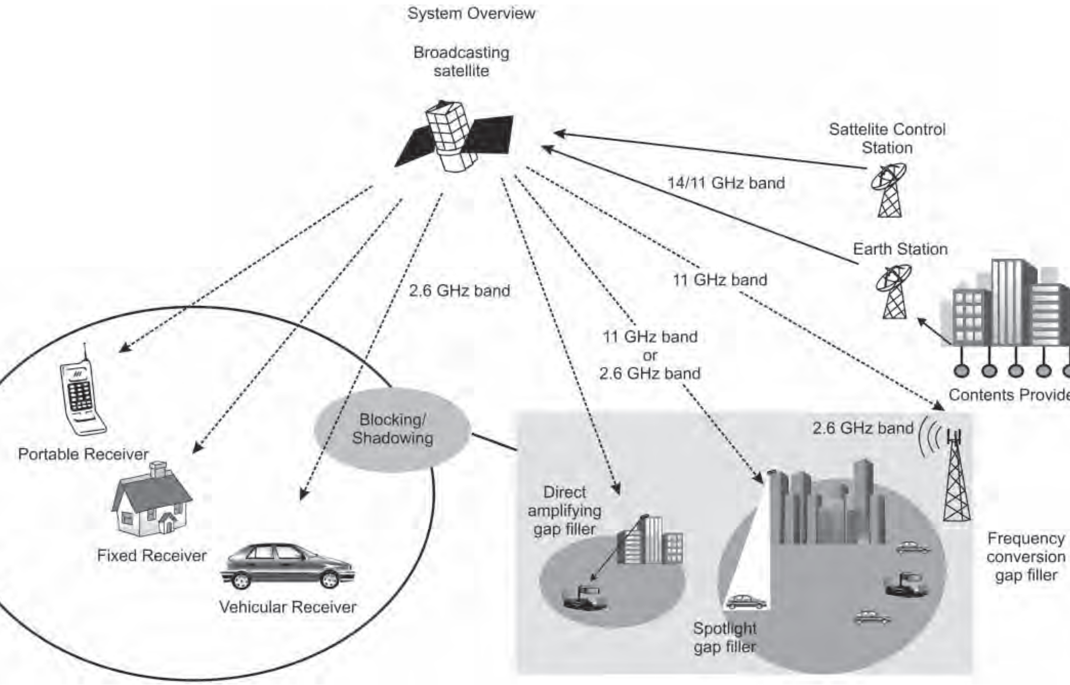


Figure 12.3 A mobile broadcast system configuration. (Source: ITU-R, 2001. Reproduced by permission of ITU.)

three (rather than two) satellites would be necessary. Assuming a single beam antenna and an omni-directional receiving antenna, compared to the geostationary system, the proposed Archimedes system required a considerably lower power (up to 30 times lower) for sound broadcasts than required by an equivalent geostationary channel (Giovagnoli, 1989). Such an architecture is used in the Sirius XM Radio Inc. (Sirius component) satellite system that provides satellite radio services in the American continent (see Table 12.2).

Other approaches to mitigate the propagation impairments are spatial diversity by multiple satellite visibility and time diversity by transmission repetition sufficiently distanced in time. The respective premises are that the probability of at least one of the satellites being in a favourable direction of the user increases and the probability of a fade lasting more than a certain duration reduces.

The space segment of the Sirius system comprises three satellites in a highly elliptical Tundra orbit, with a period of 23 h and 56 min, two of which are visible over the US with an average elevation angle of about 60° for ∼16 h each day. This configuration therefore provides a high elevation angle visibility at moderate-high latitudes of US and additionally provides diversity advantage. The improvement in link margin reduced the requirements of the number of the complementary terrestrial transmitters quite considerably when compared to an equivalent geostationary satellite system (see Table 12.2) (Layer, 2001).

Several systems have been developed and standardized based on the previously discussed requirements and solutions. ITU-R recommends up to five systems that are introduced later in the chapter.

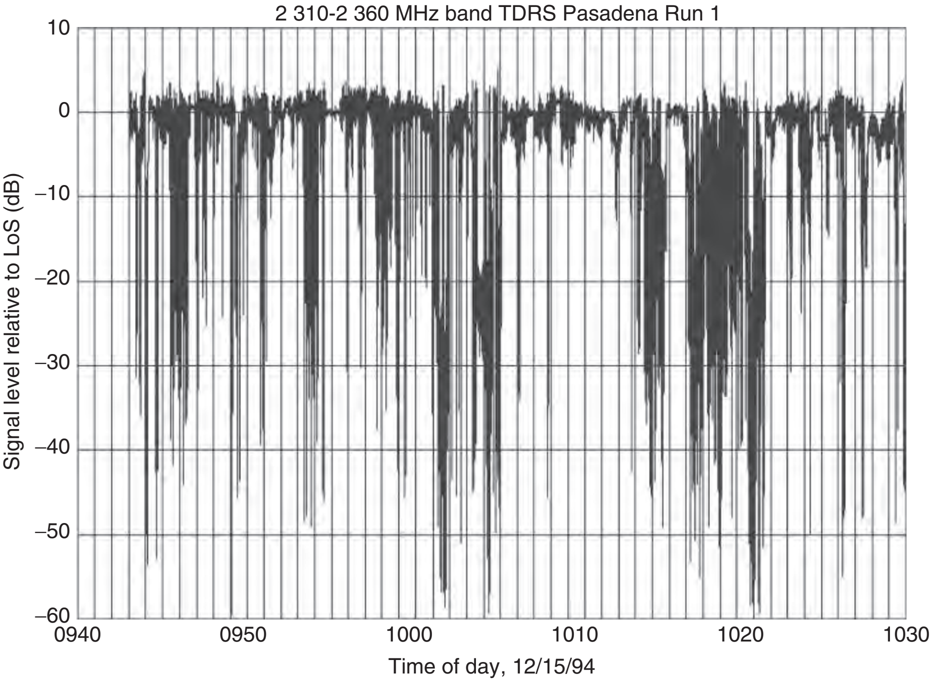
12.6 Transmission Technology

While satellite systems provide wide-area coverage spanning a continent, the EIRP available from satellites is insufficient to provide reliable coverage in cities and populated areas due to excessive shadowing. It was mentioned earlier that the trend is therefore to utilize a *hybrid architecture* comprising satellite and complemented by a terrestrial infrastructure that retransmits the satellite signals. These retransmitters, known as complementary ground segment (CGS) or Auxiliary terrestrial transmission (ATC), have become the *de-facto* norm in satellite mobile broadcast system. The architecture also allows considerable commonality in the satellite and terrestrial receiver to provide significant economies of scale to the satellite component (SC).

We noted earlier that in the *mixed-mode* systems, satellite and terrestrial transmissions are transmitted at separate frequencies but the same band while maintaining an identical transmission format to provide large commonalities in the receiver. This arrangement allows reception of both transmissions by conventional receivers through a front-end converter/translator for satellite reception.

An understanding of propagation conditions likely to be encountered in the target service area is essential in assessing the service reliability and coverage. It has been established that the difference in attenuation loss between the L and S bands is relatively marginal (e.g., Butt, Evans and Richharia, 1992). Hence the modelling approach used in the L band applies generally in the S band with minor modifications.

Consider the model used in the development of digital system B (described later). Figure 12.4(a) and (b) respectively show a sample time plot and the corresponding statistics of a narrowband measurement conducted in the Pasadena, California area over a route that



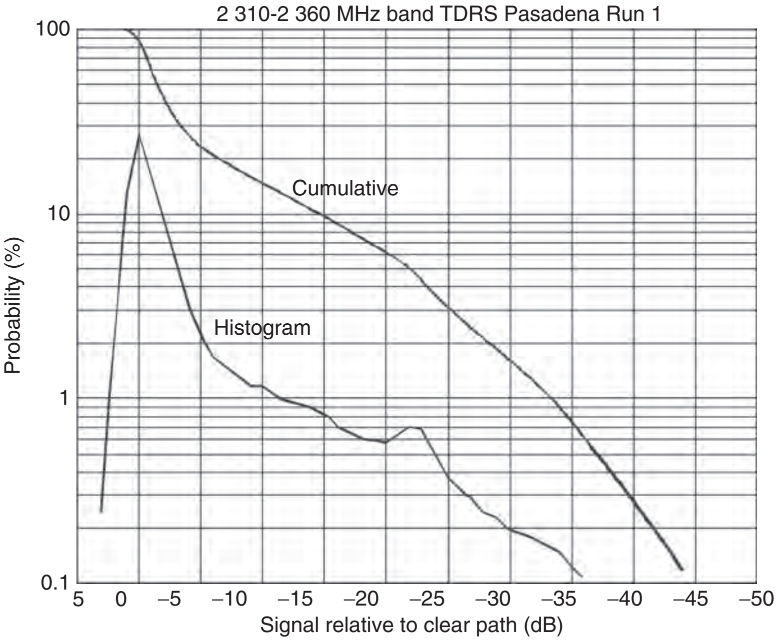


Figure 12.4 (a) A time plot of received signal on a mobile in Pasadena, USA for a 45 min run in a street with a mix of local surroundings. (b) The statistics of the time plot illustrated in part (a). (Both parts source: ITU-R, 2001. Reproduced by permission of ITU.)

included open, moderately shadowed and severely shadowed segments (ITU-R, 2001). Note the wide variation in signal level attenuating down to 60 dB demonstrating the necessity of a huge propagation margin. For example, the reliability of only 90% is achievable with a link margin of the order of 15 dB.

In a hybrid system the propagation mechanism between terrestrial and satellite systems differ due to a variety of factors. Elevation angle dependence is one such factor – whereas terrestrial transmissions are primarily affected by terrain (elevation angle ~0) satellite transmissions exhibit environment-dependent elevation angle dependence. Terrestrial transmissions are not power limited as are satellite systems. While frequency selective fades are unlikely to affect satellite broadcast systems (coherence bandwidth of several megahertz have been reported from measurements) , this is not always true for terrestrial systems where the multipath echoes could cause inter-symbol interference necessitating equalization.

Yet another consideration, particularly applicable to satellite-only broadcast systems, is the necessity to estimate building penetration loss as some of the portables are likely to be used within buildings. Data collected through ATS-6 transmissions at 860, 1550 and 2569 MHz within various types of house demonstrate that penetration loss within buildings depends on frequency, receive antenna polarization, the material used in house construction (e.g. wood siding or brick veneer), thermal insulation within the house and proximity of rooms to an outside wall. It was noted that penetration loss is independent of the elevation angle down to 5° when there is a clear line of sight to the satellite, and furthermore that the loss was normally distributed with a standard deviation of 3 dB within each type of house (Wells, 1977; Miller, 1985).

Forward error correction (FEC) coding and diversity methods are an essential part of mobile satellite links. For example, half rate FEC convolution code of constraint length 7 with a Viterbi maximum likelihood soft decision decoding algorithm can offer advantage of the order of 3–6 dB over uncoded QPSK with coherent detection in a memory-less Rayleigh fading channel for a bit error rate (BER) of 10^{-5}.

However, when the signal received at a receiver is in a deep fade, coding gain is no longer realizable, as the signal will be irretrievable. The probability of signal drop-outs by such deep fades is reduced by frequency, spatial or time diversity.

In frequency diversity, the same information is transmitted on a number of carriers, separated by more than the coherence bandwidth of the channel. Thus it is possible to retrieve information from the carriers that are unfaded at a given time. Measurements performed on simulated satellite paths demonstrate that the coherence bandwidth is narrower in urban environments than in rural environments and therefore carrier spacing has to be larger in rural environments. Frequency diversity can be achieved by a modulation scheme, called the COFDM scheme, which was introduced Chapter 4.

Spatial diversity can be used by transmitting (or receiving) from multiple antennas separated far enough from each other so that the signal variations are uncorrelated at the receiver. In the XM Radio system diversity is achieved by transmitting the same programme from two geostationary satellite orbit (GSO) satellites separated by 30° longitude. Signals from each transmission are combined to form a composite signal that has a lower fade than individual components. Measurements have shown that when multiple antennas are installed at the receiver, diversity gain is possible even when antenna separation is of the order of half wavelength or greater (Hess, 1980). Signals from diverse paths can be combined by techniques such as maximum ratio combining, which can be implemented by introducing phase-lock-loop in each path to obtain phase coherence before amplitude weighting and combining.

Although not matured at present, investigations reveal advantages in application of multiple input multiple output (MIMO) to satellite broadcast radio links – various approaches have been investigated. A dual-polarized transmit antennas fed by signals on each polarization provides a possible 2×2 configuration for satellite systems (see Chapter 14). Another possibility would be to utilize transmissions from two different satellites; in another approach, transmissions from satellite and terrestrial antennas constitute a 2×2 MIMO matrix, where the terrestrial antenna acts as the second antenna. This type of configuration can provide both diversity and multiplexing gain. Performance improvements in a DVB-SH based broadcast system were reported in a theoretical study dealing with application of MIMO in various configurations (Pérez-Neira *et al*., Pérez-Neira et al., 2008).

Time diversity is also possible in digital systems by interleaving/de-interleaving. The interleaving depth is adjusted so that a fade event remains uncorrelated between adjacent symbols. Note, however, that the extent of the depth is limited (constrained by the block size and the permissible delay in message transfer); the scheme can nevertheless be utilized to contend fades of several seconds such as caused by a receiver mounted on a moving mobile in clear or partially shadowed conditions. This scheme is best utilized in conjunction with other fade mitigation methods such as repeat transmissions mentioned earlier. L band measurements taken on a highway from Marecs transmissions in Europe demonstrate an advantage of over 7 dB at 99% probability level in the receive threshold when time difference is of 4 s. S band measurements conducted in USA over a test course of 45 min in a mixed environment demonstrate that most of the improvements occur for retransmission delay of 4 s (ITU-R, 2001).

Because of increasing spectrum demands, considerable effort is being directed towards a reduction in the bit rate of sound coders and the transmission format. A number of efficient source coding techniques have been developed with sub-band coding or transform coding algorithms to compress the bit rate by a factor of up to 32. For general audio application a technique known as perceptual encoding combined with lossless data compression technique provides an effective solution.

Video coding is based on removing the inherent redundancy between picture frames. Instead of sending the full luma and chroma signals only the frame–frame difference of each is transmitted, and while encoding the difference, abandoning the insignificant high spatial frequency components. Motion Picture Expert Group (MPEG) produces and continues to improve video and associated audio encoding standards.

MPEG-2 audio layer II (MP2 – also known as MUSICAM) encoding is used in many DAB system and is included in a digital video broadcasting standard. MP2 has an intermediate complexity with target bit rates of 128 kbps. MPEG audio layer III (MP3) encoding provides higher compression that increases the computational complexity. MP3 is in wide use for a number of applications such as personal music systems including digital satellite broadcast, as used in 1worldspace satellite audio broadcast system (Campanella, 1996), improving a compression of CD-quality sound by a factor of 12.

The bit rate of compressed standard video channel using MPEG-2 algorithm is about 3–15 Mbps, which equates to a compression in the range 10–50. MPEG-4 Advanced Video Coding (AVC) provides a greater compression ratio allowing a much higher transmission efficiency for High Definition Television (HDTV) transmissions over satellite. Hence it is widely used for HDTV transmissions over satellite.

The systems part of MPEG specifications provides rules to combine data, audio and video streams into a single stream for transmissions or storage. Such packetized elementary streams (PESs) streams are combined with others to constitute transport streams (TSs) or programme streams (PS). The PS consists of a single multiplexed stream ready for transmission over satellite or other transmission medium.

The reader interested in a concise summary of audio and video encoding and the MPEG transmission technique can refer to vast literature on the subject (e.g. Richharia and Westbrook, 2010).

12.7 OSI Architecture in a Broadcast Context

ITU-R recommendation BT.807 presents the applicability of OSI architecture as described in ISO/IEC International Standard 7498-1 in a data broadcast context (ITU-R, ITU-R, 1992). The context facilitates the description and introduction of integrated services digital broadcasting (ISDB) to broadcast audio, data and other types of information to mobile or fixed receivers. The OSI broadcast context is presented in Table 12.3 (ITU-R, ITU-R, 1992).

The *Physical layer* (Layer 1) relates to the electrical transmission of the data signal (e.g. modulation, bit rate pulse shaping and physical details of connectors). The *Data link layer* (Layer 2) deals with an individual link to include logical functions of data transmission (e.g. frame synchronization, error control procedures, data formatting and link-access). The *Network layer* (Layer 3) includes logical functions relates to multiplexing, demultiplexing and error control of data packets belonging to different communication flows (e.g. data channel addressing, data packet sequencing). It deals with the passage of data in a network. The process involved, when data passes through a broadcast network is as if data is passing over a single link when broadcasting to individual subscribers (i.e. excluding a distribution channels such as cable head end). The *transport layer* (Layer 4) is responsible for arranging the data for secure transfer by, say, segmenting data into groups, delivering them for transmission to lower layers and at the receiver reconstituting the groups and arranging them in a proper sequence. In data broadcasting, the Layer 4 protocol either operates in connectionless mode (i.e. no reverse path), or uses a different type of network for the 'reverse path' Layer 3 passes the contents of the frames, or packets to Layer 4. Although distinction between

Table 12.3 The OSI broadcast context

Layer	Principal function
7 Application	Practical use of information at application level by the end user.
6 Presentation	Conversion for presentation of information
5 Session	Selection of and access to information/data
4 Transport	Identification and grouping of data
3 Network	Identification of logical channel
2 Data link	Linkage (formatting) with logical transmission unit
1 Physical	Physical (radio) transmission

(Adapted from ITU, 1992.)

layers 3 and 4 may only be theoretical in a broadcast receiver that implements all the layers, it becomes significant when broadcast data are intended for onward transmission through a different type of network. The *session layer* (Layer 5) performs functions for the user to gain access to services (e.g. access control and service identification). This accounts for the situation when multiple services are carried over the same network and services can include additional or optional components. The *presentation layer* (Layer 6) deals with the presentation of information for each application, for example video and sound. The *application layer* (Layer 7) deals with the end-user application such as audio, video, stock market data, telemusic, etc.

Table 12.4 illustrates the interpretation of the OSI layer applied to Digital System A (ITU-R, 1995).

The *application layer* applies to the facilities and audio quality that the system provides, which broadcasters can offer to their listeners. The layer also provides different transmission modes depending on the broadcaster's requirements. The system A broadcasts multiplexed data conveying audio programme data, programme-associated data, multiplex configuration information, service information (SI) and general data services. The facilities offered to the listeners include the programme itself, text display of information related to SI, conditional access and general data service such as traffic message channel. The broadcasters can select

Table 12.4 Interpretation of OSI layer applied to digital system A

Name of layer	Features
Application layer	System facilities Audio quality Transmission modes
Presentation layer	Audio encoding and decoding Audio presentation Service information
Session layer	Programme selection Conditional access
Transport layer	Programme services Main service multiplex Ancillary data Association of data
Network layer	ISO audio frames Programme associated data
Data link layer	Transmission frames Synchronization
Physical layer	Energy dispersal Convolutional encoding Time interleaving Frequency interleaving Modulation by 4-DPSK OFDM Radio transmission

DPSK: differential PSK.
(Adapted from ITU, 2001.)

the desired signal quality (very high quality, speech quality, etc.) within the coverage area for satellite and complementary terrestrial transmissions. The system offers three transmission modes that allow the use of a wide range of transmitting frequencies up to 3 GHz with ability to counter Doppler spread and delay spread for mobile reception. Modes II and III pertain to satellite systems. Mode II is suitable to support hybrid satellite/terrestrial transmission up to 1.5 GHz; Mode III can support satellite and complementary terrestrial transmission at all frequencies up to 3 GHz.

The *presentation layer* converts the incoming signals to an appropriate format for presentation of the broadcast information. It involves audio encoding using ISO/IEC MPEG-Audio Layer II algorithms, also known as the MUSICAM system. The encoding can provide a wide range of bit rates to suit broadcasters' requirements. The decoding operation is relatively simple requiring only demultiplexing, expanding and inverse-filtering operations. The presentation of audio signal involves providing monophonic or stereophonic audio reproduction depending on the transmission. Other presentation features may include linking a programme to different language, and dynamic control of signal at the receiver through transmission of appropriate signals. Various elements of SI may be displayed at the receiver – basic programme label, time and date, programme information (e.g. the names of performers), language, programme type (e.g. news, sport, music, etc.), transmitter identifier and so on.

The *session layer* deals with the programme selection and access to broadcast transmissions. The system transmits multiplex configuration information on a fast information channel (FIC) to expedite the delivery. The user selects the desired programme, for example on the basis of available textual information. There is provision for the user to switch to an alternative programme source in case the system's reception becomes unacceptable and switch back automatically when the reception improves. The layer provides for both synchronization and control of conditional access.

The *transport* layer deals with identifying, grouping and multiplexing programme services, which typically comprises an audio service component optionally with additional audio and/or data service components. The multiplex capacity can be dedicated or shared between operators. Data of each service are individually convolutional encoded, time-interleaved and fed to the main service multiplexer where they are framed in 24 ms segments. The combined bit-stream output gives a net bit rate ranging over ~0.8 – 1.7 Mbitps, which occupies 1.5 MHz RF bandwidth. General data can be included subject to available capacity as an unstructured but synchronized stream to the multiplexer in multiples of 8 kbps. Ancillary data can be carried within the system multiplex and be introduced at different points depending on available capacity, that is at the FIC within each audio channel or as a separate service. Further, the transport layer provides description of the current and future content of the main service channel (MSC), essential service information and information linked to each sound programme and transmitter network for internal use by broadcasters.

The *network layer* identifies groups of data as programmes. Each audio frame carries data of 24 ms and programme associated data (PAD) for a single programme in a ISO/IEC MPEG-Audio Layer II (standardized) format so as to facilitate the use of a reciprocal MPEG decoder in the receiver.

The role of *data link layer* is to provide a means for receiver synchronization. Each frame is structured such as to aid synchronization, automatic gain control (AGC), automatic frequency control (AFC) and phase reference functions at the receiver in addition to providing data for the MSCs. The frame duration is either 24 ms or 96 ms, depending on the transmission mode; each audio service within the MSC is allotted a fixed time slot in the frame.

The *physical layer* deals with radio transmission (i.e. the modulation scheme, associated error protection and RF carrier transport). It deals with energy dispersal that is achieved by data scrambling of each channel. Each data channel within the multiplex is convolutionally encoded using a constraint length of 7. Greater protection is given to some source-encoded bits than others for audio channels. Depending on the desired protection the average code rates vary from one third to three quarters with a provision to vary code rates for each audio source. The code rates of general data are derived from a selection.

The convolutionally encoded data are interleaved up to a depth of 16 frames. Furthermore, to contend with frequency selective fading, the data are frequency-interleaved by distributing the digital bit-stream amongst the carriers, such that successive source samples are positioned apart in frequency to reduce the probability of being affected simultaneously by a frequency-selective fade event. The data are modulated as 4-DPSK (Differential Phase Shift Keying) OFDM that provides a robust performance on mobile, portable and fixed receivers in a multipath environment.

Figure 12.5 illustrates a conceptual schematic of the transmission section of the system in accordance to the preceding explanation of the system (hence some repetition is inevitable). The audio coder uses an MPEG-Audio layer II algorithm. The sub-band coding section of the coder uses the MUSICAM system, giving bit rates of 32, 48, 56, 64, 80, 96, 112, 128, 160 and 192 kbps monophonic channels and 2 kbps of programme-related data. General and auxiliary data services can be combined with the audio. The data are multiplexed into a frame, which consists of a synchronization channel, a FIC that has multiplex-related information, service information and other general/auxiliary data and a number of programme service channels. The frame duration is 24 ms for mode 3 transmission, best suited for satellite and complementary terrestrial transmissions up to 3 GHz. There are, in all, three modes with different frame duration, each suited to a specific propagation environment. As explained, each channel is convolution coded with a constraint length of 7; the audio channels have a variable code rate at an average coding rate between one third and three quarters; and the FIC is coded at a constant rate of one third. The convolution encoded data is interleaved to a depth of 16. The coded data is transmitted using a COFDM technique. The occupied RF bandwidth of the channel is 1.5 MHz. Each carrier of the COFDM signal is modulated by a differentially encoded PSK.

A digital system A receiver consists of a front-end responsible for receiving and converting the signals to an intermediate frequency, a demodulator, de-interleaver, convolution decoder, demultiplexer and voice/data decoder with a suitable user interface.

12.8 Prevalent Transmission Systems

A number of transmission schemes have been studied, experimented and many have been submitted to the ITU-R and other international bodies such as European Telecommunications Standard Institute (ETSI) for Recommendation and standardization. ITU-R recommends five transmission schemes respectively called digital systems A, B, D_S, D_H and E for satellite digital sound broadcasting services (BSs) to vehicular, portable and fixed receivers

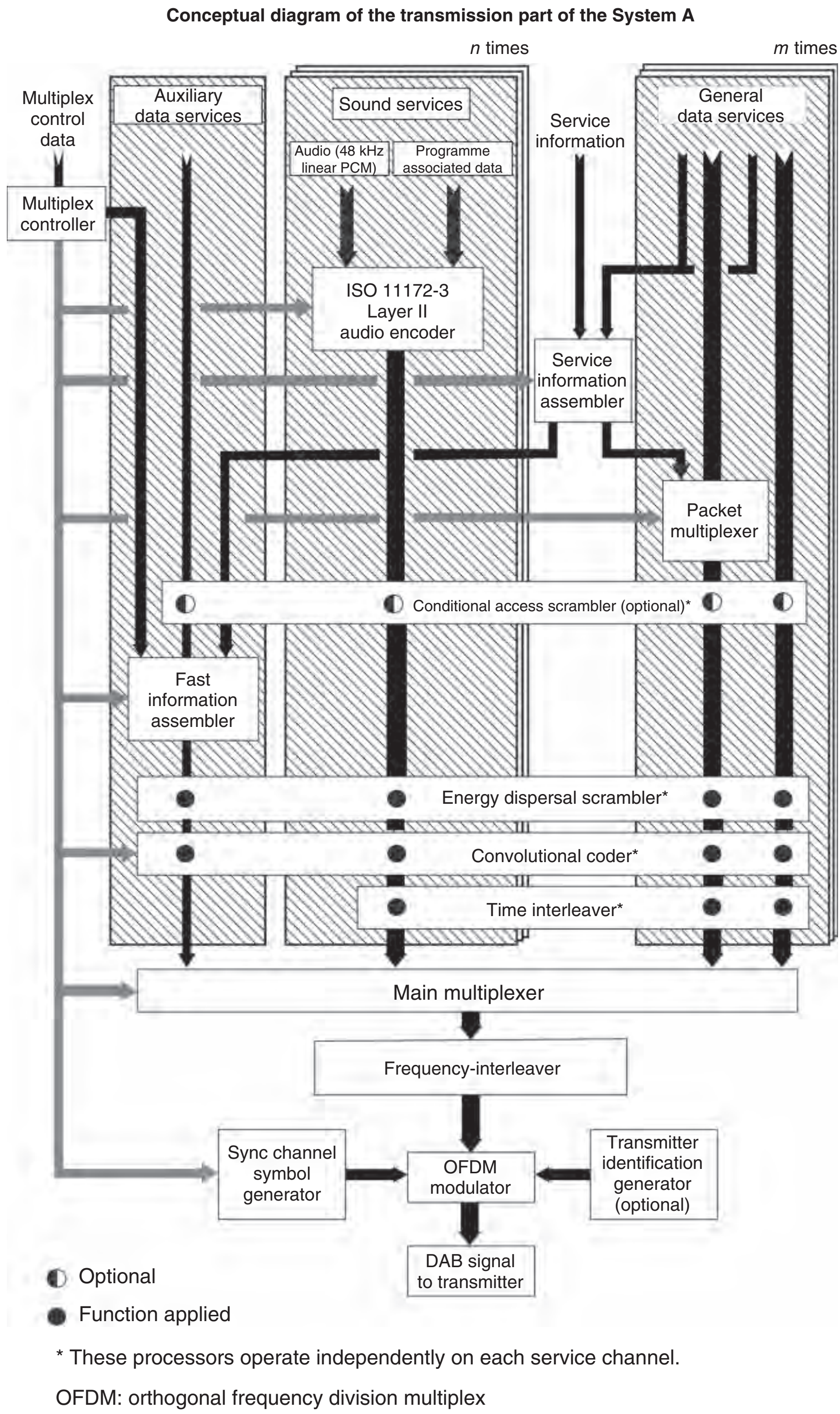


Figure 12.5 A conceptual transmission block schematic of a digital system A. (Source: ITU-R, 2001. Reproduced by permission of ITU.)

in the frequency range 1400–2700 MHz (ITU-R, 2001). These systems include the capability of data/multimedia broadcasts. System A was described in some detail in Section 12.7.

- Digital System A
  - Known as Eureka 147 DAB system, it aims to promote use of a low-cost receiver;
  - Standardized by ETSI as ETS 300 401 for BSS (sound)/broadcasting service (sound) to vehicular, portable and fixed receivers;
  - Target user terminals include vehicular, portable and fixed with low gain omnidirectional receive antennas;
  - Utilizes complementary terrestrial transmitters to improve availability in urban situations and provide EIRP flexibility to space segment;
  - Offers various levels of sound quality, comparable to recorded media at the upper end;
  - Offers data services and different levels of conditional access;
  - Includes capability of re-arranging services dynamically within the multiplexed stream.
- Digital System B
  - Aims to improve satellite efficiency;
  - Receivers reject multipath by adaptive equalizers;
  - Simple modulation scheme (offset quadrature phase shift keying, OQPSK) simplifies system design and facilitates development of low-cost receivers.
- Digital System D_S
  - A satellite-only system, it is also known as WorldSpace system (used by 1worldspace);
  - Provides digital audio and data broadcasts;
  - Optimized for 1452–1492 MHz band through a coherent TDM/QPSK scheme, concatenated block and convolutional FEC and linear amplification;
  - Provides a flexible multiplex of digitized audio sources transmitted over the TDM carrier;
  - Receivers use digital large-scale integrated circuit to achieve low-cost and high-quality performance.
- Digital System D_H
  - It is the hybrid (satellite-terrestrial) version of WorldSpace system;
  - Provides digital audio and data broadcast;
  - Supports inexpensive receivers;
  - Satellite delivery enhances features of Digital System D_S to improve reception in areas partially shadowed by trees;
  - Enhancements include fast QPSK phase ambiguity recovery, early-late time diversity and maximum likelihood combination of early-late time diversity signals;
  - Terrestrial delivery system component is based on OFDM technique.
- Digital System E
  - Also known as the ARIB (Association of Radio Industries and Businesses) system;
  - Utilizes hybrid satellite-complementary terrestrial on-channel repeater architecture;
  - Provides digital audio and multimedia data;
  - Optimizes performance for both satellite and terrestrial components in the 2630–2655 MHz band;

- Utilizes code division multiplex (CDM), based on QPSK modulation with concatenated block and convolutional error correcting coding;
- Promotes low-cost receiver production with high-quality performance.

Table 12.5 compares characteristics of the ITU-R's recommended systems on the basis of the ITU recommendation BO.789 (ITU-R, 1995), which specifies the necessary technical and operating characteristics for digital sound broadcasting systems to vehicular, portable and fixed receivers for satellite delivery. Generally all the systems (except system B) are based around the MPEG-2 transmission system and support a range of audio quality, adjusted as required. In order to contend shadowing and multipath – all the systems, except system D_s, support terrestrial gap-fillers with a provision of combining multipath signals, implying diversity advantage. Systems A, B and E support terrestrial service in the same band with the same architecture for both satellite and terrestrial receivers with variable protection through punctured convolution coding scheme in order to provide reliable services throughout the service area and can provide mixed-mode operation. All the systems utilize an OSI compliant multiplexing architecture and promote mass-production. System D_s is designed exclusively for satellite delivery.

With advancements in digital technology and users' expectations, the scope of sound broadcast systems has enhanced to embrace multimedia broadcasts including television.

The digital video broadcast (DVB) forum has developed a digital video broadcast-satellite services to hand-held (DVB-SH) transmission system standard designed to deliver video, audio and data services to hand-held devices (DVB-1, 2008). The DVB-SH utilizes a hybrid satellite/terrestrial system to operate at frequencies below 3 GHz (typically around 2.2 GHz). The system and waveform specifications have been published as ETSI standards TS 102 585 (ETSI, 2007) and EN 302 583 (ETSI, 2010)]. The system has been outlined later in this chapter as representative of the new generation mobile broadcast system.

In parallel, ETSI has been working towards evolution of a suitable European standard for SDR. A technical report ETSI TR 102 525 V1.1.1 (ETSI, 2006) developed by ETSI technical committee, Satellite Earth Stations and Systems (SES), addresses the functionalities, reference architecture and technologies related particularly to the receiver radio interface of SDR systems, including identification of candidate technologies to support the reference architecture without attempting to be definitive. The intention of the report is to provide guidelines 'for the creation of an SDR radio interface standard or standards for particular parts of the SDR system'.

A typical SDR system architecture would combine satellite broadcast and, where necessary, complementary terrestrial transmitters to ensure seamless reception. The SDR system addresses broadcasting services (multicasting being a subset) to provide digitally compressed audio contents including enhanced multimedia content, text, image, video and data content. The capability to provide interactive services at application layer level is included. ETSI's report conceives an architecture that allows the introduction of new technologies. It defines the scope of the broadcast layers, functionality of related building blocks and the interfaces. It concludes that applicable existing satellite standards such as ITU-R Recommendation BO.1130-4 system D_H and system E would not fulfil the targeted functionalities, services and market (primarily, European) and hence a new standard is needed.

Table 12.5 Representative performance of ITU-recommended digital systems on the basis of some of the recommended technical and operating characteristics of recommendation ITU-R BO.789

Characteristics as per recommendation ITU-R BO.789	Digital system A	Digital system B	Digital system D_S	Digital system D_H	Digital system E
Range of audio quality and types of reception	8–384 kbps per audio channel in increments of 8 kbps	16–320 kbps per audio channel in increments of 16 kbps	16–128 kbps per audio channel in increments of 16 kbps	16–128 kbps per audio channel in increments of 16 kbps. Each 16 kbps increment can be split into two 8 kbps services	16–320 kbps per audio channel in any increment size
	MPEG-2 Layer II audio decoder typically operating at 192 kbps	Perceptual audio codec (PAC) source encoder at 160 kbps used for most field tests	MPEG-2 and MPEG-2.5 Layer III audio coding	MPEG-2 and MPEG-2.5 Layer III audio coding	MPEG-2 AAC audio coding
	Vehicular, portable and fixed reception	Vehicular, portable and fixed reception	Portable and fixed reception	Vehicular, portable and fixed reception	Vehicular, portable and fixed reception
Performance in multipath and shadowing environments	Power summation of echoes falling within a given time interval	Attempts to maximize link margin on satellite channels	Supports direct or lightly shadowed reception via satellite	Supports diversity reception of satellite and terrestrially retransmitted signals	Summation of echoes by RAKE receiver
	Supports on-channel repeaters	Supports on-channel repeaters		Supports on-channel repeaters	Supports on-channel repeaters
					Segmented convolutional bit wise interleaver allow > 1 s shadowing

(continued overleaf)

Table 12.5 (*continued*)

Characteristics as per recommendation ITU-R BO.789	Digital system A	Digital system B	Digital system D_S	Digital system D_H	Digital system E
Common satellite-terrestrial receiver signal processing	Allows same receiver from RF front-end to audio and data output for satellite and terrestrial systems using integrated or separate antennas	Allows same receiver for both satellite and terrestrial transmission, with equalization for terrestrial delivery	Allows same receiver for fixed and mobile reception in rural environments when terrestrial augmentation (for indoor reception) is limited to micro-power gap fillers Receiver supports integrated or separate antennas	Receivers were being developed for reception in urban environments at the time the report was written	Allows same receiver, from the RF front end to the audio and data output with simultaneous reception from satellite and on-channel repeaters
Number of programme trade-off	Five levels of protection for audio and eight levels of protection for data services through punctured convolutional coding for each of the 64 sub-channels (FEC range: $\frac{1}{4}$ to $\frac{3}{4}$)	Based on an information bit rate contained in steps of 32 kbps and a variable FEC rate	Number of programme is fixed	Number of programme is fixed	Data rate can be selected from 236 to 413 kbps through punctured convolutional coding (FEC range: $\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$, $\frac{5}{6}$ or $\frac{7}{8}$)

Mixed and hybrid support (Mixed: use of same band for satellite and terrestrial) (Hybrid: on-channel repeaters)	Mixed and hybrid	Mixed and hybrid in BSS bands allocated in WARC-92	Not applicable	Hybrid in bands allocated for BSS (sound) by WARC-92	Mixed and hybrid
Compatibility of multiplex structure with OSI	Generally compliant (with an exception)	Capable, but not tested when report was written	Compliant	Compliant	Compliant with MPEG-2 System architecture
Receiver cost driver	Allows mass-production	Anticipated development of relatively low-cost consumer receivers	Optimized for an initial low complexity portable receiver	Signal processing expected to be embedded in microchips suitable for mass production	Optimized for an initial low complexity vehicular receiver deployment.

(Adapted from ITU-R, 2001.)

12.9 Receiver Architecture

The receiver architecture depends on:

- **Service requirements** – audio, video, data, multimedia or a mix.
- **Type of usage and target market** – handheld, vehicular or portable.
- **System configuration** – satellite-only, hybrid or mixed-mode, extent and type of diversity (i.e. frequency, space and time), user interactivity.
- **Transmission format** – MPEG (audio/video coding and encapsulation), programme multiplex, interleaving and FEC coding.

Here, we outline as an example the receiver architecture of Digital System D_H mentioned in a preceding section (ITU-R recommendation BO.1130-4, ITU-R, 2001). The system broadcasts digital audio and data for reception by inexpensive vehicular, fixed and portable receivers in 1452–1492 MHz band incorporating spatial (dual satellite) and time diversity. It utilizes coherent QPSK modulation with block and convolutional error coding, and linear amplification. The audio programmes are multiplexed and transmitted as a TDM signal. Receivers select the channel of user's interest from the TDM data streams to recover the desired digital baseband.

The terrestrial delivery utilizes an orthogonal frequency division multiplex technique called multi-carrier modulation, which is a commonly used in terrestrial broadcast system such as Digital System A to contend frequency-selective fades observed on terrestrial wideband channels. Table 12.6 summarizes relevant attributes of the system and Figure 12.2 illustrates its coverage scheme.

The outermost region represents the wide-area coverage where terrestrial infrastructure is absent – a receiver operating in this zone receives only satellite transmissions. A receiver in the intermediate zone receives both terrestrial and satellite signals where a receiver can either combine or select the stronger of the two signals. In the inner zone, satellite signals suffer extensive blockage and therefore only the terrestrial signals are available reliably at the receiver.

The hybrid architecture utilizes dual-satellite (spatial) diversity and time diversity. The satellites are fed with the same signal but transmission to one of the satellite is delayed by

Table 12.6 Characteristics of satellite component of digital system D_H

Audio encoding	MPEG Layer 3
TDM bit-rate	3.68 Mbps
Number of programmes per TDM channel	96 at 16 kbps
Typical TDM capability of first generation satellite	6
Modulation	Coherent QPSK
Frequency spacing between carrier	2–3 MHz
Typical satellite EIRP – primary service area	49 dBW
Channel coding	Cascaded codes – convolution inner code, an interleaver and a RS block outer code
Down-link delivery scheme	Multiple channel per carrier/TDM

(Adapted from ITU-R, 2001.)

4.28 s. A typical satellite-only receiver consists of an antenna system of 6 dBi gain, a front-end amplifier, demodulator, TDM frame synchronizer, demultiplexer to extract the selected broadcast channel, FEC decoder and audio decoder or a data decoder for extracting data services. The receiver G/T is specified as $-16.5\,\text{dB}(\text{K}^{-1})$. The antenna allows reception of both circular polarizations. A typical antenna would have a patch size of $6\,\text{cm} \times 6\,\text{cm}$ with about 100° beamwidth, therefore requiring no pointing. Using the given error protection mechanisms the target BER of 1×10^{-4} can be met with carrier to noise (C/N) of 4.5 dB.

In difficult reception environment such as experienced in secondary reception areas a tracking 12 dBi gain antenna can be used. A detachable antenna system installed with direct line of sight to satellite can provide indoor reception. High-gain antenna options include helix or a standard patch as a feed into a parabolic dish reflector.

The QPSK phase ambiguity recovery is made every 1.4375 ms. The system includes early-late time diversity feature in conjunction with maximum likelihood combination of diversity signals.

The basic satellite-only receiver architecture must be enhanced to take advantage of the terrestrial component and the diversity. Thus, it includes two additional receiver branches – one to receive the second satellite signal and the other for the reception of the terrestrial single frequency component. A single antenna can be shared for the reception of all the three components. The diversity satellite signals are received by a receiver identical to the single satellite receiver with an additional provision to combine the early satellite signal to achieve time and space diversity gain. An additional branch is necessary to receive the terrestrial MCM signals with a provision to select the best option between satellite and terrestrial components. It is possible to conceive of a scheme wherein all the signals are combined synchronously.

An example implementation of system D_H would comprise two satellites spaced apart (e.g. 20°) to provide an effective spatial diversity. To achieve time diversity, the same baseband would be transmitted to each satellite with transmission of the second satellite delayed by about 4.28 s with respect to the other satellite. The time delay is considered sufficient to combat signal blockages by bridges, short tunnels and trees along rural highways travelling at typical speeds on the basis of measurement campaigns conducted in Europe from MARECS-A satellite transmissions (15° W longitude) on a vehicle travelling at 60 kmph in a rural highway. These measurements demonstrated that 99% of fade events occur for less than 4 s (ITU-R, 2001).

Figure 12.6 shows a top-level receiver architecture of a dual-diversity (space-time) space-terrestrial hybrid system configuration. All the receiver branches share the antenna system. One branch carries early broadcast channels and the other the delayed broadcast channels. The third branch receives the terrestrial time-division-multiplexed MCM transmissions derived from the terrestrial broadcast channels. Each satellite branch comprises a satellite tuner that selects the desired TDM satellite carrier, a QPSK demodulator to recover the TDM symbol stream and a TDM demultiplexer. The early and late versions of the TDM signals are fed into a Viterbi maximum likelihood FEC trellis decoder to combine the two signals. The delay of the early signal is dealt with in the TDM demultiplexer. Viterbi decoder combining is accomplished by aligning the preambles of the early and late broadcast channel frames. The terrestrial branch operates simultaneously and independently of the satellite branches following the same sequence as in the satellite branch. Synchronization for combination

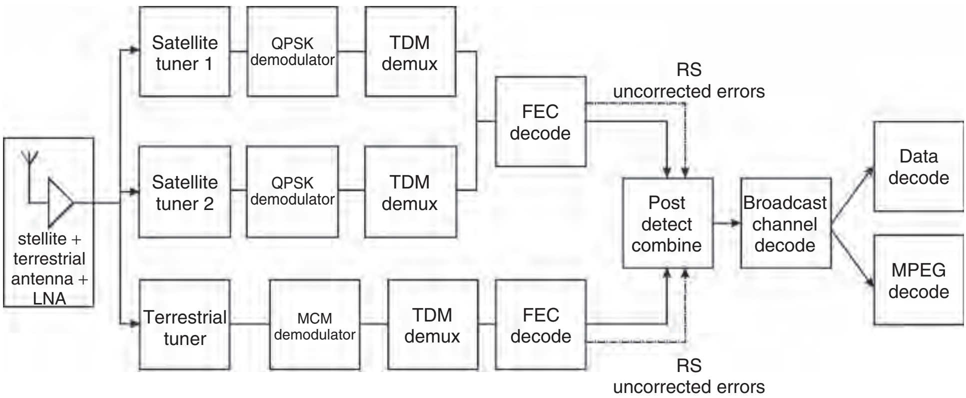


Figure 12.6 A receiver architecture applicable to a satellite-terrestrial hybrid system with space and time diversity in the space segment. (Source: ITU-R, 2001. Reproduced by permission of ITU.)

with satellite reception is achieved by aligning the preambles with the early and late broadcast channel frames. The post detector logic can be used to select the component that has the better signal quality in order to provide the desired broadcast channel.

12.10 DVB-SH System Architecture

This section presents the main technical aspects of the DVB-SH standard as representative of recent advances in technology. This standard is considered as a candidate for future implementation of multi-media broadcasts including television to mobile and small user terminals (ITU-R, 2012).

The DVB-SH transmission system standards (TS 102 584, TS 102 585 and EN 302 583) promulgated by ETSI provide waveform and system specifications for provisioning video, audio and data broadcast services – particularly mobile television – to portable, mobile and fixed user terminals at frequencies below 3 GHz over a seamless hybrid satellite-terrestrial architecture. There is a provision for the complimentary ground component (CGC) to additionally include local content if necessary (ETSI, 2007), ETSI EN 302 583 V1.1.2, (ETSI, 2010). The standard thereby allows a progressive coverage penetration beginning from wide-area coverage by a satellite and increasing the penetration into coverage gaps such as cities or tourist hot-spots by addition of CGC as necessary. The standard specifies the entire system comprising physical link and service layers.

Figure 12.7 presents a generic architecture of the system. The terrestrial air interface utilizes an OFDM scheme as used in the terrestrial broadcast systems, DVB-H and DVB-T (Digital Video Broadcast-transmissions to fixed installations).

There are two transmission modes (ETSI, 2010):

1. SH-A based on OFDM broadcast, based on the terrestrial DVB-T standard, over both satellite and terrestrial media.
2. SH-B based on TDM broadcast (derived from the DVB-S2 standard) from satellite and OFDM broadcasts in the terrestrial component.

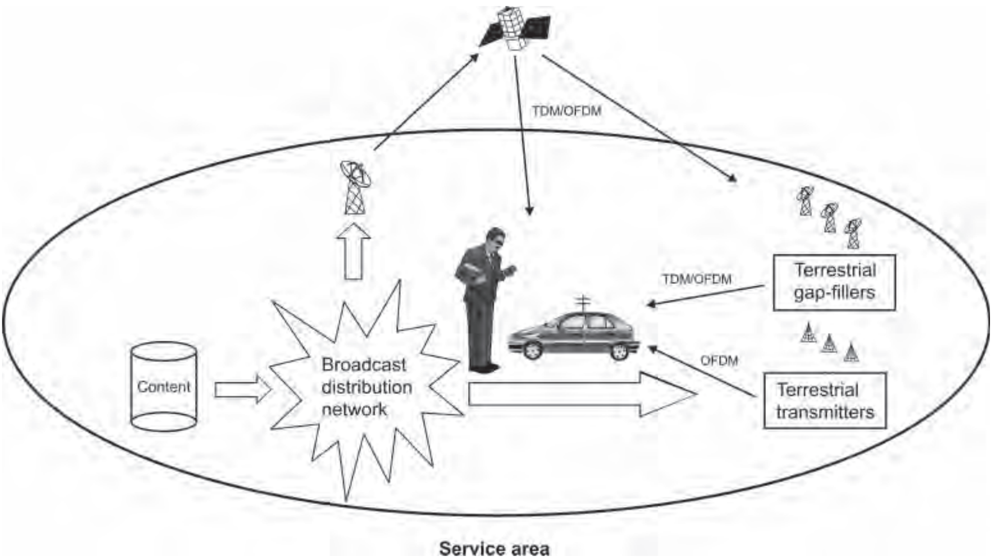


Figure 12.7 A generic architecture of DVB-SH system

The specifications provide maximal commonality between these two architectures; the commonality in higher protocol layers allows receiver architecture to be simplified. A dual-mode receiver allows the user to switch to either of the two SC configurations – OFDM or TDM.

The TDM stream can operate on QPSK, 8-PSK and 16-APSK schemes in conjunction with a variety of roll-off factors, while OFDM can operate on QPSK, 16-QAM and non-uniform 16-QAM with support of hierarchical modulation (Note: A hierarchical modulation scheme consists of two separate data streams – high-priority and low-priority; the high priority streams can be demodulated by a receiver where signal quality is poor whereas both the streams can be demodulated by receivers placed in good signal condition: DVB, 2000). To facilitate occupancy within an operator's available transmission band there is a feature to select channelization bandwidths from the set 8, 7, 6, 5, 1.7 MHz and an Fast Fourier Transform (FFT) length from amongst 8k, 4k, 2k and an additional 1k directly scaled from the 2k mode.

Depending on the target audience, three types of CGC transmitters are specified that may provision service within:

- relatively large local areas (e.g. dense cities) – these transmitters permit addition of local content;
- confined 'personal' space such as areas within a building and
- moving platforms such as railway trains; these transmitters incorporate a provision to include local content.

In order to contend weak signals and signal loss due to fading, a combination of the 3GPP2 turbo-code that supports various code rates and a flexible interleaving scheme to provide time diversity ranging from 100 ms to several seconds is recommended. Short and long physical layer interleaving are supported in addition to interleaving at link layer managed by the service layer; thus the operator selects the scheme(s) tailored to the given requirement (e.g. in terms of the quality of service, service type and the target receiver category). The same interleaver can be configured for terrestrial and satellite signals either in a common configuration (SH-A system) or as two separate configurations for the SH-B system. Pilot symbols are used in both the SC architectures for robust acquisition and reacquisition after a prolonged fade event.

Simultaneous reception of satellite and terrestrial signals allows signal diversity in single frequency network (SFN) architecture (SH-A), maximal ratio combining feature in both the architectures and code diversity is supported in the SH-B architecture via a common frame structure shared between SC (TDM) and CGC (OFDM) modes.

Based on the features incorporated in the receiver, two classes of receivers, namely class 1 and class 2, are specified in each SC architecture, as follows:

- Class 1 receivers support *short* interleaving at the physical layer complemented by the long interleaving at the link layer.
- Class 2 receivers support *long* interleaving at the physical layer complemented by the long interleaving at the link layer.

The link layer (ETSI, 2004, 2010) supports MPEG-2 TS packets with provisions to introduce generic packets, multi-protocol encapsulation (MPE), time sliced power saving with MPE-FEC compatibility and a provision for frequency and beam handover.

Work is in progress to ensure that DVB-SH can support DVB-IPDC (DVB-Internet protocol data casting), which is a set of DVB specifications to support IP data casting over DVB mobile TV networks. These DVB specifications aim to subsume broadcast and mobile telecommunications in a converged network.

Figure 12.8 shows the protocol architecture of DVB-SH (DVB-2, 2008). The ESG protocol stack (ES 102 471) formats, structures and transports the ESG to assist users in selecting the services and retrieving stored content on the receiver. The content delivery protocols (CDPs)

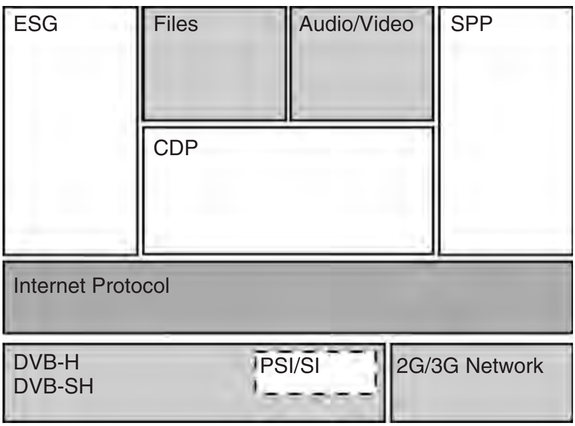


Figure 12.8 A reference architecture of DVB/IPDC. (Source: DVB-2, 2008. Reproduced by permission of the DVB Project.)

(TS 102 472) support content delivery, for example audio/video streaming and file transfer (e.g. software upgrade). The service purchase protection (SPP) layer sets out the encryption of services and signalling to support various subscription schemes. The programme-specific information/service information (PSI/SI) specifications (TS 102 472) provides a platform that allows programme extraction and service tables at the receiver and support signalling for roaming and mobility. As stated earlier, DVB-SH data broadcast protocols are fully compliant with the (terrestrial) DVB-H standard.

Figure 12.9 shows a functional block diagram of both the modes of DVB-SH transmitters (ETSI, 2008). The incoming stream is typically a MPEG-transport scheme containing the content in time-sliced bursts compliant to the terrestrial DVB-H standard. The stream undergoes a series of processes common to both the OFDM and TDM transmission modes. The output of the common processing block is next processed individually for transmissions by each mode before RF processing.

The *mode adaptation* subsystem performs a Cyclic Redundancy Check (CRC) encoding on each MPEG packet to enable error detection at the receiver and appends encapsulation signalling used for support of other input stream formats including unpacketized streams. The *stream adaptation* block encapsulates each packet by padding (appending zeroes after the data field) to provide a fixed frame size of 12 282 bits (EFRAME) for compatibility with the turbo code input specification, and also scrambles the data by randomizing each complete EFRAME. The turbo code is based on the 3GPP2 standardized code with additional code

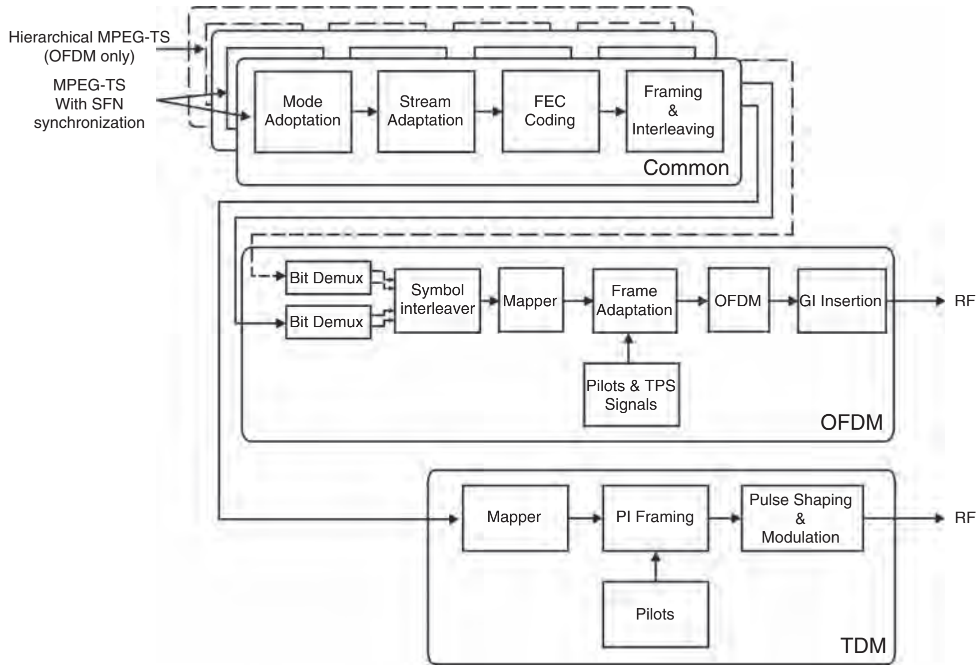


Figure 12.9 A functional block diagram of both modes of DVB-SH transmitter. (Source: ETSI, 2008. Reproduced by permission of the DVB Project.)

rates introduced for finer carrier to noise granularity and facilitate code combining between TDM and OFDM mode. The output symbols are set to equal $(L_{in}+6)/R$ where L_{in} is the number of input bits per block and R is the code rate. The channel interleaver comprises a bit-wise interleaver operating on encoded words in tandem with a convolution interleaver operating on interleaver units (IU) of 126 bits with a rate adaptation unit in between to ensure that integer IUs enter the convolution encoder. The interleaved IUs are assembled to constitute SH frames whose structure differs for TDM and OFDM. However, the framing duration of TDM and OFDM SH frames are identical to facilitate diversity combining of TDM/OFDM transmission modes. A SH-IP packet is included in the frame to facilitate synchronization between both transmitter type (OFDM and TDM). The OFDM and TDM branches beyond this stage are different.

The TDM SH frames are mapped to physical layer slots of 2176 symbols – each slot consists of 2, 3 or 4 capacity units (CU) of 2016 bits each – respectively for QPSK, 8-PSK and 16-APSK modulation schemes. The frame includes two pilot fields of 80 symbol duration in which unmodulated symbols are inserted. Before modulation the slots are scrambled by a randomizer to avoid spectral spikes in the transmissions. The QPSK, 8-PSK and 16-APSK constellation mapping follows the DVB-S2 standard. The modulated signal is filtered in a square root raised cosine filter with roll-off factor set to 0.15, 0.25 or 0.35.

The definition of OFDM signals is based on DVB-T specifications with a provision for reduced bandwidth in satellite channels. The SH frame consists of 816 CU irrespective of modulation scheme and is aligned to the OFDM frame. The length of the OFDM super frame depends on the modulation scheme and FFT mode (outlined later). The output of the channel interleaver, consists of up to two bit streams for hierarchical modulation and a single bit stream for non-hierarchical stream, which is demultiplexed into v sub-streams, where $v = 2$ for QPSK, $v = 4$ for 16-QAM. For the hierarchical stream, the high priority stream is demultiplexed into two streams and the low priority steam into v-2 sub-streams. The output of the demultiplexer is grouped into v bit words. The symbol interleaver maps these v bit words onto the 756 (1K mode), 1512 (2K mode), 3024 (4K mode) or 6048 (8K mode) active carriers per OFDM symbol. Each symbol contains data and reference information. All data carriers of each OFDM frame are modulated using QPSK, 16-QAM or non-uniform 16-QAM constellations with Gray mapping.

The transmitted signal is organized in frames each of 68 OFDM symbols, with four frames constituting a super-frame. Each symbol is constituted by a set of carriers depending on value of K and transmitted with duration TS comprising two parts: a useful part with duration TU and a guard interval. Each symbol is divided into cells, each corresponding to the modulation carried on one carrier during one symbol.

In addition to the transmitted data, an OFDM frame contains scattered pilot cells, continual pilot carriers and transmission parameter signalling (TPS) carriers. The pilots can be used for synchronizing frame, frequency, time, channel estimation, transmission mode identification and to follow the phase noise. The TPS signalling carriers provides information on system parameters such as modulation, hierarchy, guard interval; transmission mode (i.e. 1k, 2k, 4k, 8k), frame number in a super-frame, cell identification, DVB-SH mode (selector bit) and code rates. The spectrum shaping of the baseband is based on DVB-T specifications.

12.11 Multimedia Broadcast and Multicast Services (MBMS)

MBMSs meet the requirements of 'interoperability between mobile telecommunication services and interactive digital broadcasting services' but lie outside the formal broadcasting services regime (ITU-R, 2012). The MBMS is a terrestrial multicast service for mobile reception standardized by the 3GPP (third Generation Partnership Project). ETSI, a partner in 3GPP, also specifies the 3GPP specifications at a certain stage of the standards developing process. For example, ETSI TS 123.246 (alternatively, 3GPP TS 23.246), provides MBMS architecture and functional description in both the regimes.

Although the MBMS is a terrestrial service, such a service is also of interest to the MSS due to the vast areas covered by the MSS and the transport efficiency, which multicast offers when addressing a widely dispersed user population, by sharing the radio resource. To derive economies of scale and seamless service, it is imperative that satellite systems be adapted to utilize the terrestrial multicast infrastructure. The MBMS in itself reuses the core network of 3GPP telecommunication network. In this context (Febvre *et al*., 2007) describe the adaptation of MBMS to Inmarsat's 3GPP based Broadband Global Area Network (BGAN) system as a way forward (see Chapter 11 for BGAN description).

A few representative features of MBMS are mentioned here to heighten the reader's interest. The interested reader may refer to (ITU-R, 2012) for a concise summary of MBMS including a list of references. MBMS is a multimedia and broadcast point-to-multipoint bearer service for IP packets delivery in the packet switched (PS) domain. The architecture provides efficient use of radio and core network by allowing multiple users to receive the same content, reusing the existing core network components and protocol elements The MBMS routing, information and data flow occur in a core network, encapsulated within an IP multicast framework. When possible, the multicast content utilizes only a portion of radio carrier, leaving the remaining capacity for other traffic to improve radio resource efficiency and offer flexibility. The service area is scalable from small areas of a few hundred meters to wide areas. The system supports over-the-air mobile audio/video (A/V) multimedia streaming and downloads with or without acknowledgement. It is interoperable with Internet engineering task force (IETF) IP Multicast addressing. Two types of services are identified – streaming services, the basic user service and a download service to deliver binary data (file data).

Each 5 MHz wide radio bearer operating in a code division multiple access (CDMA) mode can support different bit rate, up to 256 kbps. The CDMA scheme uses three logical channels. The point-to-multipoint control channel (MCCH) contains details of on-going and upcoming MBMS mobile A/V multimedia service sessions; point-to-multipoint scheduling channel (MSCH), provides information on data scheduled on point-to-multipoint traffic channel (MTCH); MTCH carries the MBMS application data. A physical channel called the MBMS notification indicator channel (MICH) notifies the user equipment (UE) of available MBMS information on MCCH. Interleaving depths of 40 and 80 ms provides for two different time diversities to counter the fading in mobile channels.

The UE must be able to support activation/deactivation of the MBMS bearer services. UEs receive the A/V multimedia content only when attached to the network, while continuing to support functions related to other services. The MBMS Session Identifier contained in the notification to the UE enables the UE to prepare or ignore reception of the incoming multicast. The UE can be notified about a forthcoming and an on-going data transfer from other MBMS services.

A group membership function can be attached to a transmitter to attach or detach the transmitter from a multicast group for common addressing to the group.

12.12 DBS Reception on Mobile Terminals

Although the chapter primarily addresses mobile BSS systems, reception of the conventional (fixed) direct broadcast transmissions has been available on mobile platforms for over a decade. Consider, first, the differences in requirements, characteristics and features between fixed and mobile satellite broadcast systems.

- Radio propagation characteristics of these systems differ due to differences in frequency and operating environment – fixed (DBS) receivers operate on a stable radio link, whereas, mobile receivers operate in a dynamic radio environments;
- The display sizes and receiver capabilities are different for mobile receivers and hence the range and scope of contents differ;
- Usage characteristics are likely to be different (for example user on mobile terminals may prefer short video clips rather than full-length movies);
- The interaction channel requirements of mobile broadcasts differ due to mobility;
- Methods to ensure cyber security and conditional access solutions may require a different type of solution for mobile users.

Despite these differences, manufacturers have adapted the design of DBS receivers to facilitate portability, and provide reception on land vehicles, ships and aircraft in order to meet the needs of the travelling public. Over 150 000 user terminals were said to be in operation in 2012 though the outlet of one company alone (KVH, 2012). Modifications to support mobility apply to:

- Antenna systems to enable satellite tracking during motion; Phased array designs offer advantages in terms of drag, weight, maintenance and reliability and are less obtrusive when compared to mechanical or hybrid mechanical–phased array designs, but are more expensive.
- Packaging to provide ruggedness against shocks and vibrations and to enable convenient viewing.
- Frequency control circuit to correct Doppler frequency shifts when used on high speed mobiles such as aircrafts. Doppler shifts are of the order of 100 Hz (40 m/s) for a slow-moving vehicle, such as a ship, to 16 kHz for a fast-moving aircraft (400 m/s) at 12 GHz, which amounts to frequency uncertainty of 1.2×10^{-8} to 1.3×10^{-3}.

An early example of a development in this area is Boeing's Satellite Television Aeroplane Receiving System (STARS), which was developed for commercial operation in the USA to operate with a regular DBS system in the 12.2–12.7 GHz band. The system comprised a phased array system with a low noise block assembly, a G/T of 8.2 dB/K and a television receiver that provided the required signal quality with a satellite EIRP of 50.4 dBW (Vertatschitsch and Fitzsimmons, 1995). Television broadcasts during off-peak

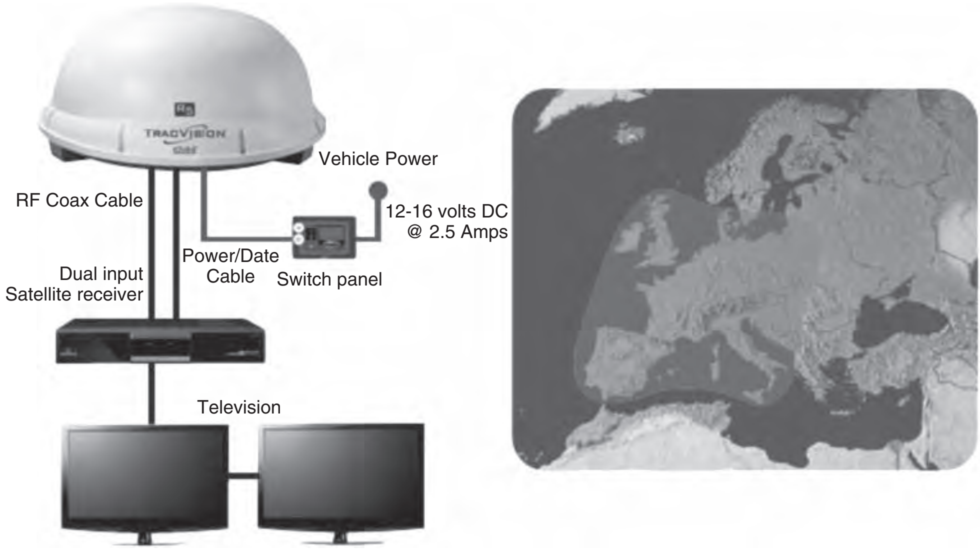


Figure 12.10 (a) Block schematic of a mobile DBS receiver. (b) Coverage of Astra 2 South 28.2E wherein mobile broadcast service is targeted. (Both parts source: KVH. Reproduced with permission of KVH.)

telecommunication traffic hours, for transmitting important events to US ships, were made regularly via Inmarsat satellites within a bandwidth of 400 kHz using a QPSK modulation scheme for a number of years in 1990s.

Figure 12.10(a) (KVH, 2012) illustrates a schematic of a mobile broadcast receiver to provide services on coach, boats and automobiles, etc., within the footprint of a geostationary DBS satellite such as Astra. Figure 12.10(b) shows the footprint of Astra-2 satellites where mobile DBS TV services are available. The K_u band tracking antenna of 80 cm diameter, depending on the antenna's tracking capability, is fitted within a radome of 30–40 cm height. The DC power, at a voltage of 12–18 V, is taken off the vehicle battery. The system can support HDTV reception. The indoor unit comprises a satellite receiver and a television screen.

Revision

1. System features of a mobile satellite broadcast service are conditioned by numerous practical considerations and constraints such as small screen size, reception on a variety of receiver types, uninterrupted and ubiquitous service, and so on. Based on such issues, suggest a set of service requirements of a mobile satellite broadcast system. Give an example of service features of an advanced mobile broadcast system.

2. Compare the characteristics of alternative mobile satellite broadcast system architecture, including an example of each. Suggest a configuration suited for broadcasts to subscribers dispersed in a mix of rural and urban environments.

3. State the considerations applied in the selection of the space segment to provide a reliable broadcast service, including the countermeasures that can be used to increase the robustness of the broadcast radio link.
4. Discuss the OSI model as applied to a satellite mobile broadcast system. Illustrate its applicability to a practical system with the help of an example.
5. State the salient features of each of the five ITU-R-recommended digital mobile satellite broadcast system and highlight the differences between them; suggest the system(s) that could be used to provide reliable broadcasts to a mid-latitude region such as Europe that harbours diverse languages and culture.
6. With the help of a block schematic discuss the functioning of a mobile receiver of a hybrid satellite system that supports spatial, time and frequency diversity.
7. Highlight the salient features of the DVB-SH system.
8. What are the main differences in requirements, characteristics and features between fixed and mobile satellite broadcast systems.
9. Discuss the feasibility of adapting a hand-held receiver for reception of K_u band direct broadcast transmissions.

References

Burger, R.A., Jacovoni, G. and Reader, C. *et al*. (2007) A survey of digital TV standards China. CHINACOM'07: Second International Conference on Communications and Networking in China 2007, Shanghai, China, August 22–24, 2007, pp. 687–696.

Butt, G., Evans, B.G. and Richharia, M. (1992) Multiband propagation experiment for narrow- band characterisation of high elevation angle land mobile- satellite channels. *Electronics Letters*, **28**(15), 1449–1450.

Campanella, S.J. (1996) The WorldSpace satellite-to-radio multimedia broadcast system: a technical overview. Proceedings 16th International Communications Satellite Systems Conference AIAA, Washington, DC, February 25–29, pp. 826–831.

DVB (2000) DVB-T, Hierarchical Modulation, www.dvb.org/technology/fact_sheets/DVB-T_Factsheet.pdf (accessed 28 August, 2013).

DVB-1 (2008) DVB-SH Fact-Sheet, www.dvb.org/technology/fact_sheets/DVB-SH_Factsheet.pdf (accessed 28 August, 2013).

DVB-2 (2008) DVB IPDC Factsheet, www.dvb.org/technology/fact_sheets/DVB-IPDC_Factsheet.pdf (accessed 28 August, 2013).

ETSI (2004) Digital Video Broadcasting (DVB); DVB Specification for Data Broadcasting, ETSI EN 301 192 (V1.4.1).

ETSI (2006) Satellite Earth Stations and Systems (SES); Satellite Digital Radio (SDR) Service; Functionalities, Architecture and Technologies, Technologies for the Radio Interface, ETSI TR 102 525 v.1.1.1, 21 September 2006.

ETSI (2007) Technical Specification Digital Video Broadcasting (DVB); System Specifications for Satellite services to Handheld services (SH) below 3 GHz, ETSI TS 102 585 V1.1.1.

ETSI (2008) DVB-SH Implementation Guidelines, DVB Document A120, May.

ETSI (2010) European Standard (Telecommunications Series) Digital Video Broadcasting (DVB); Framing Structure, Channel Coding and Modulation for Satellite Services to Handheld Devices (SH) below 3 GHz, EN 302 583 V1.1.2.

Evans, B.G. and Thompson, P.T. (2007) Aspects of satellite delivered mobile TV (SDMB). Proceeding of the 16th IST Mobile and Wireless Communications Summit, 2007, August 2007; doi: 10.1109/ISTMWC.2007.4299322.

Febvre, P., Bouthors, X., Maalouf, S. and Bath, B. (2007) Efficient IP-multicast *via* Inmarsat BGAN, a 3GPP satellite network. *International Journal of Satellite Communications and Networking*, Special issue on ASMS 2006, **25**(5), 459–480.

Galligan, K.P. (1989) Sound broadcasting applications of highly elliptical orbits. IEE Colloquium on Highly Elliptical Orbit Satellite Systems (1989/1986), May 24, Digest, pp. 7/1–7/4.

Giovagnoli, F. (1989) Land mobile communications from non-geostationary orbits (Review of Archimedes ESA study). IEE Colloquium on Highly Elliptical Orbit Satellite Systems, (1989/1986), May 24, Digest, pp. 3/1–3/5.

Hess, G.C. (1980) Land mobile satellite excess path loss measurements. *IEEE Transactions on Vehicular Technology*, **VT-29**(2), 290–297.

Hirakawa, S., Sato, N. and Kikuchi, H. (2006) Broadcasting satellite services for mobile reception. *Proceedings of the IEEE*, **94**(1), 327–332.

ITU-R (1992) Reference Model for Data Broadcasting. ITU-R recommendation BT.807 1.

ITU-R (1995) Services for Digital Sound Broadcasting to Vehicular, Portable and Fixed Receivers for Broadcasting- Satellite Service (Sound) in the Frequency Range 1400-2700 MHz, BO Series Facile. Recommendation ITU-R BO.789-2, Broadcasting-Satellite Service (Sound and Television), Geneva.

ITU-R (2001) Systems for Digital Satellite Broadcasting to Vehicular, Portable and Fixed Receivers in the Bands Allocated to BSS (Sound) in the Frequency Range 1 400-2 700 MHz, BO Series Facile. Recommendation ITU-R BO.1130-4, Broadcasting-Satellite Service (Sound and Television), Geneva.

ITU-R (2011) Broadcasting of Multimedia and Data Applications for Mobile Reception. Report ITU-R BT.2049-5, (05/2011).

ITU-R (2012) Broadcasting of Multimedia and Data Applications for Mobile Reception by Handheld Receivers. Recommendation ITU-R BT.1833-2 (08/2012).

KVH (2012) web site www.tracvision.co.uk (accessed 28 August, 2013).

Layer, D.H. (2001) Digital radio takes to the road. *IEEE Spectrum*, **38**(7), 40–46.

Miller, J.E. (1985) Satellite sound broadcasting around 1 GHz. *IEEE Journal on Selected Areas in Communications*, **SAC-3**(1), 204–210.

Narenthiran, K., Karaliopoulos, M. and Evans, B.G. (2004) S-UMTS access network for broadcast and multicast service delivery: the SATIN approach. *International Journal of Satellite Communications and Networking*, **22**(1), 87–111.

Pérez-Neira, A., Ibars, C., Serra, J. *et al.* (2008) MIMO applicability to satellite networks. 10th International Workshop on Signal Processing for Space Communications, 2008. SPSC 2008, Rhodes, Island, October 6–8, 2008, pp. 1–9.

Richharia, M. and Westbrook, L.D. (2010) *Satellite Systems for Personal Applications*, John Wiley & Sons, Ltd, Chichester.

Vertatschitsch, E.J. and Fitzsimmons, G.W. (1995) Boeing Satellite Television Airplane Receiving System (STARS) performance. IMSC 1995, The Fourth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre/Industry Canada and Jet Propulsion Laboratory/NASA, pp. 301–305.

Wells, P.I. (1977) The attenuation of UHF radio signals by houses. *IEEE Transactions on Vehicular Technology*, **VT-26**(6), 358–362.

13

Related Satellite Systems

13.1 Introduction

Distress and safety-related satellite services, satellite-aided navigation and mobile services from very small aperture terminal (VSAT) systems are similar to those available from mobile satellite systems – all are useful satellite-delivered services targeted at business, groups or individuals delivered through small, low-cost terminals and symbiotic applications. They do not, however, belong to the mobile satellite service (MSS) per se (i.e., as defined by the International Telecommunications Union (ITU)) due to differences in system characteristics, tradition and radio regulations. While each of these topics is a subject in its own right, it was felt that a chapter summarizing the features of such systems would be an appropriate and useful addition to the book. In previous chapters, we mentioned the convergence of satellite systems with terrestrial cellular system architecture(s). We have therefore included a brief section outlining the architecture of a cellular system for a better appreciation of the terrestrial technology.

13.2 Distress and Safety Systems

Radio systems have been used for distress and rescue communications since the turn of the twentieth century. Early systems used in ships were Morse telegraphy on 500 kHz. Installation of the wireless system was mandatory in ships of 1600 gross registered tonnage (GRT) and above and in all passenger ships. The primary objective was for ship–ship communication within a maximum range of about 250 nautical miles. This system had a number of limitations, such as manual alerting and a need for ships to maintain a constant watch on communications from other ships in the vicinity. There were instances of ships being lost at sea without being able to contact for help.

To address the problem, the International Maritime Organization (IMO) began investigations in the early 1960s to improve the system by use of modern terrestrial and satellite technology (O'Neil, 1992). As a first step, the IMO established the International Maritime Satellite Organization (Inmarsat; now the International Mobile Satellite Organization), which gave a distress communication capability in all maritime regions below about $\pm 76°$

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

latitude. The system in addition included digital selective calling (DSC) and direct printing telegraphy; DSC refers to telex by radio, satellite and terrestrial systems. The Cospas-Sarsat (Cospas, Cosmicheskaya Sistyema Poiska Avarinyich Sudov; Sarsat, Search and Rescue Satellite-Aided Tracking) system (described later) enhanced satellite coverage up to the poles by the use of low earth orbiting (LEO) satellites. These terrestrial and satellite systems collectively contributed to the foundation of an operational concept of a fully automatic, global distress communication system known as the Global Maritime Distress and Safety System (GMDSS). The ITU extended its cooperation by assigning exclusive frequencies for search and rescue (SAR).

Following the amendments made in 1988 to 1974 Safety of Life at Sea (SOLAS), carriage of satellite EPIRBs (emergency position indicating radio beacon) on all ships of over and including 300 tonnes became mandatory from 1 August 1993. Some nations have since made carrying of distress transmitters on aircrafts a requirement. Various countries have authorized the use of Personal Locator Beacons (PLBs) – a 406 MHz emergency beacon triggered in an emergency (discussed later) – in remote or rugged areas (Cospas-Sarsat, 2009).

The modern GMDSS system with satellite communication and other enhancements is quicker, simpler, more efficient and automatic, with the following features and goals:

- fully automatic, eliminating the need for manual watch-keeping;
- applicable to general and distress communications;
- all distress messages should be heard and addressed on shore;
- equipment should be easy to operate.

The GMDSS system provides a rapid alert to ships in the vicinity and to rescue authorities on shore in case of distress, enabling the recipients to assist in a coordinated search of the affected party. Further, the system provides urgent and safety communications, and dissemination of maritime safety information such as navigation and meteorological warnings.

There are nine essential communications to be performed by ships falling into the category, by using equipment that would meet the requirement in each region specified by the IMO. The communication functions are: distress alerting (ship−shore, shore−ship and ship−ship), SAR coordination communications, on-scene communications, signal location, transmission-reception of maritime safety information and general radio communication; the sea areas are categorized as A1 to A4 on the basis of availability of communication facility. Sea area A3 refers to the area outside of areas A1 and A2 that is covered continuously through Inmarsat's geostationary satellite system. The ship equipment, depending on sea areas, comprises 9 GHz radar transponders, a NAVTEX (Navigational telex), equipment to receive Inmarsat's Enhanced Group Call system (an international SafetyNet Service), a satellite EPIRB, which is free floating with automatic activation capability, and equipment for continuously monitoring DSC on channel 70. NAVTEX is an international service for automated delivery of maritime safety information to ships within about 200 nautical miles of the shore. DSC supports ship-to-ship, ship-to-shore and shore-to-ship radiotelephone and multiframe/high frequency (MF/HF) radio telex calls. All the GMDSS equipment must comply with the performance standard specified by the IMO. Additionally, there are equipment availability specifications as well as a requirement for every ship to include personnel suitably qualified to handle distress/safety radio communication.

13.2.1 Cospas-Sarsat Search and Rescue System

Cospas-Sarsat, a component of the GMDSS for sea area A4, is a global SAR communication system for sea, air and land, in use since 1982, to provide accurate and timely distress alert and location data to SAR authorities throughout the world to assist in rescue operations and thereby improve the probability of survival of persons in distress. The distress information is forwarded by the Cospas-Sarsat Mission Control Centre (MCC) responsible for the region to the appropriate national SAR authorities.

The Cospas-Sarsat international programme is credited with saving thousands of lives (Cospas-Sarsat, 2011). Figure 13.1(a) illustrates the number of SAR events in which the Cospas-Sarsat system participated in the period January 1994 to December 2010 demonstrating about 10-fold increase in the number of persons rescued per year (from ~ 225 to ~ 2425). Figure 13.1(b) shows the SAR events by the affected sector for the year 2010, demonstrating the event distribution for maritime, land and aviation sectors as 56, 24 and 20% respectively.

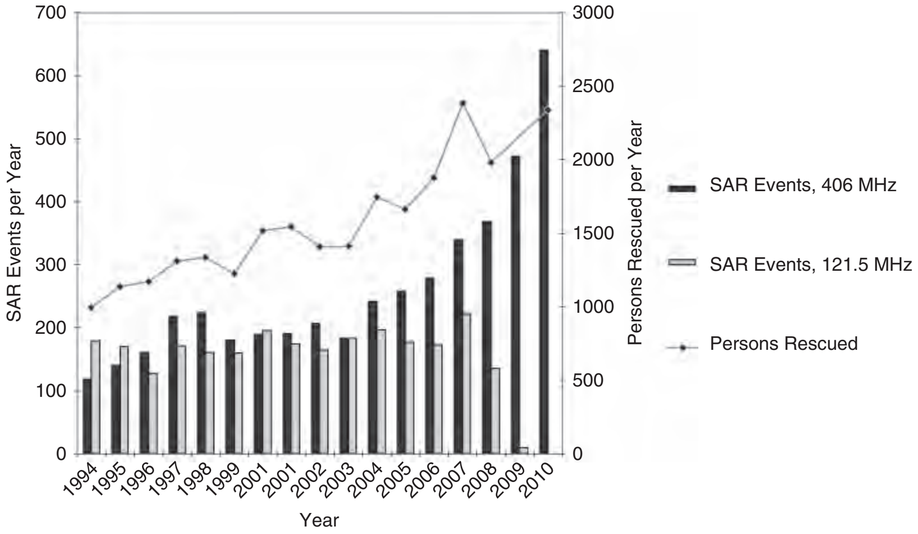
The IMO and International Civil Aviation Organization (ICAO) recommend the use of satellite beacon transmitters on ships and aircraft respectively.

The system consists of a small radio transmitter (beacon) installed on ships, aircraft, land vehicle or carried in person, which transmits distress signals in the case of a distress situation; the signals are received by a constellation of low earth orbit (LEO) satellites and retransmitted to ground stations known as local user terminals (LUT), which forwards the location information of the distress event to a MCC. The MCC initiates a SAR operation through an appropriate rescue coordination centre (RCC) or a search and rescue point of contact (SPOC).

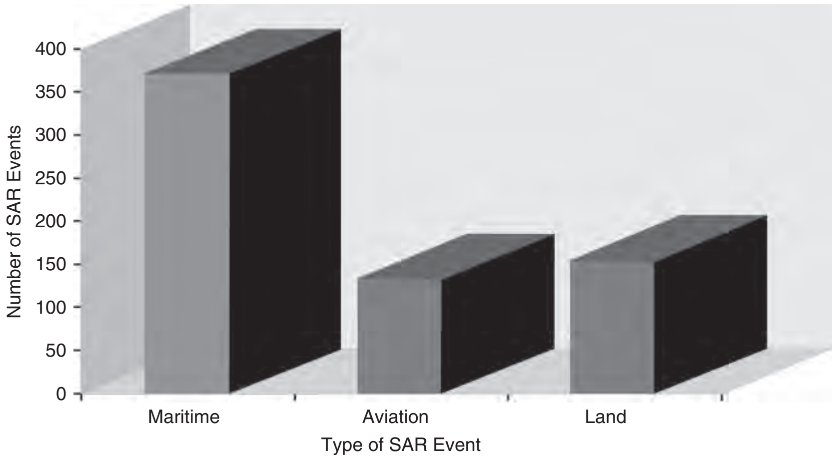
The concept of a satellite-aided SAR system, using LEO satellites, was proposed in the 1970s and proven in the USA, Canada and France, which helped the setting up of a joint experiment for Sarsat by NASA (USA), the Communications Research Centre/Department of National Defence (Canada) and the French Space Agency. Later, the former USSR joined the experiment and agreed to develop the COSPAS (Cosmicheskaya Sistyema Poiska Avarinyich Sudov – translated as 'space system for search of vessels in distress') component, which would be a compatible SAR system. The combined system became known as the Cospas-Sarsat system (King, 1999). During the demonstration phase beginning in 1982, the benefits of the system were realized when many lives were saved in various countries. Due to the success of the experiment, the system, instead of being switched off in the mid-1980s at the end of the demonstration phase, was declared a world-wide operational system in 1985 by the four founding countries. The Cospas-Sarsat secretariat was established in 1987; a formal inter-governmental agreement was signed in 1988 between the four founding countries, associating the IMO, the ICAO and the ITU, thus ensuring long-term availability of the system to all states on a non-discriminatory basis.

The implementation and operating cost is shared by member countries. The space segment is provided by the four founder countries. Users pay for purchasing the distress beacon transmitters but not for the use of the system. The programme is managed by a council consisting of representatives of member countries and supported by a secretariat located in Canada.

The main components of Cospas-Sarsat system, illustrated pictorially in Figure 13.2(a) and as a block schematic in Figure 13.2(b), consists of beacon transmitters carried by ships (called, emergency position indicating radio beacon or EPIRB), aircrafts (called, emergency locator transmitter or ELT) and individuals (called, personal locator beacon or



(a)



(b)

Figure 13.1 (a) Number of SAR Events and persons rescued with the assistance of Cospas-Sarsat Alert Data (121.5 and 406 MHz), (January 1994 to December 2010). (b) SAR events assisted by Cospas-Sarsat in the period January to December 2010. (Both parts source: Cospas-Sarsat, 2011. Reproduced with permission of Cospas-Sarsat.)

PLB); a space segment comprising satellites in LEO and geostationary orbit (GEO); and fixed earth stations called local user terminals connected to a MCC responsible for the region, which is linked to regional rescue centres.[1]

The space segment consists of four LEO satellites in a near polar orbit, called the Low Earth Orbit Search and Rescue (LEOSAR) system and a complementary geostationary

[1] Note: Cospas-Sarsat system/network details/statistics presented here are for illustrative purposes only. The reader should refer to the organization's web-site for the latest information.

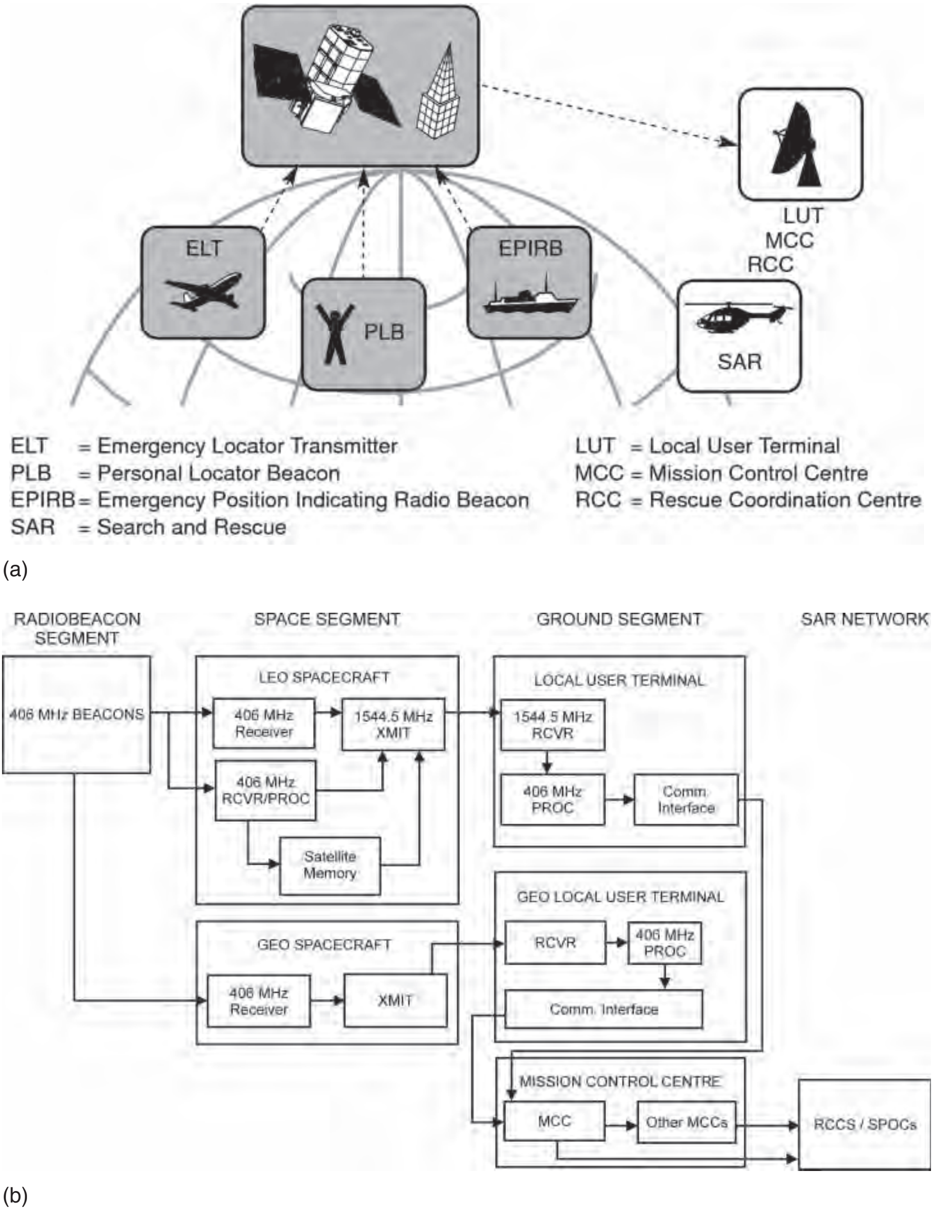


Figure 13.2 (a) Pictorial representation of the Cospas-Sarsat system. (Source: Cospas-Sarsat, 2013. Reproduced with permission of Cospas-Sarsat.) (b) A block schematic of the Cospas-Sarsat system. (Source: Cospas-Sarsat, 2009. Reproduced with permission of Cospas-Sarsat.)

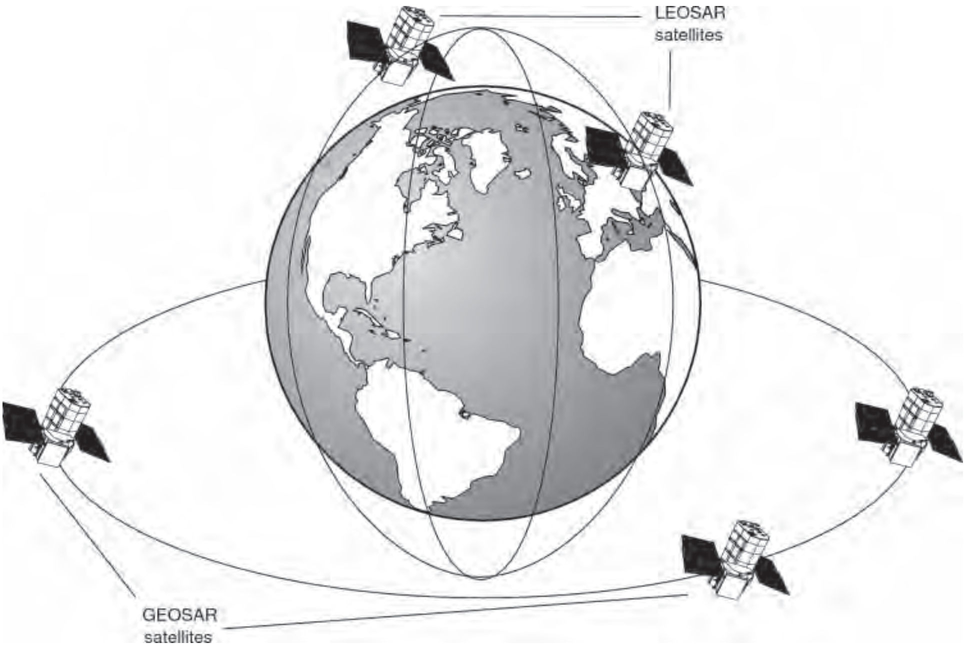


Figure 13.3 Cospas-Sarsat constellation – conceptual diagram. (Source: Cospas-Sarsat, 2013. Reproduced with permission of Cospas-Sarsat.)

system overlay, called the Geostationary Orbit Satellite Search and Rescue (GEOSAR) system. The LEOSAR system consists of an on-board processing payload and a transparent payload. GEOSAR system uses a transparent repeater. The beacon transmitter operates within 406.0–406.1 MHz band reserved exclusively for distress satellite beacons. These constellations complement each other and are illustrated conceptually in Figure 13.3. The GEOSAR system has the advantage of providing near-instantaneous alerting within about $\pm 76°$ latitude and LEOSAR provides true (but intermittent) world-wide coverage with its own position estimation capability. Furthermore, the dynamically changing Earth-Space path profile of LEOSAR satellite helps to reduce the probability of total signal blockage. By using both the sources of alerts, the probability of false alarm is reduced. Location accuracy can be improved by utilizing the Doppler information of both the segments or by combining location data supplied independently by each segment.

LEO was selected to maximize the Doppler frequency change that is used in the system for position determination, and the low path loss (relative to medium earth orbit (MEO) or GEO) that facilitates reception of transmissions from low power beacon transmitters at LEOSAR satellite. Each satellite, in a near polar orbit, can cover the entire Earth; however, due to the limited number of satellites used in the system, the coverage is intermittent. When the constellation is replenished, more than four satellites become available, until the oldest satellite(s) become unusable. Two Cospas satellites are supplied by Russia and the remaining two by the USA, piggy-backed on NOAA (National Oceanic and Atmospheric Administration) meteorological satellites. The Sarsat payload of 121.5 and 406 MHz is supplied by Canada and France. Some satellites also have a 243 MHz payload. However, due to the limitations in the use of 121.5/243 MHz their use was terminated on 1 February 2009 in favour of 406 MHz, which provides superior accuracy with worldwide coverage and includes a format to carry encoded data such as beacon identification and location data computed by beacon equipment. Cospas satellites are placed at an altitude of 1000 km and near-polar orbit and NOAA satellites are 850 km-altitude Sun-synchronous satellites at an inclination of 98°. At these altitudes, the orbital period is about 100 min and satellites cover a 4000-km-wide swathe as they move. The entire Earth is covered within 12 h. Thus with four satellites distributed in various orbital planes, a waiting time is typically less than 1 h at mid-latitudes and 2 h in the equatorial region. Visibility time from the ground is of the order of 10–15 min on each pass.

LEOSAR satellite platform is not dedicated to the SAR mission alone – it is piggy-backed on other payloads. The transponder comprises 406 MHz repeater unit for retransmission of distress signals in the local coverage mode and receiver-processor and memory units for signals received on 406 MHz for retransmission in the local and the global coverage mode. In local mode the connectivity is limited to the field of view of each satellite whereas in the global mode the alert data is stored and transmitted continually so that all the LUTs receive the alert. The repeater accepts 406 MHz inputs and retransmits them at 1544.5 MHz after frequency conversion and amplification. The receiver-processor unit demodulates the received signal, measures and time-tags the frequency. This data is included in the frames that are transmitted at 2400 bps.

The GEOSAR payload is similarly shared with other payloads. It comprises the 406 MHz antenna and receiver, and downlink transmitter where downlink frequency depends on the geostationary platform used.

Future Cospas/Sarsat space segment will utilize 406 MHz SAR repeater payload hosted on Global Positioning System (GPS), Global Navigation Satellite System (GLONASS) and Galileo medium earth orbit satellite constellations. The Medium Earth Orbit Search and Rescue (MEOSAR) system, expected to be operational in 2016–2017 time frame, will provide near-instantaneous global coverage with beacon locating capability, space segment redundancy and improved resistance to signal blockage. Figure 13.4 shows the configuration of the planned MEOSAR system. The return link will be used for receiving short messages at the user end such as successful reception of message sent by the RCC or other types of messages sent by a return link service provider on user's beacon equipment that include receive capability.

The geostationary overlay uses a number of satellites totalling six in 2013 as listed in Table 13.1.

Figure 13.5 illustrates the footprint of the GEOSAR constellation.

Due to low Doppler variations from geostationary satellites, the Doppler position determination technique is inaccurate. The position information is therefore obtained through another navigation source for operation with the GEO system or through a LEOSAR satellite when it crosses the distress site. Recall that beacon position can be derived from the LEOSAR constellation using the Doppler information. The signals transmitted by the distress transmitter are received by ground stations called geostationary local user terminal or GEOLUT.

The ELT used in aviation can be activated manually or automatically in the case of an aircraft crash; similarly, an EPIRB can float and transmit radio beacons automatically or can be activated manually in the case of a sinking ship. The radio beacons, transmitted at

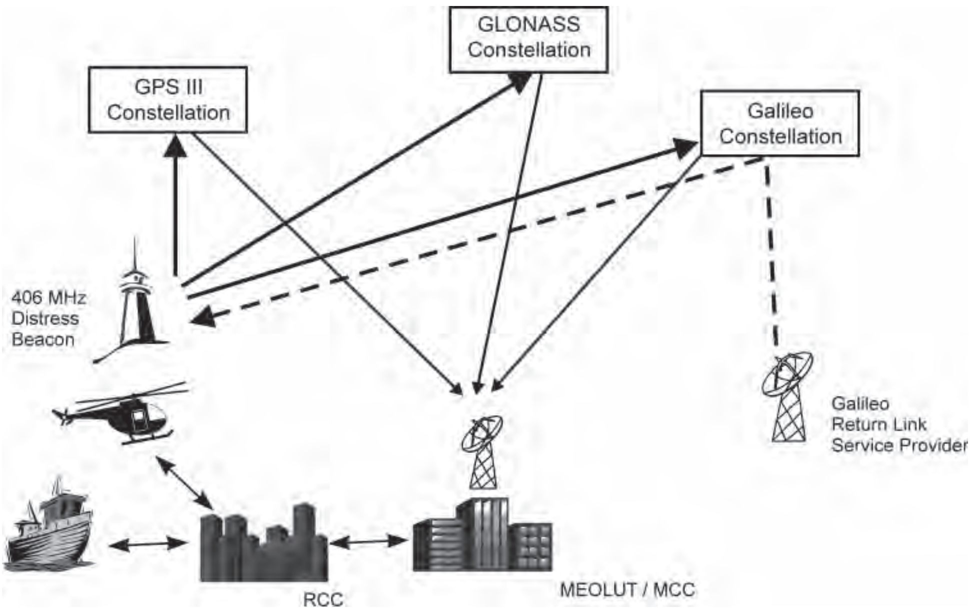


Figure 13.4 Configuration of the proposed MEOSAR system. (Source: Cospas-Sarsat, 2011. Reproduced with permission of Cospas-Sarsat.)

Table 13.1 Geostationary satellite overlay in the Cospas-Sarsat system

GOES-12 East (60°W)	Stand by	
GOES-13 East (75°W)	F	Operational GOES-East satellite
GOES-14 (105°W)	Stand by	In-orbit spare
GOES-15 (135°W)	F	Operational GOES-West satellite
INSAT 3A (93.5°E)	F	
MSG-1 (9.5°E)	F	
MSG-2 (0°)	F	
Electro-L1 (76°E)	F	

F = Functional.
(Data source: Cospas-Sarsat, 2013.)

a frequency in 406 MHz distress band, are transponded back by the receiving satellite to a large ground station known as a low earth orbit local user terminal (LEOLUT); as the LEO satellite in view of the transmitter moves rapidly with respect to the transmitter, the received signal is Doppler shifted. The position of the observer can be estimated at the LEOLUT, either independently when signals are received via LEO satellite repeaters or with assistance of partially processed data (i.e. time-tagged frequency) when using satellite's on-board processing payload, by measuring the Doppler frequency and knowledge of the

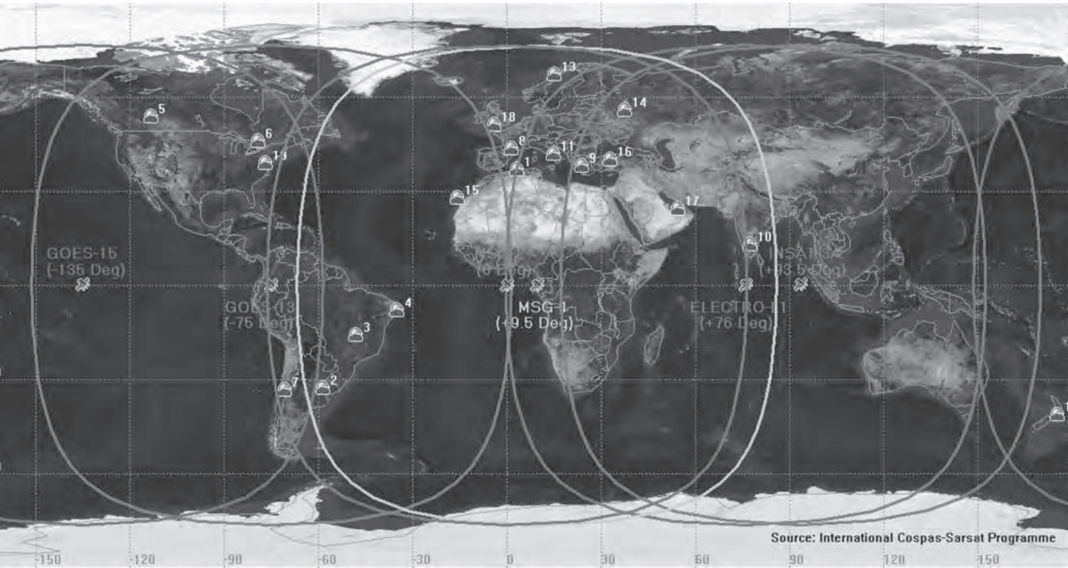


Figure 13.5 Geostationary footprint of Cospas-Sarsat. (Source: Cospas-Sarsat, 2013. Reproduced with permission of Cospas-Sarsat

satellite's orbital location. Usually, two passes were needed to resolve the position ambiguity when using the (now obsolete) older 121.5 MHz system; however, the introduction of a more stable 406 MHz beacon enables ambiguity to be removed in a single pass. To ensure reliability, consistency and compatibility, each LEOLUT has to comply with Cospas-Sarsat specifications and procedures.

Up to four channels could be supported by a LEOLUT when 121.5 and 243 MHz Search and Rescue Repeater (SARR) were in use. Since the removal of transmissions on the latter frequency bands, support of only two channels is adequate – that is, the 406 MHz Search and Rescue Processor (SARP) channel; and 406 MHz SARR channel. The SARP channel comprises partially processed data, which provide identification, transmit time and received frequency for each distress burst of 2400 bps. The SARR channel provides unprocessed transponded data.

In the local mode satellite transmissions are received by LUTs within the footprint of each LEO satellite. In the global mode the beacon signal is partially processed and stored at the satellite and transmitted for several orbits, ensuring that all the LUTs have received the signals. Since the message is received by a large number of LUTs, a high redundancy is built into the system. Most LUTs are fully automated; many are unmanned, operated remotely from a MCC. These ground stations deploy a tracking antenna that generally uses programme tracking. The transmission formats of all the LEO satellites are compatible with LUT capabilities. Each LUT must comply with the organization's system specifications.

A distress message received at an LUT is passed to the MCC, which transfers the information to an appropriate RCC or rescue points of contact (RPOC) or if outside its jurisdiction then to a better placed MCC, after removal of repetitions and false alerts. The MCC also provides system information to assist operations, including satellite ephemeris and time calibration used for beacon location, coordination of messages and status of the space and ground segments. To exchange a free flow of information, the MCCs are interconnected. MCCs have to comply with Cospas-Sarsat's performance specifications to meet the organization's reliability and system integrity standards. Regular exercises are performed to ensure correct functioning and compliance of various elements of the overall system.

The first generation 121.5 MHz transmitters, credited with saving many lives, were discontinued due to a number of deficiencies as the system was not optimized for operation with satellites. These amplitude modulated 0.05–0.1 W transmitters exhibited poor frequency stability limiting the number of simultaneous transmissions to 10; and the location accuracy was limited to ~20 km with little information about the user's identity.

The 406 MHz beacon transmitters are designed to operate with satellite systems and therefore include improved features, such as better Doppler location accuracy and ambiguity resolution, increased system capacity, high peak power output (5 W), low duty cycle (around 1/2 s bursts at 400 bps every 50 s), higher frequency stability, unique identification code for each transmitter, digital transmissions permitting storage on satellite and the use of ITU assigned distress band. Consequently, 406 MHz has a larger system capacity of 90 simultaneous transmissions, higher location accuracy (~2 km), user identification and negligible interference from aircraft emissions. The frequency range of the beacons lies within 406.025–406.037 MHz. The low bit rate message, containing information about country of origin, originator's identification and position data (optional) phase modulates the carrier. Although not mandatory, beacon equipment can include a homing transmitter at 121.5 MHz to facilitate SAR teams to locate the alert site. The beacons can continue transmissions for up to at least 24 h at a minimum temperature of −40°C or −20°C (depending on the beacon class) with the capability to operate up to +55°C (Cospas-Sarsat, 2009).

The message structure comprises a synchronization pattern, country code number, identification or identification + position data, error correcting code, supplementary data and optionally, additional data. Short message consists of 112 bits and long message consists of a 32 bit extension.

13.3 Navigation Systems

13.3.1 Background

Prior to the advent of radio systems, sailors used celestial bodies, known landmarks, lighthouses, etc. as their navigation aids. Navigation received an impetus after the invention of radio by extending the navigation range far beyond the traditional visual aids, and overcoming the limitation of visual navigation during adverse weather conditions. The first radio systems, providing continuous reference signals, were installed in 1921 for terrestrial/maritime usage, at about the same period when airborne navigation aids were installed. Navigation systems have continually improved since.

The *Loran* (*long-range navigation*) *system* was developed for marine and air-borne applications. The system is based on the time difference navigation principle, wherein the time difference in signals arriving from a reference station and a slave station are measured at a receiver. Points where signals are received with equal time difference lie on a hyperbola, and if two slave stations are used, a pair of parabolas is created. The receiver lies at the intersection of the parabolas. Loran-C operates at 100 kHz covering a range of over 3330 km from the master station. Its use continues to decline in favour of satellite navigation systems; several countries have terminated Loran-C service. Enhanced version of the system called e-Loran have been introduced in the UK to meet the needs of the maritime community around the UK and western Europe when the performance of satellite navigation system (GPS) is degraded or the satellite signals are unavailable.

The *Decca system* is named after the British company that introduced it in 1946. The system uses the same principle as the Loran system. However, the reference and slave stations transmit continuous signals instead of the bursts transmitted in the Loran system, and phase difference is measured instead of time difference. The system, operated in 70−129 kHz band, and was widely used in ships, to provide highly accurate fixes in the 160−480 km range. The system has been superseded by satellite navigation systems.

The *Omega system* is based on the same hyperbolic principle as Decca. This long-range system operated at very long wavelengths and provided world-wide coverage through eight synchronized sites, transmitting encoded signals continuously. Its use has now been terminated due to the emergence of satellite navigation systems.

A terrestrial system known as *very high frequency omnidirectional radio range* (*VOR*), evolving since 1930, continues to be in use. The system consists of a very high frequency (VHF) omnidirectional transmitter, transmitting two signals simultaneously. An aircraft measures the phase difference between the two, which are calibrated and displayed to the pilot as a bearing. VHF transmissions minimize propagation-related effects and interference. The VOR, used in conjunction with distance measuring equipment (DME), provides a basic navigation to aeroplanes. An internationally approved DME standard consists of an aircraft radio transmitter that transmits on one of the 126 approved frequencies when requiring a distance measurement. The signal is received at a ground station and transponded back. The time difference between the transmitted and received signals, excluding the ground processing delay of 50 μs, is estimated at the aircraft and used to calculate the range.

The era of satellite navigation dawned in the early 1960s when a system named Transit was launched by the USA. Following a series of developments in satellite navigation technology, the GPS was launched by the USA and Global Navigation Satellite System (GLONASS) by the former Soviet Union. These systems are expected to remain the dominant world-wide navigation system at least up to about 2015, with myriads of civilian applications entwining a multi-billion-dollar industry. Beyond this timeframe a number of emerging global and regional systems are likely to be available (see Table 13.3). The GPS system is addressed in more detail later in the section.

Table 13.2 traces the accuracy of various navigation systems, demonstrating the evolution of navigation systems (Schänzer, 1995; UN, 2010).

Satellite navigation systems are an integral part of mobile satellite systems, with symbiotic applications ranging from vehicle fleet management and terminal position reporting for mobility management to spacecraft attitude and orbit control (AOC) systems.

13.3.2 Satellite Navigation Principles

Navigation systems may be categorized as *active* or *passive* depending on the extent of user participation; and as *single* or *multi-satellite* contingent on the preferred navigation technique. In active systems, users transmit a signal that is processed at a hub for estimating user location. In passive systems, users receive signals and process them to estimate their location. For either mode single or multiple (1–4) satellites can be used.

In single satellite systems, location is fixed by measuring the Doppler shift and estimating the user position using the satellite ephemeris broadcasted by each satellite. The solution for intersection of the Doppler ‘sphere’ with the Earth provides two possible locations – the ambiguity is resolved by considering the deviation caused by rotation of the Earth. Accuracy can be improved by transmission at two frequencies, which allows application of group delay correction. The principle is used in the Sarsat and Argos (French) systems which are active systems, and was used in Transit (USA) and Cicada (former Soviet Union) systems, which were both passive systems.

Position can also be calculated by estimating the range from two satellites. The user position lies at the intersection of three spheres – those around each satellite and the Earth. This technique is used in the OmniTracs (USA) and EutelTRACS (EUropean TELecommunication and TRACking System) (Europe) systems, which are identical active systems.

In a three-satellite system, range is estimated through three satellites; the user is at the intersection of three spheres centred at each satellite and the Earth. A fourth satellite becomes necessary to resolve uncertainties in range estimation (see next section). The Navigation System with Time and Ranging Global Positioning or Navstar GPS (USA), more commonly called GPS, and the Global Orbiting Navigation Satellite System or GLONASS (Commonwealth of Independent States), and Galileo (European Commission, EC) all passive systems utilize this principle.

Table 13.2 Indicative accuracy of various types of navigation system

System name	Positional accuracy (2 standard deviation) m	Comments
Differential GPS	0.3–6	Accuracy depends on distance to reference station
Instrument landing system (ILS)	5–10	Used for aircraft landing
Microwave landing system (MLS)	5–10	Developed to replace ILS but never took-off due to cost and other reasons
Galileo: Single frequency/dual frequency – open service	Horizontal: 15/4 Vertical: 35/8	Target performance with various assumptions: user terminal, local environment: clear visibility, no interference, reduced multipath environment, mild local ionospheric conditions, no scintillation
GLONASS	12.4 (as specified for GLONASS-M)	Neglecting user clock bias and errors due to propagation environment and receiver; assuming position dilution of precision availability = 2 Average measured range error = 5.46 m over ~ 4.8 months in 2009 against specification of ≤ 6.2 m (2σ) (UN, 2010)
VOR and DME	60–180	Used for aircraft navigation; see text
GPS (C/A code)	≤ 3 horizontal/5 vertical	Civilian version of GPS in 2008
Transit/Cicada	450	Historic: Early satellite systems for military/civilian applications; see text
Omega	3600–7200	Historic: No longer in active use
Aircraft inertial navigation	300 to 20 km	Best accuracy near take-off/touch down

(Data source: Schänzer, 1995; UN, 2010 and others.)

These navigation systems are categorized generically as global navigation satellite system (GNSS). The principle to derive a fix based on ranging from three to four MEO satellites are identical in all the GNSS systems. Implementation details between systems vary in the physical layer, transmission parameters including message format, ground network topology and characteristics of the constellation. Technical details of GPS systems, presented in the next section, are representative of this class of satellite navigation system. Concept of differential GPS is included, where the accuracy of the fix provided by GPS is augmented through correction data derived at a neighbouring reference site and transmitted to users over a separate radio channel. Due to spatial and temporal correlation in error over quite large distances, the corrections can be applied by users in large areas around the reference station. Satellite-based augmentation systems (SBASs) utilizing differential principles are used in aviation and other industries to improve the reliability and accuracy of the navigation system fix by utilizing additional independent satellite ranging signals and corrections.

13.3.3 Navigation System Examples

A variety of navigation systems have developed and evolved over the years. Their applicability depends on the desired accuracy, environment, cost and to some extent historic association. Global Navigation Satellite System (GNSS) refers to a generic global architecture subsuming all the global navigation systems including the GPS, GLONASS, Galileo and other planned and upcoming global and regional systems to facilitate strengthening and improvement of safety standards. Compatibility and interoperability are the key requirements of such a concept (UN, 2010). Compatibility ensures a harmonious blend of systems either together or individually. Interoperability refers to the concept that these navigation satellite systems and their augmentations can be utilized in concert to improve the overall navigation capability. Thus, the reliability and accuracy of navigation fixes can be improved by processing signals of more than one constellation.

Following the experience and success of GPS and GLONASS systems, several systems are being introduced in various parts of the world. Tables 13.3 and 13.4 (UN, 2010) summarize salient features of a number of existing and planned navigation and SBASs.

The Transit system used the Doppler signature technique (Yionoulis, 1998). It comprised a constellation in a 1100 km circular polar orbit, each satellite transmitting continuously at about 150 and 400 MHz. The signals were used by the user to estimate the Doppler shift and rate of Doppler change by comparing the measured and the expected frequency. The Doppler signature changes sign at the closest approach to the observer at which point the user is located on a perpendicular line below the satellite path. The processing provides two results one each on either side of the orbital path and this ambiguity can be resolved by taking the Earth rotation into consideration. The system provided an accuracy of about 160 m at the sea level, adequate for ship navigation but not for aircraft navigation due to uncertainties in position fixes caused by extraneous Doppler components introduced by aircraft motion. A Doppler uncertainty of 1 m/s would introduce an error of 1 km. Moreover, the system was not available continuously – users having to wait up to 2 hours for a satellite pass. The system was superseded by GPS. Cicada, operated by the former Soviet Union, used the same principle.

The limitations of the Transit system were removed in the Navstar GPS and GLONASS navigation systems (see Section 13.3.3.1).

The Argos system and the Cospas-Sarsat are both single-satellite active navigation system. The Argos system uses navigation transponders on NOAA satellites that operate in ~ 805 km polar orbit (Argos, 2008). Users transmit messages at 90–200 s repetition in the 401 MHz band, which get relayed by the satellite to hub stations where the Doppler curve is extracted. Intersection of the Doppler 'sphere' with the Earth provides the location of the user on resolving the East–West ambiguity. The fix is available to the user through the Argos website, e-mail or manual dispatch.

13.3.3.1 Global Positioning System (GPS)

The GPS was developed by the US government for the Department of Defense. It is essentially a US military system, but offers navigation services to civilians; Position fix is obtained in passive receivers by the triangulation method, wherein estimated ranges from four satellites are used to derive the position and altitude of a point. Ranges from three satellites can provide the latitude and longitude of a point on the Earth; the addition of a fourth satellite

Table 13.3 Salient features of a number of existing and planned navigation and satellite-based augmentation systems

System	Nominal constellation size	Year of full operational capability	Number of operational satellites	Coverage	Civilian spectrum (key in Table 13.5)
GPS	24 MEO	1995	30 (January 2010)	Global	Current 2009: L1 C/A, L2C Future: L1 C/A, L1C, L2C, L5
WAAS (augmentation system)	3 GEO	2008	2	Regional (North America)	Current: L1 C/A, L5 future: L1 C/A, L5
GLONASS	24 MEO	1995 (GLONASS) 2010 (GLONASS-M)	24	(December 2010) Global	Current 2007: L1PT L2PT future: L1PT, L2PT, L3PT[a], L1CR[a], L2CR[a], L5R[b]
SDCM	2 GEO	2014 (expected)	2	Wide area (Russian federation)	SBAS L1 C/A
GALILEO	30 MEO	2014	2 medium earth orbit satellites	Global	E5 OS/SoL[c] E6 CS/PRS[c] E1 OS/SoL/PRS[c]
EGNOS (augmentation system)	3 GEO satellites	2009 for open service 2010 for safety-of-life Service	3 geostationary satellites	Regional	Current: L1 C/A
COMPASS/ BeiDou	5 GEO and 30 non-GEO	2020	6 (January 2010)	Global	1559.052 ~ 1591.788 MHz, 1166.22 ~ 1217.37 MHz, 1250.618–1286.423 MHz

(*continued overleaf*)

Table 13.3 (*continued*)

System	Nominal constellation size	Year of full operational capability	Number of operational satellites	Coverage	Civilian spectrum (key in Table 13.5)
GAGAN (augmentation system)/IRNSS	3 /7 (3 GEO, remaining in inclined GSO)	2013/2014	–	Regional	GAGAN: L5, L1 IRNSS: S L5 AND L1
MSAS (augmentation system)	2 GEO	–	2 geostationary satellites (MTSATs, Multi-functional Transport Satellite)	Asia and the Pacific	L1
QZSS[d]	1 (first phase) 3 (second phase) NGEO	–	1 (FY 2010)	Regional (Asia and Oceania)	L1 C/A, L1C, L2C, L5, L1-SAIF (L1 – sub-metre class augmentation with integrity function) LEX (L-Band experimental signal)

[a] Signal structure is under refinement.
[b] Pending final decision.
[c] See section dealing with Galileo later in this chapter.
[d] The QZSS proceeds to the second phase of public-private cooperation after the evaluation of the results of technological verifications and demonstrations of the first phase.
C/A = Course acquisition; MSAS = Multi-functional Transport Satellite Satellite-based Augmentation System; QZSS = Quasi-Zenith Satellite System; GAGAN = GPS-aided GEO-Augmented Navigation System centre frequency and bandwidth; IRNSS = Indian Regional Navigation Satellite System; NGEO = non-geostationary earth orbit.
(Adapted from UN, 2010.)

Table 13.4 Key to centre frequencies used by satellite navigation and satellite-based augmentation systems listed in Table 13.3

System	Band and centre Frequency (MHz)	Brief description
GPS	L1 = 1575.42 L2 = 1227.6 L2C = 1227.6 L5 = 1176.45	Navigation system: USA (See text)
WAAS	L1 = 1575.42 L5 = 1176.45	Satellite based augmentation system: USA (see text)
Galileo	E1 = 1575.42 E6 = 1278.75 E5a = 1191.795 E5b = 1207.14	Navigation system: EC (see text)
EGNOS	L1 = 1575.42	Satellite based augmentation system: EC (see text)
GLONASS	L1 band: 1598.06 ~ 1604.40 L2 band: 1242.94 ~ 1248.63	Navigation system: Russia (see text)
Compass	B1 = 1575.42 B2 = 1191.795 B3 = 1268.52	Navigation system: China:
MSAS	L1 = 1575.42	Satellite based augmentation system: Japan
QZSS	L1 C/A and L1-SAIF = 1575.42 L1C = 1575.42 L2C = 1227. 6 L5 = 1176.45 LEX = 1278.75	Navigation system: Japan
RNSS	SPS – L5 = 1176.45, 24 MHz wide RS – L5 = 1176.4, 24 MHz wide SPS – S = 2492.028, 16.5 MHz RS – S = 2492.028, 16.5 MHz	Indian Regional Navigation Satellite System
GAGAN	L1 – 1576.42 L5 – 1176.45	India satellite based augmentation system

RS = Restricted Service.
SPS = Standard Positioning Service.
(Adapted from UN, 2010.)

can provide a user's altitude and correct receiver clock error. It is also possible to derive the velocity of the user and precise time information originating from on-board atomic clocks, which have a drift rate of 1 s per 70 000 years. There are two rubidium and two caesium clocks aboard each first generation satellite.

Two types of pseudo-random codes are transmitted on two frequency bands, L1 (1575.42 MHz) and L2 (1227.6 MHz) – the encrypted Precise-Code (P-Code) meant for US military operation is available on both frequencies, and the unencrypted coarse acquisition code (C/A-code), available without restriction, is transmitted only in the L1 band, where it is combined with P-Code in phase quadrature. Codes have low cross-correlation, allowing transmissions from each satellite on the same frequency. Spread spectrum modulation also provides some resistance against multipath and immunity to interference. The C/A code, operating at 1.023 Mbps, is a 1023-bit pseudo-random code repeating each millisecond, and the P-code, operating at 10.023 Mbps, has a cycle of 267 days but it is reset every seven days. Each code is combined with a navigation message comprising the status of the satellite, time synchronization information for transferring from coarse to fine code, clock correction, satellite ephemeris, propagation delay corrections and approximate ephemeris and status of the constellation useful for signal acquisition.

One of the key features of GPS modernization programme has been to increase the number of transmissions with up to five transmissions on GPS III satellites. We will see later that transmissions on multiple frequencies are used for improving the accuracy of navigation fixes. The 2σ accuracy obtained from well-designed C/A code receiver is reported to be ≤ 3 m, horizontal/5 m, vertical (UN, 2010). To avoid the system being used against itself, the military can degrade the accuracy by using selective availability (S/A), which increases the error probability by satellite clock dithering and corruption of navigation message data. There is a provision to switch on or off the S/A function. Nevertheless several techniques exist for improving the accuracy of C/A code fixes.

The GPS nominal constellation comprises 21 satellites and three in-orbit spares. Satellites are in a circular orbit at an altitude of about 20 200 km (orbital period of about 12 hours). Table 13.5 (Daly, 1993) summarizes the main orbital parameters of GPS, comparing it with the GLONASS system (see Section 13.3.3.2).

The range is estimated as the product of the time taken for a signal to travel from the satellite to a receiver. To estimate range precisely, each user clock must be synchronized to

Table 13.5 Navstar and GLONASS orbital parameters

Orbital parameter	GPS	GLONASS
Type of orbit	Near circular	Near circular
Semi-major axis (km)	26 560	25 510
Period (minutes)	717.97	675.73
Inclination (degree)	63.0 – block I 55.0 – block II	64.8
Orbital separation (degree)	60	120
Ground track repeat (days)	1	8
Drift per day (minutes)	−4.06	−4.07

(Adapted from Daly, 1993.)

the satellite, which makes the receiver complex if three satellites are used and hence conflicts with the need of a simple receiver. In the GPS system, the problem is resolved by estimating the range from a fourth satellite, which then allows resolution of user clock uncertainty.

The travel time of the signal is estimated by measuring the time shift between identical codes generated at the satellite and the receiver. The code generated at the receiver is time shifted until a maximum correlation is achieved between the transmitted and receiver codes; the time-shift provides an approximate range or 'pseudo' range, which includes numerous errors listed next. The user must have knowledge of code to be able to use the system; this feature permits the military to use the higher accuracy P-Code. Since the received signal is around 22 dB below the receiver noise code acquisition and de-spreading is necessary prior to clock recovery.

Satellite clock offset is relatively small in comparison to receiver clock and moreover the correction is available in satellite's message. The receiver's clock offset is estimated by solving the navigation equation.

In order to fix a position on the Earth a terrestrial reference system is necessary. In such a system, the Z-axis coincides with the earth rotation axis; the X-axis is associated with the mean Greenwich meridian; and the Y-axis is orthogonal to the other two axes to complete the right-handed coordinate system. A number of earth reference systems are used based on their approximation to fit the shape of the Earth. These include World Geodetic System 84 (WGS 84) the terrestrial reference system used in the GPS; and International Terrestrial Reference Frame (ITRF) established by Central Bureau of the International Earth Rotation Service (IERS). The GLONASSS system uses an earth reference frame known as Parametry Zemli 1990 (Parameters of the Earth 1990) (PZ-90). The Galileo system uses the reference Galileo Terrestrial Reference Frame (GTRF) (Sanz Subirana, Juan Zornoza and Hernández-Pajares, 2011). The difference between these reference systems depends on the reference station used to establish the model. There are regular updates to the parameters of these models and the parameters for transformation between them (Princeton University, 2013).

Pseudo-range R_p is given as:

$$R_p = R_a + c\Delta t_e + \mathrm{c}(\Delta t_r \pm \Delta t_s) \tag{13.1a}$$

Where, R_a = Actual range

$$R_a = \sqrt{(X_s - X_r)^2 + (Y_s - Y_r)^2 + (Z_s - Z_r)^2} \tag{13.1b}$$

Where subscript s and r respectively represent the Cartesian coordinates X, Y and Z of satellite and receiver respectively.

Δt_e=propagation delay inclusive of various error components
Δt_s=satellite clock-offset with respect to GPS time
Δt_r=receiver clock offset with respect to GPS time
c=velocity of light

The true range is estimated from the pseudo-range by solving a set of four simultaneous equations called navigation equation obtained by substituting Equation 13.1b) into Equation 13.1a). The matrix is populated by pseudo-range measurements from four satellites.

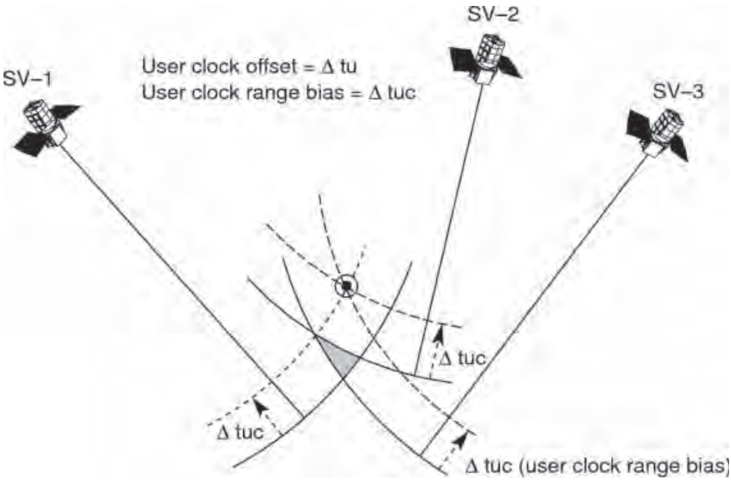


Figure 13.6 Concept of position fix and impact of errors in two dimensions. (Source: Milliken and Zoller, 1978. Reproduced with permission of The Institute of Navigation.)

The solution can be explained qualitatively by observing that the receiver lies on the intersection of three sphere of radius R_{pi}, R_{pj} and R_{pk} where subscripts i, j and k represent pseudo-range distances from satellites i, j and k respectively. The fourth measurement is used to resolve clock accuracy of the receiver. Various algorithms are used to solve these equations algebraically and numerically.

The speed with which position is calculated can be improved by using only three measurements, and traded off against accuracy and receiver complexity. The transmissions can also be used for determining user velocity by extracting the Doppler information or indirectly by taking the time derivative of distance travelled.

Figure 13.6 represents the concept of a two-dimensional navigation solution including range errors. The initial position is estimated within a zone of uncertainty marked by hatched portions. When uncertainties are removed, the solution converges to a single point, assuming that the resulting error is negligible.

There are a number of sources of error in range estimation:

- **Satellite clock offset relative to GPS system time and ephemeris errors**: GPS system time is maintained by the GPS master control station (MCS) through a set of highly accurate caesium clocks; the clock offset of satellites is measured daily and transmitted to each satellite by the MCS for retransmission to receivers, which apply the correction algorithmically; satellites themselves incorporate highly stable atomic clocks on-board. There is always some inaccuracy in estimating satellite ephemeris. Certain components in ephemeris errors cannot be isolated from satellite clock offset errors and are therefore combined with the satellite clock error in the error budget (see Table 13.6).
- **User clock offset from GPS system time**: As mentioned in the preceding text, user offset can be removed by solving the range equation.

Table 13.6 Estimated one standard deviation range errors

Error source	Estimated 1 standard deviation residual range error (m)
Satellite clock and ephemeris	1.5
Atmospheric delays	2.4–5.2
Satellite equipment group delay	1.0
Multipath	1.2–2.7
Receiver noise and resolution	1.5
Root sum square	3.6–6.3

(Adapted from Milliken and Zoller, 1978.)

- **Error due to propagation delay**: An error in range estimates is caused by delay introduced by ray bending and velocity reduction while traversing the ionosphere. The delay is approximately inversely proportional to the square of the frequency and the correction can be derived by comparison at two frequencies, 1227.6 and 1575.42 MHz. In the GPS system, it is derived at the MCS and transmitted to users. Tropospheric delays are independent of frequency and can be estimated at the receiver by applying an elevation-angle-dependent correction.
- **Group delay** error is caused by processing delay on a satellite; its value is estimated in ground tests and transmitted to users along with other corrections.
- **Multipath** errors are caused by signals arriving at the receiver from different paths.
- **Receiver noise and resolution degradation** are caused by hardware and software limitations in the receiver. Receiver motion can introduce additional errors. Use of well-designed receivers and filtering algorithms such as the Kalman filter reduces the impact of the error.

Each component of error is corrected at the receiver to the extent feasible. Table 13.6 summarizes the extent of residual errors after corrections have been applied (Milliken and Zoller, 1978). These estimates have continued to improve with improvements in technology.

The overall accuracy of a fix is determined by a combination of ranging error and the geometry of the satellites used in range estimation; accuracy improves when satellites are wide apart. The effect of geometry is measured as geometric dilution of precision (GDOP); a lower GDOP represents a better geometry. For example, the error is multiplied by 1.5 when satellites are spaced far apart, whereas the multiplier could be five or more for closely spaced satellites. Terminals use an algorithm based on the volume of the tetrahedron formed by the vector joining the user to the satellites for selecting the best satellite combination.

Figure 13.7 shows the architecture of the GPS system. The MEO constellation is controlled by a MCS in Colorado Springs (plus an alternate MCS), a network of four ground antennas and monitoring stations dispersed around the world. The MCS estimates the orbital parameters of each satellite from ranging data collected by the monitoring stations, formats and transmits them to each satellite. The monitoring stations receive GPS satellite transmissions and forward this data to a MCS located at Schriever Air Force Base, Colorado, USA, where this data is analysed, GPS and universal standard times compared and corrections

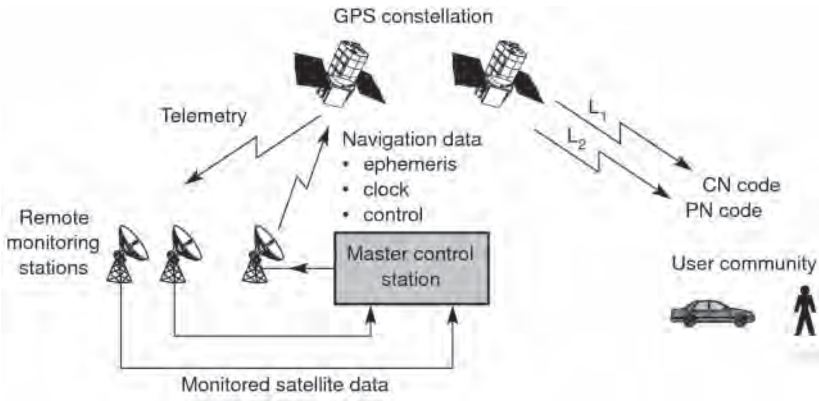


Figure 13.7 The GPS system architecture

for satellite ephemeris and clock data prepared and then relayed to the satellites for user assimilation.

A number of improvements were introduced in the next generation satellites and system. These include:

- **Increase in satellite** EIRP (effective isotropic radiated power): With 50 W satellite transmissions, the level received at the receiver is of the order of −160 dBW for C/A code and −166 dBW for the P-Code, making the system susceptible to extraneous noise. Higher power and frequency diversity improve robustness.
- **Provision of three additional frequencies**: Next generation satellites, called Block III, additionally transmit C/A code in band L2C. Subsequent satellites will also transmit in L5 band (960–1215 MHz) currently used by the Aeronautical Radio Navigation Service and a transmission in band L1C. The advantages of these additional transmissions are outlined later in this section.
- **Auto-navigation feature on GPS satellites**: Satellites will cross-link in space to generate their own ephemeris, enabling them to provide ephemeris data for up to 180 days without ground uploads. At present, the data uploaded from the ground can be saved from 14–180 days (depending on satellite generation), but in the absence of regular refresh from the ground the accuracy deteriorates with time.
- **Mitigation of multipath by use of directional antennas**: Multipath noise degrades the accuracy of fixes, which is not acceptable when very high accuracy is needed (see Table 13.6). Directional antennas tend to reduce the multipath and therefore their applicability is under investigation.

The L2C transmissions, in conjunction with L1 C/A, enable ionospheric correction to improve accuracy, speed up signal acquisition on existing dual-frequency receivers. Its modulation scheme makes reception of signals under trees and even indoors easier and facilitates miniaturization of low-power GPS chipsets for mobile applications. The L5 frequency is in the exclusive aviation safety and radio navigation satellite services band aiming to support safety-of-life (SoL) transportation and other high-performance applications. Thus in combination with L1 band the L5 band will provide an ionospheric correction feature and robustness through redundancy. By additionally including L2C band, the receivers improve signal robustness and enable sub-metre accuracy. The L1C transmissions to be transmitted on GPS III satellites have been designed to support interoperability of GPS with other satellite navigation systems such as Galileo.

As mentioned, the accuracy of a fix depends on a number of factors, such as the prevailing atmospheric and ionospheric conditions. The errors are corrected to the extent possible by signal processing and downloaded correction data. It has been observed that ephemeris and measurement errors are correlated in time and spatially. Spatial correlation decreases with distance rather slowly – correlation is reasonably close even for a distance of a few tens of kilometres. Similarly temporal correlation decreases rather slowly with time. Propagation errors, and S/A errors are quite well correlated in a 5–10 s time span. An accurate estimate of such errors can be obtained when the coordinates of a location are known precisely; if these corrections are transferred to users in the vicinity of the measurement site, and the user applies such corrections the accuracy of the fix improves considerably. Application of this technique is known as differential global positioning system (DGPS), where errors in the navigation solution are derived at a reference site and transmitted over a radio link to GPS receivers present in the neighbouring areas. Reference sites can be located from less than a kilometre to over 1000 km from the users, the accuracy of the fix obviously improves with a reduction in distance of a user to the reference station. Measurements and simulation demonstrate errors ranging from tens of centimetres when the reference site is a few kilometres, to 5 m for a distance of 1000 km. A number of commercial and other bodies responsible for safety operate DGPS systems in various parts of the world. Service is offered through a radio carrier relayed through satellite or terrestrial transmitters. Commercial users include offshore oil platform operators, fleet managers in the trucking industry, and others. Safety-related services are offered, for example by the US Coastguard, which provides differential corrections at 285–325 kHz in coastal areas free of charge (Hall, 2013). Potential beneficiaries include agriculture, forestry, the emergency services, etc.

The accuracy of fix estimates is reduced in the case of failure or unavailability of satellite(s) in the constellation, possibly without the user being aware. An external system that provides such an integrity monitor can be very useful. Geostationary satellites transmitting signals identical to GPS and GLONASS (or, another navigation system as the case may be), containing the integrity of satellites and error estimates, improve the accuracy.

Such an augmentation when applied to civil aviation can revolutionize air traffic control. At present, a majority of aircraft navigation relies on ground-based transmitters or inertial navigation systems when aircraft are over oceans, which can cause them to drift several kilometres off course. Prior to landing, VOR and DME systems (see Section 13.3.1) provide non-precision navigation aid to pilots, which is particularly useful in adverse weather conditions; pilots then use either visual information or an instrument landing system (ILS) for landing. Clearly, GPS or an equivalent satellite system offers promising solutions for in-flight and the approach part of landing, in particular when the cost of VOR systems is taken into consideration. Another limitation of the existing navigation aid is that it requires planes to fly wide apart – separation of 5–8 km requires use of radars, which are not available for transoceanic flights, causing inefficient use of the air corridor; VOR-based systems require separation of up to 8 miles (13 km). The solution being developed is called Automatic Dependent Surveillance-Broadcast (ADS-B), where an augmented GPS receiver on aircraft

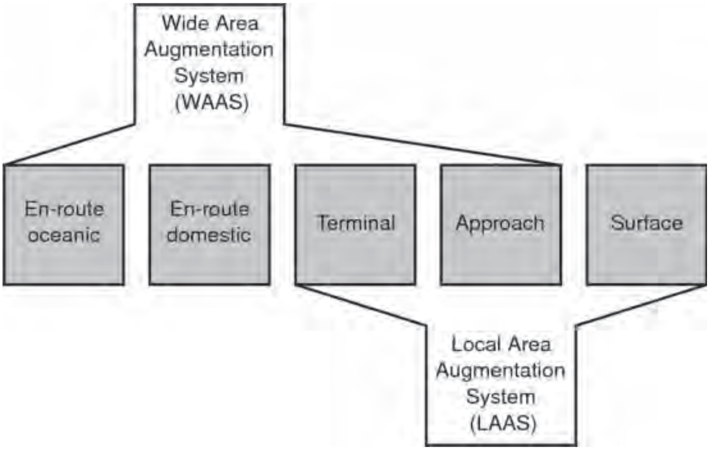


Figure 13.8 An integrated WAAS-LAAS concept. The WAAS system is intended to provide pilots with GPS based navigation information during flight, approach and departure, while LAAS is intended to provide navigation for surface as well as approach

broadcasts its own position to the ground receiver, thereby removing the need for radar, permitting separation to be reduced considerably. The technique is also being developed for the next generation traffic-alert and collision-avoidance system (TCAS); the existing systems are based on radar monitoring.

A DGPS-based navigation solution, known as the Wide Area Augmentation System (WAAS), is represented in Figure 13.8. WAAS provides GPS-based information for en route, departure and approaches under conditions where ceiling and visibility are 200 ft (60.96 m) and ½ mile (804.67 m) respectively (Bretz, 2000). The measured 95% horizontal and vertical accuracy within CONUS (Continental United States) (Grand Forks) during the interval 1 July to 30 September 2012 is reported as 1.46 m horizontal and 1.7 m vertical respectively (FAA, 2012).

The FAA also intends to deploy the Local Area Augmentation System (LAAS), which will provide GPS navigation down to the surface, allowing near-blind landing and airport surface navigation using ground receivers placed in areas within 30–45 km to derive the corrections and transmit them to the aircrafts via terrestrial radio link.

SBAS systems are either being developed or are operational in Europe, India, Japan and Russia (see Tables 13.3 and 13.4 for status up to 2010). The System for Differential Corrections and Monitoring (SDCM) SBAS is a component of GLONASS (SDCM, 2013).

GPS is used widely in mobile communications for location-based applications, such as used for vehicle tracking, network mobility management and spot beam identification. It is also used on yachts and ships and by explorers, mountaineers, scientists, surveyors, etc.; specific examples include delivery or pick-up points for an accurate delivery/pick-up service; real-time response to marine hazards such as oil spills; AOC of LEO/MEO satellites. An insurance company monitors the usage of cars for insurance charges; a journey at midnight is more expensive than during the day, as people are more alert during the day! Low power GPS receivers can now be installed in watches and personal digital assistants; GPS time is in use for synchronizing code division multiple access (CDMA) cellular base stations (essential for functions such as soft handover); medical services use it for responding to life-threatening emergencies.

13.3.3.2 GLONASS

The GLONASS system, developed by the former Soviet Union for military application, was inaugurated in 1982. Table 13.5 compares their orbital parameters. The full constellation comprises 21 satellites with three in-orbit spares at an altitude of about 19 150 km inclined at 64.8°. Satellites transmit on two frequencies, at about the same power level as GPS satellites, but use the frequency-hopped spread spectrum technique. For civil applications, a single 511-bit code, repeating every 1 ms, is used with differential coding in an RZ (Return-to-Zero) format at 50 bps (Daly, 1993; Dale, 1989); the military code operates at 10 times the civilian code rate. Frequencies are 1242.94–1248.63 MHz and 1598.06–1604.40 MHz, the channel spacing for these two bands is respectively 0.4375 and 0.5625 MHz, and the number of channels is 24 in each case. A binary phase shift keying (BPSK) modulation scheme is used in a bandwidth of 1 MHz for civilian applications and 10 MHz for military applications. Navigation data and the principle of position-fixing are similar to GPS. However, there are a number of differences that cause difficulty in integrating them, but nevertheless integration is technically feasible. Some notable comparative system features include:

- There is a large degree of compatibility in orbital geometry.
- The time references of these two systems are different.
- Both systems benefit from a geostationary overlay.

13.3.3.3 Galileo

The European Union is developing a system known as Galileo in order to reduce dependency on the other satellite navigation systems and foster the European navigation market. Studies conducted in Europe estimate the global navigation market to exceed €200 billion by 2020 (EC, ESA and Galileo Joint Undertaking, 2003). The Galileo system alone is predicted to deliver direct and indirect revenue of about €90 billion to the EU economy over the first 20 years of operations. Direct revenues are expected to arise from the space, receivers and applications industries and indirect revenues will result through improvements such as improvements in effective transport systems, rescue operations, and so on (EC, 2013). The Galileo system provides a number of services targeting various user groups:

- A free-of-charge open service *(*OS*)* available to all, which provides positioning and synchronization information capable of an accuracy of ≤ 15 m in the horizontal direction and ≤ 35 m in the vertical direction on a single frequency receiver, improving to ≤ 4 m (horizontal) and ≤ 8 m (vertical) on a dual-frequency receiver.
- A *SoL service for* users for whom safety is essential, guaranteeing a specified service continuity, availability, accuracy and capable of integrity failure alerting.
- A *commercial service* that may include a limited capacity data broadcasting.
- A *publicly regulated* encrypted service for 'Government-approved use in sensitive applications that require a high level of robustness'.
- A *search-and-rescue service* to support the Cospas-Sarsat system.

The SoL service provides a global integrity service for critical safety applications by complementing the dual frequency utilization (E1-E5b) of the OS (explained later). This service complies with the ICAO LPV2007 definition (precision approach with vertical guidance and 200-ft decision height).

The Galileo free-of-charge Cospas-Sarsat SAR service includes a return channel supported on Galileo's OS receivers. The localization accuracy is expected to be ≤ 100 m when used with Galileo OS receivers and 5 km with legacy Cospas-Sarsat receivers.

The nominal Galileo constellation consists of 27 evenly distributed satellites, in three 56° inclined circular orbits – thus there are nine operational satellites per orbital plane. The circular Earth orbits have a nominal radius of about 30 000 km and an approximate revolution period of 14 h.

Galileo transmission bands are listed in Table 13.4. E1 transmissions that include navigation messages provide the *OS* and the *public regulated service*. Integrity message for the *SoL service* is also included in the OS signal. E6 signals support *commercial service* and public regulated service signals both of which include navigation message and encrypted ranging codes. The E5 signal is a wideband carrier, which comprises two independent side bands – the lower sideband called E5a signal provides a second signal to facilitate dual frequency reception for the OS and SoL services including navigation data messages. The upper sideband called E5b signal, containing integrity and navigation messages, supports a SoL service.

Galileo satellites transpond search-and-rescue signals emanating from distress beacon to LUT in the frequency range between 1544 and 1545 MHz to support the Cospas-Sarsat system explained earlier in the chapter.

The European Geostationary Navigation Overlay Service (EGNOS) is a satellite based augmentation service at present (2012) supporting the GPS standard positioning service, where the augmentation signals are transmitted on GPS L1 band. The system provides correction and integrity information to improve positioning navigation services over Europe. The system utilises geostationary satellite transponders for transmissions of integrity information.

13.4 Mobile Very Small Aperture Terminals

VSAT systems have been widely dealt in the literature. Keeping within the scope of the book, we will only summarize technical features of mobile very small aperture terminals (MVSAT) systems. MVSAT systems provide broadband connectivity in the air, sea and land (generally) in K_u and K_a fixed satellite service (FSS) and MSS bands. The commercial emergence of MVSAT systems was driven by the need and expectations of affordable ubiquitous broadband connectivity both in business and consumer telecommunication sectors, which could not be supported at L and S MSS bands due to limited availability of MSS spectrum in them. The MSS spectrum allocation to support comparable broadband services is available in the K_a band but neither the technology nor the traffic volume incentivized the utilization of this MSS band until recently.

The first generation MVSATs use C and K_u FSS bands where sufficient bandwidth and coverage is available. The demand for affordable broadband led to introduction of high capacity K_a band second generation MVSAT systems both in the FSS and the MSS bands. Various technical problems arise in K_u and K_a bands due to terminal mobility and restricted

Table 13.7 A comparison of representative MVSAT and MSS maritime products

	L band MSS (low-end)	MVSAT K_u band (low-end)	L band MSS (upper-end)
Antenna unit diameter (cm)	27	39	69
Antenna unit Weight (kg)	5	11	23
Maximum return data rate (kbps)	284	128	432
Maximum forward data rate (kbps)	284	2 Mbps	432
Price differential	1	1.3	1.5

(Data source: KVH, 2012.)

antenna mounting space. At these frequencies impairments due to the troposphere, particularly rain, are superimposed over those caused by mobility. Shadowing on the radio link gets more pronounced due to a reduction in diffraction advantage and an increase in penetration loss compared to the L and S bands (Butt, Evans and Richharia, 1992). A restricted antenna size leads to lower antenna gain and higher off-axis transmissions.

13.4.1 Rationale

VSAT systems provide high throughput cost-effective communication services to fixed sites on small and affordable user terminals. They suit high volume cost-sensitive users such as a remote office staff and residents in remote locations without telecommunication infrastructure.

The VSAT technology has evolved to the extent that terminal size, service and air-time costs are now affordable by small businesses and individuals. The demand for data services continues to escalate and, as such, users expect services in the mobile sector comparable to those in the fixed network. The traditional L and S MSS bands are congested so that wideband capability is severely curtailed and expensive. The fixed satellite service authorized for VSAT transmissions in K_u and K_a bands provide sufficient bandwidth to support broadband at low cost/bit on VSATs small enough to be mounted on mobiles. The VSAT industry aware of the trend has, therefore, introduced MVSAT systems as an alternative to the L and S band mobile satellite system. These systems are in now regular use in all the mobile telecommunications sectors – air, land and sea. As explained later, the mobile environment poses various technical challenges for VSATs to operate in these frequency bands (Arcidiacono, Finocchiaro and Grazzini, 2008).

Table 13.7 (KVH, 2012a) illustrates a comparison between contemporary L band MSS and MVSAT broadband products for maritime applications (2013 base). Notice the throughput advantage offered by K_u band MVSAT and the insignificant differential in MSS and MVSAT user terminals prices. Later in the chapter we will see that airtime charges of MVSATs are significantly lower.

The reader will recall that Chapter 12 covered the applicability of VSAT-like Ku band terminals for the reception of direct broadcast satellite signals and Chapter 8 discussed the extension of digital video broadcasting/return channel by satellite (DVB/RCS) standard to a mobile environment (recall, DVB/RCS+M). The mobile DVB standard thus provides a platform to support the emergence of MVSATs as a viable commercial proposition for the mobile broadband sector.

13.4.2 Issues

Various aspects differentiate L and S band MSS technologies and fixed VSAT technology from those of the MVSATs; these differences set a datum for MVSAT systems.

1. VSAT systems belong to the FSS, which implies that regulatory matters pertain to fixed installations.
2. VSAT radio link performance, optimized for static links and rain fading in K_u and K_a bands, is unreliable in a mobile environment where the links undergo shadowing and multipath degradations; VSAT antennas are fixed and with relatively relaxed antenna mounting restrictions whereas MVSAT antennas require satellite tracking, compliance to stringent mounting space and must also comply aero-dynamic constraints.
3. MVSAT systems for high speed vehicles, that is, railway and aircraft, must include Doppler counter-measures.
4. MVSAT transmitters should be highly efficient in order to reduce terminal size and minimize power drain.
5. VSAT system coverage in the C band is available globally but K_u band coverage is generally tailored for fixed land applications on spot beams and therefore there may be gaps in connectivity on inter-continental routes, although this level of coverage may be adequate for less-demanding users. As an illustration, Figure 13.9(a) (Rodrigue, Comtois and Slack, 2012) shows the major worldwide maritime routes and Figure 13.9(b) (Statfor, 2010) illustrates the Functional Airspace Blocks (FAB) used by the European Organization for the Safety of Air Navigation (EUROCONTROL) in forecasting Instrument Flight Rules (IFR) compliant flight movement across Europe (FAB are 'blocks of airspace based on operational requirements regardless of the States

Figure 13.9 (a) Global maritime routes. (Source: Rodrigue, J-P et al., 2012. Reproduced with permission of Rodrigue, J-P.)

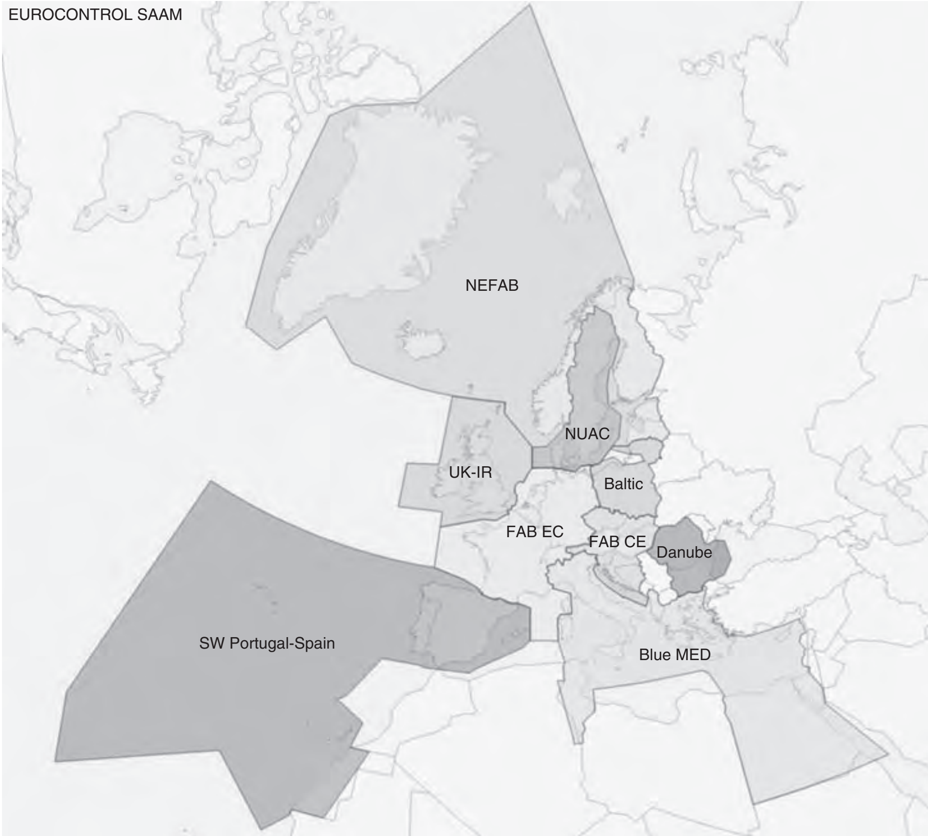


Figure 13.9 (b) The Functional Airspace Blocks (FAB) used by EUROCONTROL to forecast Instrument Flight Rules (IFR) compliant flight movement across Europe. (Source: Statfor, 2010. Reproduced by permission of EUROCONTROL.)

boundaries' and IFR refers to 'Rules and regulations established by the FAA to govern flight under conditions in which flight by outside visual reference is not safe').

6. Due to the large size of VSAT antennas they are better suited for shared and captive environments such as on ships, railway and aircrafts; although man-pack MVSATS are available for special applications, L band user terminals remain the primary choice for global personal and portable usage.
7. MVSAT systems can provide a more efficient use of the orbital arc due to the higher directivity available from MVSAT antennas when compared to L band systems designed to support a mix of user terminals including hand-held and small terminals deploying low-directivity antennas.
8. Due to the size restrictions of MVSAT antennas, the side lobe levels are higher than those of fixed VSATs and thus to comply to radio regulatory requirements the power spectral density (PSD) of return links should be reduced so as to remain within the specified mask while the forward link must be robust enough to reject interference from other satellite systems received at MVSATs.

9. When FSS (rather than MSS) allocations are used, obtaining regulatory clearance is arduous due to sharing constraints imposed in FSS bands. Trans-border communication is not permitted in large parts of the world. In contrast, MSS bands have primary and exclusive allocations throughout the world and therefore the trans-border risk is significantly lower.
10. MVSAT system architecture needs to be robust with spares and back-up provisions to achieve the reliability offered by L band systems.

13.4.2.1 Regulatory Matters

Fixed VSATs require special provisions when interworking with large and medium FSS earth stations because the VSAT antennas are small in diameter (1–3 m) causing and receiving high levels of interference; the situation gets compounded with mobility. The geometry of user terminals with respect to other satellite systems varies with movement and thus interference level are variable; while antenna side lobe performance degrades due to the restricted size of the antennas and tracking requirements. Therefore the return link EIRP should have to be curtailed and/or PSD reduced to comply with permissible off-axis radiation mask.

The FSS and MSS frequencies of interest lie in the K_u, K_a and X bands (military). There are numerous regional and by-country variations and sharing constraints within these bands as listed in the frequency allocation tables (Article 5) of the Radio Regulations (ITU, 2008). Licensing and other regulations vary between countries and regions. Some countries reserve bands specifically for VSAT operations in order to promote the technology and segments of K_u band spectrum permit VSAT mobility. When applicable, cross-border emissions compliance of the neighbouring countries has to be taken into consideration.

ITU-R (International Telecommunication Union – Radiocommunication sector) recommendation S.1857 (ITU-R, ITU, 2010) provides methodologies to estimate off-axis transmissions from vehicle mounted antennas including those attributed to pointing errors in the 14 GHz band. ITU-R recommendation S.728-1 (ITU-R, 1995) specifies the maximum permissible off-axis angle EIRP density from VSAT operating in the 14 GHz frequency band. At any angle (Φ) off the main-lobe axis of an earth-station antenna, the maximum EIRP in any direction within 3° of the geostationary satellite orbit (GSO) should not exceed the values specified in Table 13.8. The cross-polarized component permissible mask is listed in the Table 13.9 (ITU, ITU-R, 1995).

Table 13.8 Off-axis transmission mask for VSAT

Angle off-axis	Maximum EIRP in any 40 kHz band (dBW)
$2° \leq \Phi \leq 7°$	$33-25 \log \Phi$
$7° < \Phi \leq 9.2°$	12
$9.2° < \Phi \leq 48°$	$36-25 \log \Phi$
$\Phi > 48°$	-6

(Data source: ITU-R, 1995.)

Table 13.9 Cross-polar off-axis transmission mask for VSAT

Angle off-axis	Maximum EIRP in any 40 kHz band (dBW)
$2° \leq \Phi \leq 7°$	$23 - 25 \log \Phi$
$7° < \Phi \leq 9.2°$	2

(Data source: ITU-R, 1995.)

13.4.3 Technology

The key enhancements to VSAT systems to incorporate mobility apply to antenna and the transmission system. These topics have been discussed in other chapters and therefore here we summarize the salient requirements.

Antenna systems should be efficient, compact with tracking and low side-lobes to minimize harmful interference into adjacent satellite systems in compliance to the radio regulation. The small size of the antenna conflicts with low side lobe requirements and hence the emitted PSD is reduced which, therefore, affects the transmission design.

The antenna system requirements and capability depends on the physical restrictions imposed by – environment, antenna mounting space, stability of the mobile and speed of travel (see Chapter 5). Mounting space restrictions in medium/large ships is relatively relaxed but the pitch and roll movements require correction; high speed of aircrafts requires tracking agility with low drag; the limited mounting space on road vehicles restricts the size and profile of the antenna; the limited clearance space of tunnels and high speed require low profile, agile antennas for railways. Phased arrays provide the desired tracking speeds and profile; however their performance tends to degrade at low elevation angles. Hybrid solutions utilize a combination of mechanical and electronic steering

An efficient scheme to minimize interference in to adjacent satellite systems is a reduction in the PSD of signals by spreading the signal. Spread spectrum with CDMA is therefore the preferred choice in many systems; low order phase shift keying (PSK) schemes BPSK/QPSK (QPSK, quadrature phase shift keying) are preferred over high order schemes where spectral spikes would exceed the PSD mask. In order to optimize radio links in fading environments, schemes based on adaptive modulation/coding schemes such as in DVB-S2 system offer advantage. In such schemes the impact of forward link interference from adjacent satellites and fades can be negated on individual links through modulation and code rate adaptation.

MVSATs operating on land undergo the heaviest shadowing and multipath degradations where railways and road vehicles exhibit different environmental characteristics. Aeronautical and maritime channels have stable links for a majority of time barring instances of shadowing caused by the mobile's structure (e.g. fuselage) and when operating at low elevation angle. However, the dispersive nature of channels can manifest under some environmental conditions. Although the mechanisms impacting atmospheric propagation are identical for each user terminal, the manifestation at the receiver depends on the speed and direction of movement. Rain cells move with wind and typically have a size of only a few kilometres – Khamis, Din and Rahman (2005) report that the rain cell size ranges from less than 1 km to over 20 km in a tropical region with about 70.5% of the cells of 1 km and less. Therefore rain fades at a receiver get conditioned by the velocity of the mobile as rain-affected portion of the radio path varies.

A prolonged blockage lasting tens of seconds or minutes such as while traversing a tunnel can break an IP connection and result in a loss of receiver synchronization. The Auxiliary Terrestrial Component (ATC) scheme as proposed for the MSS and satellite radio broadcasts is an effective solution in such challenging conditions.

Implementation issues include the selection of an appropriate space segment and frequency band compliant to the radio and other regulations in the target service area, network architecture, quality-of-service management scheme at the physical layer and at the user level, selection of a transmission format commensurate with the mobile environment – i.e., aeronautical, land-vehicular, railways and maritime. These issues are identical to those applicable for MSS, including DVB/(RCS+M) system discussed elsewhere. Figure 13.10 illustrates an example architecture of a standalone maritime MVSAT system (KVH, 2012b). Notice that the architecture does not provide the resilience available from established L and S band MSS systems that incorporate robust ground and space segment redundancies, advanced network features such as handover, global technical and logistic support, multiple service providers and equipment vendors, connectivity with GMDSS system for SAR support, and so on. The preferred architecture of the system is driven by implementation aspects and commercial decisions. Note that operation in the frequency bands where MSS has a

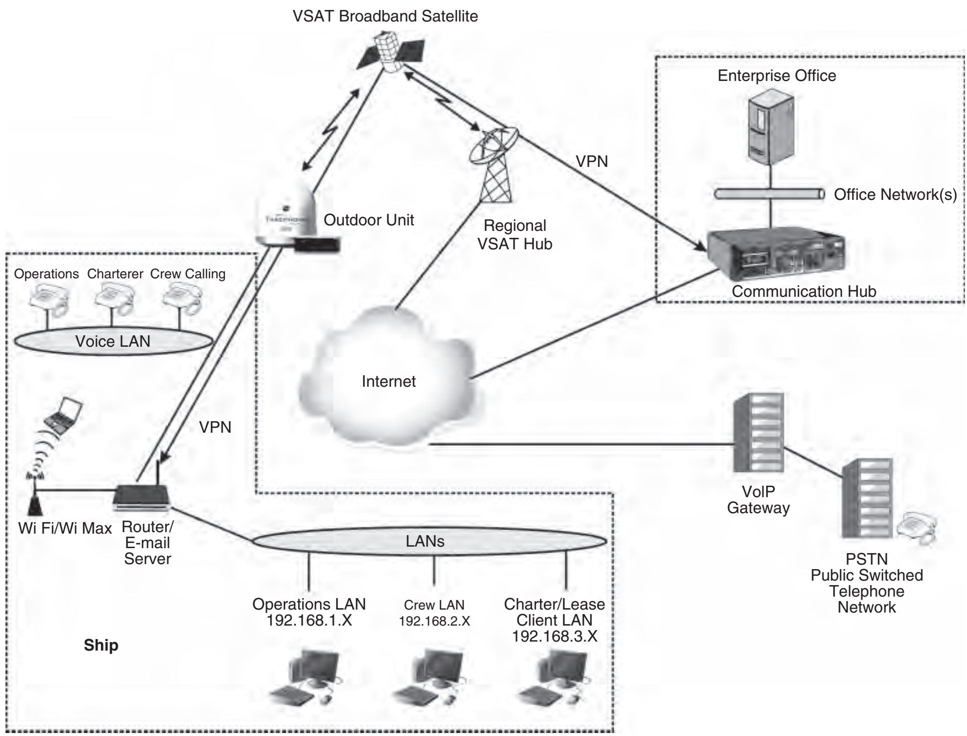


Figure 13.10 Example of a maritime MVSAT system architecture. (Source: KVH, 2012-1. Reproduced by permission of KVH.)

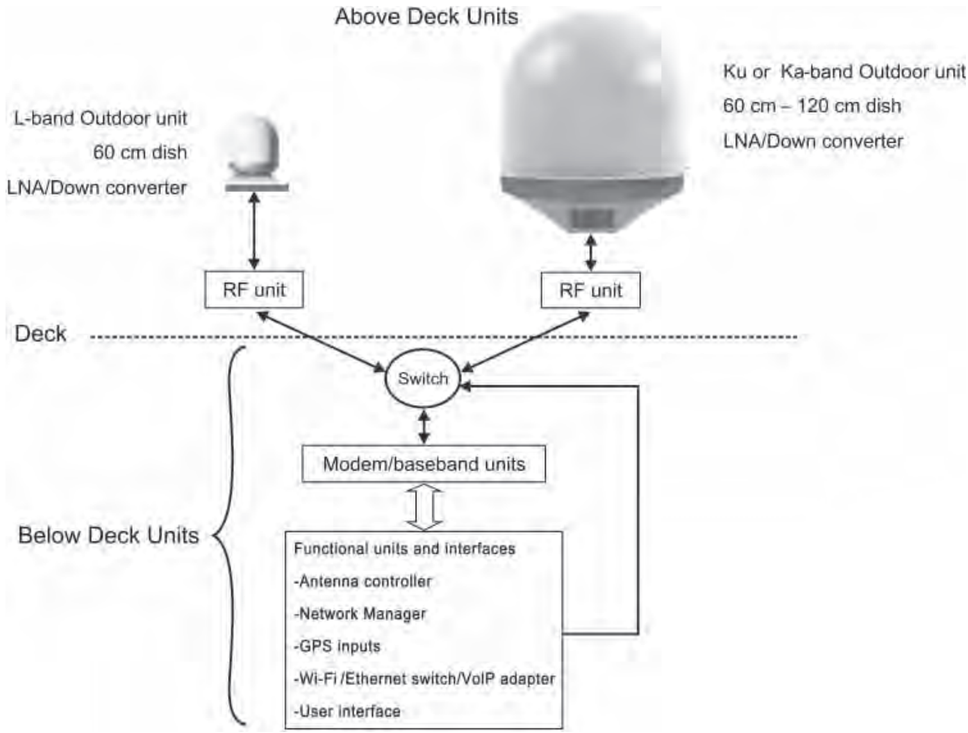


Figure 13.11 A conceptual block schematic of an L-K_u-Ka user terminal. (Source: Adapted from KVH, 2012-2. Reproduced by permission of KVH.)

primary status (e.g. 29.9–30.0 GHz Earth to space/20.1–20.2 space to Earth) reduces the regulatory constraints in relation to mobility. For this reason next generation MSS systems provide MVSAT services on these frequencies.

Figure 13.11 illustrates a conceptual block schematic of a hybrid K_u-K_a-L band system. The above-deck unit comprises an antenna system and a radio frequency unit containing a low noise amplifier (LNA) and a down converter for each frequency band. The antenna system may comprise separate units as illustrated or a single compact multi-band unit. The switch controlled by the network manager seamlessly switches between the systems. The below-deck unit comprises modem and baseband units that perform various receiver related and network functions with various interfaces such as voice over IP adapter, WiFi connection, antenna control, etc.

13.5 Terrestrial Cellular System

In this section, we introduce features of a typical cellular mobile system due to its relevance to the mobile satellite communications. We have noted that a majority of MSS systems share the core network of terrestrial mobile systems to integrate seamlessly with the terrestrial mobile system infrastructure to benefit from economies of scale by reuse of cellular components while enhancing users' communication reach. There is abundant literature available on terrestrial cellular systems and therefore this section is included for completeness highlighting those aspects of architecture that are relevant to MSS.

The basic attributes of a terrestrial cellular service are:

1. large subscriber capacity;
2. efficient use of spectrum;
3. national/international compatibility;
4. widespread availability;
5. adaptability to traffic density (in-building, urban, rural, etc.);
6. service to vehicle and portables;
7. value-added services;
8. high quality (as good as the fixed network);
9. affordability.

The necessity of increasing capacity continually with minimum spectrum allocation led to the development of the cellular concept. The cellular architecture is based on the principle of frequency reuse, cell splitting and mobility management including inter-network roaming with access to the public networks. We will not consider the details of the terrestrial radio interface as they are rarely shared with the satellite air interface because of differences in their respective radio environments and optimization criteria (see Chapter 8 for system G, which, in fact, reuses a modified version of terrestrial air interface).

13.5.1 System Architecture

Figure 13.12 portrays a generic architecture of a terrestrial mobile system to illustrate the primary entities:

- mobiles;
- a network of cells each served by a base station; notice that on the contrary all the satellite cells are served centrally;
- one or more mobile switching centres (MSC), depending on the size of coverage area, interfaced with base stations at one end and the public network at the other.

The radio link between mobile units and a base station consists of two types of physical channels – a control channel and traffic channels. The control channel transfers system messages, and traffic channels carry traffic and supervisory signals during a call. The reader should notice here the similarity with an MSS's broadcast channel. A base station is connected, usually by land lines or backhaul satellite links, to an MSC with a group of voice trunk and data links for exchange of information and process calls. The analogy between a base station and MSS Earth stations should be evident. The MSCs are connected to the public networks.

Typically, each mobile telephone is assigned a 'home area'. The home MSC maintains a location database (Home location register or HLR) in which it keeps the most recent position of mobiles registered with the MSC. This information is used for routing calls to the appropriate section of the network. Each mobile automatically registers its location with

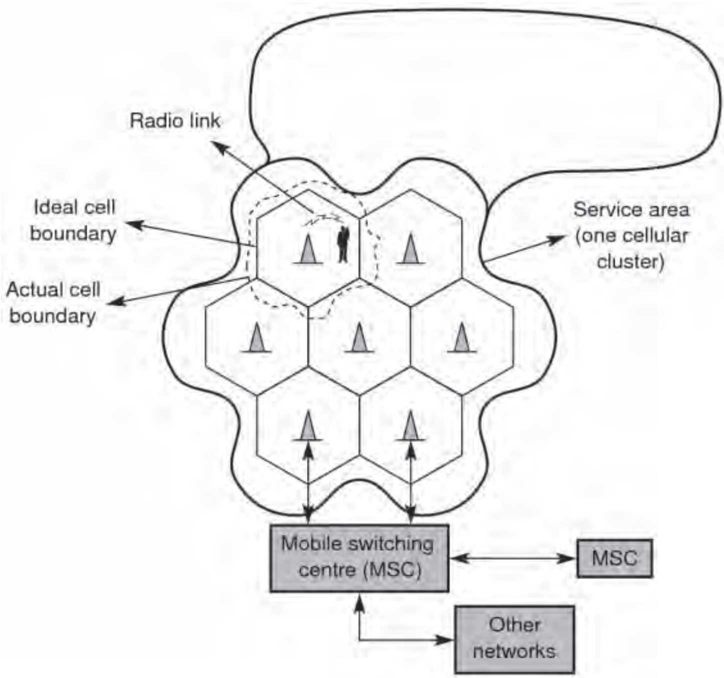


Figure 13.12 The architecture of a terrestrial cellular system. Triangular entities depict base stations

the visiting MSC whenever the mobile migrates outside its home MSC. The visited MSC transfers the information to the mobile's home MSC and keeps it within its visitor location register (VLR). Whenever a mobile telephone is switched on, it scans all system control channels and locks to the strongest channel. This operation is continuous, ensuring that the mobile always operates with the strongest signal. The control channel is used for two types of messages: general system information, which contains network identity, available channels, area code, other facilities/requirements; and mobile control information, which consists of paging messages to notify a mobile and channel assignment messages used to set up calls. If the mobile signal quality degrades during a call, the call is handed over to the cell that can provide a better link quality. The handover involves signalling between the base stations involved and the MSC but it is transparent to a user. There are a large number of terrestrial systems in use, which differ in detail. To harmonize growth throughout the world, therefore, various standards such as Global System Mobile (GSM), General Packet radio Service (GPRS), Universal Mobile Telephone System (UMTS), CDMA-2000, etc. The fourth generation systems are rolling out now as work for the 4.5 and fifth generation progresses (see, for example Andrews, 2008).

13.5.1.1 Frequency Reuse

Radio channels are reused at a distance large enough to reduce co-channel interference to acceptable limits much as spatial frequency reuse by spot beams in MSS. The coverage area is partitioned into small units, or cells, of a few tens/hundreds of metres to tens of kilometres, depending on the traffic – compare this to the smallest beam sizes of MSS (400–7500 km). Each cell is served by a base station, which establishes a connection on a frequency selected from a demand assigned pool. The frequency set is reused in other cells in accordance with a predefined reuse pattern. By reducing the transmitted power on each channel, cell size and hence reuse distance can be reduced arbitrarily, thereby increasing system capacity per unit area. Thus, we have the concept of cell splitting and macro (several kilometres), micro (hundreds of metres to kilometres) and pico/femto cells (tens of metres). A similar cell splitting process in MSS has yet to be used in the satellite systems. However, as mentioned in Chapter 9, cell sizes can be altered by certain operational procedures and thus the traffic capturing size of spot beams can be altered within limits to accommodate changes to traffic pattern, while steered spot beams can be moved over to a region in need of additional capacity. These features are used in Inmarsat network for managing traffic variations. Software reconfiguration of spot beams was used on Inmarsat-4 satellites to accommodate maritime and aeronautical traffic and the capability of beam steering to hot spot areas, as built into Inmarsat-5 satellites, are techniques unique to satellite systems.

13.5.1.2 Cell-Splitting

Consider a simplistic scenario: assume that T RF traffic channels are divided equally between M base stations. Then each base station has a pool of T/M channels. Eventually, certain base stations reach their capacity limit. Further growth can then be accommodated by a process called cell-splitting. The process involves reducing the transmitted power of

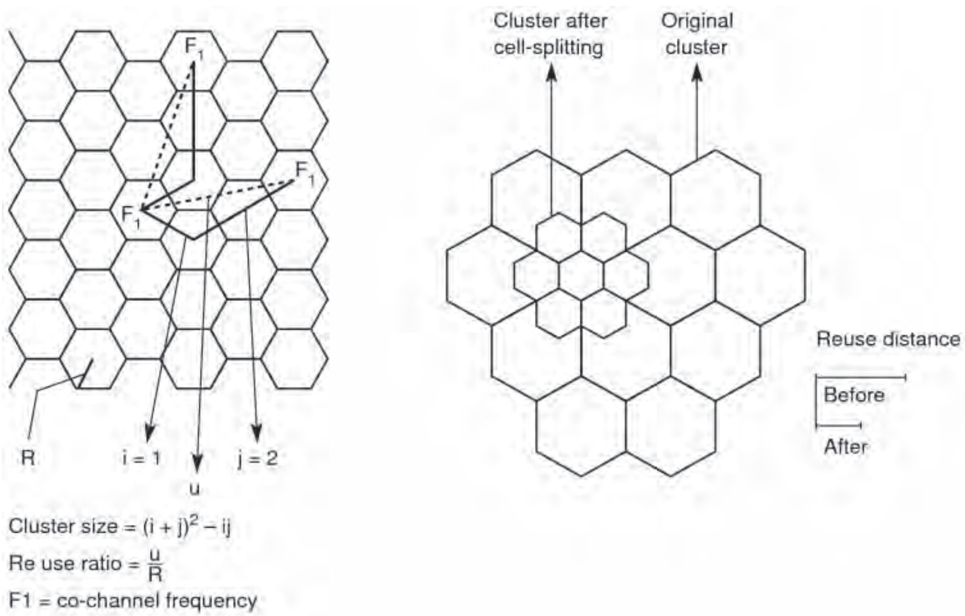


Figure 13.13 The cellular geometry and cell splitting process.

base stations, which has the effect of reducing the cell size (reuse distance) followed by increasing the number of cells (and hence RF channels) serving the region, effectively increasing the spectrum utilization in the geographical area. In other words, cell-splitting increases the spectrum efficiency per unit area to match the demands.

Figure 13.13 shows cell-splitting in advanced stages of deployment. In practice, cell splitting is complicated by logistics and other practical considerations. The cluster size C is given as:

$$C = (i + j)^2 - ij \tag{13.2}$$

Reuse ratio, R_u is

$$R_u = u/R \tag{13.3}$$

i, j, u and R are defined in Figure 13.13.

Revision

1. Highlight the salient features of Cospas-Sarsat system. State the limitations of the existing space segment and the measures being undertaken to mitigate these.
2. Outline the navigation principles used in systems based on pseudo-range measurements including the error mechanisms and methods to mitigate these errors.
3. Explain the principle of differential GPS with an example application.
4. Describe the characteristics of the Galileo system and compare them to GPS and GLONASS systems.
5. What are the issues affecting mobility of VSATs. Suggest solutions to mitigate such mobility-associated problems.
6. Compare the architecture of a terrestrial cellular system with a MSS spot beam system stating similarities and differences.

References

Andrews, J.G. (2008) 3G and 4G Cellular Standards, The University of Texas, Austin, TX, November 5, users.ece.utexas.edu/~jandrews/pubs/WirelessStandardsSummary.pdf (accessed 28 August, 2013).

Arcidiacono, A., Finocchiaro, D. and Grazzini, S. (2008) Broadband mobile satellite services: the K_u-band revolution. *Satellite Communications and Navigation Systems, Signals and Communication Technology*, 573–588 www.springerlink.com/index/q56654469771w917.pdf (accessed 28 August, 2013.

Argos (2008) Argos User Manual, www.sspi.com.br/portal/eventos/calendario-2012/vsat-day-2012/apresentacoes (accessed 28 August, 2013).

Bretz, E.A. (2000) X marks the spot, maybe. *IEEE Spectrum*, **37**(4), 26–36.

Butt, G., Evans, B.G. and Richharia, M. (1992) Multiband propagation experiment for narrow- band characterisation of high elevation angle land mobile-satellite channels. *Electronics Letters*, **28**(15), 1449–1450.

Cospas-Sarsat (2009) Introduction to the Cospas-Sarsat System, C/S G.003, Issue 6, October 2009, www.cospas-sarsat.org (accessed 29 July 2013).

Cospas-Sarsat (2011) Handbook on Distress Alert Messages for Rescue Coordination Centres (RCCs), Search and Rescue Points of Contacts (SPOCs) and IMO Ship Security Competent Authorities, Issue 1 – Revision 3, October, www.cospas-sarsat.org (accessed 29 July 2013).

Cospas-Sarsat (2013) www.cospas-sarsat.org/ (accessed 28 August, 2013).

Dale, S.A., Daly, P. and Kitching, I.D. (1989) Understanding signals from GLONASS navigation satellite. *International Journal of Satellite Communications*, **7**(1), 11–22.

Daly, P. (1993) Navstar GPS and GLONASS: global satellite navigation systems. *Electronics and Communication Engineering Journal*, **5**(6), 349–57.

EC (2013) web site, http://ec.europa.eu/enterprise/policies/satnav/galileo/index_en.htm (accessed 28 August, 2013).

EC, ESA and Galileo Joint Undertaking (2003) Business in Satellite Navigation: An Overview of Market Developments and Emerging Applications.

FAA (2012) Wide-Area Augmentation System Performance Analysis Report. Report #42, FAA/William J. Hughes Technical Center, NSTB/WAAS T&E Team, Atlantic City International Airport in New Jersey, Reporting Period: July 1 to September 30, 2012 October, www.nstb.tc.faa.gov/, www.nstb.tc.faa.gov/reports/PAN79_1012.pdf (accessed 28 August, 2013).

Hall, G.W. (2013) USCG Differential GPS Navigation Service, www.uscg.mil/directives/ci/16000-16999/CI_16577_2.pdf (accessed 28 August, 2013).

ITU (2008) *Radio Regulations*, Vol. **1–4**, ITU, Geneva.

ITU (2010) Methodologies to Estimate the Off-Axis E.I.R.P. Density Levels and to Assess Interference Toward Adjacent Satellites Resulting from Pointing Errors of Vehicle-Mounted Earth Stations in the 14 GHz Frequency Band. Recommendation ITU-R S.1857.

ITU-R (1995) Maximum Permissible Level of Off-Axis E.I.R.P. Density from Very Small Aperture Terminals (VSATs). Recommendation ITU-R S.728-1.

Khamis, N.H.H., Din, J. and Rahman, T.A. (2005) Analysis of rain cell size distribution from meteorological radar data for rain attenuation studies. Proceedings of Asia Pacific Conference on Electromagnetic, December, 20-21, 2005, Johor Bahru, Johor, Malaysia, http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=1607788 (accessed 28 August, 2013).

King, J.V. (1999) Overview of the Cospas-Sarsat satellite system for search and rescue. IMSC'99, Sixth International Mobile Satellite Conference, Ottawa, Canada, Co-sponsored by Communications Research Centre and the Jet Propulsion Laboratory, pp. 31–36.

KVH (2012a) Comparing mini-VSAT Broadband to Legacy Solutions. Technology Report, KVH Industries, Inc., www.kvh.com (accessed 29 July 2013).

KVH (2012b) *The Benefits of Maritime SATCOM – the Enterprise Ripple Effect*, SATCOM 101 Webinar Series, KVH Industries, Inc. www.kvh.com (accessed 29 July 2013).

Milliken, R.J. and Zoller, C.J. (1978) Principle of operation of NAVSTAR and system characteristics. *Navigation Journal*, The Institute of Navigation, **25**(2, summer), 95–106.

O'Neil, W.A. (1992) GMDSS. Inmarsat Ocean Voice.

Princeton University (2013) Global Navigation Satellite System (GNSS), www.princeton.edu/~alaink/Orf467F07/GNSS.pdf (accessed 28 August, 2013).

Rodrigue, J.-P., Comtois, C. and Slack, B. (2012) The Geography of Transport Systems, Hofstra University, Department of Global Studies & Geography, http://people.hofstra.edu/geotrans (accessed 28 August, 2013).

Sanz Subirana, J., Juan Zornoza, J.M. and Hernández-Pajares, M. (2011) Reference Frames in GNSS, www.navipedia.net/index.php/Reference_Frames_in_GNSS#Galileo_reference_frame_.28GTRF.29 (accessed 28 August, 2013).

Schänzer, G. (1995) Satellite navigation for precision approach: technological and political benefits and risks. *Space Communications*, **13**, 97–108.

SDCM (2013) www.sdcm.ru/smglo/staticpages?version=eng&site=extern&title=about (accessed 28 August, 2013).

Statfor (2010) EUROCONTROL Long-Term Forecast – Flight movements 2010-2030, 17 December 2010, The European Organisation for the Safety of Air Navigation (EUROCONTROL).

UN (2010) Current and Planned Global and Regional Navigation Satellite Systems and Satellite-based Augmentation Systems, United Nations, New York, Document ST/SPACE/50, www.icgsecretariat.org (accessed 29 July 2013).

Yionoulis, S.M. (1998) The transit satellite GEODESY program. *John Hopkins APL Technical Digest*, **19**(1), 36–38, www.jhuapl.edu/techdigest/td/td1901/yionoulis.pdf (accessed 28 August, 2013).

14

The Future

14.1 Introduction

This chapter introduces a representative sample of technologies being researched and refined through a wide range of initiatives spanning the world. The purpose is to heighten the awareness of the reader to these developments; references to source publications should provide the interested reader with a useful start to pursue topics of interest. It is reiterated that the true MSS (mobile satellite service) perspective is multi-faceted involving technology, economics, global and regional regulations, politics and societal trends covering a variety of economies. Keeping to the theme of the book the treatment is limited to technology.

The MSS evolution is driven by a combination of commercial and technical innovations in conjunction with general trends in other technologies and societal needs. Since an MSS is a part of a significantly larger telecommunications network, its evolution is influenced by those of the public as well as government and military telecommunications sectors – particularly the terrestrial mobile communication segment. The anticipated integration of broadcasting, computing and telecommunications technologies is well underway – for instance, consider the interactive satellite television service, mobile Internet and seamless integration of mobile satellite systems with the terrestrial mobile network. The benefits of aligning the research and development initiative, commercial goals and service evolution of MSS with terrestrial mobile streams and complementary technologies are widely recognized.

The chapter begins with a brief review of the prevailing market forecasts leading to the end of the present decade. The following sections introduce the technologies that offer the potential to improve the efficiency and performance of existing systems to foster the desired growth. We discuss a number of capacity enhancement technologies – multi-user detection, advanced frequency planning techniques, cross-layer optimization of radio resources and the potentials of the cognitive radio system (CRS) application to satellite systems. Next we trace the evolution of terrestrial systems to place the role of satellite in the mobile communication landscape. With the rollout of terrestrial 4G system the adaptability of air interface to the satellite environment is an attractive option; we introduce a few initiatives towards this goal.

In relation to the development trends in terrestrial mobile technology, we address the applicability of *ad-hoc* networks to mobile satellite communications. A hybrid or dedicated satellite *ad-hoc* network can extend the service beyond the satellite coverage, including

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

situations where satellite coverage is unreliable and as a provision for a rapidly deployable emergency network.

In the next part of the chapter we outline a sample of enabling concepts and technologies at the forefront of the present research and development activities. These deal with characterization of various types of propagation channels in the K_a band and MIMO (multiple input multiple output) channels. Topics include a summary of the trends in modulation and coding techniques, the status of MIMO application to satellite systems and the potential of software defined radio (SDR) in the future networks. We conclude the chapter on a speculative note presenting an utopian vision of a futuristic network topology and speculate on the long term capabilities of satellite communication.

14.2 Market Projections

Despite the underlying growth in user numbers and usage in absolute terms, the mobile satellite communication industry was perceived to have suffered a severe downturn around the turn of the millennium after the financial collapse of a few high-profile MSS ventures. The MSS industry has since matured and stabilized with positive financial results reported by a majority of the MSS operators in recent years.

A number of market forecasts predict a moderately positive growth of the MSS industry independently in the decade leading to 2020 anticipating 4.8–7.8 million subscribers by then (Futron, 2010; Euroconsult, 2011; NSR, 2012). The land mobile growth is said to be driven by the low-end/low cost units such as used for machine-to-machine (M2M), SCADA (supervisory control and data acquisition), asset tracking, etc., which account for nearly 90% of in-service units, although these units are estimated to provide only 15% average revenue per user (ARPU) relative to those of the broadband platforms (NSR, 2012). In spite the low numbers of aeronautical terminals, they are anticipated to provide a high return in percentage terms – that is $870 million of revenues in 2011 increasing to $2.3 billion at the end of 2021. The overall growth is attributed to a shift in usage from voice to data and an increase in broadband and low speed data applications. The role of the next generation space segment is considered as an enabler. The revenue loss attributed to terrestrial services and FSS (fixed satellite service) mobile VSATs (very small aperture technologies) are factored in the predictions. Due to the uncertainties and assumptions involved in long term predictions, these estimates can be considered as indicative but nevertheless the upward trend is evident.

The forecasts recognize the uncertainty in assessing the ‘impact of diverse external forces and specific market variables’ and thus provide low, baseline and high range in forecasts. Differentiators include economic conditions, government policies, rate of service penetration and the ability of the satellite operators to absorb unexpected demands. Since forecasting so far ahead into the future of a rapidly changing mobile communication ecosystem is fraught with uncertainty due to dynamics of the marketplace and other variables, such forecasts are taken as an indicative guideline and deviations are factored as the technology and markets evolve. Consider as an illustration the key assumptions used in one of the forecasts:

- MSS growth is driven by data applications;
- Availability of new generation MSS systems facilitates development of new application and a broader user base;

- The emerging regions contribute to a healthy growth;
- Government and military use remain key markets;
- VSAT competition is likely to erode high-end MSS market to an extent;
- K_a band MSS VSAT system is likely to influence the impact of FSS MVSAT (mobile very small aperture terminal);
- Prices are likely to reduce due to competition;
- Synergistic solutions of MSS with other technologies provide growth opportunity.

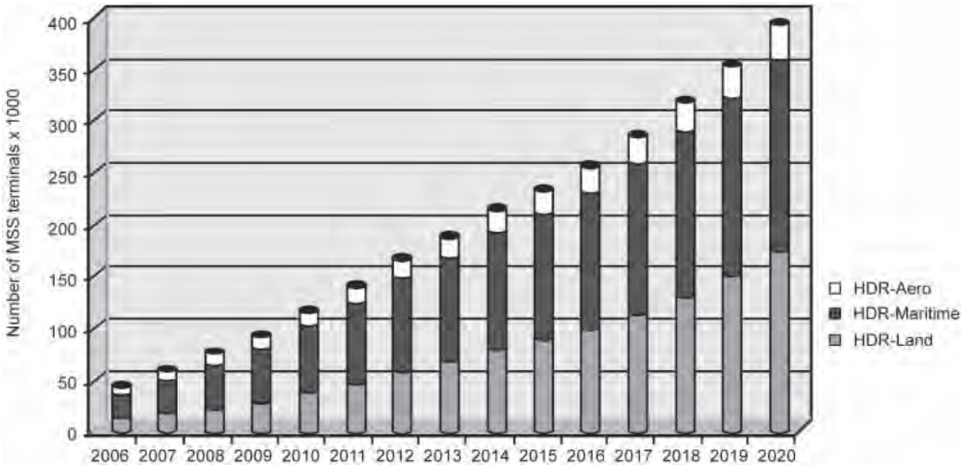
It is evident that the accuracy of forecasts would depend on the numbers applied to the assumptions. Hence in order to provide a realistic working platform, it is usual to consider pessimistic, average and high growth rate scenarios.

The land mobile market is expected to be a major growth area despite severe competition from terrestrial systems. The L and S band are likely to be the dominant modes of transmission in the land mobile (including hand-held) sector due to their superior transmission characteristics for mobility when compared to the K_u and K_a bands. M2M application is anticipated as a major contributor to terminal growth. Hybrid satellite-terrestrial architecture offers an attractive alternative. K_u and K_a band VSAT platforms could take up in excess of 50% of the maritime market by 2020 and similarly K_u and K_a high throughput solutions will make headway in the aeronautical sector posing competition for L band systems. It is envisaged that high throughput K_u/K_a band MSS will utilize the L band as a back-up and for emergency usage.

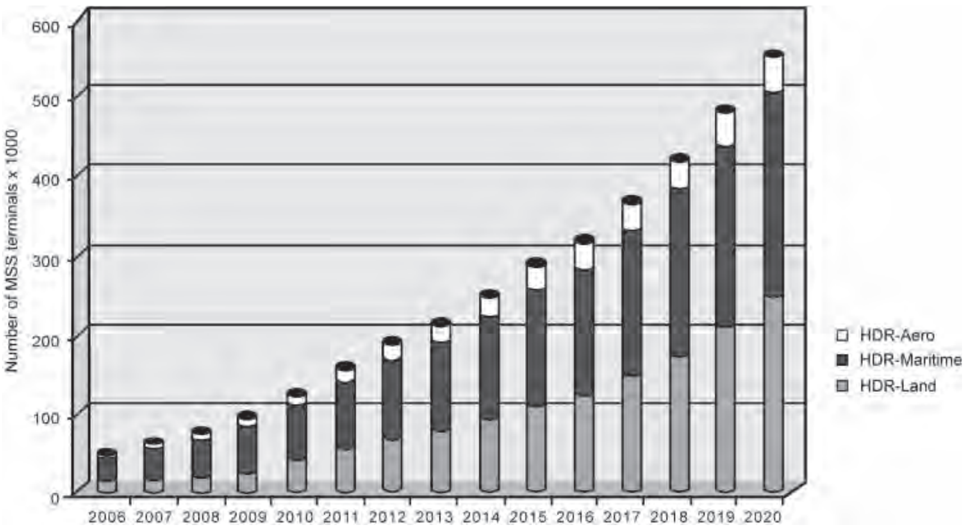
These predictions are consistent with the growth in the mobile telecommunications across the world (see Figure 1.1 in Chapter 1) (ITU, 2012). Cellular mobile growth in developing countries has reached double digit figures in recent years while the growth in developed regions continues at single figures. The rapid growth in the developing countries has been enabled by affordable equipment and service charges in conjunction with the availability of prepaid services, which mutually suit the service provider and low income population. Attributed to the introduction of 3G (and further) technologies, the number of active mobile-broadband subscriptions grew by 40% between 2010 and 2011 globally surpassing fixed broadband subscriptions by a factor of 2. The growth is expected to continue as these technologies roll out globally.

An ITU (International Telecommunications Union) report (ITU, 2006) predicts the number of MSS subscribers operating at up to 144 kbps for high mobility applications by 2020 as about 5.9 million assuming an average growth – ranging from 4.7 million for low growth to 8.8 million for high growth; while a follow-on study (ITU, 2011) covers predictions for broadband at data rates up to 4 Mbps. The broadband study assumes that technology would continue to evolve linearly to provide data rates from 56–492 kbps upwards to 4 Mbps. The figures are considered reasonable by the industry. Figure 14.1 parts (a) and (b) respectively show the broadband growth up to 2020 for low and high traffic growth scenarios for the three prime MSS market segments – aeronautical, land and maritime – independent of frequency band demonstrating that the numbers of worldwide MSS terminals by 2020 are 380 000 and 540 000 in the low and high growth rate scenarios respectively.

One of the assumptions is that underpinning commercial and technological innovations would provide the impetus to sustain the projected growth. Commercial innovations include more efficient ways to buy, sell and optimize the use of space segment capacity, foster



(a)



(b)

Figure 14.1 (a) Low growth and (b) high growth evolution scenario of prime broadband MSS irrespective of frequency band assuming a linear growth in throughput from 56 kbps to 4 Mbps. (Both parts source: ITU, 2011. Reproduced by permission of ITU.)

symbiotic partnerships across all the sectors including satellite manufacturers, equipment, providers, resellers and customers; and, create new opportunities in service provision, market and customer base. The technological innovations deal with provision of efficient techniques of delivering the service. The remaining parts of the chapter will outline a sample of concepts and technologies, which are considered potential enablers for the next generation MSS systems.

14.3 Spectrum Forecast

The ITU has been at the forefront of the work towards the projection of future spectrum requirements and its equitable allocation. As mentioned in the preceding section, ITU-R report M.2077 (ITU, 2006) presents the estimate of spectrum needed for the satellite component of International Mobile Telecommunication (IMT) in the frequency range 1–6 GHz, suitable for high mobility applications at data rates up to about 144 kbps on portable/hand-held type devices. The report concludes that taking into consideration the 2007-base, global MSS allocation of 2×86 MHz, additional allocation requirements in the space-to-Earth direction range from 14 to 33 MHz (low traffic scenario-high traffic scenario) in the year 2010 to 114–257 MHz by the year 2020. The estimate includes 30 and 90 MHz required for multimedia (MM) distribution for the years 2010 and 2020, respectively. If the MM distribution element is removed forecasts scale down respectively to 0–54 and 3–137 MHz respectively. The corresponding requirements in Earth-to-space direction range respectively 0–19 and 0–90 MHz.

These estimates were developed for high mobility application as a satellite component of IMT-2000 and beyond. ITU-R report M.2218 (ITU, 2011) extends the spectrum forecasts for *broadband services* including the terrestrial mobile backhaul services (e.g. in a captive mobile environment such as an aeroplane) in the frequency range 4–16 GHz for satellite component of IMT as well as systems outside of IMT. The data rates are considered to be of the order of several Mbps (e.g. 4 Mbps) and it is anticipated that directional antennas would be used. These estimates were made in the context of a WRC 12 (World Radio Conference, item 1.25) agenda related to 4–16 GHz range. Typical applications comprised broadband Internet access, video-conference, voice backhauling, broadband security and safety, broadband corporate network, real-time satellite news gathering (SNG), telemedicine, tele-education, and so on in addressable land, maritime and air markets. It was recognized that some of the applications could be addressed through FSS MVSATs. The estimates did not include the impact of K_a band MSS services. Nevertheless, it was observed that some high-end traffic would move into these broadband systems and that would relieve high mobility low-end needs. The estimate ranged 240–335 MHz equally in both the directions assuming packet-switched traffic and that all the services are multiplexed on the same carrier. The assumption is optimistic as the traffic would be distributed between several operators causing an inefficiency attributed to practical features such as fill-factor differences between operators, spectrum-hoarding and guard band losses.

14.4 Capacity Enhancement Techniques

The MSS allocations in the L band have become congested to the extent that broadband services (several megabits per second) cannot be accommodated within these bands with the current technology and hence, broadband mobile systems are migrating to K_u and K_a bands. Considerable effort is, therefore, being expended to improve the spectral efficiency in order to maximize the utilization of the spectrum and to reduce the cost/bit. This section reviews a sample of developments and promising technologies being pursued towards this end. The topics include utilization of multi-user detection for satellite communication systems, methods to optimize static and dynamic frequency plans, cross-layer optimization techniques and the applicability of cognitive radio in an MSS context.

14.4.1 Multi-User Detection

Multi-User Detection (MUD) techniques have been investigated extensively for terrestrial CDMA (code division multiple access) (Verdú, 1998, 2000). In a CDMA scheme, all the users transmit in the same spectrum space causing multiple access interference (MAI) that sets a bound on channel capacity. Each user is identified through a unique code assigned. In a single detection system the detector extracts the wanted signal by correlating the signal with a replica of the code; the cross-correlation between the codes used in a system is low but not zero, therefore the cross-correlated signals appears as MAI. A more efficient alternative, known as multi-user detection (also called joint-detection and interference cancellation), detects the users jointly to the benefit of all the users by minimizing the MAI thereby improving system spectral efficiency.

A large number of joint detection schemes have been proposed for Direct-sequence-code division multiple access (DS-CDMA), a commonly used multiple access technique (Moshavi, 1996). Such schemes can be generalized as belonging to one of the categories – linear detector or subtractive interference cancellation detector. A linear detector applies a transformation to the soft outputs of a conventional detector to produce an improved set. A subtractive interference cancellation detector estimates interfering signals and subtracts them from the wanted signal to improve the results.

A desirable prerequisite in MUD is that all the codes in use be known at the receiver. In a cellular or satellite system this would be possible at the base station but not at the mobile and therefore the capacity increase (with the existing technologies) would only be feasible in the return direction. The application of this technique in the forward direction is not possible at present due to the limitation in the computing power of the mobile and this topic, therefore, remains a subject of research.

The applicability of the concept has been investigated to both wideband CDMA satellite systems and narrow-band TDMA (time division multiple accessing) schemes. Neri *et al.* (2007) present the applicability of two MUD schemes – successive interference cancellation (SIC) (Wang and Poor, 1999) and turbo spatial minimum mean squared error-interference cancellation (turbo SMMSE-IC) (Reynolds and Wang, 2001) respectively to the return link of S-UMTS (satellite component of the Universal Mobile Telecommunication System) wide-band CDMA system (ETSI, 2004) and narrow-band GMR-1 (Geo Mobile Radio) TDMA (ETSI, 2001) systems. Channel Impairments in performance evaluation included mobile high power amplifiers (HPAs) non-linearity, correlated Ricean fading and residual parameter estimation errors. The results demonstrate that for the S-UMTS scheme the number of users increased substantially while full frequency reuse in adjacent beams could be feasible for the GMR-1 system. For the S-UMTS system eight users transmitting with a spreading factor equal to 4 were investigated at mobile terminal speeds equal to 50 kmph, and the terminal HPA was assumed to operate at an input back-off equal to 0.5 dB. For the GMR-1 scenario, three co-channel users were placed in adjacent beams thus creating a high interference scenario in each case.

Richharia and Trachtman (2006a) report interference mitigation achievable in an L band narrowband TDMA system by an iterative interference cancellation MUD in a number of operational environments: co-channel interference arising from spatial spectrum reuse; and adjacent channel interference (ACI) caused by tighter packing of carriers. A scheme using iterative joint detection and independent decoding was selected from a list of alternatives. The bearers used in the study belong to Inmarsat's Mobile Packet Data Service (MPDS) and Broadband Global Area Network (BGAN) that utilize 16 QAM (quadrature amplitude modulation) with turbo-coding. High complexity inherent in optimal decoding was reduced by separating the constraints introduced by channel and coding. The receiver makes joint decisions on each received symbol in a multi-user detector, ignoring the constraints due to coding; the resulting symbol streams are then independently decoded using conventional techniques. By reconstructing from the decoded output each user's transmitted waveform and feeding back to a joint detector where waveform estimates are used to cancel interference caused to each other nonlinearly to reduce multiple-access noise leading to a more accurate estimate by the decoder. The process is reiterated until desired performance goals are met. This class of method has been applied to both CDMA systems and narrowband systems (Alexander, Grant and Reed, 1998; Moher, 1998).

Figure 14.2 shows the MUD architecture. The structure with minor modifications offers encouraging results in all the applications, including the case where the MAI is substituted by ACI.

The MUD produces soft estimates of the coded and modulated symbols. It takes as input the incoming signals, and the current soft estimates (initialized to zero at the first MUD iteration) of each users' contribution to the received signal. The soft estimate for each user is updated by subtracting the current soft estimates of all the interfering users. Channel De-interleaver (De-int)/Interleaver (Int) is an MPDS/BGAN functionality to re-order bits. The soft demodulator derives posterior probabilities for each symbol taken from the 16-QAM constellation. A soft turbo decoder refines the probabilities of the coded bits and the soft modulator produces the average coded and modulated symbols that, after being conditioned by an estimate of the channel, are fed back to the multi-user detector for the next MUD iteration. The channel estimates are quite central to the convergence of the system. Thus if enough number of training symbols are present, or the signal to noise ratios is high,

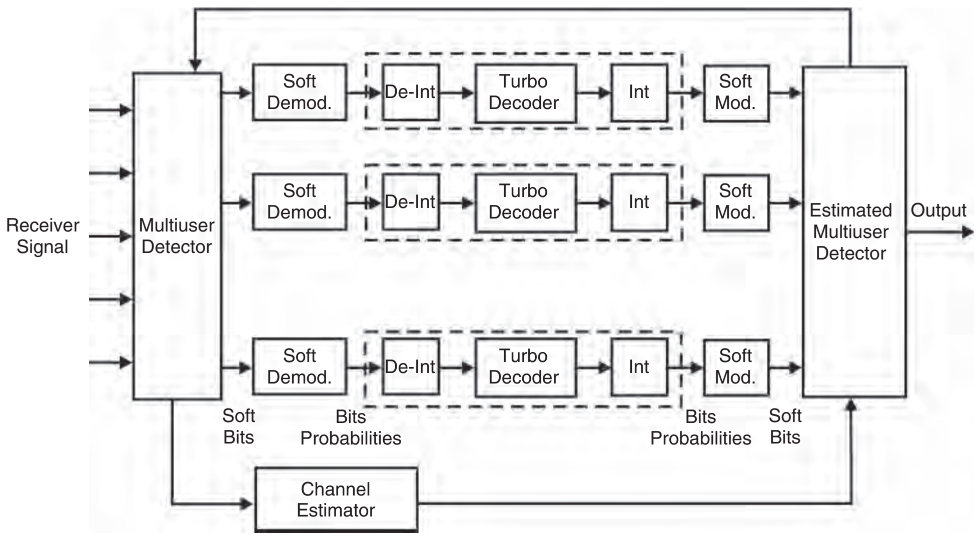


Figure 14.2 Iterative cancellation MUD architecture. (Source: Richharia and Trachtman (2006-1).)

the number of MUD iterations would reduce accordingly. The algorithmic complexity of this structure is linear in terms of number of users. A prototype receiver was developed as a proof of concept (POC) to assess practicality of hardware implementation; and develop rules and bounds for MUD receiver operation in an operational environment.

Various interference cancellation scenarios were tested:

- Co-channel interference cancellation in the return direction service link (Earth-space) to improve spatial spectrum reuse efficiency.
- Adjacent channel cancellation in the forward and return direction to improve spectral efficiency through tighter packing of carriers.

The tests included numerous operational environments typical of Inmarsat, including some stress tests that included up to seven co-channel high-level interferers. Realistic operational conditions included interference matrix as applied on Inmarsat-3 constellation, land and maritime multi-path fading – slow (0.7 Hz) and fast (20 Hz fading), mixed bearers (5 and 20 ms bursts); unknown interferers, empty slots, frequency and time offsets and other impairments as specified in system definition manuals. It was observed that certain operational constrains could be necessary in its application to an operational environment but excluding these the results were encouraging.

In a multi-beam satellite system, co-channel interference arises because of frequency reuse in spot beams. On Inmarsat third generation satellites (I3) that deploy five to six spot beams, a typical reuse would apply to the north-east/north-west spot beams; and north-central/south-east beams where isolations are respectively of the order of 18 and 12 dB. The lower isolation between NC and SE beam of I3 implies that co-channel frequency reuse is not possible and hence only partially overlapping co-channel reuse is feasible. The interleaved frequency arrangements in Inmarsat-3 north-central and north-eastern beams requires 65 kHz of adjacent channel spacing without MUD; with MUD adjacent carriers could be spaced end-end, that is at a spacing of 45 kHz saving 20 kHz or 31% per carrier.

Investigations on a sample of BGAN carriers showed promising results. The BGAN air interface is considerably more complex due to its adaptability to varying channel conditions by adaptive modulation and coding that result in over a 100 types of bearer type-code rate combination and hence the interference scenario is formidable. Supportable modulations include 16-QAM and variants of QPSK (quadrature phase shift keying) each operating with code rates ranging from around 0.3 to 0.9. Investigations on a limited set concluded with promising results leading to a follow on work mentioned later in the section (Moher, *et al*., 2010).

An adjunct investigation addressed the feasibility of applying the MUD architecture to ACI cancellation and in particular to a problem related to an evolutionary BGAN air interface to support an omni-directional variant (Note: This variant was not implemented in the following BGAN release). The work had exposed the problem of spectral spreading by class-C amplifiers at the output of user terminals leading to unacceptable levels of ACI. MUD cancellation would be able to retrieve the spectrum excess caused by spectral growth. Furthermore, the ACI technique could be extended to the forward direction to compress the inter-carrier spacing of traffic carriers that could allow higher throughput up to 1 Mbps within the same spectrum in the next generation products.

This led to the following investigations:

- Feasibility of reducing return link adjacent carrier spacing of the proposed O-QPSK (offset-quadrature phase shift keying) omni-directional air interface in order to recover bandwidth lost due to spectral spreading in user terminals.
- Application of ACI cancellation to forward and return-link high-rate bearers (turbo-coded 16 QAM at a symbol rate of ~ 151 ksps) where a number of these bearers serve the same terminal, in order to increase bits per hertz.

The following significant conclusions were drawn from simulations that included BGAN specified channel impairments and fading as appropriate for aeronautical, maritime and land-mobile terminals.

It was observed that:

- The adjacent carrier spacing reduces from ~ 45 to ~ 30 kHz (about 33% reductions); the MUD performance is bound by acquisition and synchronization threshold.
- In fading channels, losses with MUD may be up to 1 dB more than the specified allowance for ACI under worst-case conditions. This is due to dynamic variations in ACI magnitude.
- The technology is transparent to the mobile terminals.

For the high rate forward link bearer it was observed that three high rate forward bearers (designated F80T4.5X) can be transmitted in a bandwidth equivalent to that currently occupied by two such bearers; Return bearers (R20T4.5X) exhibit a similar behaviour again demonstrating a 33% reduction.

It should be noted that the investigations were constrained by numerous practical considerations as otherwise these schemes could not be introduced into an operational system without incurring prohibitive cost penalties.

The encouraging results led to further refinements (Moher, *et al.*, 2010), which demonstrated a reduction in channel spacing from 40 to 25 kHz on a hardware-prototype MUD receiver with an implementation loss comparable to the combined implementation loss and ACI allowance for the nominal arrangement without MUD.

Millerioux *et al*. (2006) present performance comparison of various non-linear MUD algorithms utilizing iterative parallel interference cancellation with semi-blind channel estimation, designed to mitigate co-channel interference on a convolutionally coded DVB-RCS (Digital Video Broadcast/Return channel via satellite) reverse K_a band link of an on-board processing multi-beam satellite system where the MUD unit is placed on-board the satellite. The results demonstrate that irrespective of the algorithm, the unit improves the performance compared to an implementation without MUD.

14.4.1.1 MUD for Polarization Reuse

The encouraging results on MUD enabled interference cancellation opens the possibility of using the technique for dual-polarization frequency reuse in the L and S MSS service links. While dual-polarization is commonly used in feeder links (typically, C, K_u and K_a bands), L band MSS systems do not utilize both polarizations, that is right- and left-hand circular

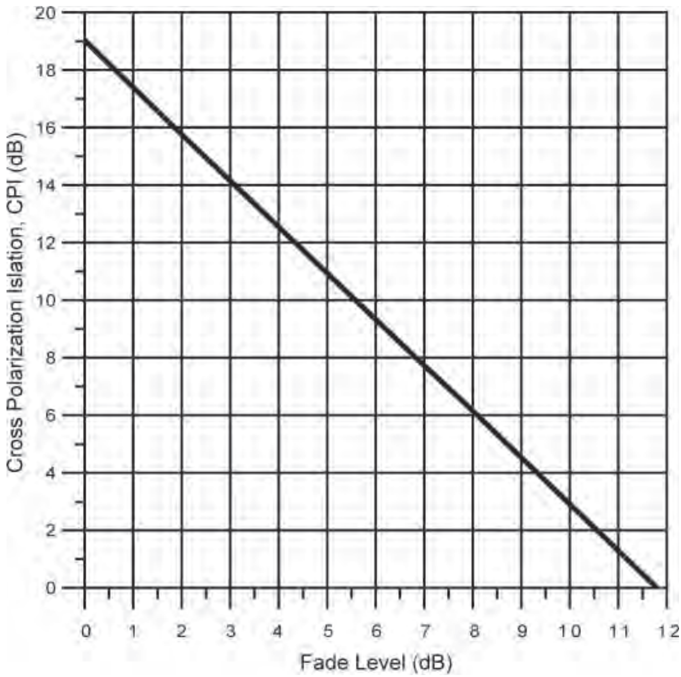


Figure 14.3 Variation in cross-polar isolation (in dB) at a given probability (P) obtained from in L-band (1.5 GHz) land-mobile campaign conducted in Australia at a satellite elevation of angle of 51°. (Source: Goldhirsh and Vogel, 1998. Reproduced with permission of NASA.)

polarization due to an unacceptable cross-coupling of signals between the polarizations due to polarization reversals from scattering surfaces resulting in interference. Figure 14.3 demonstrates the variation in cross-polar isolation (in dB) at a given probability (P) in the L band (1.5 GHz) obtained from land-mobile campaign conducted in Australia at a satellite elevation of angle of 51° (Goldhirsh and Vogel, 1998). The two components were measured on separate runs. The cross-polar isolation $CPI\,(P)$ is defined as:

$$CPI(P) = \frac{COPS(P)}{CRPS(P)} \tag{14.1}$$

where $COPS(P)$ and $CRPS(P)$ represent the co-polarization and cross-polarization signal levels at the equi-probability value P. The variation was modelled as:

$$CPI = -1.605A + 18.94 \tag{14.2}$$

where A is the co-polarized fade between 0 and 11.8 dB.

Note that the isolation of about 19 dB drops to about 14 dB at 3 dB fade and 9.3 dB at a fade level of 6 dB, making a cross-polarization system unusable. The proposed MUD technique was anticipated to cancel the cross-polar component and thus providing the possibility of frequency reuse by polarization reuse in the service link.

14.4.2 Static and Dynamic Frequency Planning

Geostationary MSSs allocations are shared between operators and hence they are quite fragmented. In particular, the L and S bands are in heavy demand and hence congested. To maximize the utilization of the available spectrum, therefore it is paramount for operators to utilize their allocation share efficiently. The reader is referred to Chapter 9 for a brief review of research activities related to this topic. For example, the work conducted for the Inmarsat system, which at present has perhaps the most complex and dynamic frequency planning environment, revealed the advantages of utilizing optimization algorithm for frequency planning using only a subset of algorithms and thus there is scope to conduct further algorithmic research including implementation aspects such as ease-of-use and speed in a heterogeneous environment comprising a mix of carrier types, a variety of user terminals and spot beams of various sizes and isolation (Richharia *et al.*, 2006b).

14.4.3 Cross-Layer Optimization

The optimization addressed in the preceding section refers to arrangement of carriers in the spectrum space without considering issues such as the quality of service (QoS). The vigorous growth in data communication has been driven by IP (Internet protocol) enabled solutions that are based on a rigid hierarchical protocol architecture where each layer communicates with its immediate neighbours and its peer on a pre-defined fixed interface and rules, without cross-layer interaction. The rationale in the emerging discipline of cross-layer optimization is that in wireless communications cross-layer protocol interactions can lead to a more efficient performance of the transmission protocol stack (Kota *et al.*, 2005). By sharing information dynamically, each layer can readjust performance in a coordinated manner such that the overall resource utilization is used more effectively. This approach can potentially enhance the efficiency of radio resource, particularly of wireless systems, which are prone to propagation-induced impairments typical of mobile satellite systems.

The vision of the next generation mobile systems is to provide ubiquitous always optimally connected service to the end user. It is envisaged that optimization in the utilization of radio resources while complying the QoS is broadly achievable at two levels – horizontally and vertically. The *horizontal* strategy would account for the best connectivity to the user at the network level from a choice of radio access network (RAN) – for instance between satellite and terrestrial networks. To provide an optimum network level utilization of radio resources, one conceives of a global manager whose task would be to provide an optimum routing to a connected service through an appropriate RAN (satellite, cellular, WiFi, satellite, etc.). The *vertical* scheme would involve an interaction between the protocol layers of the connected network such that the specified QoS is provided with the best feasible utilization of the radio resources (Giambene, 2007).

In an IP network each layer executes a predefined function and transfers the service to the adjacent layer in a top-down manner, that is application layer downwards, until transmission by the physical layer. The performance of each layer has been optimized independently. The argument is that in such a hierarchical scheme while the service needs are identified at the top-most layer, the performance is mainly governed by the lowest layer (e.g. due to packet loss caused by shadowing and fading). Vital information may be lost as the packets cascade down the layers – hence, a better efficiency is achievable if the optimization is achieved

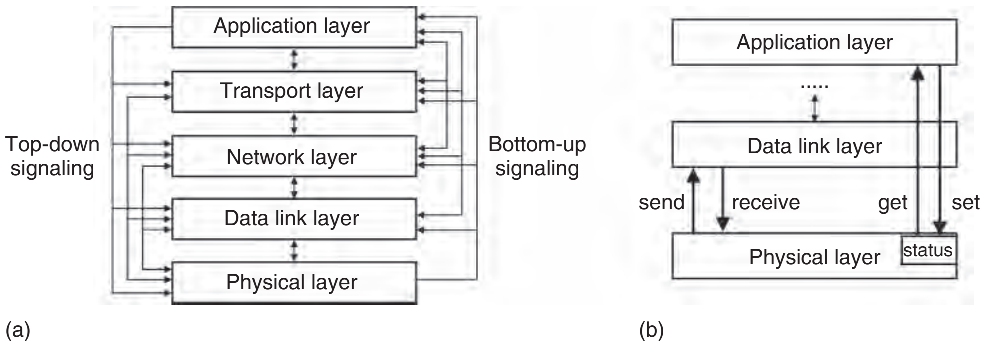


Figure 14.4 (a) Signalling paths for cross-layer communication. (b) Example commands to obtain and transfer the state of protocols in a cross-layer interactive arrangement compared to the *send* and *receive* primitives used in the conventional layered architecture. (Both parts source: Giambene and Kota, 2006. Reproduced by permission of John Wiley & Sons, Ltd.)

jointly through an interaction between the layers. The fixed protocol architecture is said to lack the adaptability needed for hybrid architectures such as a 3G-MSS network.

In an *implicit* cross-layer optimization the inter-layer optimization is performed at the design stage and therefore signalling across layers is not performed in real-time whereas in *explicit* optimization there is real-time interaction between layers thus necessitating additional interfaces. In a hybrid scheme an implicit optimization is supplemented by an explicit optimization. Figure 14.4(a) shows an OSI/IP (Open System Interconnect) hybrid representation illustrating top-down and bottom up signalling paths for cross-layer interaction. Figure 14.4(b) illustrates notional 'get' and 'set' interactive functions for an upper layer to receive internal state of a non-adjacent layer (lower layer in this example) and to make relevant changes to its state. The figure includes classical 'send' and 'receive' primitives used to exchange information between adjacent layers in the conventional layered architecture. Various methods have been proposed to acquire, manage and optimize the protocol states. Such issues and those related to effective implementation of such schemes and methods to comply negotiated service agreements remain areas of active research. Examples scenarios where cross-layer schemes could be advantageous are as follows (Giambene and Kota, 2006).

In an adaptable modulation and coding air interface as in DVBS2 (digital video broadcasting second generation) and BGAN the modulation and coding schemes are controllable at a granularity of frame depending on the prevailing propagation conditions. In a fading condition the lower level modulation and coding scheme reduce the effective data rate. An interaction with the medium access control (MAC) scheduler can provide a more intelligent and efficient flow control.

Consider a handover process between satellites or beams. The delays in routing and reconfiguration during a handover process can throttle an established TCP (Transmission Control Protocol) connection. The delay can be minimized if the mobility manager (network layer) can instruct the scheduler (MAC layer) to prioritize resource allocation to a connection being handed over.

Consider Network-MAC layer interaction in relation to scheduling of Integrated Services (IntServ), which deals with single flows and Differentiated Services (DiffServ), which deals with aggregated flows. Given the service distinction and the desired QoS the MAC layer can schedule the flows to attempt compliance with the QoS. The reference (Giambene and Kota, 2006) gives various examples of improvement feasible by cross-layer interactions.

14.4.4 Cognitive Radio

14.4.4.1 Background

The existing methods of spectrum allocation are known to be rigid, bureaucratic and time consuming. Several bands of allocated spectrum remain under-utilized; even if such allocations are in use, there are periods during a day that the spectrum remains idle. Thus there is considerable interest in promoting dynamic use of such vacant spectrum on an *ad-hoc*, opportunistic basis using agile reconfigurable transceivers. The technology involves the users to sense status of spectrum usage and reconfigure accordingly, and for the network to support such an ad hoc dynamic access. This type of dynamic spectrum access can be implemented through CRSs.

It has been recognized in developed countries like the USA as a 'potential centrepiece in a national strategy to maximize the utility of the finite spectrum resource' with the focus progressing towards solutions to impediments in relation to implementation and subsequent demonstrations of the technology (Marshall, 2011). The US national broadband plan policy document recommends making '500 megahertz of spectrum newly available for broadband within 10 years, of which 300 megahertz should be made available for mobile use within five years' (FCC, 2009). One of the provisions in order to enable this goal is through the use of dynamic spectrum access by utilizing the cognitive radio technology. The so-called TV white space regulations being pursued in the USA and Europe (including the UK) provides for unlicensed use of empty TV spectrum.

Issue in this respect relate to countermeasures against fraudulent use of the technology (e.g. a secondary user may emulate as a primary user), mechanisms to ensure coexistence of secondary user in a heterogeneous radio environment, mechanisms to manage marketing of such spectrum, a framework for co-existence of primary and secondary users compliant to the respective target QoS and standardization of cognitive radio technologies. Considerable efforts towards experimentation and standardization are in progress. Several terrestrial IEEE standards and 4G air interface have incorporated elements of cognitive radio.

Its applicability to satellite systems and to mobile satellite systems opens a new vista of research. For example, it opens the possibility of spectrum trading between cooperating operators with adequate safeguards to ensure non-interference to the primary users; or the possibility to utilize unused spectrum of one service by another in a heterogeneous mix of radio access technologies within the jurisdiction of a single operator.

We will briefly introduce the concept of the CRS. The interested reader can find vast amount of information in the literature (e.g. IEEE, 2008, 2011).

A CRS includes capabilities to:

- acquire knowledge of its operational radio environment, applicable geographical environments and policies, its internal state and usage;
- adjust dynamically its operational parameters, including the protocols;
- learn.

The operational radio environment includes current spectrum usage in the system's surroundings, identity of operational radio systems at the present location, interference levels in available parts of the frequency bands, etc. Geographical information typically consists of CRS's own location; and orientation, range, coverage and radio characteristics of the radio systems of the CRS's interest. The internal state of the CRS includes usage requirements, available frequency band, frequencies and protocols in use, etc. The policies refer to the applicable terms and conditions for usage of the available primary spectrum such as interference threshold and time restrictions. Usage considerations include users' immediate needs such as desired bandwidth, characteristic of the traffic, usage trends of the primary system, and so on.

A CRS can acquire the knowledge to perform dynamic access through a variety of sources and means – spectrum sensing, assessment of its internal status and database, access to a central database, access to a CRS broadcast channel where available, trending acquired data, location-fix from a satellite or terrestrial navigation system and interference measurement. The required knowledge depends on the design methodology of the CR (Biglieri, 2012). In the *interweaved or spectrum sensing* approach the secondary user utilizes the available empty frequencies of the primary user; in the *underlay* approach the user utilizes the spectrum of the primary user such as not to cause harmful interference to the primary user; in the *overlay* approach the secondary user has and utilizes the full knowledge of the primary users' transmission scheme such that the primary user can tolerate the interference caused by the secondary user, while the secondary user contends the interference of the primary user (e.g. by interference cancellation). In a *database driven* approach the CRS and the primary system share relevant information – for example the secondary system obtains the status of the primary system relevant to its current location in order to develop a suitable transmission technique or accesses a radio environment map containing real-time profiles of primary and secondary users. Table 14.1 (adapted from Sharma, Chatzinotas and Ottersten, 2012) summarizes the salient features of these techniques in terms of the required cognitive information, QoS constraints, the issues and examples of applicable techniques.

Spectrum sensing and either pre-existing or acquired knowledge of permissible accessible spectrum enable a cognitive system to identify candidate spectrum areas to access. Spectrum sensing techniques include energy sensing where signal is assumed to be present when the energy exceeds a threshold; coherent sensing that detects presence of a primary user through correlation, being fully aware of characteristics of the primary user; exploiting the inherent periodicity in primary users' transmissions; other methods that have been proposed to improve detection reliability and speed include autocorrelation detection multitaper estimation, wavelet transforms, Hough transform and time−frequency analysis, and other advanced techniques (Biglieri, 2012).

The CRS utilizes the knowledge to dynamically adjust the radio link parameters, that is; transmission frequency, power, access protocols, etc., respecting the specified regulations and specifications laid out to ensure link reliability of both the primary user and itself.

In the learning mode, the CRS utilizes previous information to refine its performance.

Several types of CRS configurations are envisaged and are being standardized (Filin *et al.*, 2011). A heterogeneous CRS network comprises a number of RANs using either the same or different accessing schemes; some of the RANs would be the conventional type operating within a fixed frequency regime whereas others would be reconfigurable operating to

Table 14.1 Salient features of radio cognitive systems

Cognitive techniques	Spectrum sensing (SS)	Underlay (UL)	Overlay (OL)	Database (DB)
Cognitive information	Power spectral density (PSD)	Primary user's channel state and interference threshold	Primary user's channel state and primary user data	Geolocation data, usage of primary and secondary system, usage constraint
QoS constraints	Probability of detection, probability of false alarm	Interference threshold	Rate limit threshold, access time	Database management, access time
Issues	Weak signal detection, management of sensing over wide areas	Resource optimization, interference control	Interference mitigation, efficient coding technique	Database construction and management
Examples of applicable technique	Energy detection, matched filter detection, cyclo-stationary feature detection, spectral shaping, polarization sensing, collaborative Sensing, compressive sensing, others	Power and radio resource allocation, beam-forming and scheduling techniques, multi-antenna and diversity techniques	Superposition coding, rate splitting, relaying, known interference pre-cancellation	Efficient database access techniques

(Adapted from Sharma *et al*., 2012.)

reconfigurable terminals. In another configuration the RAN population of reconfigurable capability shares a band servicing their respective reconfigurable terminals.

14.4.4.2 CRS in Satellite Communication Systems

The CRS systems have been researched extensively in the recent years. Due to congestion in MSS bands and more generally in the popular satellite communication bands (i.e. at present the L, S, C, X and K_u bands), innovative methods are being encouraged to maximize the utilization of these bands. Unfortunately the rigid ITU allocation procedures paradoxically leave parts of congested bands underutilized. Advances in signal processing capabilities of VLSI (very large-scale integration) devices have made reconfigurable receivers a reality and considerable headway has been made towards realization of reconfigurable payloads. A cognitive radio based approach to mitigate the congestion therefore appears attractive for satellite systems. A study conducted by the European Space Agency (ESA) concluded that the concept can be applied to satellite communications in a number of ways. Unique features of satellite systems in relation to CRS are:

- Satellite downlink signals (space to Earth) spread over a wide area and hence the satellite system cannot operate in a secondary mode in a terrestrial-satellite hybrid CRS as shared satellite signals can potentially interfere with one or more of the terrestrial primary users resident within the beam footprint; whereas in the up-link (Earth to Space) since the emitted power of earth stations are geographically confined, it is feasible to construct a satellite system in a secondary CRS mode; A majority of the reported work on hybrid architecture consider spectrum sensing of satellite system only in the uplink direction;
- Mobility associated to MSS user terminals adds another dimension to the CRS system design;
- Whether the CRS in question pertains solely to satellite systems or hybrid satellite-terrestrial systems; and for the latter, which of the two is the primary user;
- For hybrid systems satellite elevation angle adds a reusability dimension beyond the spatial-frequency-time space;
- Effects of propagation (such as fading) attributed to troposphere and mobility must be considered in spectrum sensing;
- Satellite-satellite CRS is feasible through a shared central dynamic spectrum database.

Yun and Cho (2009) suggest joint optimization of secondary transmitter-receiver pair based on minimization of mean square error at the output of the secondary receiver constrained by the requirement that interference at the output of the primary receiver be negligible. The authors demonstrate that when primary and secondary transmitters use linear modulation and the receiver front-end is linear, the secondary user achieves the additive white noise bound (or close to it) by reducing the secondary user's symbol rate below that of the primary and the approach significantly improves the spectral efficiency.

Sharma, Chatzinotas and Ottersten (2012) investigate interference power levels between a terrestrial base station and satellite user terminals for a multi-beam satellite illuminating Europe and propose cognitive techniques applicable to satellite communications for high and low interference. Furthermore, they summarize the current state of cognitive satellite communications. The results demonstrate the sensitivity of the elevation angle to the cognition process and confirm the intuitive expectation that satellite terminals receive more interference at low elevation angles with locations north of 52°N becoming more susceptible. They further conclude that:

- Spectrum sensing and database methods seem to provide the best performance in high interference region;
- Interference to the primary system can be suppressed by an underlay technique (e.g. beam forming, spectral spreading or power control) in low or medium interference regions.
- Overlay technique appears to be suitable only for integrated systems with very high level of interaction both for high interference and low interference scenario and by advanced coding and suitable modulation scheme(s).
- Further work was needed to quantify aggregate interference in a framework of multiple secondary users to assess its impact on the primary users.

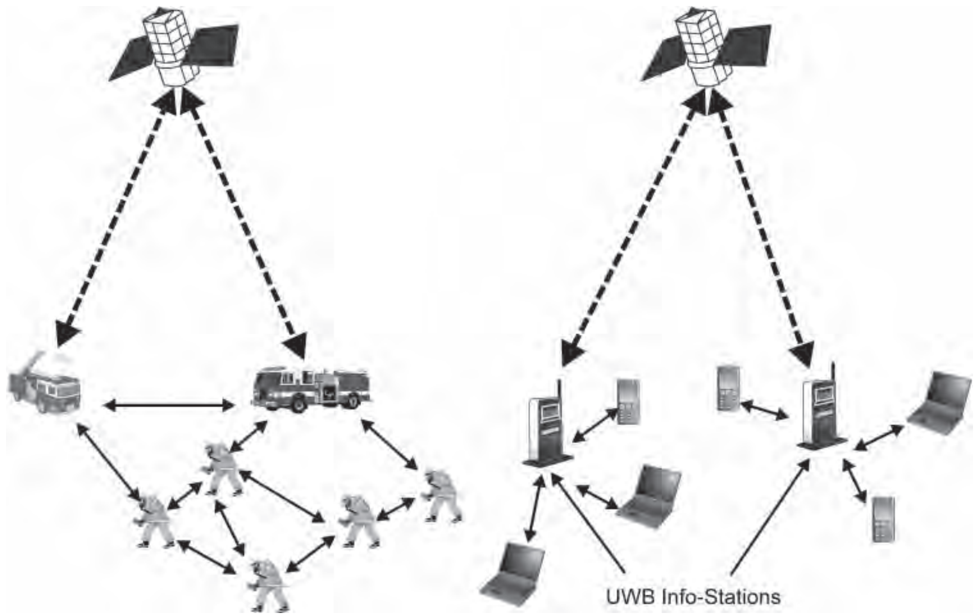


Figure 14.5 Example of a hybrid satellite-terrestrial cognitive radio system. (Source: Kandeepan, *et al*, 2010. © 2010 IEEE. Reprinted with permission.)

Kandeepan *et al*. (2010) present a futuristic cognitive satellite terrestrial radio (CSTR) concept in a hybrid satellite terrestrial system taking as examples a hybrid satellite-ultra-wideband (UWB) and satellite-IEEE 802.22 CR enabled wireless regional area network (WRAN) hybrid network as illustrated in Figure 14.5.

The satellite-UWB network would be a short range personal area network and the satellite-WRAN would be a long range communication system. CR functionalities are adapted by uplink of the satellite service link and the terrestrial radio. The authors suggest the use of elevation angle dependent (3D-spatial) reuse taking advantage of isolation available from the elevation slant. The hybrid satellite WRAN configuration consists of user terminals connected to the base stations with only the base stations communicating to satellites whereas the down-link transmissions are received by base stations and the CSTR. High rate (500 Mbps) short distance and low rate (25 Mbps) UWB systems combined with ranging and positioning operating within 3.1–10.6 GHz range were considered as potential candidate for cognitive solutions given the requirements of such a system for coexistence with other wireless systems. The UWB devices are supported on satellites via relay station in this arrangement.

Figure 14.6 shows the general architecture of the cognitive satellite terrestrial transceiver. The cognitive radio elements reside on top of the core receiver providing the necessary cognition functions for both the terrestrial and satellite components. The dynamic radio environment is shared by a number of the cognitive receiver. The receiver side performs a number of cognitive functions utilizing the radio environment map and feeds them back for transmissions to take place on appropriate channels, while the coherent receiver performs the usual detection function.

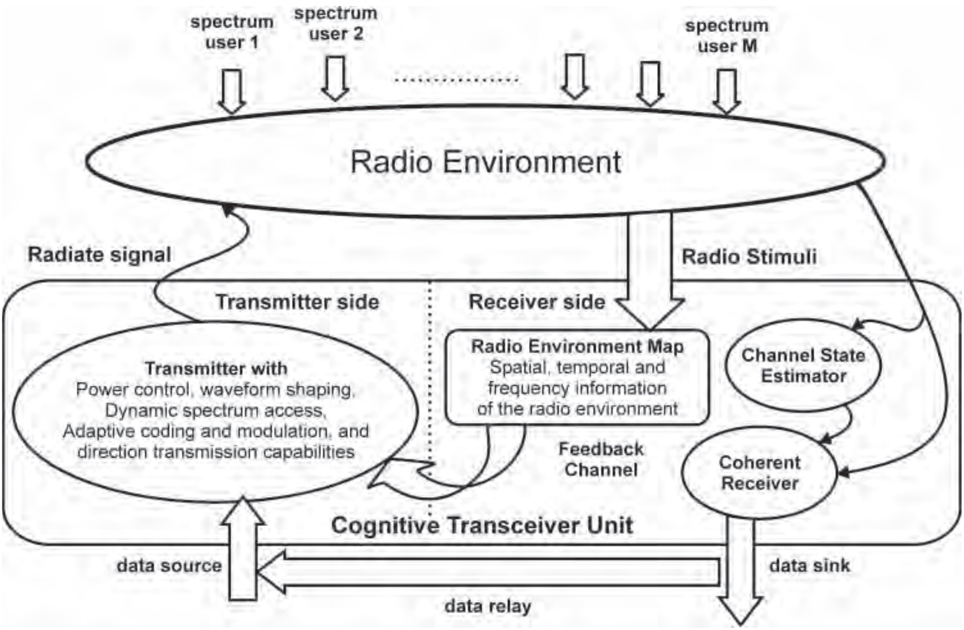


Figure 14.6 General architecture of a cognitive satellite-terrestrial transceiver. (Source: Kandeepan, *et al*, 2010. © 2010 IEEE. Reprinted with permission.)

The busiest of MSS bands exhibit unfilled spectral gaps lasting several hours on a typical working day often exceeding 12 h and extending full day during the weekends. Thus there is considerable scope in utilizing such off-peak hours for non-urgent communication on a CR basis either in a blind or cooperative spectrum assessment mode. It is thus feasible to utilize the spare spectrum either in an intra-system cognitive mode or inter-system cognitive arrangement through cooperation between the systems. A central database could facilitate in identifying the available channel for *ad-hoc* dynamic access. Such a scheme (in an intra-system scenario) has been utilized for leasing spare capacity on a pre-emptible basis by Inmarsat (Richharia and Trachtman, 2006a).

14.5 System Architecture

14.5.1 Terrestrial System Progression

Figure 14.7 (Agilent, 2011) shows the progression of the major terrestrial mobile systems beginning from second generation, leading to the fourth generation – wherein satellite components were incorporated into the second and third generation systems.

The throughput and spectral efficiency continues to increase progressively. The evolution of IEEE 802.11 standards (maintained by IEEE standards committee) for wireless local area network evolved independently under the WiFi brand, catering to the needs of wireless connectivity of devices such as personal computers in office, homes and public areas. The

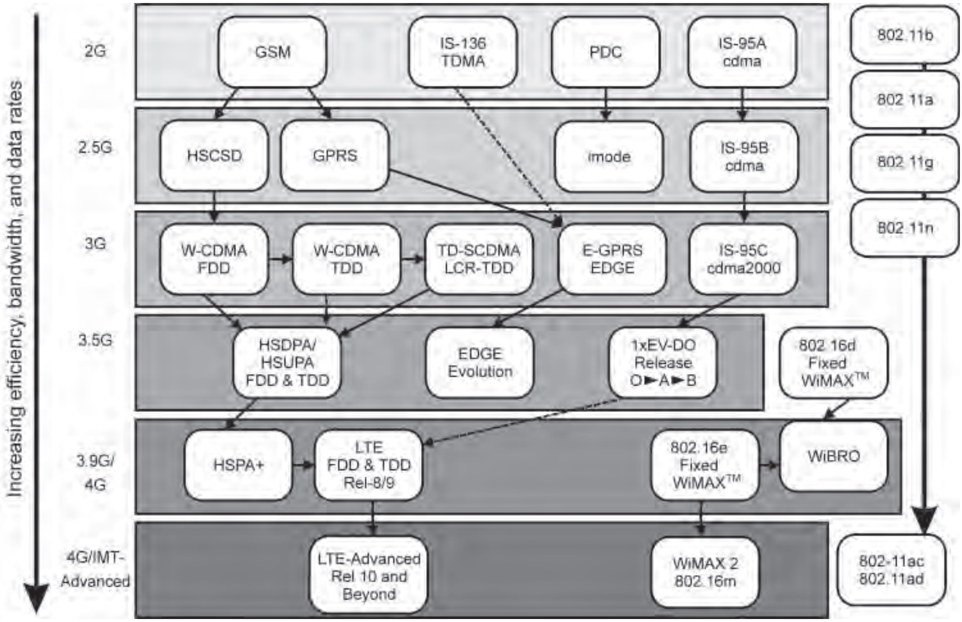


Figure 14.7 Progression of major terrestrial mobile systems. (Source: Agilent, 2011. Reproduced by permission of Agilent.)

most recent versions 802.11ad can theoretically support up to 7 Gbps. The IEEE has also developed 802.16 family of standard for fixed and mobile wireless metropolitan area network commercialized as Worldwide interoperability for Microwave Access (WiMAX) by an industry alliance.

In contrast to previous generations, which comprise a number of standards, the fourth generation systems has converged to two proposals – Long term Evolution (LTE) Advanced (Release 10 and beyond) and IEEE 802.16 m (known as Mobile WiMAX). Both systems are compliant to ITU requirements of the IMT 2000-advanced for the fourth generation terrestrial mobile systems (ITU-R, 2008a). These two systems differ in detail but the underlying technologies are similar in many respects. At the physical layer both the systems utilize OFDMA (Orthogonal Frequency Division Multiple Access), MIMO schemes, fast link adaptation and frequency scheduling. At MAC layer each system utilizes multicarrier access supporting a variety of cell types, that is macro cells, femtocells (e.g. within a building) and relays (Etemad and Riegel, 2010).

General requirements for IMT 2000-advanced include worldwide commonality in functionality, compatibility of services within IMT and with fixed networks, capability to interwork with other radio access systems, support of high-quality mobile services with worldwide roaming and peak data rates targeting 15 b/s/Hz in the downlink (towards mobile) and 6.75 b/s/Hz in the uplink using MIMO arrangements 4 × 4 and 2 × 4 respectively, to support user-friendly applications and services. ITU-R, 2008 sets minimum requirements for cell spectral efficiency, peak spectral efficiency and scalable bandwidth with a maximum consideration up to 100 MHz (non-contiguous band), cell-edge user efficiency, latency, mobility, handover restrictions and VoIP capability. Mobility classes are defined as:

- Stationary: 0 kmph.
- Pedestrian: >0–10 kmph.
- Vehicular: 10–120 kmph.
- High speed vehicular: 120–350 kmph.

Mobility environments includes indoor, microcellular, base coverage urban and high speed.

Consider as an example the main attributes of LTE – one of the systems compliant to 4G requirements set by the ITU-advanced (Rumney, 2013):

- Single-channel peak data rates of up to 300 Mbps in the down-link and 75 Mbps in the up-link;
- Improved spectral efficiency, particularly for the up-link;
- Full integration of FDD (Frequency Division Duplex) and TDD (time division duplex) access modes;
- Packet-based evolved packet core network.

Some of the enhanced technologies include:

- OFDMA and Single channel-FDMA (frequency division multiple access) for the down-link and up-link air interfaces to enable narrowband scheduling, efficient support of spatial multiplexing and lower peak-to-average ratio (PAR) to conserve battery life in mobile devices;
- Support for six channel bandwidths from 1.4 to 20 MHz to enable high data rates and efficient spectrum reuse for narrowband legacy systems;
- Baseline support for spatial multiplexing (using MIMO) of up to four layers on the down-link;
- Faster physical layer control mechanisms to provide lower latency.

14.5.2 Adaptation of Terrestrial 4G Air-Interface

ITU-R has commenced work towards definition of satellite air-interface for IMT 2000-advanced (ITU-R, 2010, 2012). Independent studies have also been conducted to assess the feasibility and adaptation of the proposed 4G standards over satellite channels.

Ansari and Dutta (2009) propose adaptation of a WiMAX system for satellite operation (S-WiMAX) in an ATC (auxiliary terrestrial component) system where the terrestrial component is served by Mobile WiMAX. Such an arrangement reduces the size and weight of dual-band handsets through commonality in hardware and the protocol stack, while maximizing economies of scale through a shared core network. The changes to the WiMAX air interface were attributed to a number of differences between the two medium leading to a greatly reduced link margin, reduced spectrum/power availability and larger and variable propagation delay of satellite networks. These differences required adaptation to WiMAX physical and MAC layers and in particular to the sub-channelization scheme, frame synchronization method and ranging technique. The S-WiMAX sub-channelization scheme was made more granular than their terrestrial counterpart to comply to resource restrictions on satellite; the TDD of the terrestrial TDD mode was made to operate in half frequency duplex mode necessitating time synchronization of the up-link and down-link to avoid overlapping at the mobile, and the ranging process was altered to account for larger differential delay and preserve bandwidth.

Taking into considerations the limitation of satellite system with respect to LTE system (3GPP, 2008 and subsequent releases) Bastia *et al.* (2009) propose solutions to adapt the 4G air interface to operate over satellites to maximize the commonality. The satellite propagation channel differs from terrestrial channels in respect of time delay, fading and system non-linearity. The authors proposed the adaptation on the premises that in comparison to terrestrial systems satellite systems exhibit a significantly larger time delay, and thus cannot utilize the terrestrial time adaptation approach; whereas terrestrial systems exhibit frequency-selective fading, satellite channels tend to have flat fading that requires a different approach from the terrestrial scheme to improve link reliability and satellite channels exhibit an inherent non-linearity, which is not amenable to high peak to average power ratio (PAPR) of the OFDM systems used in LTE. For example, in the terrestrial system the time and frequency dispersion of fades is countered through a scheduling scheme that selects the slots capable of providing better quality channels from those available. The large time delay of satellite cannot provide the agility in terms of channel quality as is possible with terrestrial system. The proposed solutions include introduction of an inter-TTI (Transmission Time Interval) interleaving technique using forced retransmission, that is able to break the channel correlation in slowly varying channels by exploiting inbuilt feature of the LTE physical layer; introduction of PAPR reduction methods to minimize the impact of system non-linearity; adaption to random access channel used by mobiles; and a scheme to improve link reliability by the use of upper layer FEC (Forward error correction). Another initiative towards adaptation of LTE air interface to the satellite media is available in (ETSI, 2008).

14.5.3 Hybrid Architecture

Satellite signals coverage extends over vast expanses. Due to the large distance the signals travel and limited EIRP (effective isotropic radiated power) of satellites, the received signals are vastly lower than their terrestrial counterparts. Users resident in urban areas or within buildings are therefore generally serviced by the terrestrial infrastructure, whereas those in remote and sparsely areas such as remote terrain, ships or inter-continental flights are best serviced by satellite systems. An adaptable, hybrid architecture operating in the same or adjacent frequency band utilizes the strengths of each segment to provide a seamless, ubiquitous service to the user.

Satellite-terrestrial hybrid architecture is a necessity for mobile broadcasts and hence an essential feature of a majority of mobile satellite broadcast systems and standards. With the provision of reusing the satellite transmission terrestrially the architecture has been adapted for communications on a limited scale. High altitude platforms (HAPs) offer an alternative to terrestrial transmitters with the advantage of servicing large areas from a single platform; while at a typical altitude of 17–22 km where they are positioned, the received signal strength is significantly larger than those of satellites so as to offset the limitations of satellites. Mohammed and Pillai (2009) propose a satellite-HAP hybrid architecture for provision of broadband IP service to WiMAX-LAN and WiFi-LAN (Wi-Fi-Local Area Network) user devices based on WiMAX payload mounted on the HAP platform in each of the following scenarios:

- HAP standalone platform;
- HAP platform linking users to the core network via a GEO (Geostationary orbit) satellite;
- HAP-platform linked via a constellation of low and GEO satellites.

The WiMAX (802.16) technology was chosen as it could provide the broadband link data rates up to 120 Mbps supportable on the HAP platform and compliance to link symmetry needs of the future. Performance of each scenario is compared in terms of delay, response time, throughput and traffic capacity on an OPTNET simulation platform. Although HAP standalone system provides the least delay, the hybrid architecture could provide the service within acceptable delay threshold with the added advantage of linking the IP backbone to such a remote system.

14.5.4 Satellite-Enabled ad hoc Networks

A mobile *ad hoc* network (MANET) establishes connectivity between users without a fixed network infrastructure by routing communication to the destination through a series of participating mobile terminals in a fixed or dynamically changing mesh network topology. Examples of *ad-hoc* enabled standards include IEEE 802.11 and Bluetooth. A MANET is well placed to set up an emergency communication network where the local infrastructure has been destroyed or absent, and for specific applications such as ship to ship or aircraft to aircraft communication. Although the use of *ad hoc* networks dates back to the 1970s, various issues remain in the realm of research. One of the fundamental issues relates to efficient and reliable routing in a dynamically changing environment where the users may not always be willing to accept undue burden of relaying messages; a further area of research relates to security, including that of unwarranted intrusion into a close *ad-hoc* network; assessment of radio link reliability depending on the frequency and user mobility is yet another topic of investigation.

It is conceivable to integrate a system composed of a terrestrial MANET to provide telecommunication services in short ranges (say a few tens of kilometres) and a satellite enabled MANET to cover large distances extending from several tens to hundreds of kilometres in scenarios where a fixed infrastructure is absent such as in an emergency situation in response to a natural disaster or in a war theatre. Figure 14.8 (Luglio *et al.*, 2007) shows an example of such architecture.

The MANET terminals are small low-powered mobile or fixed device to service small distances including indoors interfaced with the *ad-hoc* network operating on a different technology platform constituting medium sized satellite terminals that provide interconnectivity service over long distances via a LEO (low Earth orbit) MSS network and to the core IP network via a geostationary satellite network. These classes of user terminals are suitably interfaced. The system attempts to improve the voice quality of *ad hoc* user terminals through optimized protocols as usually the *ad hoc* terminals are better optimized for data; the system additionally attempts to overcome the TCP/IP performance limitations of wireless networks.

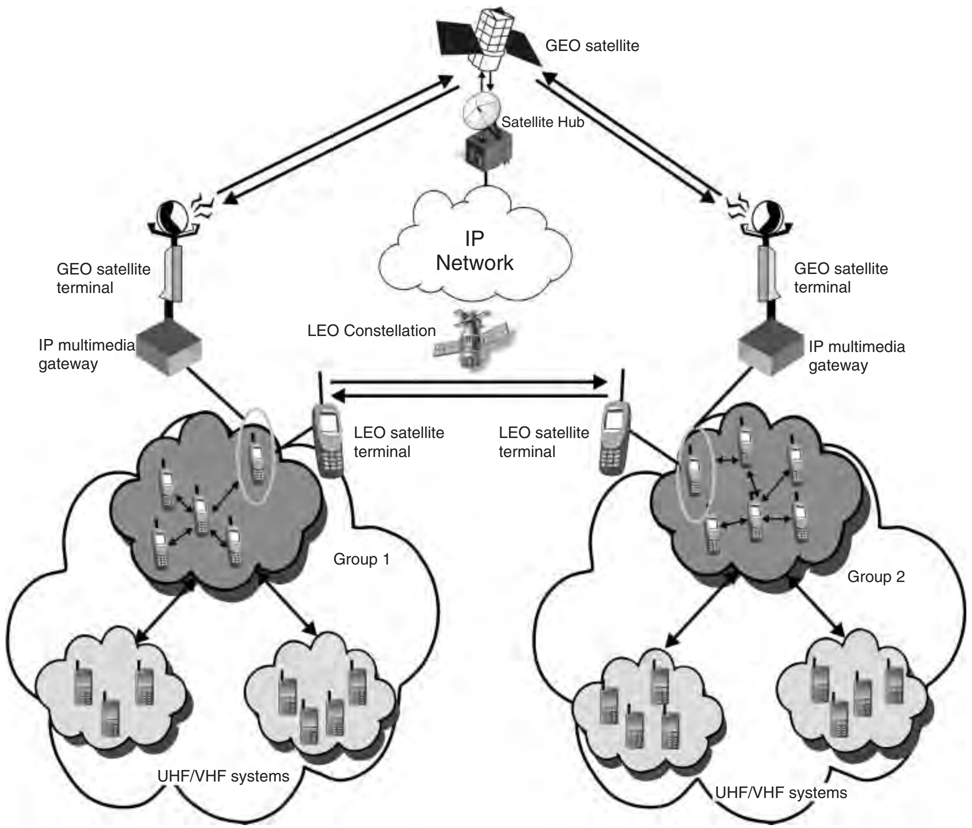


Figure 14.8 Reference architecture of the hybrid *ad-hoc* network. (Source: Luglio *et al.*, 2007. Reproduced by permission of John Wiley & Sons, Ltd.)

By simple interfaces of *ad-hoc* links to satellite links the reliability performance of long distance connections is improved when compared to terrestrial long distance connections.

The outdoor areas are partitioned into three categories depending on range and propagation environment on basis of inputs from local fire fighters in Italy. Teams typically comprising a dozen users in each unit are deployed over a wide area. Teams belonging to the same unit as well as other units are interconnected to each other via a LEO MSS satellite link and each group is interconnected with remote locations on an IP backbone via a GEO satellite link interfaced to the fixed core network. Test results of the mesh *ad-hoc* network and interconnectivity with satellite highlight the characteristics and feasibility of the network.

Gayraud and Berthou (2008) propose a network architecture to provide remote connectivity to an *ad-hoc* network with mobility management over a geostationary satellite system using DVB-RCS air interface for robustness.

Efficient routing protocols are fundamental to the functioning of an *ad-hoc* network. The performance of the protocol is based on reliable real-time signalling reports related to link states, congestions, and so on. In a terrestrial *ad-hoc* network, radio links are often disrupted due to a variety of reason including inclement weather, inhospitable surroundings, node

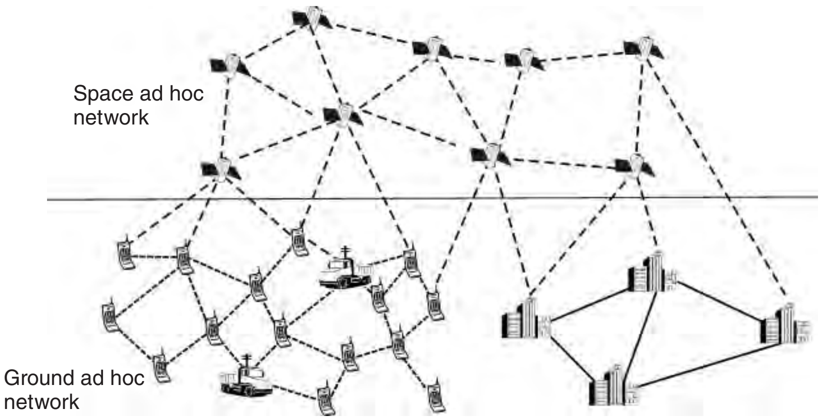


Figure 14.9 Architecture of space and ground *ad-hoc* networks. (Source: Shen *et al*., 2002. Reprinted, with permission, from Shen *et al*, 2002.)

congestion, dynamic topology, and so on leading to degradation in the performance of the network. It has been demonstrated through computer simulations that by transferring the signalling through a geostationary satellite link the packet delivery ratio can be improved by 10% in certain scenarios and the performance improves as hop distance between source and destination increase; and that the performance of routing protocols depends on the mobility model. The proposed future work would assess the feasibility of reducing load on satellite enabled nodes by combining the satellite and terrestrial signalling (Rodriguez *et al*., 2010).

The concept of the *ad-hoc* network has been extended to support communication in a non-synchronous dynamic LEO satellite constellation comprising an *ad-hoc* space network for scientific missions (Shen *et al*., 2002). In particular this work proposes and evaluates a routing concept where the satellites interrogate each other to derive details on the network topology and capacity constraints to make intelligent routing decision. Figure 14.9 shows the concept of the space-based architecture supporting an *ad-hoc* ground network.

In a dynamically changing LEO constellation efficient routing of packets requires a number of considerations – non-homogeneous traffic distribution across the satellites, dynamic changes in location and range, varying degree of congestion, power and processing limitations of space craft, amongst others. Routing can be based on selecting a path that is likely to invoke the least number of handovers on the premise that dropout probability worsens during handovers. Routing schemes are also based on: minimizing the number of hops to destination or propagation delay; balancing the traffic load across the network to avoid congestion on busy routes on basis of best QoS. Various schemes have been proposed and this topic remains an area of active research (Alagöz, KorakÖ and Jamalipour, 2007).

14.6 Enabling Concepts and Technologies

A number of physical layer concepts and technologies are being refined as the drive for higher efficiency, simpler user terminals, broadband and ubiquitous service continues. A sample of such areas is introduced here.

14.6.1 Propagation

In addition to the dynamic shadowing and multipath effects caused by near-far surroundings around an MSS user terminal, radio wave propagation in the K_u and K_a bands is affected by tropospheric impairments such as fading, scintillation and depolarization. The impact of these impairments have been researched extensively for fixed locations; however when atmospheric impairments are conditioned by mobility depending on the location and velocity of the terminals and the underlying mechanism responsible for the impairment. With the advent of mobile VSATs operating in K_u and K_a bands in maritime, land and aeronautical environments there is interest in characterizing such channels more accurately.

A majority of the reported propagation work in the aeronautical environment relate to narrow band characterization of mobile satellite channels. However, while flying over sea and large bodies of water, or more generally, over smooth terrain with good reflective property the channels can become dispersive necessitating the need of an equalizer for channels as wide as 200 kHz (Richharia, Trachtman and Fines, 2005a). The authors estimate the multipath power and delay profile of the satellite channel theoretically by integrating the spatial scattering of the power as received on directive antennas mounted on aircrafts flying at typical cruising altitude. A Doppler model was developed from propagation geometry for the given flight scenario. Different antenna patterns and environments were used to provide results for representative flight phases. The propagation characteristics with three classes of antennas, expected to be used in Aeronautical BGAN, were derived for calm and rough sea conditions at low (5°) and high (20°) elevation angles. The model was used to characterize the performance of BGAN bearers for each type of antenna in a typical aeronautical environment. The results demonstrated that the performance of the high data rate bearers was susceptible to the power-delay profiles necessitating channel equalization at the receiver.

It is argued that at typical cruising altitude the tropospheric effects get diluted and scattered power from the Earth would be further attenuated by the atmosphere. However, it is observed that many domestic flights operate at altitudes where atmospheric affects are still dominant and that scattered power are not adequately attenuated for a vast majority of time. Nevertheless further work in the K_u/K_a frequency bands should assist in optimizing the performance of aeronautical systems.

It is well known that multipaths below about 20° are detrimental to the performance of broadband communication in the L band. In order to mitigate the impact of multipath at low elevation angles, Inmarsat's BGAN system incorporates an elevation angle dependent fading margin over and above the dynamically adaptable modulation and coding scheme. The characterization of the multipath at low elevation in conjunction with rain induced fading and tropospheric scintillation appear to be an interesting area of further work.

There is a heightening interest in the application of MIMO technology to satellite communications, spurred by its successful application in the 4G terrestrial mobile system. The performance of MIMO systems is sensitive to the propagation environment, and, since only a limited amount of data are available for the purpose (e.g. King 2007), further work in this area should therefore assist in assessing the applicability of the technology to the MSS environment.

Thus we note that a better characterization of MSS propagation channels is central in optimizing performance of future broadband systems. A set of representative scenarios are listed next:

- Railway environment in L/S/K_u/K_a bands where only limited data, confined to the K_u band in the European regions, are available (Scalise, 2006);
- K_u and K_a band maritime and aeronautical environments particularly at elevation angles below $\sim 20°$ where multipath scattering begins to adversely impact performance of high order spectrally efficient modulation/coding schemes (Richharia *et al*., 2005b);
- Characterization of MIMO channels.

The ITU MSS allocations are available in a number of segments of the EHF (extra high frequency) band (30–300 GHz) that are as yet unutilized due to a lack of cost-effective technology and severe propagation impairments caused by the atmosphere. World-wide MSS allocations below 100 GHz (space to earth) are available in the upper K_a band at 39.5–40 GHz, and in the V and W bands at 40–40.5 GHz, 43.5–47 GHz, 66–71 GHz and 81–84 GHz (ITU, 2008b). Impairments in these bands are caused by gaseous absorption, absorption by hydrometeors in clouds/fog, attenuation due to rain, and cross-polarization due to rain, ice and tropospheric scintillation. [Note: Q band, often mentioned in the literature, belongs to the waveguide frequency band standard WR-22; it is not a part of IEEE Standard Letter Designations for Radar-Frequency Bands; the Q band overlaps the upper end of IEEE designated K_a band and the lower end of the V band]. About 16 FSS/MSS system proposals were submitted to the FCC by US companies for licenses to operate in these bands in the 1990s (Evans and Dissanayake, 1998); however, most (if not all) have remained at a conceptual stage ever since. It is anticipated that demand for these frequency bands will increase as lower frequency bands fill up – much in the same manner as the technology has progressed from C to the K_a band.

Several rain attenuation models for fixed links have been developed; for example Crane *et al*., 1997 estimate the attenuation distribution using various theoretical models for a location in USA to assist the development of ACTS (Advanced Communications Technologies and Services) propagation instrumentation at 27.5 GHz for the Norman, OK location. Paraboni *et al*., 2007 present eight years of propagation measurements from ITLSAT satellite transmissions at 39.6 GHz as measured at Spino d'Adda in Italy at an elevation of 37.8°. Figure 14.10 shows a sample of the cumulative distribution of fade including the mean and the best and worst years.

Evans and Dissanayake (1998) present theoretical predictions of all-inclusive attenuation for an elevation angle of 20° on two locations – one at Clarksburg, Maryland in the US representing a mid-latitude moderate rainfall climatic zone and the other at Singapore in the equatorial region that has heavy rainfall, representing the worst case; and depending on the relative humidity (10–100%), respective losses due to gaseous absorption at 40 and 50 GHz range between 0.8–2.2 dB and 4.5–6.5 dB. Figure 14.11 illustrates the estimated cumulative for Clarksburg and Singapore locations.

Approximate values of attenuation obtained from the report are summarized in Table 14.2 for illustrative purpose to demonstrate feasibility of communication in Q and V bands for $\sim 90\%$ of time with achievable link margins, even in the locations with high rain rate. Thus it could be surmised that broadband communication utilizing the wide bandwidth available in the Q and V band is conceivable. A multi-band system would step down the frequency progressively with a corresponding reduction in throughput in compliance to the available bandwidth in a V-K_a-K_u and L band sequence depending on the intensity of rain. Such a scheme has been envisaged for Inmarsat's K_a band Global express system where it would be feasible to step down from K_a down to the L band in order to maintain communication continuity.

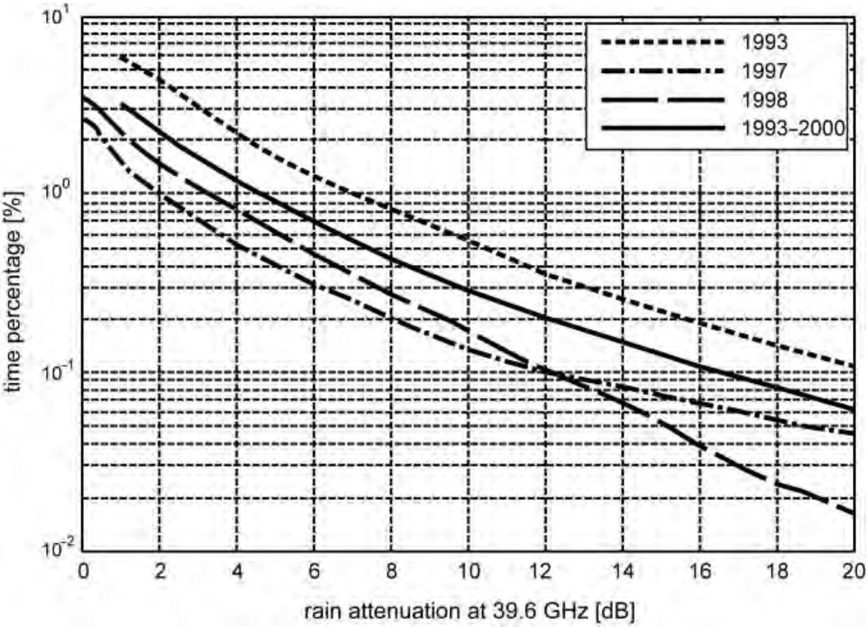


Figure 14.10 A sample of cumulative distribution of fade at a Spino d'Adda in Italy showing the mean and the best and the worst year obtained from 39.6 GHz transmissions of ITALSAT satellites. (Source: Paraboni *et al.*, 2007. Reproduced by permission of IET.)

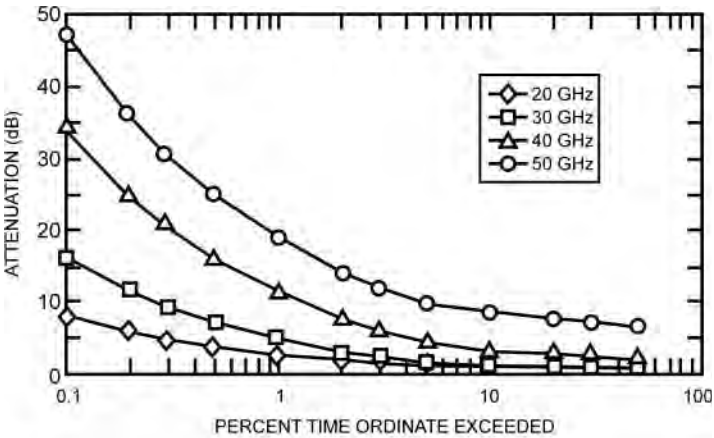
14.6.1.1 EHF Band Status

A number of Q, V and W band experiments have been proposed in the past for future personal communication applications. Table 14.3 summarizes some initiatives for experimentation in this band.

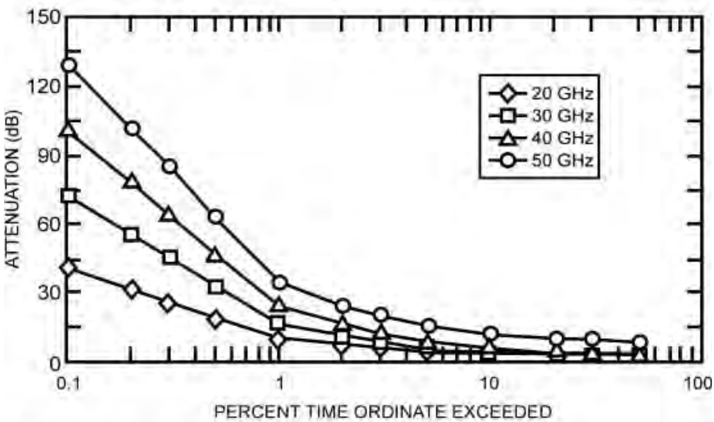
In addition, the ESA intends to include a K_a/Q/V beacon payload in Alphasat satellite mission to gather data in preparation for future missions (Paraboni *et al.*, 2007).

Notable amongst these is the Data Audio and Video Interactive Distribution (DAVID) experiment proposed by the Italian space agency, originally planned for 2002 (Ruggieri *et al.*, 1998). The mission planned use of the 94 GHz band, which lies above the oxygen absorption band (50–70 GHz), for the service link to a LEO satellite. The LEO satellite is connected via an ISL (intersatellite link) operating in the 23–27 GHz band to ESA's ARTEMIS GEO satellite and then on to a gateway and backbone packet switching network, as illustrated in Figure 14.12.

The DAVID satellite is planned to deploy a regenerative transponder with data rate adaptation for matching the service data rate to the intersatellite data rate. The mission will provide data for channel characterization, new service trials and is expected to support an Italian Antarctic mission. Its application is envisaged for the next generation Internet service; the system is planned to comprise a hybrid constellation. Some of the salient features of such a network for Internet services are stated as follows:



(a)



(b)

Figure 14.11 Cumulative distribution of fade for Clarksburg (a) and Singapore (b) corresponding to the rain regions of moderate and heavy rain. (Both parts source: Evans and Dissanayake, 1998. Reproduced with permission from Evans and Dissanayake, 1998.)

- LEO satellites may be deployed gradually, as required, avoiding the need for a fully interconnected constellation from the outset.
- GEO K_a band satellites already deployed by then may be used as an overlay.
- The 94 GHz band and a LEO constellation can provide enormous space segment capacity.
- Relatively lower availability caused by propagation outages is expected to be acceptable for Internet services.
- Such a system is robust to space segment and network failures due to in-built redundancy in the architecture.

Propagation loss at EHF bands is high, which must be countered by high satellite EIRP. For example, 6 mm/h rain rate fade at 40 and 50 GHz are of the order of 11 and 16.5 dB

Table 14.2 Approximate link margins at corresponding to various link reliabilities to illustrate the frequency and climatic rain intensity dependence (Mod = moderate rain zone (US); Heavy = heavy rain zone (Singapore))

	Link availability	99.9%		99%		98%		90%	
	Rain zone	Mod	Heavy	Mod	Heavy	Mod	Heavy	Mod	Heavy
Link margin (dB)	20 GHz (Evans and Dissanayake, 1998)[a]	8	~ 40	2.5	~ 7.5	2	~ 6	~ 1	~ 3.5
	30 GHz (Ibid)	16	~ 70	5	~ 15	2.5	~ 8	~ 1	~3.5
	40 GHz (Ibid)	34	~ 100	12	~ 22.5	7.5	~ 15	~ 2.5	~ 3.5
	50 GHz (Ibid)	47	~ 128	18	~ 37.5	14	~ 22.5	~ 6	~ 7.5
	27.5 GHz (Crane *et al*., 1997)-US	6–19		2.5–5.5		1–4		<1.0	
	39.6 GHz (Paraboni *et al*., 2007)[b] -Italy	12–20		1.9–7		0.6–4.5		<0.6	

[a]Includes all types of atmospheric impairments.
[b]Best-worst variability over 8 years of measurement.

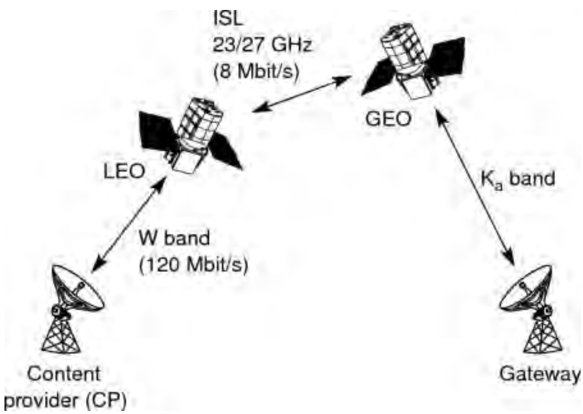


Figure 14.12 DAVID network architecture. (Source: Gargione *et al*, 1999. © 1999 IEEE. Reproduced by permission.)

and the corresponding atmospheric loss around 2.1 and 0.5 dB respectively. As EHF power amplifiers have a limited capability, satellites must use very narrow spot beams and include a capability to distribute power to each spot beam on demand to maximize the use of the satellite power resource.

14.6.2 Modulation and Coding

PSK schemes up to QPSK provide robustness in weak signal to noise and interference prone environments and hence are dominant. Recent advances have enabled higher order modulation schemes up to 16 QAM and 32 APK (Amplitude Phase Keying), the latter providing a

Table 14.3 A summary of Q, V and W band initiatives proposed for future personal communications applications

Frequency band (GHz)	Goal	Spacecraft	Comments	Country/year	References
38–43	New frequency bands; Intersatellite link (ISL) development; encourage millimetre wave device development; experiment personal communication system.	ETS-VI	Frequency selected as a trade-off between achievable technology level and atmospheric attenuation; elliptical orbit used for experiments due to launch failure; Ground terminal used for Simulating ISL.	Japan/1994	Gargione *et al*. (1999)
44–47 and 21–31	High bit rate personal communications experiment.	COMETS	Elliptical orbit (Apogee 17, 700 km, Perigee 500 km) used due to launch failure; 21–30 GHz and 44–47 GHz regenerative transponders interconnected at IF and baseband	Japan/1998	Wakana *et al*. (1998)
40–50 and 20–30	Channel characterization; land and aeronautical field trials	ITALSAT		Italy	
94	Personal wideband communication (120 Mbps user data rate) experiments.	Data audio and video interactive distribution (DAVID) mission; LEO satellite with ARTEMIS GEO satellite	LEO-GEO system with ISL	Italy/ESA (was planned 2002; status not available)	

IF = intermediate frequency.

better resistance to system non-linearity when applied with pre-compensation at the transmit end. The state-of-art in coding has converged to turbo-code and low density parity code, which through iterative decoding provide capacity within a fraction of the theoretical limit. In another development applied to the DVB/RCS + M standard, an FEC scheme introduced at an upper layer, helps to protect packet loss. It is feasible to operate at very low carrier to noise ratio with these codes; unfortunately such low signal levels place a heavy burden on carrier and timing recovery. Moreover, frame synchronization, and code acquisition and tracking, particularly after a prolonged shadowing event, are problematic at such low signal-to-noise ratios. Various advanced techniques are under investigation in these respects – these include iterative techniques based on turbo decoding scheme and insertion of pilot symbols for initial estimate of a channel (Vanelli-Coralli *et al*., 2007).

14.6.3 Multiple-Input Multiple-Output

MIMO technology has been successfully deployed in the terrestrial systems for over a decade testified by their adaptation to various terrestrial based wireless standards, LTE and 802.16 m amongst others. Abundant literature dealing with MIMO techniques – mostly addressing applicability to terrestrial wireless system – is available for the interested reader (e.g. Paulraj *et al*., 2004; Mietzner *et al*., 2009). There is a growing interest in adapting the technology to benefit performance of satellite systems. And hence, several investigations have addressed MIMO application over satellite channels (Arapoglou *et al*., 2011).

MIMO techniques provide performance gain through the use of the spatial domain without incurring bandwidth penalty. A *single user multiple input multiple output* (SU-MIMO) configuration refers to a single user point-to-point connection where transmissions from n_t transmit antennas are received by n_r receive antennas commonly referred to as the $n_r \times n_t$ configuration. Using space-time coding an SU-MIMO scheme enhances reliability through diversity gain and alternatively, improves channel throughput by spatial multiplexing. Figure 14.13 represents a SU-MIMO configuration representing transmissions from two

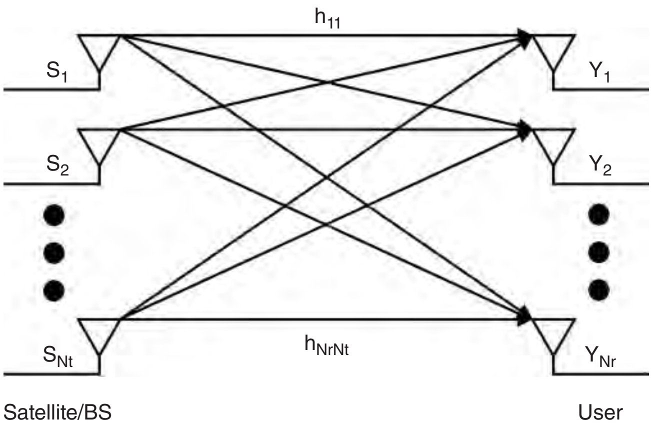


Figure 14.13 A SU-MIMO configuration representing transmissions from two antennas and reception by two antennas. (Source: Arapoglou *et al* (2011). © 2011 IEEE. Reproduced with permission.)

antennas and reception by two antennas that is potentially implementable in a satellite system.

Considerations in evaluating suitability of a MIMO include:

- Availability (or not) of channel state knowledge at transmitter and receiver; in fading condition with zero mean circularly symmetric complex Gaussian distribution and full knowledge of the channel at the receiver the capacity increases as min($n_{r,}$ n_t), where n_r and n_t are respectively numbers of receiver and transmitter antenna;
- MIMO systems benefit from an increase in spatial decorrelation between the signals received on each arm, and hence high correlation are detrimental to MIMO efficiency;
- Multiple polarisation transmission from a single antenna can override the space limitations attributed to hoisting of separate MIMO antennas (a typical restriction of a satellite system) and yet provide MIMO gain; this is a key finding for implementation on satellite channels;
- Space time codes (STCs) provide a practical way of achieving MIMO gain with channel state information at the receiver only, which can be readily obtained by provision of pilot symbols; the Alamouti scheme is said to be a simple and ingenious scheme that can be used without channel state information at the transmit end (Alamouti, 1998; Arapoglou *et al.*, 2011); various other schemes have been proposed (e.g. Golden STC block code);
- For diversity gain the STC can transmit one or less than one independent symbol per symbol period over the MIMO antennas; for spatial multiplexing T independent symbols are transmitted per symbol;
- Various types of MIMO receivers have been proposed to support MIMO techniques over satellite. For instance successive interference canceller receiver, based on iterative interference cancellation, offers a trade-off between complexity and performance.

Multiuser multiple input multiple output (MU-MIMO) networks utilize the degrees of freedom available spatially from multiple users (transmit antennas) to enhance the system capacity through spatial multiplexing by scheduling multiple users to share the channel in the spatial domain. Unlike traditional scheme such as TDMA, MU-MIMO achieves capacity enhancement without a bandwidth penalty. In addition to stream multiplexing in the forward direction MU-MIMO offers multi-user multiplexing advantage in the return direction depending on the numbers of users and base station antennas; note that the scheme supports capacity gain at the base station without the necessity of multiple antennas at the user end (Gesbert *et al.*, 2007). The MU-MIMO requires knowledge of each channel at the transmission end that becomes difficult due to multiplicity of supported users. Coding-decoding is more complex in comparison to SU-MIMO as the antennas (i.e. users) are not collocated. A technique known as precoding is used in the forward direction to cancel out interference from other users based on the knowledge of channel state and other transmissions. Figure 14.14 illustrates a schematic of the MU-MIMO configuration. The forward direction is the broadcast channel and the return direction, the multiple access channels.

Since there are differences in the channel characteristics of terrestrial and satellite systems and there are various arrangements of MIMO configurations, the research focus has been to select the best configuration suited to satellite communication systems while including the economic impact of deploying MIMO antennas on satellites and economics and

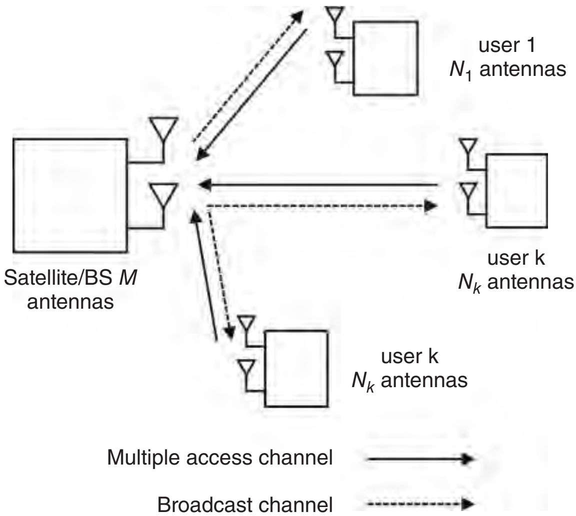


Figure 14.14 A MU-MIMO configuration representing transmissions from two antennas and reception by two antennas. (Source: Arapoglou *et al* (2011). © 2011 IEEE. Reproduced with permission.)

antenna mounting constraints of mobile terminals. The MIMO applicability would further be governed by the service (i.e. multicast, broadcast, point-to-point, fixed, mobile), orbit (GEO or NGEO (non-geostationary earth orbit)), application (e.g. delay tolerant or not) and the frequency band (e.g. L, S, K_u, K_a).

The applicability of MIMO technique to land mobile channels in the L and S bands and to FSS frequency bands above 10 GHz MIMO as well as MIMO propagation measurements and models have been investigated by several authors (Arapoglou *et al*., 2011; Liolis, Andrikopoulos and Cottis, 2008; King, 2007). To circumvent the antenna mounting space restriction both on satellite and on mobile, dual polarized antennas are favoured to provide a compact 2×2 MIMO configuration (see Figure 14.15a). Figure 14.15(b) shows a representation of a dual-satellite single polarization 2×2 MIMO configuration and Figure 14.15(c) illustrates a dual-satellite dual-polarized 2×4 MIMO configuration. The relatively greater interest in MIMO application to MSS in comparison to FSS is attributed to the fact that MIMO gains would be expected to be higher due to the greater de-correlation between signal paths in MSS channels.

Pérez-Neira *et al*. (2008) evaluate the applicability of MIMO techniques to a DVB-SH MIMO satellite broadcast system where satellite and user receivers deploy dual-polarized antennas for MIMO diversity and multiplexing gain. The performance of OSTBCs (Orthogonal Space-Time Block Codes) technique and STTCs (Space-Time Trellis Codes) technique were evaluated. It was demonstrated that the proposed STTC and OSTBC techniques can improve the BER (bit error rate) in comparison to independent stream multiplexing on each polarization and single input single output (SISO) transmissions. By performing a joint distributed OSTBC, the spectral efficiency increased. A hybrid satellite-terrestrial network was also considered for MIMO transmission. It was demonstrated that the spectral efficiency can be multiplied by 4 when joint encoding of satellite and terrestrial signals was performed.

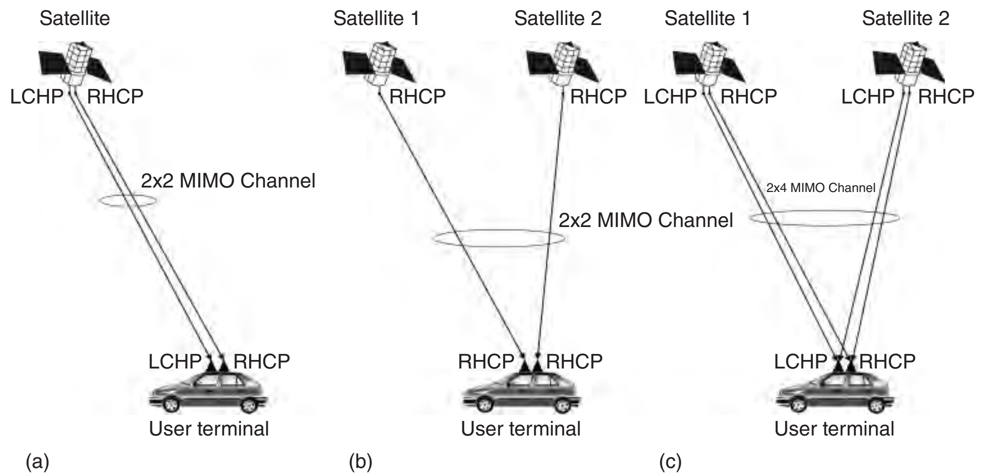


Figure 14.15 (a) A single satellite dual-polarized 2 × 2 MIMO, (b) a single polarized dual-satellite 2 × 2 MIMO and (c) a dual-satellite dual-polarized 2 × 4 MIMO. (All parts source: Arapoglou *et al* (2011). © 2011 IEEE. Reproduced with permission.)

Table 14.4 A comparison of capacity achievable by SISO and 2 × 2 MIMO techniques

Environment	Capacity (bits/s/Hz): 50% of channels		Increase in capacity (C_M/S_S)[a]
	SISO	MIMO	
Major road (UK)	>0.39	>0.96	>2.46
Suburban	>0.8	>1.35	>1.69
Urban	>0.27	>0.67	>2.48

[a]C_M = MIMO Capacity; S_S = SISO Capacity
(Data source: King and Stavrou, 2006.)

King and Stavrou (2006), demonstrate the theoretical capacity gain achievable by a 2 × 2 MIMO single satellite dual-polarized channel in comparison to a SISO scheme on the basis of measured propagation data at 2.45 GHz over an averaged 15° elevation simulated path in three types of land-mobile satellite environment (main road, suburban and urban) of about 2 km each around Guildford, UK. Table 14.4 summarizes a part of the results to highlight the theoretical bound in improvements for the measured channel, assuming that the underlying theoretical assumptions are applicable to the measured data. The table shows the capacity (bits/s/Hz) exceeded by 50% of the channels demonstrating an increase by a factor of 1.69–2.48.

The theoretical capacity improvements were derived by assuming that the transmitted signal vector is composed of n_T statistically independent equal power components each with a Gaussian distribution and the transmitter has no knowledge of the channel state (and hence transmits equal power on each channel), using the capacity bound proposed by Foschini and Gans (1998):

$$C = \log_2 \det[\boldsymbol{I}\mathbf{n_r} + (\rho/n_T).\boldsymbol{H}\boldsymbol{H}^{\dagger}] \text{ bps/Hz} \tag{14.3}$$

where, $\boldsymbol{I}\mathbf{n_r}$ is $\mathrm{n_r} \times n_t$ identity matrix; ρ is the signal to noise ratio; $H^{\dagger}$ is the conjugate transpose of H, the channel matrix (i.e. normalized channel power transfer characteristics); n_r and n_t are respectively the numbers of antennas at the receiver and the transmitter.

Various aspects of the MIMO technology require consolidation in order to better scope its applicability to MSS systems. A wider exposition to channel modelling is necessary in order to accurately quantify the advantage. Experimental data to validate theoretical studies are scarce. The dependency of MIMO performance on channel behaviour indicates variability in performance when applied to different mobile channels – rural, suburban, urban, railways, maritime and aeronautical environments – and elevation angle. Further investigation is therefore necessary to include the full range of MSS environments. A majority of the reported work considers the L and S band channels; with the advent of the K_u and K_a bands for MSS; therefore, MIMO characterisation at these higher frequencies.

Although Golden code in a dual polarization/single satellite scenario can be considered as a feasible near-term solution, its decoding becomes complex to exploit full potential of MIMO. The preferred solution would therefore be an STC capable of providing high performance with lower complexity to enable operation in a dual polarization/single satellite scenario (Arapoglou *et al.*, 2011). Occurrence of cross-polar interference in circularly polarized L and S band systems arising due to polar reversal from scattering is well known; its adverse impact (or indeed, possible benefits) on the performance of MIMO systems appears not to be addressed adequately in the literature.

A majority of the reported work address the GEO satellite systems, and hence, investigations on the applicability of the MIMO techniques to the vast variety of NGEO (nonstationary earth orbit) systems warrants further work. System issues in terms of link budgets, implementation aspects, spectrum availability and economic viability would have to be traded off against throughput advantage and link robustness before MIMO can be considered for practical implementation.

14.6.4 Software Defined Radio

Wireless technologies continue to evolve at a phenomenal rate, making hardware platforms obsolete within a very short span incurring heavy costs in the development of new platforms. Incorporating upgrades to an existing wireless technology incurs a similar development cycle. Software radio (SR) technology offers the potential of a cost-effective, flexible and software upgradeable platform that can be reconfigured to suit the changes/upgrades to operating platforms. The technology can perform on all the layers of communication architecture to enable a number of standards to be supportable on the same hardware platform, allowing rapid changes from one communication standard or functionality to another. For example, a multi-mode satellite terminal could be reconfigured in real time to select the appropriate medium. Similarly the infrastructure can be upgraded rapidly at a fraction of the cost of a total refurbishment.

An ideal SR operates end-end covering the entire platform – that is from antenna downwards; however, in practice the front ends are usually analogue and therefore a practical incarnation of the SR is called as software defined radio (Jondral, 2005). An antenna receives a wide band of radio frequencies that would require digitization of a vast bandwidth and hence digitization is more efficient at intermediate frequencies. The SDR utilizes hardware devices that support reprogrammable firmware and software technologies to implement the desired functionalities. Such devices include field programmable gate arrays (FPGAs), digital signal processors (DSPs), general purpose processors (GPP), programmable system on chip (SoC) and other application-specific programmable processors.

The SDR technology offers numerous benefits (Wireless Innovation Forum, 2012):

- Reconfigurable platform architecture facilitates rapid and cost-effective development of new products and service multiple markets;
- It offers remote reprogramming capability, allowing software upgrades and addition of new features while the unit is in service and in a remote location;
- The SDR platform future-proofs network functionalities, services and products within limits of the platform technology;
- It reduces end-user costs and enhances user experience through ubiquitous communication availability through reconfigurable user terminals.

Although SDR has been used since the 1980s, its implementation has increased markedly in recent years. Figure 14.16 compares the shipments of base station equipment of software controlled and software defined technologies to illustrate the high penetration of the technology in the terrestrial mobile infrastructure.

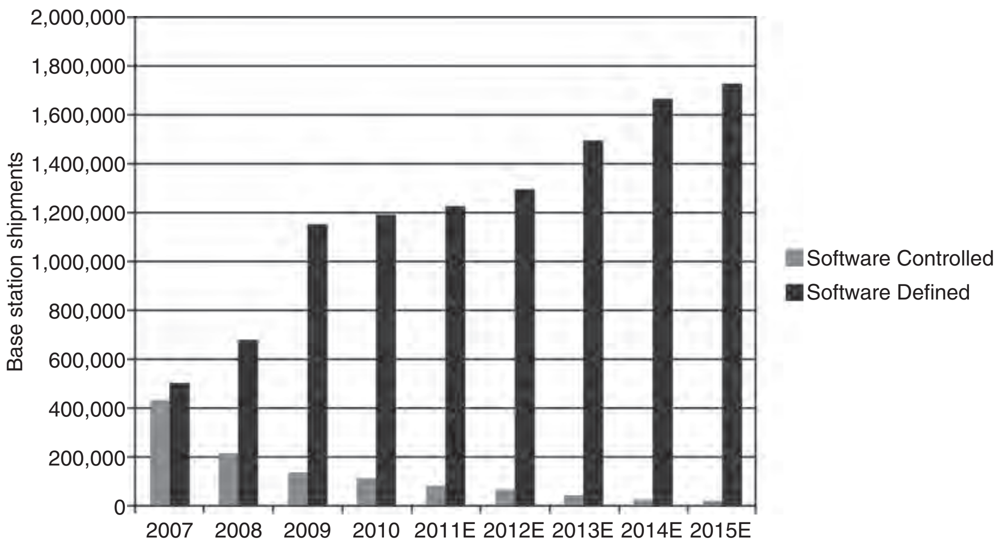


Figure 14.16 A comparison of base station equipment shipment of software controlled and software defined technologies (E = Estimate). (Source: Wireless Innovation Forum, 2012. Reproduced by permission of Mobile Experts and SDR Forum.)

Reprogrammable units are in regular use for modems and baseband units for satellite communications (Wireless Innovation Forum, 2012). Inmarsat's BGAN user terminal has been implemented on a Software Communications Architecture (SCA) version 2.2.2 compliant platform, a non-proprietary specification sponsored by the Joint Tactical Radio System (JTRS) programme of the US Department of Defence (explained later in this section). The platform can operate as a multimode radio integrated with any military-grade SDR capable of transmitting and receiving in the BGAN L band frequency range (1.5–1.6 GHz) and also enables upgrade of BGAN terminals functionality such as inclusion of new or higher data rates (Gatehouse, 2011). An SDR system was developed under ESA's jurisdiction (ESA, 2008) to demonstrate the feasibility of a reconfigurable regenerative spacecraft that would be upgradable and capable of reconfiguring itself to support existing and emerging standards (ESA, 2008). The concept of cognitive radio mentioned earlier in the chapter is achievable primarily by SDR technology. Depending on the requirement, an SDR can be configured as multi-band, multi-standard, multi-service, multi-channel or multi-mode system.

Figure 14.17 shows a conceptual diagram of a generic SDR transceiver; the main blocks of the unit are the front end, the baseband processing unit and the data processing. The functionality of the unit is controlled by changing parameters through the control bus (Jondral, 2005).

Consider an example of a platform useful in the development of open interchangeable SDR software (as opposed to propriety). In order to provide a common platform for the development of JTRS, the future communications architecture of the US joint forces, based on extensive use of SDR to support multiple wireless waveforms, the responsible entity for the system *Joint Programme Office* (JPO) recommends the use of a tool known as *software communications architecture* in order to ensure that the industrial partners and developers may develop components of the system independently and maintain interoperability and exchangeability (JTRS Standards, 2012). It is an open framework based on SDR philosophy that prescribes the interaction between hardware or software blocks within the JTRS. It provides for portability of applications software between different SCA implementations

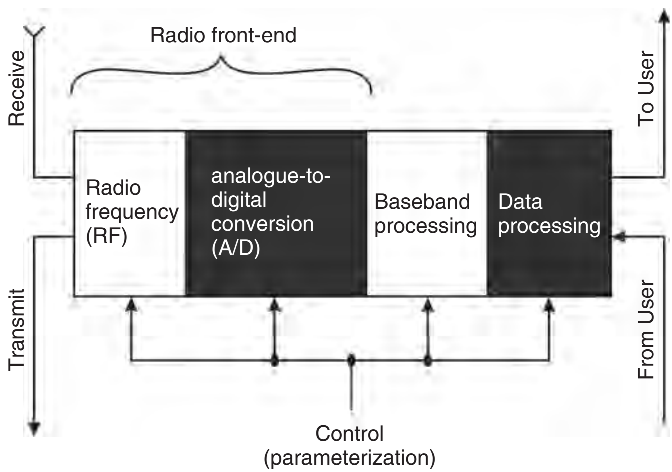


Figure 14.17 A generic SDR transceiver. (Source: Jondral, 2005. Reproduced with permission of EURASIP.)

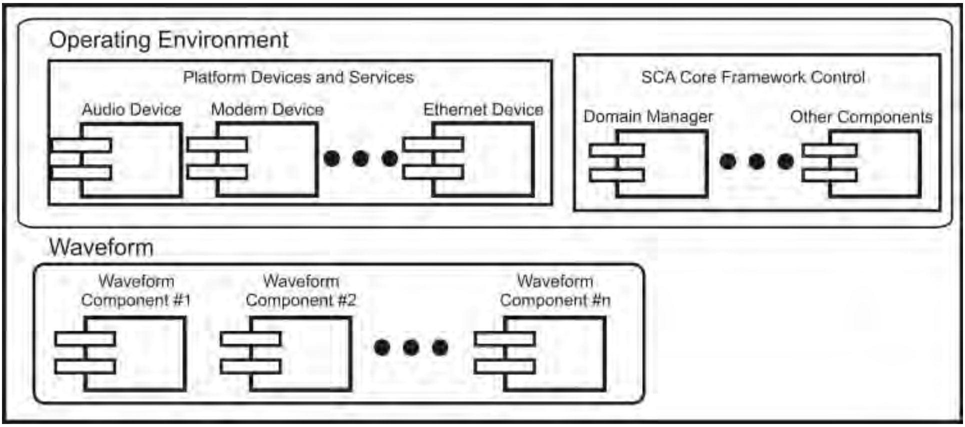


Figure 14.18 Composition of the SCA system. (Source: JTRS Standards, 2012.)

and thus reduces software development time through the ability to reuse design modules while benefiting from evolving commercial frameworks and architectures. The SCA is an open standard that can be utilised for military as well as commercial application developed through industrial participation and thus can provide a solid platform for future development in SDR technology. It is ‘not a system specification but an implementation independent set of rules that constrain the design of systems to achieve the objectives listed above’.

Figure 14.18 shows the architecture of the SCA platform illustrating its constituents – an Operating Environment (OE) responsible for hosting and access to system resources to one or more ‘waveforms’, that is application (JTRS Standards, 2012).

SCA devices are those software components that provide access to the system hardware resources. Software components for use by applications are generically referred to as services – the SCA does not specify an interface for these components.

14.7 Little-LEO Systems

There is an array of applications that can be served by low bit rate/low cost systems. In this respect little-LEO systems have a role in the future MSS landscape. Little-LEO satellite systems operate in a different band, use low-cost simple satellites and deliver small volumes of data to small low-cost user terminals. The markets include pipeline monitoring, reading utility metres in remote areas, fleet management, hand-held messaging, assisting rescue services, machine-to-machine communication, etc. The market for such systems is driven by a number of complementary devices such as hand-held computers, embedded GPS (global positioning system) chips, cellular voice and data, Internet/e-mail, amongst many others. One of the key features of such systems is the low-cost space segment comprising simple small satellites and a basic infrastructure, resident on low-costs including the Internet.

As geostationary satellites become larger and more complex, little-LEO satellites get smaller, smarter and cheaper with the evolution of space-hardened VLSI chips. On the extreme end of this shift, researchers are pursuing radically different technologies that could enable satellites to shrink to palm-size, weighing around 15 g and using control systems based on the principles of animal neurons (Page, 1998). These satellites do not use microprocessors for control or fixed algorithms for operation. The so-called nervous net technology 'works like the neurons in animal nervous systems which put out spiked pulses that hold information in the timing between the pulses'. The gradients of the Earth's magnetic fields are used for stabilization and the light gradient from photo-sensors for orienting the satellite with respect to the Sun. A combination of short and long range communications establish intersatellite (*ad-hoc*) links for connectivity. Such concepts, applied to communication satellites, can bring a significant reduction in the cost of LEO systems in the long term. Low-cost LEO satellite technologies due to their cost-effectiveness continue to develop rapidly in a number of companies, institutions and universities.

14.8 Mobile Satellite Systems in Future Networks

14.8.1 Aeronautical Systems

We will introduce a brief case study pertaining to the role of satellite in future aeronautical networks for its topical interest. According to a global market forecast undertaken by Airbus, the passenger and freight traffic will grow at a compound annual growth rate (CAGR) respectively of 4.7 and 4.9% to the year 2031 resulting in an increase in global passenger carrying aircraft fleet rising from 15 556 in 2011 to 32 551 in 2031 while freighter aircraft fleet will increase from 1615 in 2011 to 2938 during the same period. The air traffic is said to approximately double every 15 years as judged from the historic data (Airbus, 2013).

This level of growth will also drive up the passenger and air traffic management (ATM) communications requirements. It is envisaged that the existing infrastructure will not be able to support this level of traffic and the desired performance quality.

EUROCONTROL, the organization responsible for safety of air navigation in Europe and its US counterpart, the Federal Aviation Administration (FAA), both striving to provide safe and the efficient aerospace system worldwide, initiated a joint study to identify potential future communications technologies to meet safety and regularity of flight communication requirements for Air Traffic Services (ATS) and Aeronautical Operational Control (AOC). One of the tasks within this joint framework was to assess the new technologies to meet the requirement of the future radio system (FRS) that was defined as a part of the work in a requirements document called Communications Operating Concept and Requirements (COCR) (EUROCONTROL/FAA, 2007). Figure 14.19 shows the conceptual diagram of such an FRS (QinetiQ/EUROCONTROL/FAA, 2007).

The communication requirements of an aircraft from the originating airport to the destination depend on its location, which are identified in the COCR as follows:

- Airport Surface
- Airport Zone
- Terminal Manoeuvring Area (TMA)
- En-Route
- Oceanic Remote and Polar

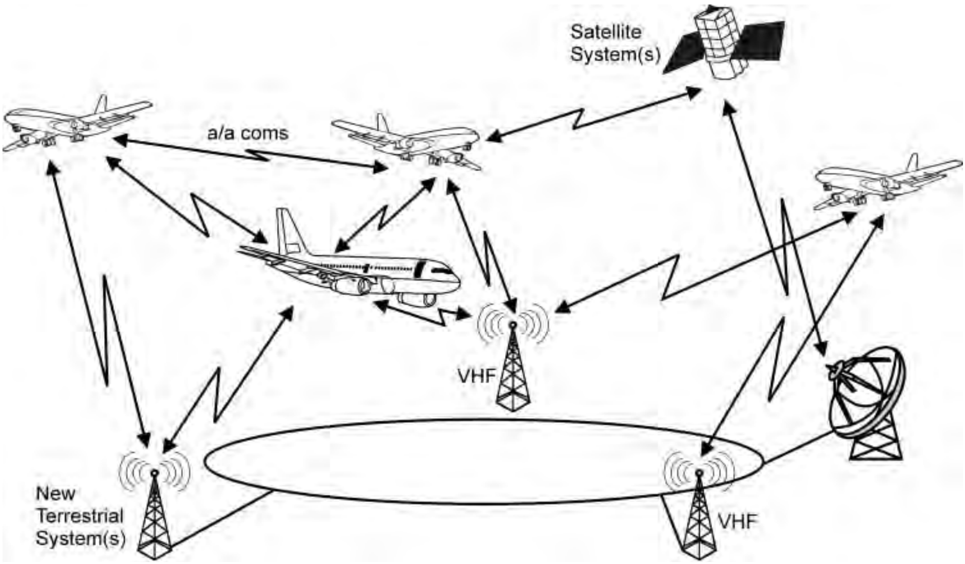


Figure 14.19 The concept of future radio system for aircraft ATS/AOC communication. (Source: QinetiQ/EUROCONTROL, 2007. Reproduced by permission of EUROCONTROL.)

The study concluded that satellite communication will continue to play an important role in the FRS for oceanic and remote areas, with a possibility of its introduction to high-density airspace such as continental Europe to complement or perhaps replace terrestrial systems provided the required performance can be achieved (QinetiQ/EUROCONTROL, 2007). Consequently, investigations were undertaken into new and existing satellite systems that may be able to meet the requirements of future higher density airspace.

Inmarsat's Swift broadband (SBB) platform, which is one of a family of BGAN service that operate on Inmarsat's fourth generation satellite, although not fully compliant with the International Civil Aviation Organization (ICAO) Standards and Recommended Practices (SARPs) at the time the study was conducted, was considered as a possible interim solution until about 2020 when Inmarsat's fourth generation satellites will begin to enter the last phase of their lives. However, since its availability with the necessary upgrades could not be guaranteed, Inmarsat's classical (legacy) services (see Chapter 11) and those of other operators such as MTSAT (Multi-functional Transport Satellite) compliant to ICAO SARPs were considered as possible interim alternatives.

The ESA initiated a project known as Iris within the framework of its Advanced Research in Telecommunication Systems (ARTES) Programme to define the role of satellites in the future aeronautical data links with an aim 'to supply a validated satellite-based communication solution for the European Air Traffic Management System' (ESA, 2013). The proposed solution reuses concepts of the classical geostationary aeronautical system, incorporating recent technological development to mitigate the limitations of the legacy systems, and it is tailored to support ATM communications in all types of aerospace (including dense urban) and aircrafts. The system will be standardized to enable participation by multiple service providers on an identical platform. It is designed to operate in those protected parts of L-band MSS spectrum where Aeronautical Mobile Satellite (Route) Service (AMS(R)S) communication has a priority over other types of mobile communications; and to support smaller light weight aeronautical earth stations (AESs) allowing its use by long and short haul (single aisle aircrafts). The AMS(R)S spectrum lies between 1646.5 and 1656.5 MHz (Earth-to-space) and 1545–1555 MHz (space-to-Earth) and refers to in-flight communication. The protection applies to communication and priorities defined in Article 44.1 and 44.2 of the radio regulations in the order as follows:

1. Distress calls, distress messages and distress traffic.
2. Communications preceded by the urgency signal.
3. Communications relating to radio direction finding.
4. Flight safety messages.
5. Meteorological messages.
6. Flight regularity messages.

The system offers a number of improvements over the classical system namely:

- Improved QoS in terms of availability and integrity performance through implementation of QoS mechanisms.
- Lighter Terminals: minimal antenna size and drag, reduced transceiver size and weight using, for example passive omni-directional antenna, passive cooled HPA, etc.
- Improved spectrum efficiency using improved channel coding and access techniques.

The system offers a highly reliable packet data service with guaranteed QoS and is designed to integrate with the global aeronautical communication network with a provision to integrate systems owned and operated by different entities in order to provide global connectivity. The system can operate in a multi-star configuration allowing multiple operators to provide the ATM service to specific sets of aircrafts. It provides bi-directional, unidirectional and broadcast services and includes a packet-based voice communication capability. Two air interface options are proposed – a baseline air interface based on CDMA and an interface based on DVB-S2 for the forward link using a single GES (ground Earth station) in a centralized architecture.

Figure 14.20 shows the proposed architecture illustrating the robustness. The space segment comprises of two separate geostationary satellites with transparent transponders to provide space segment redundancy covering the region by a single or multiple spot beams in the AMS(R)S band. The ground segment comprises a set of GESs and two Network Management Stations (in nominal and backup arrangement), installed at separate locations. Each network management station can serve each satellite independently. The support segment comprises a System Management Network, and Monitoring and Control Facility. The user segment comprises a variety of AES classes to serve the full range of aircraft.

Iridium NEXT is the next generation Iridium system, which is expected to be available beyond 2020; it is anticipated to be capable of providing the ATM compliant service.

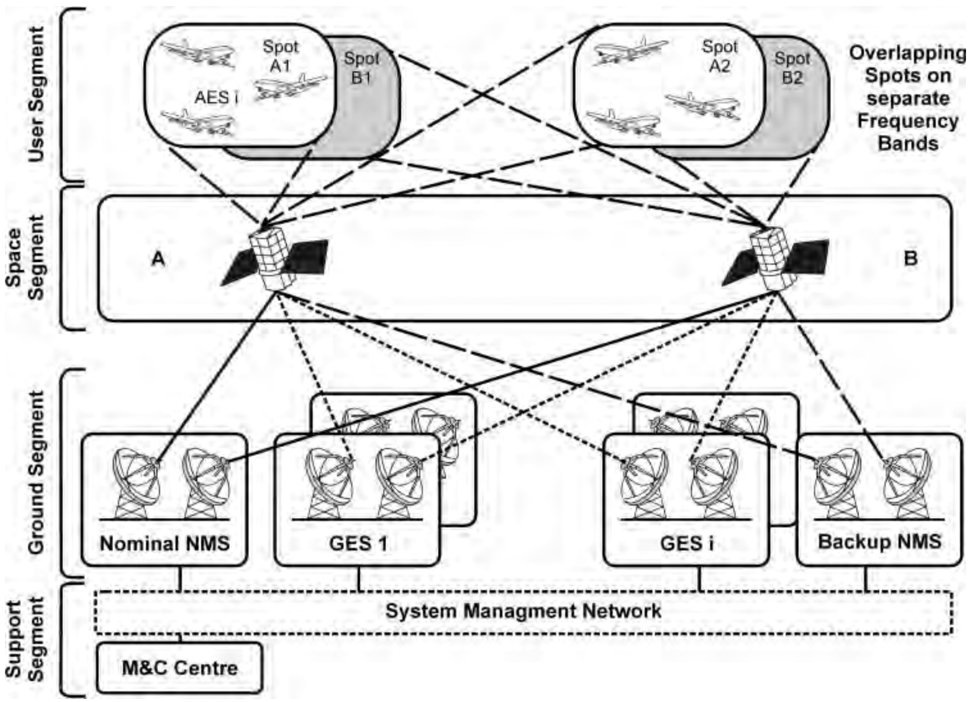


Figure 14.20 A proposed architecture to support ATM communications. (Source: ESA, 2013.)

14.8.2 Speculative Vision

A vision of global telecommunications is a Global Information Infrastructure (GII) – a high capacity backbone network. The concept comprises a plethora of wired and wireless systems, which together meet service needs economically and with the desired flexibility and mobility (Pelton, 1994). Within such a meshed broadband network comprising optical fibres and terrestrial wireless transmission systems, satellite systems are considered as one of the weakest links in terms of bandwidth, throughput, QoS and latency. For example, a 25 Gb link operating at a BER of 10^{-11} and transmission delay of < 100 ms is clearly not a norm in the fixed satellite system, let alone the MSS.

Is there a fundamental limitation due to laws of physics that precludes this level of performance? Take the Teledesic FSS LEO (Tuck *et al.*, 1994) satellite system as an example; the system was to provide fibre-optic-like performance through a large constellation of LEO satellites. Technology deployed in the network was stretched to the extent feasible. By simple reckoning, it would appear unlikely that such a technology is ultimate in satellite communications, judging by the fact that in the past 20 years satellites have become 10 000 times more powerful, significantly more cost-effective, last 10–15 times longer and large LEO constellations are operating successfully. Nevertheless, there are major regulatory and technical hurdles that have to be surmounted before satellite technology can provide digital rates of hundreds of gigabytes per second through portable and mobile terminals.

Consider the limitations of the current MSS systems:

- Excessive transmission delay for GEO and MEO (medium earth orbit) systems when compared to optical fibre systems;
- Severe shortage of spectrum in bands where technology is mature;
- Inability to provide consistently high signal quality;
- Susceptibility to rain fades at mm waves, scintillation and shadowing.

Consider a few possibilities – the material has been extracted from the literature (e.g. Pelton, 1994; Purchase, 1995), and includes the author's views and interpretation.

The problem of latency can be solved by using LEOs, stratospheric platforms or their combination. In such a system, terrestrial and stratospheric platforms cover the densely populated areas, while thinly populated areas are served by a LEO system. For applications where latency is not critical, an umbrella of GSO system provides the solution. It is not necessary for the entire system to be owned by a single entity or organization if interfaces are standardized so that the system can be owned and operated by different organizations and systems added as required, emulating the Internet model; Figure 14.21 portrays an impression of a MSS system consisting of such a conglomerate.

To offer gigabytes of information rate, several gigahertz of spectrum will be necessary, which can be available above 35 GHz, coupled with large investments to develop the necessary hardware; large segments of continuous spectrum could provide optical-fibre-like

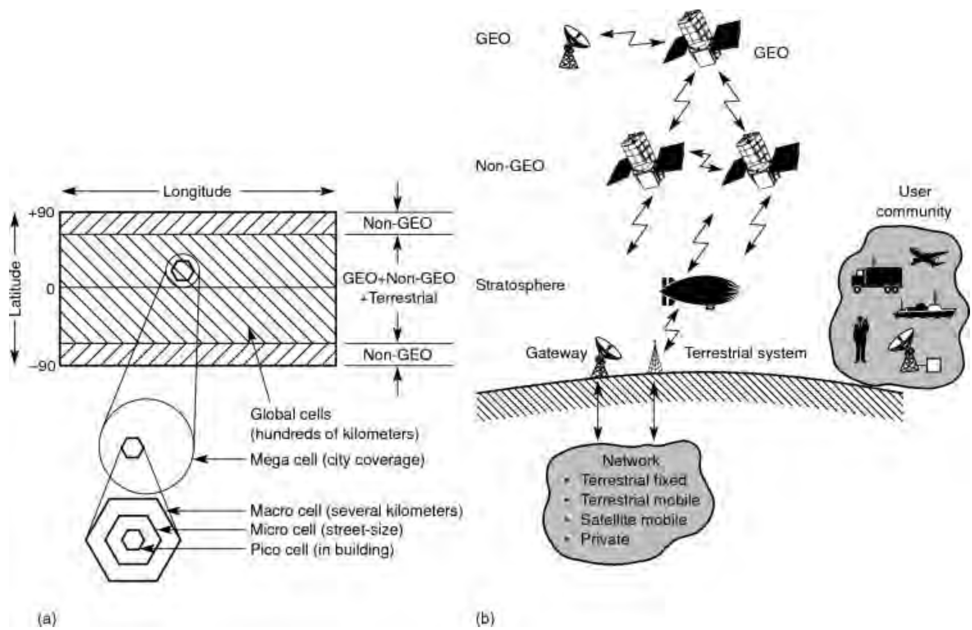


Figure 14.21 A hybrid mobile communication system using a terrestrial cellular system, stratospheric platforms, LEO and GEO satellite constellations to support future broadband transmissions: (a) coverage and (b) entities

throughputs and this is where regulatory decisions may be necessary. Coupling this paradigm with the advantages of on-board processing satellites, advanced spot beam technology and improved coding techniques, it is estimated that up to a 100-fold frequency reuse is possible, giving several hundred gigahertz of spectrum in the 35–50 GHz band. The problems of severe rain fades and scintillation conflict with usage of high efficiency modulation schemes that are highly susceptible to flat and frequency selective fading, and thus powerful coding and fade resistant modulation schemes become desirable. A design for generating up to 3500 spot beams capable of 500-fold frequency reuse was created by Douglas Lockie of Endgate Inc. decades ago (Pelton, 1994) and on-board processing technology has made significant inroads such that all modern MSS satellites utilize. Dynamic power control can mitigate the problem of rain fades and shadowing by diverting large amounts of power to fading channels. With more powerful transmitters and antennas deploying thousands of micro spot beams, link margins of 15–20 dB are conceivable, giving extremely high and stable BER. In order to alleviate spectrum congestion – in addition to the use of Q, V, W bands – regulatory procedures and approach will require changes in order to maximize spectrum inefficiencies – for example, by the use of cognitive radio technology and dynamic allocation, waiving the distinction between MSS and FSS in a controlled manner (e.g. by the use of location-based cognitive-radio enabled allocations). We have already noted the surge towards the convergence of MSS and VSAT technologies in captive land portable, maritime and air scenarios. Converged network architecture would allow a user to roam between a home ‘gigabyte’ dish and a nomadic personal communicator while away.

Consider the shortcomings in satellite radio link reliability particularly for mobile on-the-move user terminals; the ATC technology provides a cost-effective solution for filling coverage gaps in populated areas. Much as true of the extension in coverage of cellular systems, the terrestrial ATC infrastructure could evolve as the technology spreads, conserving the satellite power for areas where the infrastructure is uneconomic. *Ad-hoc* networks enabled user terminals could provide a layer of connectivity in absence of an infrastructure, particularly in remote areas. Furthermore, it should be expected that satellite EIRPs will increase with the use of micro spot beams and more efficient power so as to shrink the coverage gaps.

Revision

1. State the assumptions used in the recent market projections outlining the risks associated with each assumption.

2. Assume the total number of in-service units globally at the end of the decade as 5 million distributed between land mobile, maritime, hand-held and aeronautical platforms respectively as 50, 25, 22 and 3%. Further, assume that the users are distributed uniformly and service is provided globally by a three region geostationary satellite system with a single satellite covering each region. Estimate the spectrum required in each region to provide a packet-mode service. State each assumption qualifying the basis.

Hint: Assumption may include: Aggregate busy hour traffic per satellite, traffic symmetry in forward and return direction, the frequency band such as L band, number of spot beams per satellite and reuse factor, spectral efficiency in bits/Hz, etc.

3. Outline the concept of an iterative interference cancellation multi-user detector.
4. State the difference between frequency plan and cross-layer optimization schemes? How are these two interrelated in view of the fact that each strives to enhance spectrum utilization.
5. Describe the principle of cognitive radio with an example of a potential application in MSS.
6. Outline the principles and advantages of a software defined radio. State the advantage in using a common development platform such as software communications architecture.
7. Comment with sound reasoning on the assumptions and arguments presented in relation to the speculative vision of the future MSS network as presented in the chapter.

References

3GPP (2008) 3rd Generation Partnership Project; Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8), v8.2.0, March 2008. 3GPP TS36.211.

Agilent (2011) Introducing LTE-Advanced, Application Note, Printed in USA, March 8, 2011, 5990-6706EN, www.agilent.com (accessed 26 July, 2013).

Airbus (2013) 'Navigating the Future', Airbus Global Market Forecast 2012–2031, www.airbus.com/company/market/forecast/ (accessed 28 August, 2013).

Alagöz, F., Korak, Ö. and Jamalipour, A. (2007) Exploring the routing strategies in next-generation satellite networks. *IEEE Wireless Communications*, **14** (3), 79–88.

Alamouti, S. (1998) A simple transmit diversity technique for wireless communications. *IEEE Journal on Selected Areas in Communications*, **16** (8), 1451–1458.

Alexander, P., Grant, A. and Reed, M. (1998) Iterative detection in code–division multiple–access with error control coding. *European Transactions on Telecommunications*, **9** (5), 419–426.

Ansari, A., Dutta, S. and Tseytlin, M. (2009) S-WiMAX: adaptation of IEEE 802.16e for mobile satellite services. *IEEE Communications Magazine*, **47** (6), 150–155.

Arapoglou, P-D., Liolis, K., Bertinelli, M., Panagopoulos, A.,Cottis, P. and Gaudenzi, R.D. (2011) MIMO over satellite: a review. *IEEE Communication surveys and Tutorials*, **13** (1), 9(First quarter).

Bastia, F., Bersani, C., Candreva, E.A. *et al*. (2009) LTE adaptation for mobile broadband satellite networks. *EURASIP Journal on Wireless Communications and Networking*, **2009**, 13. doi: 10.1155/2009/989062, Article ID 989062.

Biglieri, E. (2012) An overview of Cognitive Radio for satellite communications. IEEE First AESS European Conference on Satellite Telecommunications (ESTEL), October 2–5, 2012, pp. 1–3.

Crane, R.K., Wang, X.D., Westenhaver, B. and Vogel, W.J. (1997) ACTS propagation experiment: experiment design, calibration, and data preparation and archival. *Proceedings of the IEEE*, **85** (6), 863–878.

ESA (2008) web site, http://telecom.esa.int/telecom/www/object/index.cfm?fobjectid=18310 (accessed 28 August, 2013).

Etemad, K. and Riegel, M. (2010) Topics and updates on 4G technologies. Series editorial. *IEEE Communications Magazine*, **48** (8), 38–39.

ETSI (2001) GEO-Mobile Radio Interface Specifications; Part 1: General Specifications; Sub-part 3: General System Description. GMR-1 01.202 (ETSI TS 101 376-1-3).

ESA (2013) ARTES 10-Iris, http://telecom.esa.int/telecom/www/area/index.cfm?fareaid=56 (accessed 28 August, 2013).

ETSI (2008), Satellite Earth Stations and Systems (SES); Satellite Component of UMTS/IMT-2000; Evaluation of the OFDM as a Satellite Radio Interface. ETSI TR 102 443 v1.1.1, August 2008.

ETSI (2004) Satellite Earth Stations and Systems (SES); Satellite Component of UMTS/IMT-2000; Evaluation of the W-CDMA UTRA FDD as a Satellite Radio Interface. ETSI TR 102 058 v1.1.1 (2004–11).

Euroconsult (2011) web site www.euroconsult-ec.com/research-reports/satcom-reports/mobile-satellite-communications-markets-survey-39-35.html (accessed 28 August, 2013).

EUROCONTROL/FAA (2007) Communications Operating Concept and Requirements (COCR) – v2.0, April 2007.

Evans, J.V. and Dissanayake, A. (1998) Prospects for commercial satellite services at Q- and V-bands. IEEE Military Communications Conference, 1998, Proceedings MILCOM 98, Boston, MA, October 18–21, 1998, Vol. 1, pp. 1–7.

FCC (2009) Connecting America: the National Broadband Plan – Executive Summary, http://download.broadband.gov/plan/national-broadband-plan-executive-summary.pdf (accessed 28 August, 2013).

Filin, S., Harada, H., Murakami, H. and Ishizu, K. (2011) International standardisation of cognitive radio systems. *IEEE Communications Magazine*, **49** (3), 82–89.

Foschini, G.J. and Gans, M.J. (1998) On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communications*, **6** (3), 311–335.

Futron (2010) www.futron.com/2010_Futron_Forecast_Summary_Findings.xml, (accessed 28 August, 2013).

Gargione, F., Iida, T., Valdoni, F. and Vatalaro, F. (1999) Services, technologies, and systems at K_a band and beyond – A survey. *International Journal on Satellite Communications*, **17** (2), 133–144.

Gatehouse (2011) www.gatehouse.dk/en-US/News-Events/News/Satellite-news/Show-news/Approval-of-the-GateHouse-SDR-Waveform.aspx?Action=1&PID=285 (accessed 28 August, 2013).

Gayraud, T. and Berthou, P. (2008) Evaluation of ad-hoc and mobility architectures using satellite communications. 26th AIAA International Communications Satellite Systems (ICSSC 2008), San Diego, CA, 2008, p. 10.

Gesbert, D., Kountouris, M., Heath, R.W. Jr.,, Chae, C.-B. and Sälzer, T (2007) From single user to multiuser communications: shifting the MIMO paradigm. *IEEE Signal Processing Magazine*, **24** (5), 36–46.

Giambene, G. (ed) (2007) *Resource Management in Satellite Networks Optimisation and Cross-Layer Design*, Springer Science + Business Media, New York.

Giambenel, G. and Kota, S. (2006) Cross-layer protocol optimization for satellite communications networks: a survey. *International Journal of Satellite Communications and Networking*, **24**, 323–341.

Goldhirsh, J. and Vogel, W.J. (1998) *Handbook of Propagation Effects for Vehicular and Personal Mobile Satellite Systems: Overview of Experimental and Modelling Results*, 2nd edn, NASA Reference Publication, p. 1274, www.utexas.edu/research/mopro/index.html (accessed 28 August, 2005).

IEEE (2008) Cognitive radio communications and networks. *IEEE Communications Magazine*, **46** (4), 30–85.

IEEE (2011) 'Topics in radio communications' and advances in standards and test beds for cognitive radio networks: part II. *IEEE Transactions on Wireless Communications*, **49** (3), 30–109.

ITU (2006) Traffic Forecasts and Estimated Spectrum Requirements for the Satellite Component of IMT-2000 and Systems beyond IMT-2000 for the Period 2010 to 2020, Report ITU-R M.2077, ITU, Geneva.

ITU (2008a) Requirements Related to Technical Performance for IMT-Advanced Radio Interface(s). Report ITU-R M.2134, International Telecommunication Union Radio (ITU-R).

ITU (2008b), Radio Regulations, Article 5, International Telecommunication Union, Geneva.

ITU (2010) IMT-Advanced-Satellite Submission and Evaluation Process, www.itu.int/ITU-R/index.asp?category=study-groups&rlink=rsg4-imt-adv-sat/en (accessed 28 August, 2013).

ITU (2011) Traffic Forecasts and Estimated Spectrum Requirements for Future Development of the Mobile-Satellite Service in the Range 4-16 GHz. Report ITU-R M.2218, (10/2011), M Series Mobile, Radio Determination, Amateur and Related Satellite Services, ITU, Geneva.

ITU (2012) Measuring the Information Society, www.itu.int/ITU-D/ict/publications/idi/index.html (accessed 28 August, 2013).

Jondral, F.K. (2005) Software-defined radio - basics and evolution to cognitive radio. *EURASIP Journal on Wireless Communications and Networking*, **2005** (3), 275–283.

JTRS Standards (2012) Software Communications Architecture Specifications, Version: 4.0 28 February 2012 Joint Program Executive Office (JPEO) for the Joint Tactical Radio System (JTRS), San Diego, CA.

Kandeepan, S., Nardis, L.D., Benedetto, M.-G.D., Guidotti, A. and Corazza, G.E. (2010) Cognitive satellite terrestrial radios. IEEE Globecom 2010 Proceedings, p. 6.

King, P.R. (2007) Modelling and measurement of the land mobile satellite MIMO radio propagation channel. PhD thesis. University of Surrey.

King, P.R. and Stavrou, S. (2006) Capacity improvement for a land mobile single satellite MIMO system. *IEEE Antennas and Wireless Propagation Letters*, **5**, 98–100.

Kota, S.L., Hossain, E., Fantacci, R. and Karmouch, A. (2005) Cross-layer protocol engineering for wireless mobile networks: part 1. *IEEE Communications Magazine*, **43**, 110–111.

Liolis, K., Andrikopoulos, I. and Cottis, P. (2008) On statistical modelling and performance evaluation of SIMO land mobile satellite channels. 4th Advanced Satellite Mobile System, ASMS 2008, Bologna, Italy, August 26–28, 2008, pp. 76–81.

Luglio, M., Monti, C., Roseti, C., Saitto, A. and Segal, M. (2007) Interworking between MANET and satellite systems for emergency applications. *International Journal of Satellite Communication and Networking*, **25**, 551–558. doi: 10.1002/sat.890, Published online in Wiley InterScience (www.interscience.wiley.com) (accessed 12 August 2012).

Marshall, P. (2011) The maturation of dynamic spectrum access from a future technology, to an essential solution to immediate challenges, Guest editorial. *IEEE Communications Magazine*, **29** (3), 30.

Mietzner, J., Schober, R., Lampe, L., Gerstacker, W.H. and Hoeher, P.A. (2009) Multiple-antenna techniques for wireless communications - a comprehensive literature survey. *IEEE Communications Society for Tutorials and Surveys*, **11** (2), 87–105, Second Quarter.

Millerioux, J.P., Boucheret, M.L., Bazile, C. and Ducasse, A. (2006) Iterative multiuser detection performance evaluation on a satellite multibeam coverage. 14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4–8, 2006, Copyright by EURASIP.

Mohammed, H.A. and Pillai, P. (2009) Performance Evaluation of a WiMAX Enabled HAPs-Satellite Hybrid System, www.cms.livjm.ac.uk/pgnet2009/Proceedings/Papers/2009017.pdf (accessed 28 August, 2013).

Moher, M. (1998) An iterative multiuser decoder for near-capacity communications. *IEEE Transactions on Communications*, **46**, 870–880.

Moher, M., Wei Zhang, Febvre, P. and Rivera-Castro, J (2010) Multi-user detection for Inmarsat's BGAN system Advanced Satellite Multimedia Systems Conference (ASMS) and the 11th Signal Processing for Space Communications Workshop (SPSC), 2010 5th, September 13–15, 2010, Conference Publications, pp. 135–140.

Moshavi, S. (1996) Multi-user detection for DS-CDMA communications. *IEEE Communications Magazine*, **34** (10), 124–136.

Neri, M., Casadei, M., Vanelli-Coralli, A. and Corazza, G.E. (2007) Multiuser detection for S-UMTS and GMR-1 mobile systems. *International Journal of Satellite Communications and Networking*, **25** (5), 529–550, Special issue on ASMS, Article first published online: 19 September 2007 September/October 2007.

NSR (2012) Mobile Satellite Services, 8th Edition, Report Briefing, www.nsr.com/upload/research_reports/nsr_mss8_brief_1.pdf (accessed 28 August, 2013).

Page, D. (1998) Paradigm shifting without a clutch. *Satellite Communications*, **21**, 70–72.

Paraboni, A., Vernucci, A., Zuliani, L., *et al*. (2007) A new satellite experiment in the Q/V band for the verification of fade countermeasures based on the spatial non-uniformity of attenuation. Proceeding of the Second European Conference on Antennas and Propagation, EuCAP 2007, November 11–16, 2007, pp. 1–6.

Paulraj, A.J., Gore, D.A., Nabar, R.U. and Bolcskei, H. (2004) An overview of MIMO communications – a key to gigabit wireless. *Proceedings of the IEEE*, **92** (2), 198–218.

Pelton, J.N. (1994) Low Earth Orbit Satellites: best hope for advanced satellite communication and the global information infrastructure. *Space Communications*, **12**, 233–247.

Pérez-Neira, A., Ibars, C., Serra, J., del Coso, V., Gómez, J. and Caus, M. (2008) MIMO applicability to satellite networks. 10th International Workshop on Signal Processing for Space Communications, 2008, SPSC 2008, Rhodes, Island, October 6–8, 2008, pp. 1–9.

Purchase, J.F. (1995) Establishing the satellite-based Infobahn in a multicultural world. *Via Satellite*, **10** (7), 18–22.

QinetiQ/EUROCONTROL (2007) Future Communications Infrastructure – Step 2: Technology Assessment Results, Edition: 1.0, 29 October 2007.

Reynolds, D. and Wang, X. (2001) Low complexity turbo-equalization for diversity channels. *Signal Processing*, **81** (5), 989–995.

Richharia, M. and Trachtman, E. (2006) Capacity enhancement techniques in mobile satellite systems. 3rd Advanced Satellite Mobile Systems Conference (ASMS), Herrsching am Ammersee Germany, May 29–31, 2006, p. 8.

Richharia, M., Trachtman, E. and Fines, P. (2005a) Broadband global area network air interface evolution. 23rd AIAA International Communications Satellite Systems Conference (ICSSC 2005), September 25–28, 2005, p. 12.

Richharia, M., Trachtman, E., Westbrook, L. *et al*. (2006) Frequency planning approaches in mobile satellite systems. 3rd Advanced Satellite Mobile Systems Conference (ASMS), Herrsching am Ammersee Germany, May 29–31, 2006, p. 9.

Richharia, M., Kaluvala, N., Fines, P., *et al*. (2005b) Aeronautical channel model for broadband L-band satellite communication. 23rd AIAA International Communications Satellite Systems Conference (ICSSC 2005), September 25–28, 2005, p. 12.

Rodriguez, C.G., Franck, L., Baudoin, C. and Beylot, A-L. (2010) Mobile ad hoc network assisted by satellites. Proceeding of International Communications Satellite Systems Conference (AIAA ICSSC 2010), At Anaheim, CA, p. 7.

Ruggieri, M., Vatalaro, F., Paraboni, A. and Bonifazi, C. (1998) DAVID: a small satellite mission for data distribution. Proceedings CNES-98, Antibes, Juan les Pins, France, September.

Rumney, M. (ed) (2013) Looking towards 4G: LTE-advanced, in *LTE and the Evolution to 4G Wireless: Design and Measurement Challenges*, Chapter 8, 2nd edn, ISBN: 978-1-1199-6257-1, John Wiley & Sons, Inc., Hoboken, NJ.

Scalise, S., Mura, R. and Mignone, V. (2006) Air interfaces for satellite based digital TV broadcasting in the railway environment. *IEEE Transactions on Broadcasting*, **52** (2), 158–166.

Sharma, S.K., Chatzinotas, S. and Ottersten, B. (2012) Satellite cognitive communications: interference modeling and techniques selection. 6th Advanced Satellite Multimedia Systems Conference (ASMS) and 12th Signal Processing for Space Communications Workshop (SPSC), p. 8.

Shen, C., Borkar, G., Rajagopalan, S. and Jaikaeo, C.(2002) Interrogation-based relay routing for ad hoc satellite networks. Global Telecommunications Conference, 2002. IEEE GLOBECOM'02, November 17–21, 2002, Vol. 3, pp. 2920–2924.

Tuck, E.F., Patterson, D.P., Stuart, J.R. and Lawrence, M.H. (1994) The calling network: a global wireless communication system. *International Journal of Satellite Communications*, **12**, 45–61.

Vanelli-Coralli, A., Corazza, A., Karagiannidis, G.E., *et al*. (2007) Satellite communications: research trends and open issues. Proceedings of International Workshop on Satellite and Space Communications, 2007, IWSSC 07, September 13–14, 2007, pp. 71–75.

Verdú, S. (1998) *Multi-User Detection*, Cambridge University Press, New York.

Verdú, S. (2000) Wireless bandwidth in the making, *IEEE Communications Magazine*, **38** (7), 53–58, www.ece.mtu.edu/faculty/ztian/ee5560/BW-Verdu-2000.pdf (accessed 28 August, 2013).

Wakana, H., Tanaka, M., Kozono, S. *et al*. (1998) The COMETS experiments for advanced mobile satellite communications and advanced satellite broadcasting. Proceeding International Symposium Space Technology and Science, 98-h-05, Omiya, Japan, May.

Wang, X. and Poor, H. (1999) Iterative (Turbo) soft interference Cancellation and Decoding for Coded CDMA. *IEEE Transactions on Communications*, **47** (7), 1046–1061.

Wireless Innovation Forum (2012) www.wirelessinnovation.org/wireless_resources (accessed 28 August, 2013).

Yun, Y.H. and Cho, J.H. (2009) An orthogonal cognitive radio for a satellite communication link. IEEE 20th International Symposium on Personal, Indoor and Mobile Radio Communications, September 13–16, 2009, pp. 3154–3158.

Appendix

A.1 Coverage Snapshot of Representative Non-Geostationary MSS Systems

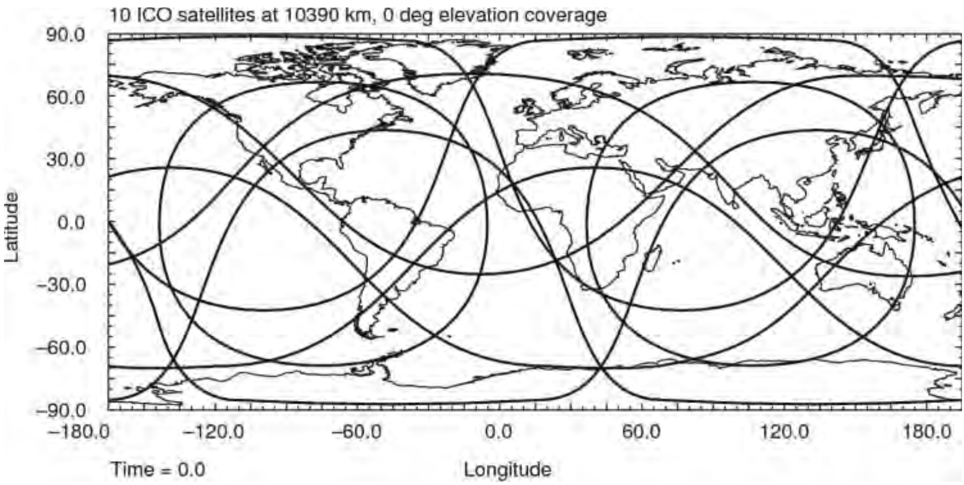


Figure A.1 A snapshot of ICO coverage (Bains, 1999; see Chapter 11 for details). (Source: Bains, 1999. IMSC '99, The Sixth International Mobile Satellite Conference, Ottawa, 1999, co-sponsored by the Communications Research Centre and the Jet Propulsion Laboratory.)

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com

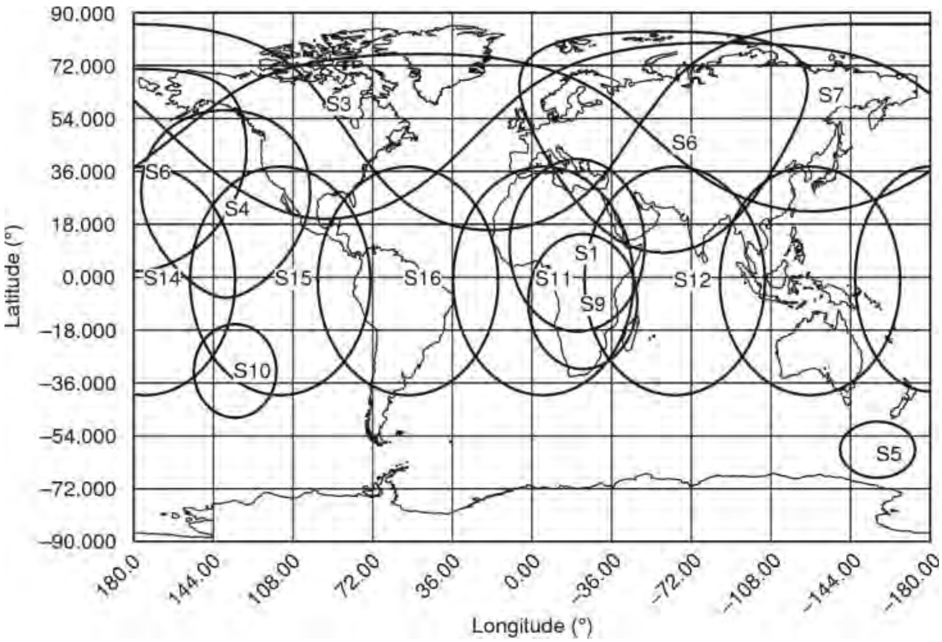


Figure A.2 A snapshot of ELLIPSAT coverage (see Chapter 11 for details). (Source: Krewel and Maral, 1998. Reproduced by permission of John Wiley & Sons, Ltd.)

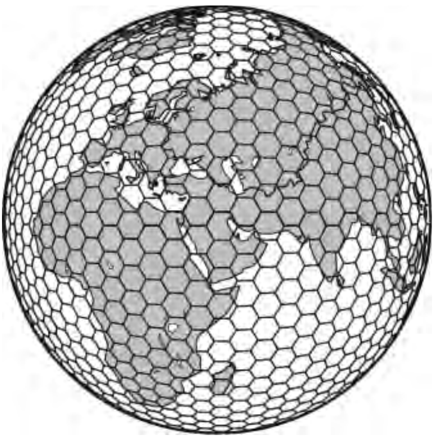


Figure A.3 A snapshot of Iridium coverage (see Chapter 11 for details). (Source: Wu *et al*., 1994. © IEEE. Reproduced with permission.)

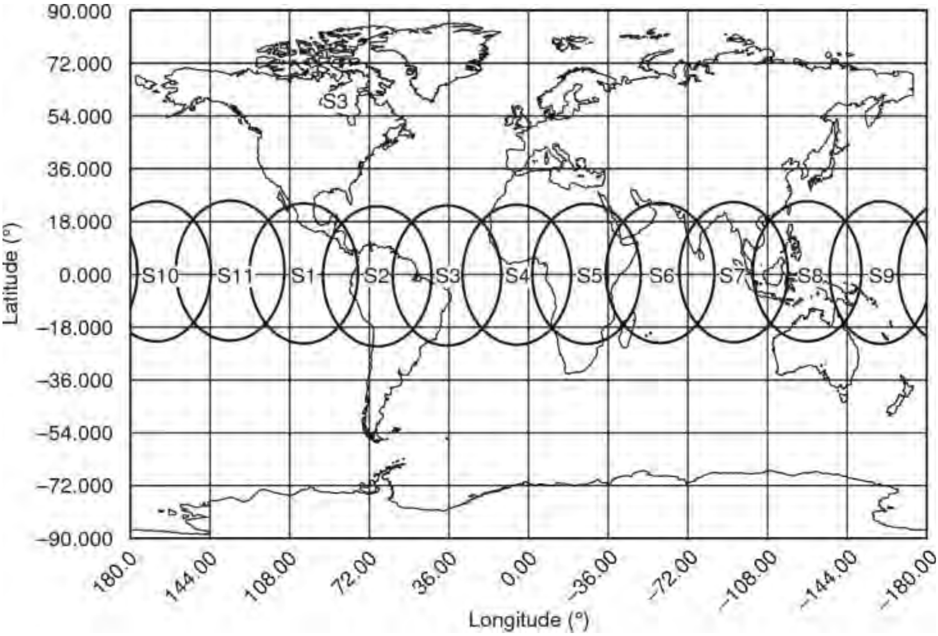


Figure A.4 Snapshot of ECCO equatorial coverage (46 satellites, 2 planes inclined at 0° and 62°; 11 satellites in equatorial plane). (Source: Krewel and Maral, 1998. Reproduced by permission of John Wiley & Sons, Ltd.)

A.2 A List of Useful Formulas

A.2.1 Geostationary Orbit

For a preliminary system analysis simple geometric solutions involving spherical and planar geometry are useful.

Satellite elevation and azimuth angles are used to locate the position of a satellite in space and are useful in radio link design, earth station siting, etc. Figure A.5(a) and (b) shows the geometry used in developing the relationships.

A.2.1.1 Elevation

Satellite elevation angle η from a point on the Earth is the angle between the line from the given point to the satellite and the tangent drawn towards the satellite at the point, as illustrated in Figure A.5(a). Elevation angle is estimated as follows:

$$= \arctan\left(\frac{\cos\psi - \sigma}{\sin(\psi)}\right) \tag{A.1}$$

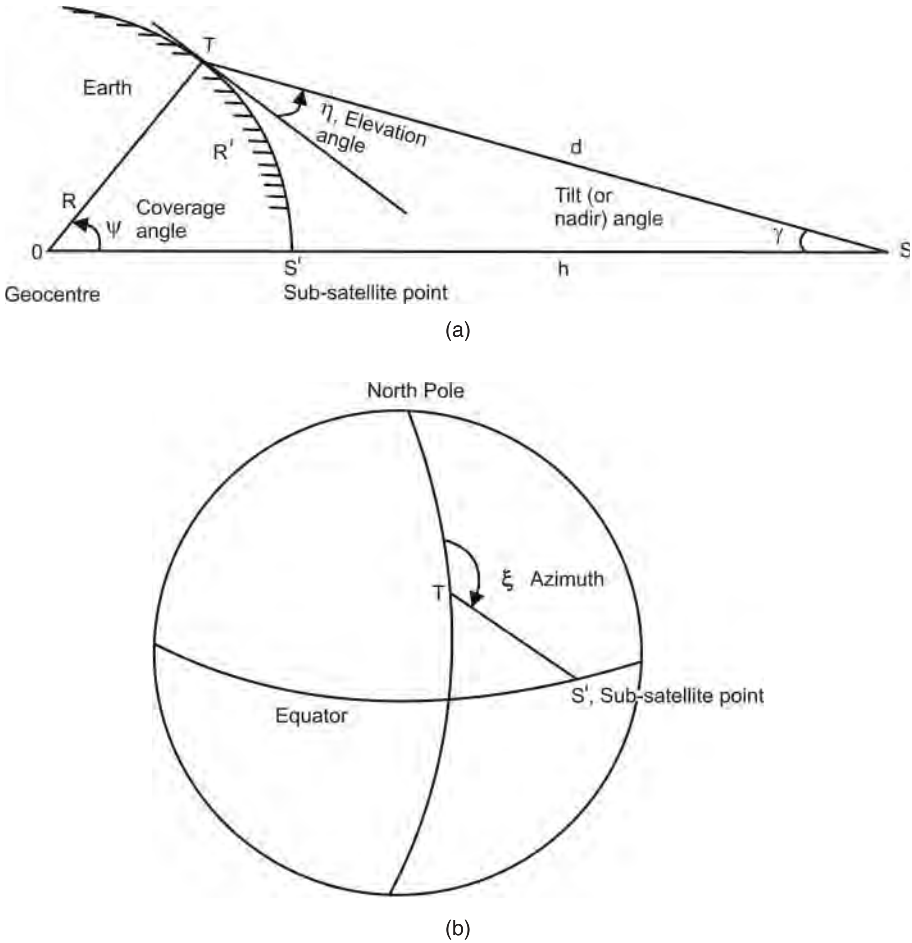


Figure A.5 Earth-geostationary satellite radio link geometry illustrating (a) elevation angle and (b) azimuth angle from a point T on the Earth.

ψ is the coverage angle, given as,

$$\Psi = \text{arcos}\,(\cos(\theta_e)\cos(\Phi_{es})) \tag{A.2}$$

$$\Phi_{es} = (\Phi_e - \Phi_s)$$

$$\sigma = \frac{R}{(R+h)} \approx 0.151$$

θ_e = latitude of Earth station
Φ_e = longitude of Earth station (°E)
Φ_s = longitude of sub-satellite point (°E)
R = radius of the Earth (6378 km)
h = satellite height above the Equator (35 787 km)

The tilt angle (sometimes, known as nadir angle), γ, is given as,

$$\gamma = \arctan\left(\frac{\sin(\psi)}{6.6235 - \cos\psi}\right) \tag{A.3}$$

A.2.1.2 Azimuth

Satellite azimuth, ξ, from a location on the Earth is the angle between the true north and the direction of the satellite from the location, as illustrated in Figure A.5(b) for a point T. There is a slight difference between the true north and magnetic north known variously as *magnetic variation*, *magnetic deviation or magnetic declination*. Its value can be obtained through the local geological organization such as the British Geological Society in the UK. The value is added to the magnetic north when the variation is to the west of true north and subtracted when the variation is to the east. Azimuth can be estimated by determining a factor, A, dependent on the latitude and differential longitude of the given location and factoring it in to the applicable quadrant as follows:

$$A = \left|\arctan\left(\frac{\tan\left(\Phi_{es}\right)}{\sin(\theta_e)}\right)\right| \tag{A.4}$$

In the Northern Hemisphere,

$\xi = 180° + A°$; when the satellite is to the west of the location.
$\xi = 180° - A°$; when the satellite is to the east of the location.

In the Southern Hemisphere,

$\xi = 360° - A°$; when the satellite is to the west of the location.
$\xi = A°$; when the satellite is to the east of the location.

A.2.1.3 Range

Range is used to estimate earth station-satellite path loss and propagation delay. A number of geometric relationships are available.

$$d = 35\,786[1 + (0.4199\,\{1 - \cos(\psi)\}]^{1/2}\text{ km} \tag{A.5}$$

Range in terms of Earth radius R is given as,

$$d_{er} = 6.6235\frac{\sin(\psi)}{\cos(\eta)} \tag{A.6}$$

where

$$d_{er} = \frac{d}{R}$$

A.2.1.4 Eclipse

Satellites draw power from solar cells; during an eclipse since power from this source is not available, storage batteries are used. Eclipses caused by the Earth always occur within about ±22 days of the equinox. Eclipses caused by the Moon are less frequent and their occurrence is irregular due to non-static Sun-Moon geometry.

A.2.1.5 Eclipse Caused by the Earth

The movement of the Sun relative to the equatorial plane can be approximated as (Maral and Bousquest, 1987)

$$\mathrm{i}_s = 23 \sin\left(\frac{2\pi t_d}{T_y}\right) \tag{A.7}$$

where

i_s = Sun inclination in degrees
t_d = time in days, referenced to an equinox
T_y = 365 days

The equatorial plane, the Earth and the Sun lie on the same plane on the day of the equinox, causing an Earth-induced solar eclipse at GEO (geostationary earth orbit). Since the Earth's shadow covers about 17.4° of the geostationary orbital arc, its shadow begins to fall on the equatorial orbit when the Sun's inclination is ≤8.7° on either side of the Equator. The duration of eclipse increases each day until it reaches a maximum on the day of the equinox; it then subsides gradually as the Sun moves to the other end of the inclination range. The number of eclipse days is twice the number obtained by substituting, $i_s = 8.7°$ in Equation A.7, approximating 44 days. Duration of an eclipse, t_e, is the period that a satellite remains within the shadow. It can be obtained by the equation,

$$t_e = \frac{\omega_e}{360} T_{geo} \tag{A.8}$$

where

ω_e is the angle subtended by the Earth's shadow on geostationary orbit (degrees).

On substituting $\omega_e = 17.4°$ corresponding to the equinox, we obtain the duration to be about 69.4 min.

T_{geo} is the orbital period of geostationary orbit (23 h, 56 min, 4.09 s).

The estimate does not take into consideration the penumbra region caused by the diffraction effects of the Earth.

A.2.1.6 Eclipse Caused by the Moon

The technique uses Sun and Moon position data available in a publication called the *Nautical Almanac*. The Almanac provides positions of well-known celestial bodies such as the Sun and Moon yearly from the Earth in terms of declination and GMT hour angle.

A lunar eclipse occurs when the look angles of Sun and Moon from a geostationary location are coincidental or close enough to cause shadowing of the Sun by the Moon. A technique to determine the azimuth-elevation of these celestial bodies has been proposed by Siocos (1981) and summarized in Richharia (1999). We will not include this here due to lack of space. We will endeavor to host it on our website in due course.

A.2.1.7 Occurrence-time of Eclipse

An eclipse at GEO occurs at the mid-night of the sub-satellite point (i.e. the point on the Earth directly below the satellite); at other longitudes the time T_l at which eclipse occurs can be approximated as (Mertens, 1976):

$$T_l = 23.38 - \frac{1}{15}(\Phi_{\mathrm{s}} - \Phi_{\mathrm{lz}}) \tag{A.9}$$

where

Φ_{s} = longitude of the satellite
Φ_{lz} = longitude of the time zone assuming a single time zone.

A.2.1.8 Solar Interference

The Sun is a strong radiator of electromagnetic radiation. Solar interference can occur when the Sun moves within the beam-width of an earth station's receiving antenna. An approximate increase in antenna noise temperature caused by the Sun, ΔT_{as}, can be obtained with the assumption that gain of an earth station antenna remains uniform within its half-power beam-width and negligible outside (Mohamadi *et al.*, 1988). Then,

$$\Delta T_{as} = \mathrm{p} T_s \xi D_s^2 \tag{A.10}$$

where

p = polarization factor to account for random polarization of noise
 = 0.5
T_s = Sun's equivalent noise temperature
 = $120\,000\mathrm{f}^{-0.75}$; where f = frequency in gigahertz
ξ = antenna efficiency
$D_s = \dfrac{\textit{Optical diameter of sun}}{\varphi}$; where φ = half-power beam-width of antenna
$D_s = 1$, when $\varphi < 0.48°$

A.2.1.9 Number of Days of Solar Conjunction

A solar conjunction occurs within ±22 days of the equinoxes when Sun declination and an earth station's angle relative to the equator are coincident such that the Sun is captured within the beam-width of the earth station antenna. Declination is the elevation of an object measured in the northward direction from the celestial equator in a right-ascension-declination coordinate system. The centre of the coordinate system is the geocentre but can be chosen as another point.

The number of days, D_{max}, when solar conjunction occurs is given as:

$$D_{\max} = \frac{\varphi + 0.48}{0.4}\ \text{days} \tag{A.11}$$

where

φ is the antenna half-power beam-width.
0.4 represents an approximate north-south motion of the Sun around the equinox.

A.2.1.10 Peak Duration of Solar Conjunction

The maximum duration of solar conjunction, T_{max}, that is, on the day when the event peaks, can be approximated as,

$$T_{\max} = \left(\frac{\varphi + 0.48}{0.25}\right)\ \text{min} \tag{A.12}$$

where, 0.25°/min represents east-west motion of the Sun viewed from ground.

A.2.1.11 Solar Conjunction Occurrence Prediction

A relatively straightforward method of predicting occurrence of conjunction is based on conversion of the satellite's azimuth-elevation angle from the given earth station to an ascension-declination coordinate system; and looking-up from the *Nautical Almanac* the days on which the Sun's position lies within an earth station antenna's beam-width. Note that the *Almanac* data is tabulated in an ascension-declination coordinate system.

Declination, D is determined as follows:

$$\mathrm{D} = \sin^{-1}(\sin\theta\sin\eta - \cos\theta\cos\eta\cos\xi) \tag{A.13}$$

where

θ = Earth station latitude
η = satellite elevation
ξ = satellite azimuth (should be positive when the denomination is west)
D is positive when the denomination is north

The ascension, α of an Earth station relative to satellite location is given as

$$\alpha = \sin^{-1}\left(\frac{\cos\eta\sin\xi}{\cos\ D}\right)\ \text{Hour Angle} \tag{A.14}$$

α is positive when it lies to the west.

The value of α should be taken with respect to the Greenwich Meridian to make it compatible to *Nautical Almanac* tabulations. The conversion is given as:

$$HA_G = \Phi_e - \alpha$$

where

HA_G = hour angle with respect to Greenwich Meridian
Φ_e = earth station longitude

A.2.2 Propagation

A.2.2.1 Ricean Probability Distribution

A majority of MSS (mobile satellite service) systems are designed to operate with a direct line of sight to the satellite. The signal comprises a direct component with a steady mean and Doppler shift and multipath components caused by scattering from uniformly distributed scatters around the vehicle. Multipath components are assumed to be Rayleigh distributed. The probability density function p(x) of the received signal is Ricean under such conditions (ETSI, 2003), given by:

$$p(x) = (1 + k)\, e^{-x(1+k)-k} I_0[2\,(xk(1 + k)]^{1/2} \quad \text{for } x > 0 \tag{A.15}$$

where

k (Ricean factor or parameter) = power in direct component/power in diffuse components,
I_0 = zero-order modified Bessel function.

When the signal gets shadowed, $k = 0$, and the distribution reverts to Rayleigh.

A.2.2.2 Doppler Caused due to Mobile Motion

Doppler shift of the direct signal component is represented by the equation,

$$\Delta f_d = \left(\frac{v_m}{c}\right) f\,(\cos\eta) \tag{A.16}$$

where

Δf_d = Doppler shift of direct signal
c = velocity of light
f = transmission frequency
η = elevation angle

Doppler Shift of Multipath Component

Doppler shift of nth component arriving at an angle β_n relative to mobile motion is determined by the equation:

$$\Delta \mathrm{f_n} = \Delta f_d \cos \beta_n \tag{A.17}$$

Due to the presence of a large number of random components the multipath signal can be approximated as a Gaussian band-pass noise

The power spectral density S(f) of the multipath spectrum is then given by the following equation set (ETSI, 2003):

$$\mathrm{S(f)} = (\sigma^2)/[\pi * (1 - (\mathrm{f}/\Delta \mathrm{f_d})^{1/2})] \quad \text{when } |\mathrm{f}| < \Delta f_d \tag{A.18a}$$

$$\mathrm{S(f)} = 0 \quad \text{when } |\mathrm{f}| > \Delta f_d \tag{A.18b}$$

σ^2 = average power of the Gaussian band-pass equivalent process.

The corresponding auto-correlation function can be obtained from the equation:

$$\mathrm{r}(\tau) = \sigma^2 \mathrm{J_0}(2\pi \Delta \mathrm{f_d}) \tag{A.19}$$

where J_0 is a zero-order Bessel function of the first kind.

Received Signal Variation at Beam Edge

In assessment of systems supporting beam-beam handover, the rate of change of signal near the beam edge is a useful parameter. It can be used to determine whether a handover is necessary for the given mobile class, and if so, the tolerable time of handover completion. The rate of change of signal is the slope of antenna radiation pattern. For practical purposes, the shape of the main lobe (in dB) can be approximated as Gaussian. The loss in gain, $\Delta G(\theta_o)$, at an offset θ_o relative to the boresight for an antenna of beamwidth φ can, therefore, be approximated by the expression,

$$\Delta G(\theta_o) = 12\left(\frac{\theta_o}{\varphi}\right)^2 \tag{A.20}$$

The rate of change of gain is obtained by differentiating $\Delta G(\theta_o)$ with respect to θ,

$$\Delta G(\theta_o)' = \frac{24}{\varphi^2} \tag{A.21}$$

The beam-edge need not correspond to the half-power beam-width; in such a case if the offset gain is $G_{eoc,}$ the corresponding angular offset, θ_{eoc} can be obtained by the expression,

$$\theta_{eoc} = \sqrt{\frac{G_{eoc}}{12}}\,\varphi \tag{A.22}$$

The corresponding rate of change in gain can be obtained by substitution,

$$\Delta G(\theta_{eoc})' = \frac{4\sqrt{3G_{eoc}}}{\varphi} \tag{A.23}$$

The useful angular coverage then is given as $2\theta_{eoc}$.

A.2.3 Receiver

A.2.3.1 Receiver Noise

Receiver noise comprises a sum of antenna noise, noise introduced by the transmission line and the noise originating in the receiver electronics. The receiver's system noise temperature, T_{rs}, is, therefore,

$$T_{rs} = \frac{T_a}{L} + T_i(1 - 1/L) + T_r \tag{A.24}$$

T_a = Antenna noise temperature
L = loss factor of transmission line, connector, and other miscellaneous components
T_i = Ambient temperature in K (usually, 290K)
T_r = receiver noise temperature given as

$$T_r = T_1 + \frac{T_2}{G_1} + \frac{T_3}{G_1 G_2} + \ldots + \frac{T_n}{(G_1 G_2 \ldots G_n)} \tag{A.25}$$

T_n and G_n are respectively the effective noise temperature and the gain of the nth receiver stage.

A.2.3.2 Sensitivity Trade-Off

Given the required energy per bit to noise power spectral density ratio $\left(\frac{E_b}{N_0}\right)$, data rate, R_b, receiver power, C is given by the relationship (ETSI, 2005):

$$C = \frac{\frac{E_b}{N_0} k A_f R_b}{\frac{G}{T}} \tag{A.26}$$

where

$\frac{G}{T}$ = specified sensitivity of the receiver
k = the Boltzmann constant (1.38×10^{-23} J/K)

$$A_f = G_a - \left(\frac{G}{T}\right).T_a \tag{A.27}$$

where

G_a = Receiver antenna gain
T_a = Receiver noise temperature

A_f defines the amount by which the receiver temperature can be increased relative to the nominal case of $G_a = 1$ and $T_a = 0$, while maintaining the same terminal G/T.

References

Bains, N. (1999) The ICO system for personal communications by satellite. IMSC '99: The Sixth international Mobile Satellite Conference, Ottawa, Canada, 1999, Co-sponsored by the Communications Research Centre and the Jet Propulsion Laboratory June 16–18, pp. 88–93.

ETSI (2003) GEO-Mobile Radio Interface Specifications; Part 5: Radio Interface Physical Layer Specifications; Sub-part 5: Radio Transmission and Reception; GMR-2 05.005, ETSI TS 101 377-5-5 V1.1.1 (2001-03). Technical Specification.

ETSI (2005) GEO-Mobile Radio Interface Specifications (Release 1); Part 5: Radio Interface Physical Layer Specifications; Sub-part 5: Radio Transmission and Reception; GMR-1 05.005, ETSI TS 101 376-5-5 V1.3.1 (2005-02), Technical Specification.

Krewel, W. and Maral, G. (1998) Single and multiple satellite visibility statistics of first-generation non-geo constellations for personal communications. *International Journal of Satellite Communications*, **16**, 105–125.

Maral, G. and Bousquet, M. (1986) *Satellite Communications Systems*, John Wiley & Sons, Ltd, Chichester.

Mertens, H. (1976) EBU 3320. *Satellite Broadcasting Design and Planning of 12 GHz Systems*, EBU Technical Review, EBU, Tech, March.

Mohamadi, F., Lyon, D.L. and Murrel, P.R. (1988) Effects of solar-transit on K_u band VSAT systems. *International Journal of Satellite Communications*, **6**, 65–71.

Richharia, M. (1999) *Satellite Communication Systems: Design Principles*, 2nd edn, Macmillan Press Ltd, Basingstoke and London.

Siocos, C.A. (1981) Broadcasting satellites power blackouts from solar eclipses due to moon. *IEEE Transactions on Broadcasting*, **BC-27**(2), 25–28.

Wu W.W., Miller E.F., Pritchard W.L. and Pickholtz R.L. (1994) Mobile satellite communications, *Proceedings of the IEEE*, **(82)**, 1431–1449.

Index

Mobile Satellite Communications: Principles and Trends, Second Edition. M. Richharia.

Companion Website: www.satellitesandyou.com