Mobile Cellular Communication

Mobile Cellular Communication

Gottapu Sasibhushana Rao
Professor and Head
Department of Electronics and Communication Engineering
Andhra University College of Engineering
Visakhapatnam

PEARSON
Chennai • Delhi • Chandigarh

Associate Editor—Acquisitions: S. Shankari
Associate Editor—Production: M. R. Ramesh

ISBN 978-81-317-7361-1

First Impression

Published by Dorling Kindersley (India) Pvt. Ltd, licensees of Pearson Education in South Asia.

Head Office: 7th Floor, Knowledge Boulevard, A-8(A), Sector 62, Noida 201 309, UP, India.
Registered Office: 11 Community Centre, Panchsheel Park, New Delhi 110 017, India.

Compositor: AcePro India Ltd
Printed in India.

Contents

About the Author

Prof. Gottapu Sasibhushana Rao is Professor and Head of the Department of Electronics and Communication Engineering, Andhra University College of Engineering, Visakhapatnam, India. A seasoned academician with 27 years of R&D, industrial, and teaching experience, he has published more than 275 research papers in various journals of national and international repute and participated in scholastic conferences. He is a senior member of IEEE, fellow of IETE, member of IEEE Communication Society, Indian Geophysical Union (IGU), and International Global Navigation Satellite System (IGNSS), Australia.

Prof. Rao was also the Indian member in the International Civil Aviation Organisation (ICAO), Canada Working Group, for developing the SARPS. His current research areas include mobile cellular communication, SONAR, and acoustic signal modelling. A recipient of the Best Researcher Award and Dr Survepalli Radhakrishnan Award for Best Academician, he is the Chief Editor of *Journal of Communication, Navigation and Signal Processing*, Andhra University College of Engineering. He is also an Editorial Member of *International Journal of Positioning* (*POS*), Scientific Research Publishing, USA, and Journals of Scientific & Academic Publishing, USA.

Preface

Wireless communications has become essential part in our day to day life. During recent years there has been a significant improvement in the field of wireless communications technology and has rapidly evolved from **First Generation (1G)** to **Fourth Generation (4G)** systems. The rapid growth of cell phones, which principally carry voice, is now being used widely for communicating data and images. Wireless LANs, for example Wi-Fi, and recently the wireless Internet, are propelling the whole world towards greater integration. Over the years, various multiple access techniques and spread spectrum techniques have been developed to improve the performance of the wireless communication system. To understand this technology, it is important to know in detail, a number of concepts associated with wireless communication.

Objective

The increasing importance of wireless communication and networking in consumer electronics, professional applications, academics, domestic applications and industry has led to the realization of this well-organized book, which can be used as a comprehensive reference for specific wireless standards, architectures, technologies and protocols.

About the book

This book covers mobile cellular communications, wireless communication networks emerging wireless networking technologies, mobile propagation channel models, design of cellular networks besides dealing with fundamental aspects of antenna principles, cellular and base station antenna designs, various coding and modulation techniques, hand off techniques and latest multiple access techniques.

This book helps students to develop background knowledge and expertise in wireless communications and networking, which is one of the fastest growing technology in the world. It is easy-to-read, logical, and has a step-by-step approach that enables the reader to follow new and complex ideas clearly.

Spread across 30 chapters, the sections of each chapter are structured in a modular fashion. The book is designed to serve as text book for UG and PG students, and research scholars of Electronics and Communication Engineering, Electronics and Telecommunication Engineering, Computer Science and Engineering, Information Technology and Instrumentation Technology. It is also suitable for students and professionals in the fields of Data Processing and Data Communications and can be used as a good reference for practising professionals and scientists.

Chapter organization

Chapter 1 gives an introduction to mobile and cellular communication systems, cellular standards, evolution of cellular systems from 1G to 4G and their limitations, the geometry of the cell, different performance methods such as frequency reuse, handoff, and cell splitting. It also includes the principle of cellular radio systems, a brief overview of the analog and digital cellular mobile systems and some examples of existing mobile communication technologies.

Chapter 2 describes the frequency reuse concept and the importance of hexagonal cell shape in cellular systems in detail. Different methods used to increase capacity of cellular system like cell splitting, sectoring and repeater for range extension and zone microcells are also discussed.

Chapter 3 deals with various elements of cellular system design such as frequency reuse, co-channel interference, and cell splitting. The derivations for co-channel interference, reduction factor for design of cellular system in normal- and worst-case scenarios using omnidirectional antennas are presented.

Chapter 4 discusses different types of interferences that limit the capacity of a cellular system and explains about the estimation of interference from different sources.

Chapter 5 focuses on the need for an interference model and the general features of geographical and statistical models. The six interference models for evaluating interference and few methods for reduction of interference are described.

Chapter 6 introduces teletraffic engineering and its objectives. It presents the concepts of trunking, blocking, cell capacity, grade of service and their relevant expressions. Some operational techniques to improve the efficiency of the system are also discussed.

Chapter 7 unfolds the basics of the antenna theory and characteristics, properties of antennas and different types of antenna arrays used in wireless and microwave communication. These are useful for a better understanding of mobile antennas and cell site antennas, which are explained in the next chapter.

Chapter 8 expounds on the principle of operation of mobile antenna and different types of antennas used for mobile communication. The radiation patterns of these antennas are also discussed.

Chapter 9 delineates the design of base station antennas and their types of operation along with the considerations to be made for reducing interference.

Chapter 10 elucidates the basics of mobile radio channel propagation. A number of propagation models used to predict the path loss and small scale propagation effects are covered in this chapter.

Chapter 11 concentrates on cell site antenna heights, signal coverage cells and overview of near and long distance propagation.

Chapter 12 introduces frequency management and fixed or dynamic channel assignments. It also includes relevant simulation process and results.

Chapter 13 reveals various analog and digital modulation techniques which are used in cellular systems like Frequency Modulation, Frequency Shift Keying (FSK), Phase Shift Keying (PSK), Minimum Shift Keying (MSK), Gaussian MSK, M-ary Quadrature Amplitude Modulation (QAM), M-ary Frequency Shift Keying and Orthogonal Frequency Division Multiplexing (OFDM).

Chapter 14 relates the various principles of diversity techniques, equalization, speech and channel coding, coding schemes and their importance in detail.

Chapter 15 deals with the concepts of various spread spectrum techniques which are CDMA-specific and gives a performance comparison of different spread spectrum techniques.

Chapter 16 details the functionality and implementation of various multiple access techniques, their advantages and disadvantages. Packet radio protocols and space division multiple access technique are also discussed.

Chapter 17 outlines the various parameters used to compare the efficiency of different multiple access techniques. The necessary capacity and efficiency computations for various multiple access techniques are explained with examples.

Chapter 18 introduces handoff, handoff parameters, classification of handoffs based on the nature of handoff, handoff protocols and their purpose. It also concentrates on cellular structures and procedures for implementation of handoffs.

Chapter 19 deals with GSM and its architecture, specifications, operation and its maintenance systems, GSM channels, protocol stack configuration and its basic call flow.

Chapter 20 illustrates different types of orbits for satellite systems with an introduction on Global Mobile Satellite Systems like Iridium and Global Star. It also gives an account of the features and functional architecture of personal access communication system. An overview of rake receivers, mobile network signaling and its applications is presented.

Chapter 21 has a flow similar to Chapter 1. It traces the evolution of wireless technologies from 1G to 4G and describes their standards, advantages and limitations in detail.

Chapter 22 presents the principles of WCDMA air interface and CDMA2000 air interface, their architecture, design concepts and salient features, with necessary computations for capacity. It also gives a comparison between WCDMA and CDMA2000.

Chapter 23 depicts the history of evolution of cellular 4G technology and its advantages over 3G. The hardware and software aspects of 4G technology are explained in detail. A comparison between 4G and 3G is presented in terms of its performance.

Chapter 24 deals with wireless local area networks and its IEEE802.11 standards.

Chapter 25 is a continuation of the previous chapter and describes ZigBee, IEEE 802.15.4, WiMAX and IEEE 802.16 technologies and their architectures, devices, uses etc.

Chapter 26 begins with a discussion on the various generations of wireless networks, differences in wireless and fixed telephone networks and then describes the various traffic routing techniques and Wide Area Network technologies such as frame relay, ATM and ISDN.

Chapter 27 lays out the principles of Bluetooth technology, its architecture and baseband protocols, and radio specifications. It also gives an introduction to WLL Technology and compares the techniques of Bluetooth and Wi-Fi Wireless LAN.

Chapter 28 gives an account of the latest services which are supported by mobile phone GPRS system. It provides an overview and deals at length with the architecture, different entities and major functionalities of the interfaces. It also contains the details of PDP context procedures and combined RA/LA update procedures.

Chapter 29 gives an insight into Mobile IP, WAP and its application layer and describes their operations and architectures.

Chapter 30, the last chapter of the book, analyzes mobile data networks and describes important mobile packet data services such as SMS, GPRS and their technologies in terms of their main technical characteristics and architectural aspects.

Each chapter has a rich pedagogy that includes example problems, multiple-choice questions, review questions and open book questions. A list of abbreviations and glossary of terms are included at the end of the book for the student's benefit. In addition, there are eight appendices that examine, in turn, the decibel, frequencies for communication, fundamentals of antenna radiation, Weiner filter, mathematical formulas, computer networks, OSI model, and the Erlang-B table. These topics facilitate effortless understanding of the book's main concepts with minimal reference to alternative sources of information.

Comments and suggestions to enhance the contents of this book are welcome.

Acknowledgements

I express my gratitude to KVVS Reddy, MNVS Santosh Kumar, G. Sateesh Kumar, Rajkumar Goswami and R. Madhu for their enthusiasm in reviewing the manuscript and providing necessary suggestions to improve the text.

I thank S. Deva Prasad and B. Anurag for giving their best to add the pictorial representations to the text, which greatly enhanced the value of the book.

I am obliged to I. Silpa, D. Eswara Chaitanya, Ch. Vasavi Sridevi, T. Srisudha, A. L. Narayana and Sattar Ahmed for their effort in verifying the calculations and minimizing the errors.

I am indebted to all my research scholars, M.Tech students, S. Shankari and M. R. Ramesh from Pearson Education for their sincere work in correcting and editing the text, which helped greatly in bringing out this book in record time.

Finally, I would like to thank my family members wholeheartedly, without whose cooperation and constant encouragement this milestone could have not been achieved.

Gottapu Sasibhushana Rao

Introduction to Mobile and Cellular Communication Systems 1

1.1 Introduction

A mobile phone is a portable telephone that does not use a wired connection. It is also known as a *wireless phone, cell phone*, or *cellular telephone*. In many developing countries, mobile technology is a substitute for traditional fixed services. Worldwide the number of cellular phone users in the years 1984, 1994, and 1997 were 25,000, 16 million, and 50 million, respectively. In 2000, the number of wireless users became equal to the wired users and this number increased to 1.9 billion worldwide in the year 2005. The number of mobile users increased to 3 billion by 2007, which is almost half of the world's population. By 2011, the estimated mobile phone subscriber base in India will be 298 million and India will become the second largest country in the world, next to China. The resulting cellular penetration rate is 23.9 per cent of the nation's population. Mobile technology extends access to formerly unreserved population groups such as the urban poor and rural users. In addition to the standard voice function, current mobile phones also support latest services such as short message service (SMS), general packet radio service (GPRS), and multimedia service (MMS) for sending and receiving photos and videos, e-mail, packet switching, wireless access protocol (WAP), and Bluetooth. The cellular concept was developed by Bell Labs in 1960s–1970s.D

> *The evolution of cellular communication systems is commonly known by the **1G, 2G, 3G,** and **4G** designations.*

We are currently in the third generation (**3G**) cellular communication systems. The cellular network provides wireless connection between mobile phones or between a mobile phone and landline phone using radio waves. These mobile phones connect to the cellular networks which are further connected to the public switched telephone network (PSTN).

> *The cellular network uses a number of low-power transmitters called base stations (BSs) and each BS covers a unit area called a "cell".*

The cellular network concept is against the use of a single high-power transmitter with antenna mounted on a tall tower as is the case in the early mobile radios (shown in Fig. 1.1(a)) to cover a large area. The difficulty in the early mobile radio systems was the reuse of same frequencies throughout the system resulting in significant interference and lot of bandwidth being dedicated

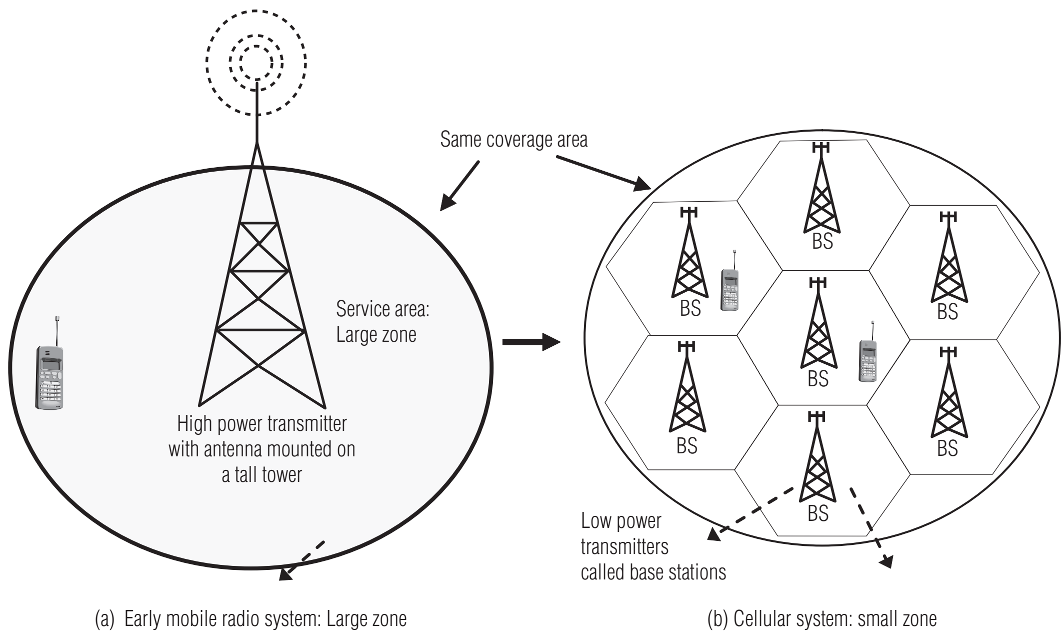


Figure 1.1 Early mobile radio system and cellular system

to a single call. The cellular system shown in Figure 1.1(b) uses a number of low-power transmitters called BSs to cover same area and to avoid the above difficulties.

This chapter introduces the different generations of cellular systems and their limitations, the geometry of the cell, different performance improvement methods such as handoff, frequency reuse, cell splitting, and so on, the principle of cellular radio systems, and also includes a brief overview of the analogue and digital cellular mobile systems.

1.2 Generations of wireless mobile systems

Wireless communication is basically transmitting and receiving voice and data using electromagnetic waves in open space. The origin of wireless communications can be traced back to the year 1857, when the behaviour of electromagnetic waves was explained mathematically using four equations by James Clerk Maxwell. Maxwell's four equations describe the electric and magnetic fields arising from varying distributions of electric charges and currents, and how those fields change with time.

The equations were the mathematical representation of decades of experimental observations of the electric and magnetic effects of charges and currents. According to Maxwell, an accelerated charge creates a magnetic field in its neighbourhood which in turn creates an electric field in the same area. A moving magnetic field produces an electric field and vice versa. These two fields vary with time, so they act as sources of each other. Thus, an oscillating charge having non-zero acceleration will emit an electromagnetic (EM) wave and the frequency of the wave will be same as that of the oscillation of the charge. Though the electromagnetic waves were first discovered as a communication medium at the end of the nineteenth century, these were put in use for the masses later.

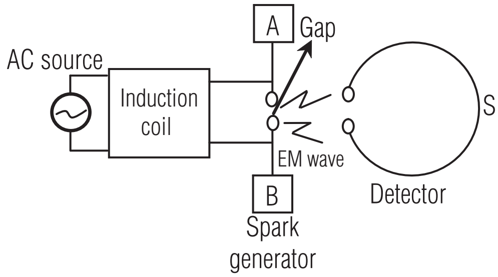


Figure 1.2 EM wave generation

Twenty years later, in the period 1879–1886, after a series of experiments, Heinrich Hertz came to the conclusion that an oscillatory electrical charge $\boldsymbol{q = q_0 sin\ t}$ radiates EM waves and these waves carry energy. Hertz was also able to produce EM waves of frequency 3×10^{10} Hertz. The experimental setup is shown in the Figure 1.2. To detect EM waves, he used a loop S, which is slightly separated from EM wave generator as shown in the figure.

Hertz's radio system consists of a switch and an induction coil to generate a spark across two electrodes. The receiver was a loop(S) made from a copper wire around 35 cm in diameter, with a small gap in the loop. When the "transmitter" generated a spark, a small spark was seen to jump the gap in the receiving coil. This is the basis for the antenna theory. In 1895–1897, Jagdish Bose also succeeded in generating EM waves of very short wavelength (~25 mm).

The father of today's mobile radio systems is G. Marconi, born in Italy in 1874. He demonstrated the first radio-based wireless transmission successfully using electromagnetic waves in 1901 over a distance of 1 mile. However, the bandwidth of these transmission systems was very small; the transmission of information was very slow. Over the next couple of decades, Marconi was a leading pioneer in establishing long-range wireless communication standards and his efforts led to the deployment of first radio-based telephony system for conversations between ships in 1915.

A technological breakthrough was the development of frequency modulation (FM) in the mid-1930s by Edwin Howard Armstrong. The Second World War battlefields were a major test bed for portable two-way FM radio technologies. The first systems offering mobile telephone service (MTS) (car phone) were introduced in the late 1940s in the United States and in the early 1950s in Europe. These single cell systems were severely constrained by restricted mobility, low capacity, limited service, and poor speech quality. Also, the equipment was heavy, bulky, expensive, and susceptible to interference.

In 1964, Bell Laboratories introduced the improved mobile telephone service (IMTS) which added full-duplex features to the old MTS. In 1968 and 1970, the Federal Communication Commission (FCC) realizing the huge potentials of mobile telephony, reallocated the frequency spectrum (40 MHz band in the 800 to 900 MHz frequency range) for cellular use. The cell covers a service area where a group of mobiles or terminals (referred to as users) are served primarily by one BS – usually located at the centre of the cell.

In addition, AT&T proposed a cellular mobile scheme in 1968 to the FCC which was approved in 1974. The cellular concept of using the same frequency at different places was introduced by MacDonald in 1979. The next evolutionary steps begin in the early 1980s with the deployment of the *first generation* (1G) analogue networks based on frequency division multiplexing (FDM).

In addition to the original 40 MHz band, an additional 10 MHz band was allocated in 1988 and called as the expanded spectrum (ES). Cellular communication is full duplex and the frequency band is divided between both communications: 25MHz is allocated to the forward path or downlink, which is the path for BS to mobile unit and the other half is for the mobile to BS. The paths are separated by a 45 MHz guard band in order to avoid interference between transmission and reception channels.

The increased demand for mobile communication led to the evolution of *second generation* (2G) digital networks in the early 1990s. The introduction of *time-division multiplexing* (TDM) on top of the existing FDM, an essential feature of 2G, increased the number of served subscribers per geographical area. In addition, voice quality was improved as well, with the introduction of newer voice coding algorithms. Finally, at the beginning of the third millennium, the first 2.5G networks, which are an upgrade of 2G and the 3G networks were implemented in most countries worldwide. And now the research is on the next-generation mobile technology with more advanced features, that is 4G, which is expected to be available in the market by 2012–2015. For clear understanding of the evolution of analogue and digital cellular technology, their broad features are illustrated in Figure 1.3.

1.2.1 First generation (1G)

The first 1G mobile phone system was introduced in 1980 in the United States. Before 1G, "0G" refers to pre-cellular mobile telephony technology, such as radio telephones that we had in cars before the advent of cell phones.

Analogue circuit-switched technology is used for this system, with frequency division multiple access (FDMA), as an air channel multiple access technique, and worked mainly in the 800–900 MHz frequency bands. The 1G mobile phone had only voice facility.

Examples of 1G system are analogue mobile phone systems (AMPS) and total access communication systems (TACS). The AMPS was implemented in North America and the TACS was used in Europe.

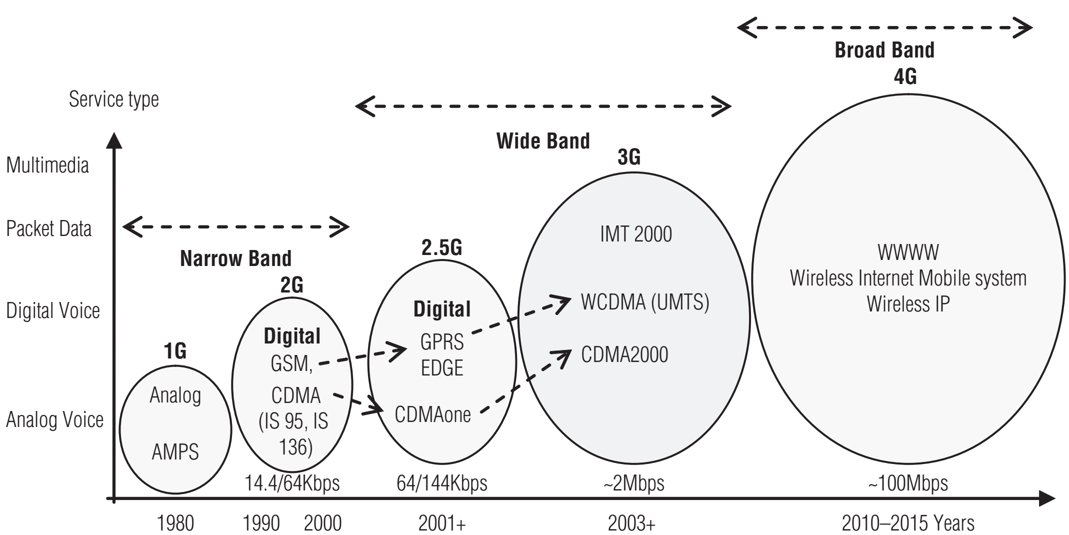


Figure 1.3 Evolution of cellular systems

In AMPS, two 25-MHz bands are allocated. One 25-MHz band is for communication from BS to mobile unit and the other for communication from mobile unit to BS.

The following are the limitations of 1G:

- Supports only speech
- Low traffic capacity
- Unreliable handover
- Long-call setup time and frequent call drops
- Inefficient use of bandwidth and poor battery life
- Poor voice quality and large phone size
- Allows users to make voice calls in 1 country only

1.2.2 Second generation (2G)

The need for more user capacity per cell led to the development of 2G technologies. 2G systems are *digital cellular* systems and were introduced in the late 1980s and were in use till the late 1990s. 2G technology supports data, speech, FAX, SMS, and WAP services. The frequency bands used by GSM are 890–960 MHz and 1710–1880 MHz. In the 890–960 MHz frequency band, the band at 890–915 MHz is dedicated to uplink communications from the mobile station (MS) to the BS, and the band at 935–960 MHz is used for the downlink communications from the BS to the MS. 2G digital technology is divided into two standards: time division multiple access (TDMA) and code-division multiple access (CDMA).

Global system for mobile (GSM) was the first commercially operated digital cellular system and uses TDMA/frequency-division duplexing (FDD)

IS-95 is commonly referred to as CDMA one standard and is used in North America and some parts of Asia

The following are limitations of 2G:

- Provides low data rates ranging from 9.6 kbps to 28.8 kbps.
- Circuit-switched network, where the end systems are dedicated for the entire call session. This causes reduction in usage of bandwidth and resources.
- Too many 2G standards globally (e.g. GSM, CDMA, PDC, and PHS)

1.2.3 Interim generation (2.5G)

The need for increased throughput data rates in data transfer (such as web browsing and e-mail) led to the evolution of 2.5G which is between 2G and 3G.

The mobile technology using GPRS standard has been termed as 2.5G.

The 2.5G was started in 1998 with added GPRS and enhanced data rates for GSM evolution (EDGE). In addition to the hyper text transfer protocol (HTTP), it supports the wireless access protocol (WAP) through which web pages can be viewed on the small screen of a mobile phone or a handheld device, which led to mobile commerce (m-commerce).

1.2.4 Third generation (3G)

The need for high-speed internet access, live video communications, and simultaneous data and voice transmission led to the development of 3G cellular networks. The 3G technology has added multimedia facilities to 2.5G phones. 3G operates in the frequency band of 1710–2170 MHz. It provides high transmission rates from 348 Kbps in a moving vehicle to 2 Mbps for stationary or mobile users.

The aim of 3G systems is to provide communication services from person-to-person at any place (global roaming) and at any time through any medium with guaranteed quality of service.

Examples of 3G system are universal mobile telecommunication systems (UMTS) and international mobile telecommunications at 2,000 MHz (IMT-2000).

UMTS are designed to provide different types of data rates, based on the following circumstances: up to 144 kbps for moving vehicles, 384 kbps for pedestrians, and 2 Mbps for indoor or stationary users. UMTS will integrate all the services offered by different mobile communication systems such as mobile phone, cordless telephone, and satellite radio in one service. Japan was the first country to introduce 3G system IMT-2000 network nationally, and in Japan the transition to 3G was completed in the year 2006.

Figure 1.4 illustrates the mobile phone samples of 1G, 2G, and 3G cellular network generations. Figure 1.4(a) is the first handheld device from Motorola Company which was available in 1984 in 1G network. Figure 1.4(b) is Ericsson GH218 which was introduced in 1994 and operated in 2G networks. Figure 1.4(c) is the LG U8110 that was introduced in 2004 and is operating in 3G networks.

The following are the drawbacks of 3G system:

- High bandwidth requirement
- High spectrum licensing fees
- Expense and bulk size of 3G phones
- Lack of 2G mobile user buy in for 3G wireless service
- Lack of network coverage because it is still a new service
- High prices of 3G mobile services in some countries

1.2.5 Fourth generation (4G)

Even though the 3G networks have been deployed since 2001, the true broadband access will be achieved with the 4G mobile phones. The 4G mobile communications will have transmission rates up to 20 Mbps higher than that of 3G.

4G technology is expected to provide very smooth global roaming universally with lower cost. Theoretically, 4G is set to deliver 100 Mbps to a roaming mobile device globally, and up to

(a) 1G (b) 2G (c) 3G

Figure 1.4 Samples of mobile phones from the three generations

1 Gbps to a stationary device. 4G will bring almost the perfect real world wireless internetworking called "WWWW: **WorldWide Wireless Web**"

With the expected features in mind, 4G allows for video conferencing, streaming picture-perfect video (e.g. tele-medicine and tele-geo processing application) and much more. Since the 4G is a research item for the next-generation wide-area cellular radio, the technology is expected to be available around 2012–2015. The following modulation techniques are proposed to be used in the 4G cellular phones.

- Variable spreading factor-orthogonal frequency and code division multiplexing (VSF-OFCDM).
- Variable spreading factor code-division multiple access (VSF- CDMA).

A short history of cellular evolution from 1G to 4G cellular systems is shown in Table 1.1. From Table 1.1 we can observe that the 4G is not a single defined technology or standard, but rather a collection of technologies and protocols aimed at creating fully packet-switched networks optimized for data. Another major difference between 3G and 4G is that unlike the 3G networks which are a combination of circuit-switched and packet-switched networks, 4G will be based on packet switching only. This will allow low-latency data transmission.

Table 1.1 History of 1G, 2G, 3G, and 4G technologies

Technology	Various generations				
	1G	**2G**	**2.5G**	**3G**	**4G**
Design began	1970	1980	1985	1990	2000
Implementation	1984	1991	1999	2002	2012–2015
Service	Analogue voice	Digital voice	High-capacity packets, MMS	High-capacity broadband data	Higher capacity, completely IP, Multimedia
Multiple access	FDMA	TDMA, CDMA	TDMA, CDMA	CDMA	OFDMA
Standards	AMPS, TACS, NMT	CDMA, GSM, PDC	GPRS, EDGE	WCDMA, CDMA2000	Single standard
Bandwidth	1.9 kbps	14.4 kbps	384 kbps	2 Mbps	200 Mbps
Core network	PSTN	PSTN	PSTN, Packet network	Packet network	Internet

1.3 Cellular geometry

The main reason for defining cells in a cellular land mobile radio system is to outline areas in which specific channels and specific cell sites are used. However, designers have realized that visualizing all cells as having the same geometrical shape helps to ease the design of cellular systems, not only in locating transmitter sites relative to one another and making economical use of equipment, but also in making the adaptation to traffic much easier. From this point of view, cellular geometry helps to ease the assessment of spectral efficiency of various cellular systems, in particular to calculate the significant co-channel interference (CCI) in the system.

1.3.1 Cell shapes

There are only certain patterns of cells or tessellations which can be repeated over a plane: the regular hexagon, the square, the circle, and the triangle.

The regular hexagon is favoured by system designers for the following reasons:

- It provides the best approximation to the circular omni-directional radio patterns achieved in practice.
- It is more economical to use since a hexagonal layout requires fewer cells and hence fewer stations.
- It combines ease of geometry with the practical realization of overlapping circles.
- For a given distance between the centre of a polygon and its farthest perimeter points, the hexagon has the largest area, and it almost approximates a circular radiation pattern.

Hexagons are generally used to represent the cells due to geometry considerations and calculation purposes.

For example, in Figure 1.5 hexagons closely approximate the circle, which is used as a coverage area by a BS that has transmission radius (range) R. More details are given in Chapter 2 about cell and sector shapes by comparing the hexagonal and circular cell shapes.

1.4 Introduction to cellular concept

The cellular concept was developed in response to the limitations of conventional mobile radio services. The main limitations of the previous mobile communication systems are as follows:

1. High-power transmitters were used to cover very large area.
2. Inefficient use of allocated radio spectrum.
3. If a user leaves the coverage area, they had to reinitiate the call on a different frequency channel.

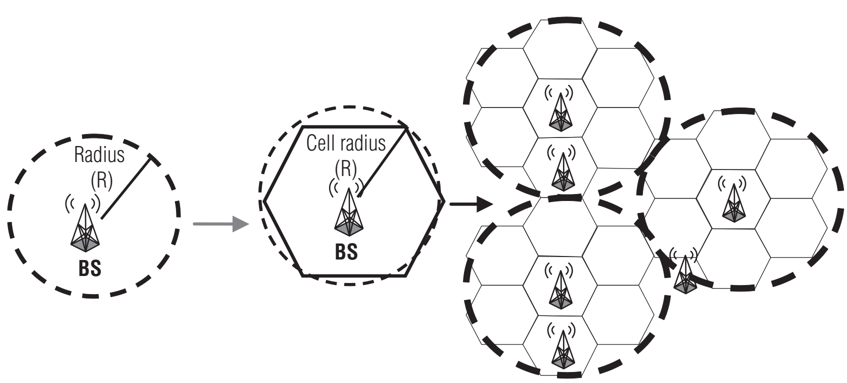


Figure 1.5 Circle to hexagonal cell shape approximation

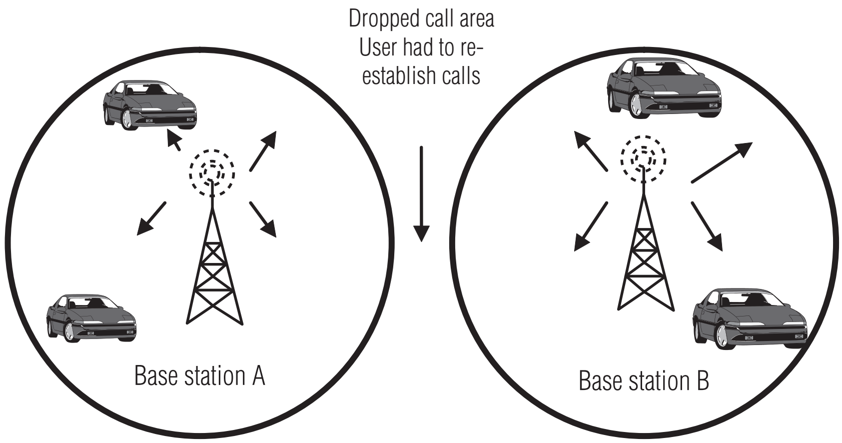


Figure 1.6 Conventional mobile radio service

In the beginning, there were no handoffs and the cellular system's size depended on how much power the centralized BS could transmit and receive. Users who stepped out of range of one system had to re-establish the call in the next system (Fig. 1.6). The capacity of these systems was severely limited because only a small number of radio channels (available bandwidth) were available for mobile systems. Therefore, they had to find a way to reuse radio channels in order to carry more than one conversation. Repeatedly reusing the radio frequencies over a given geographical area provides number of simultaneous conversations. The basic idea of the cellular concept is ***frequency reuse***.

1.4.1 Frequency reuse

Frequency reuse refers to the use of radio channels on the same carrier frequency to cover different areas that are separated from one another by sufficient distances.

Since the users in different geographical areas (cells) may simultaneously use the same frequency, this technique maximizes the number of mobile phones served in a given geographical area and spectrum efficiency. Frequency reuse causes CCI which is a trade-off link quality versus subscriber capacity. This concept is shown in Figure 1.7 and is explained in greater detail in Chapter 2. Cells with the same letter (A) use the same set of frequencies. A cell group or cluster is outlined in bold and replicated over the coverage area. In Figure 1.7, the cluster size (N) is 7 and the frequency reuse factor is 1/7 since each cell contains 1/7 of the total number of available channels.

1.4.2 Handoff

Notion of handoff is a crucial component in cellular concept. The mobile users by definition are mobile i.e. they can move around while using the phone. Hence the network should be able to provide them continuous access as they move. This will not be a problem if the user is moving within the same cell. But when the user moves from one cell to another, a ***handoff*** is required.

Handoff is one of the important concepts in the cellular mobile communications. The mobile user can move around while using the mobile phone, which is the main advantage of mobile

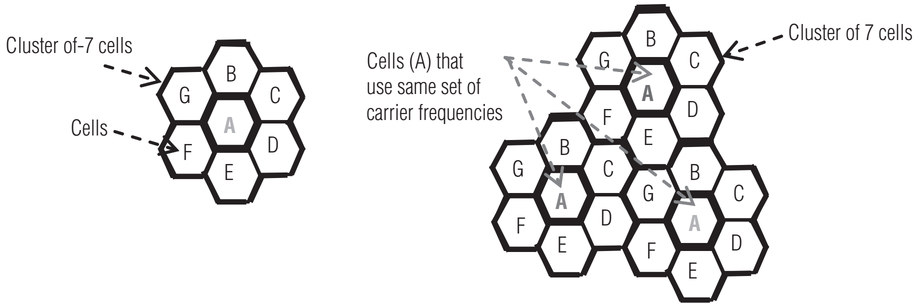


Figure 1.7 An illustration of the cellular frequency reuse concept

phones. Even when the mobile user is moving, the access to the network should be continuous. This problem does not arise if the user is moving within the same cell, but when the user is moving from cell to cell, a ***handoff*** is needed.

Handoff is the process of transferring an active call from one cell to another as the mobile unit moves from the first cell to the other cell without disconnecting the call.

When a mobile moves into a different cell while the call is in progress, the mobile switching centre (MSC) automatically transfers the call to a new channel belonging to the new BS. Handoff operation involves identifying a new BS along with the allocation of voice and control signals.

Example of a handoff process is given in steps with reference to mobile phone moving from one BS to another as shown in Figure 1.8:

- A user is transmitting and receiving signals from a given BS (say BS1).
- Assume the user moves from the coverage area of one BS into the coverage area of a second BS (BS2).
- BS1 notices that the signal from this user is degrading.
- BS2 notices that the signal from this user is improving.

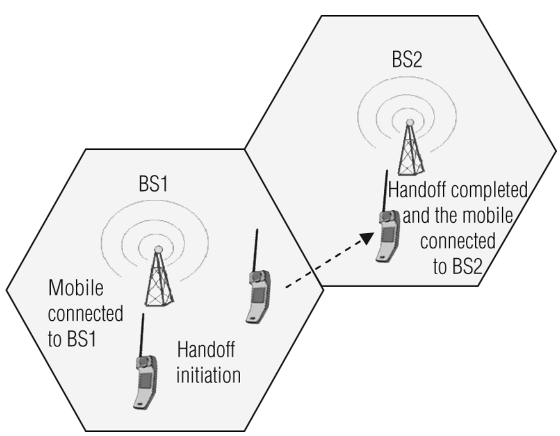


Figure 1.8 Call handoff process

- At some point, the user's signal is weak enough at BS1 and strong enough at BS2 for a handoff to occur.
- Specifically, if messages are exchanged between the user, BS1, and BS2 then the communication to/from the user is transferred from BS1 to BS2.

Figure 1.9(a) depicts an improper handoff scenario between two BSs (i.e. BS1 and BS2). When the mobile user in a car is at point A in the coverage area of BS1, then the received signal strength (RSS) is above the threshold level as shown in Figure 1.9(a). But if the mobile user in the car is moving towards point B in the coverage area of BS2, then the RSS received by the mobile due to the BS2 is dropped below the minimum acceptable threshold level and the call is terminated. Figure 1.9(b) depicts the proper handoff scenario that has taken place when the mobile user's car is moving from BS1 to BS2. In this case, when the mobile user is at point A under the coverage area of BS1, then the RSS is above the threshold level as shown in Figure 1.9(b). If the mobile user in the car is moving towards point B in the coverage area of BS2, then the RSS received by the mobile due to the BS2 is well above minimum acceptable threshold level, and therefore, the handoff is successful.

1.4.2.1 Types of handoff

Handoffs are broadly classified into two categories:

- Hard handoff
- Soft handoff (SHO)

Hard handoff is "break-before-make", meaning that the connection to the old BS is broken before a connection to the new BS is made. Hard handoff occurs when handoff is made between disjointed radio systems, different frequency assignments, or different air-interface characteristics or technologies.

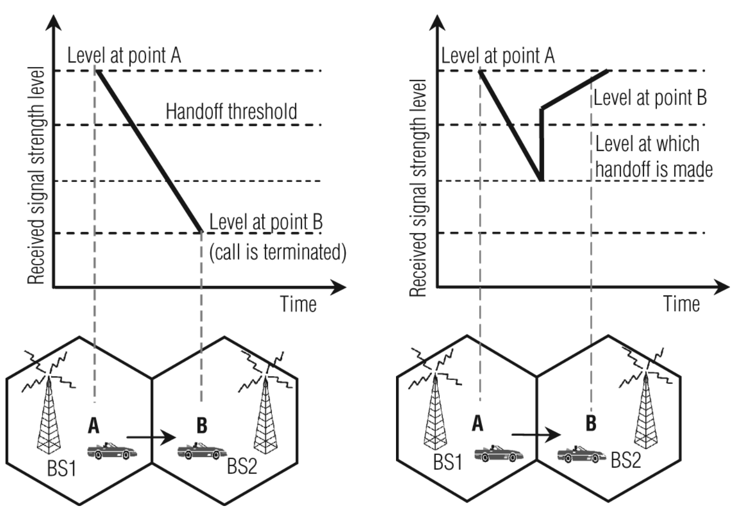


Figure 1.9(a) Improper handoff **Figure 1.9(b)** Proper handoff

Usually, the hard handoff can be further divided into two different types: intracellular and intercellular handoffs.

A handoff made within the currently serving cell (e.g. by changing the frequency) is called an intracellular handoff. A handoff made from one cell to another is referred to as an intercellular handoff.

SHO is "make-before–break", meaning that the connection to the old BS is not broken until a connection to the new BS is made. In fact, more than one BS is normally connected simultaneously to the MS. There are different types of SHO. When sectors of the same BS are involved in communication with the MS, the handoff is called softer handoff. When one sector from each BS is involved, the handoff is called soft handoff. When multiple sectors of one BS and one or more sectors of another BS communicate with the MS, the resulting SHO is called softer-soft handoff.

1.4.2.2 Handoff strategies

Mobility in network is managed by two different handoff strategies, namely horizontal handoff and vertical handoff.

In case of horizontal handoff, handoff is between two network access points or BSs that use the same wireless network access technology. The handoff is purely due to mobility of the MS. In case of vertical handoff, handoff is between two network access points or BSs that use the different wireless network access technology.

In the 1G analogue cellular systems, the RSS measurements are made by the BS and are supervised by the MSC. In the 2G systems that use TDMA technology, mobile-assisted handoff (MAHO) is used. In MAHO, every mobile phone measures the RSS from the surrounding BS and continuously reports the RSS values to the corresponding BS. Further details of the handoff mechanism are presented in Chapter 18.

1.4.3 Co-channel interference

For each cell, a set of frequencies is allocated.

Cells that use the same set of frequencies are denoted as **co-channel cells** and the interference received from co-channel cells is called **co-channel interference**.

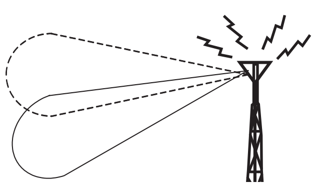


Figure 1.10 CCI reduction using beam tilting

The CCI occurs mainly due to reusing an identical frequency channel. This has become a major problem in the mobile cellular network. To reduce the CCI, minimum frequency reuse distance must be used. If all cell sizes are fixed, CCI is independent of the transmitted power of each cell. One method to reduce the CCI is by tilting down the BS antenna beam as shown in Figure 1.10 due to which the power outside the cell causing CCI reduces. CCI in the cellular system is described in detail in Chapter 3.

The following five types of approaches are followed in cellular communications to increase the user capacity.

- *Adding new channels:* New channels are added between mobile unit & base station.
- *Frequency borrowing*: Frequencies are taken from adjacent cells by congested cells.
- *Cell splitting*: Cells in areas of high usage can be split into smaller cells.
- *Cell sectoring*: Cells are divided into a number of wedge-shaped sectors, each with their own set of channels.
- *Microcells*: BS antennas move to buildings and lamp posts.

1.4.4 Cell splitting

Cell splitting is the process of dividing the radio coverage of a cell site into two or more new cell sites. Cell splitting is performed to provide additional capacity (number of channels) within the region of the original cell site by increasing the number of BSs.

Splitting a cell provides more number of cells,
reduction in the cell size, and
corresponding reduction in the antenna height and transmitter power

More number of cells gives more number of clusters, resulting in more number of channels and high user capacity.

The cell radius (R) reduction by a factor of f reduces the coverage area and increases the required number of BSs by a factor of $\boldsymbol{f^2}$.

Figure 1.11 illustrates how cells can be divided if higher capacity is needed in a spot. We need to go locally to smaller cluster size (N). Figure 1.11 consists of three clusters and each cluster

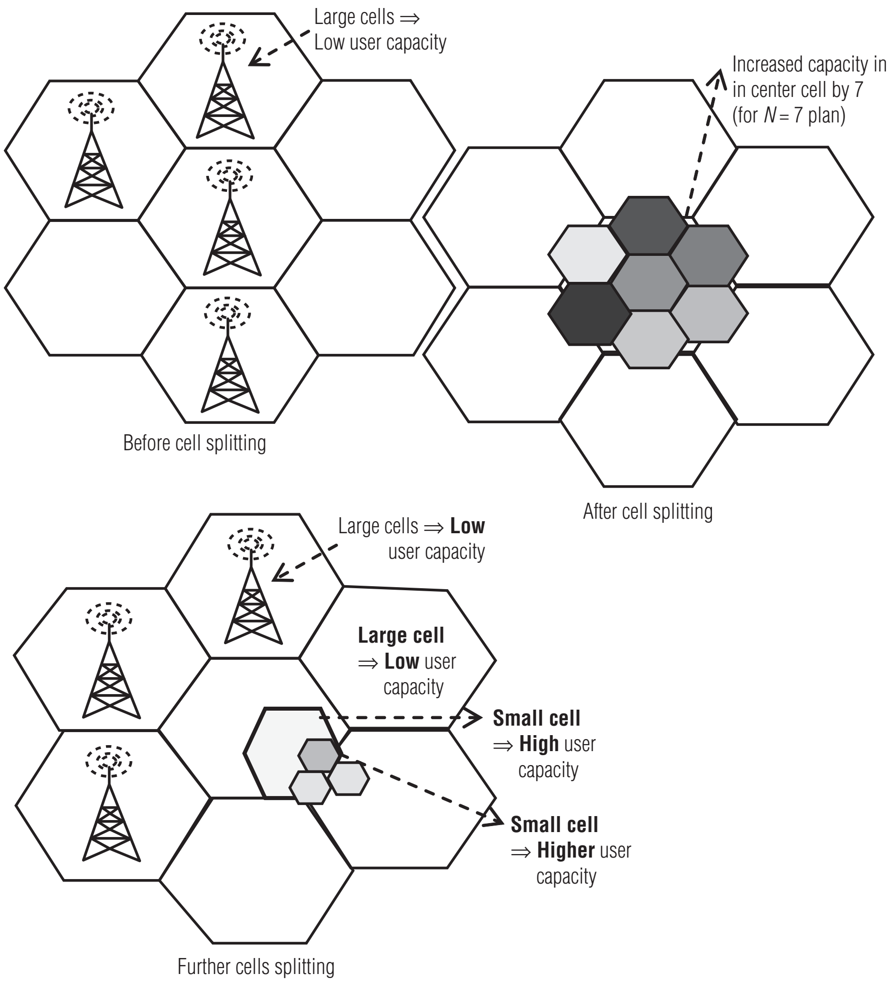


Figure 1.11 Cell splitting before and after in 3 clusters of size 7

is a group of seven cells. (http://www.wirelessdictionary.com/Wireless-Dictionary-Cell-Splitting-Definition.html). To cover a smaller area, the radio coverage area of large cells sites are split by adjusting the power level and/or using reduced antenna height. The radio coverage area of the cell site can be reduced by changing the RF boundaries of the cell site. This is similar to placing the cells farther apart and permitting new cells to be added. To use the cell resources efficiently the smaller cells can be either activated or deactivated according to the traffic patterns. More details on cell splitting are discussed in Chapter 3.

1.5 Principle of operation of a cellular mobile system

The most common example of a cellular network is a *mobile phone* (*cell phone*) *network*. A mobile phone is a portable *telephone* used to receive or make calls through a *cell site* (BS), or transmitting tower. Electromagnetic waves are used to transfer signals to and from the cell phone.

Modern mobile phone networks use cells because radio frequencies are limited, shared resource. Cell-sites and handsets change frequency under computer control and use low-power transmitters so that a limited number of radio frequencies can be simultaneously used by many callers with less interference. Figure 1.12 illustrates the cellular mobile network.

A cellular network is used by the mobile phone operator to achieve both coverage and capacity for their subscribers. Large geographical areas are split into smaller cells to avoid line-of-sight (LOS) signal loss and to support a large number of active phones in that area. All cell sites are connected to telephone exchanges (or switches), which in turn connect to the *public telephone network*.

Coverage area of cells: In cities, each cell site may have a range of up to approximately 1/2 mile, while in rural areas the range could be as much as 5 miles. It is possible that in clear open areas, a user may receive signals from a cell site 25 miles away.

1.5.1 Components of a cellular mobile network

A cellular network is formed by connecting the following five components as shown in Figure 1.12.

1. Mobile station (MS)
2. Base station (BS)
3. Mobile switching centre (MSC)
4. Base station controller (BSC)
5. Public switched telephone network (PSTN)

The function of each network component is described in the following:

Mobile station (*MS*): MSs are usually a mobile phone. Each mobile phone contains a *transceiver* (transmitter and receiver), an antenna, and control circuitry. Antenna converts the transmitted RF signal into an EM wave and the received EM waves into an RF signal. The same antenna is used for both transmission and reception, so there is a duplexer switch to multiplex the same antenna.

Base station (*BS*): One of the important components in the cellular network is the BS.

BS provides direct communication with mobile phones and it defines the cell. When cells are grouped together, a cluster is formed. Within a cluster, no channels are reused.

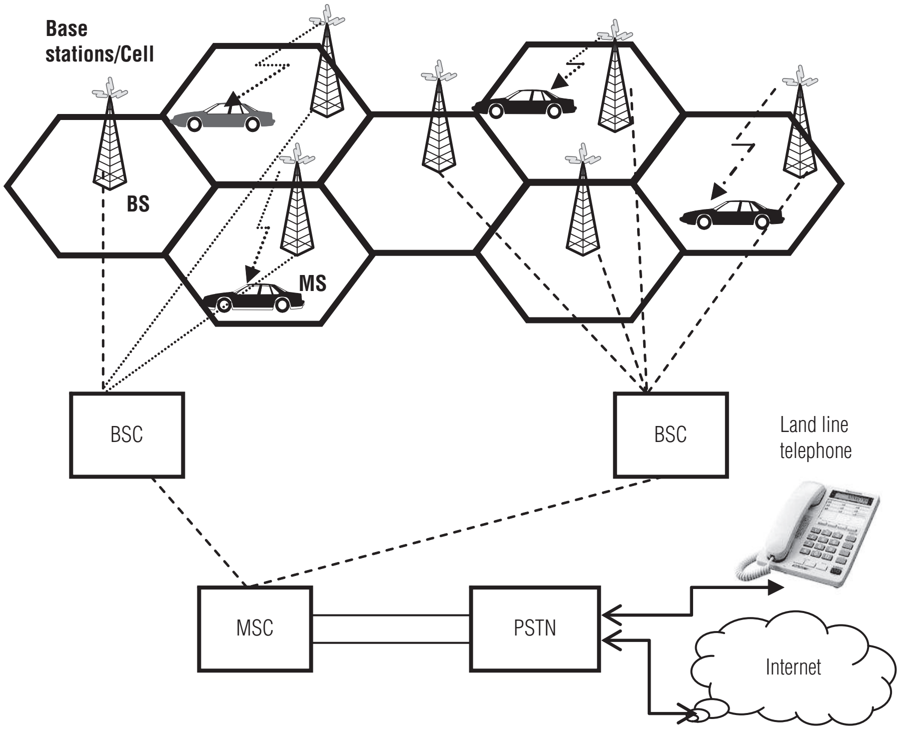


Figure 1.12 Cellular network

Two frequencies are required to establish communication between MS and BS: one from mobile phone (MS) to BS (uplink channel) and inverse (downlink channel) as shown in Figure 1.13.
A group of BSs are in turn connected to a BSC. The BS is a transceiver station or system and consists of a number of different elements.

- The first part of the BS is electronics section normally located in a container at the base of the antenna tower. The various electronic devices for communicating with the mobile handsets include RF amplifiers, radio transceivers, RF combiners, control, communication links, and power supplies with backup.
- The second part of the BS is the antenna and the feeder to connect the antenna to the base transceiver station itself. These antennas are visible on top of the masts and tall buildings enabling them to cover the required area.
- It is important that the location, height, and orientation are all correct to ensure that the required coverage is achieved.
    - If the antenna is too low or in a poor location, there will be insufficient coverage, leaving a coverage “hole”.
    - If the antenna is too high and directed incorrectly, then the signal will be heard well beyond the boundaries of the cell. This may result in interference with another cell using the same frequencies.

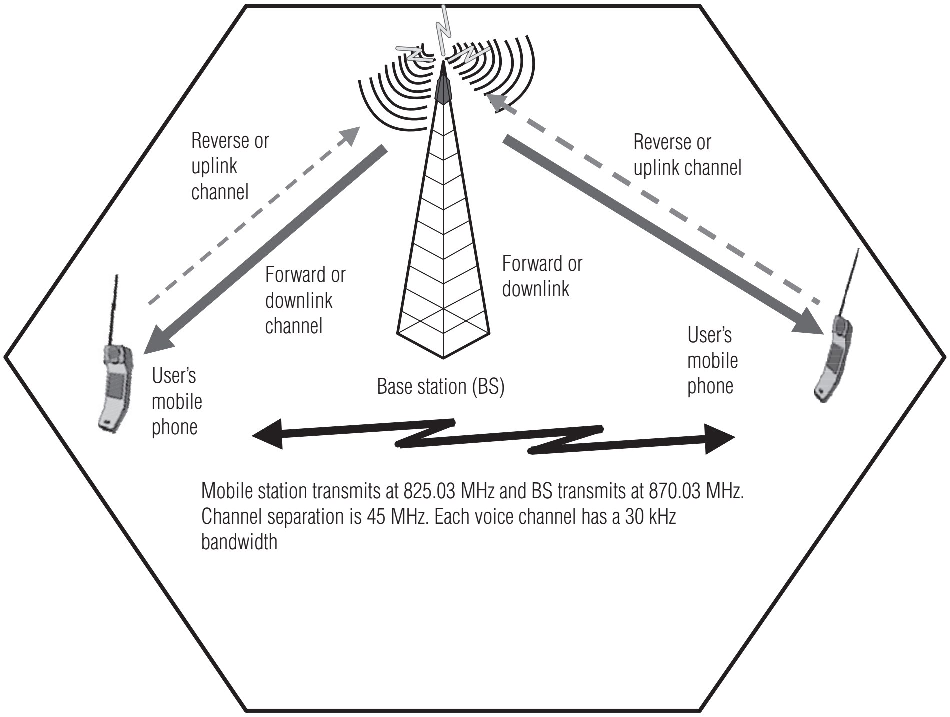


Figure 1.13 Downlink and uplink channels

BS or cell site antenna: Either omni-directional or directional antennas are used as BS antennas in the wireless industry. The typical directional antenna is shown in Figure 1.14. The cell site mast generally has three "faces" each with several frequency agile, directional antennas. Each face covers approximately 120° of the cell and each face uses a different subset of the cell's assigned frequencies. Usually, the antenna tower is at the centre of the cell.

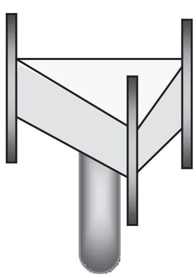


Figure 1.14 3 face directional antenna as base station antenna

Omni-directional antenna: Today, the omni-directional antenna shown in Figure 1.15a at BSs exists only in rural areas for the most part. This is because of the lower subscriber densities in rural areas and the lack of requirement for the increased capacity that is afforded by using directional antennas and sectorized BSs. Omni-directional BSs are noted for their use of omni-directional antennas which are slender, long, and tubular. There are always two receive antennas at every BS, which are known as *receive zero* (Rx0) and *receive one* (Rx1). The purpose of having two receive antennas at every BS is to provide for what is known as *space diversity*. Space diversity, also known as *receive diversity*, compensates for Rayleigh fading in the uplink to the BS. Space diversity is a tool used to optimize the signal received by a BS (transceiver); it counteracts the negative effects of

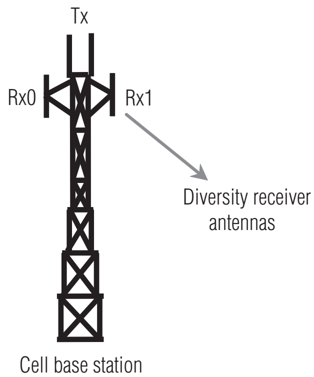


Figure 1.15a Horizontal view of tower mounting of omni directional BS antenna

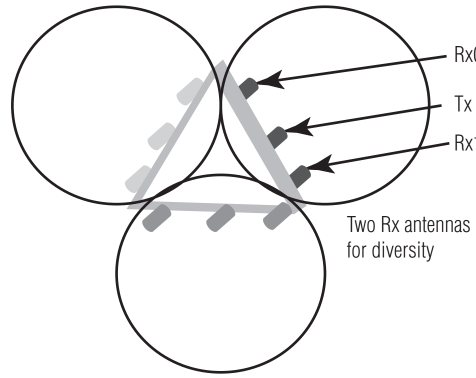


Figure 1.15b Typical antenna arrangement

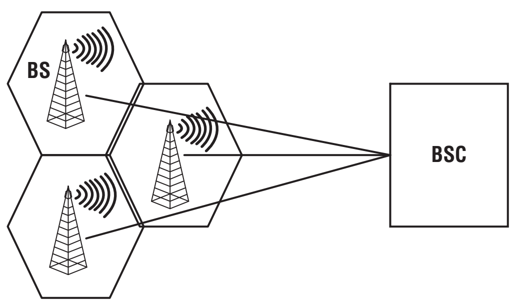


Figure 1.16 Interconnection of BS and BSC

Rayleigh fading. It ensures that the best possible receive signal is used to process all wireless calls. A typical antenna arrangement is shown in Figure 1.15(b).

*Base station controller (BSC)***:** A number of BSs are connected to a BSC as shown in Figure 1.16. An important function of BSC is that it manages the "handoff" from one BS to another as a subscriber moves from cell-to-cell. The BSC contains logic to control each of the BSs. Also, a group of BSCs are in turn connected to a MSC via microwave link or telephone lines.

Mobile switching centre (MSC): The MSC is the control centre for the cellular system. The MSC is also known as mobile telephone switching office (MTSO). It coordinates the actions of the BSs providing overall control and acts as a switch and connects into the PSTN. Various functions performed by a MSC are as follows:

- It communicates with the BSs, routing calls and controlling them as required.
- It contains databases detailing the last known locations of the mobiles.
- It also contains facilities for authentication centre allowing mobiles onto the network.
- It contains facilities to generate billing information for individual accounts.

For this purpose, the MSC makes use of the three major components of the network subsystem (NSS), that is HLR, VLR, and AUC.

Home location register (HLR): The HLR contains the information related to each mobile subscriber, such as the type of subscription, services that the user can use, the subscriber's current location, and the mobile equipment status. The database in the HLR remains intact and unchanged until the termination of the subscription.

Visitor location register (VLR): The VLR comes into action once the subscriber enters the coverage region. Unlike the HLR, the VLR is dynamic in nature and interacts with the HLR when recoding the data of a particular mobile subscriber. When the subscriber moves to another region, the database of the subscriber is also shifted to the VLR of the new region.

Authentication centre (AUC): The AUC (or AC) is responsible for policing actions in the network. This has all the data required to protect the network against false subscribers and to protect the calls of regular subscribers. There are two major keys in the GSM standards: the encryption of communications between mobile users and the authentication of the users. The encryption keys are held both in the mobile equipment and the AUC and the information is protected against unauthorized access.

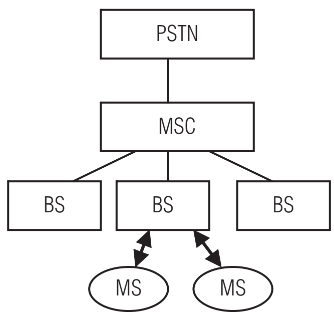


Figure 1.17 PSTN to mobile station connectivity

PSTN is a cellular network that can be viewed as an interface between mobile units and a telecommunication infrastructure (Fig. 1.17). Therefore, the PSTN network is nothing but the land-based section of the network. It is necessary that the BSs are to be connected to a switching network and that network is to be connected to other networks such as the PSTN, so that calls can be made to and from mobile subscribers.

1.5.2 Common air interface

Communication between the BS and the mobiles is defined by a standard common air interface (CAI) that specifies four different channels. The channels used for voice transmission from the BS to mobiles are called forward voice channels (FVC) and the channels used for voice transmission from mobiles to the BS are called reverse voice channels (RVC). The two channels responsible for initiating mobile calls are the forward control channels (FOCC) and reverse control channels (RECC). Control channels are often called setup channels because they are only involved in setting up a call and moving it to an unused voice channel. Control channels transmit and receive data messages that carry call initiation and service requests, and are monitored by mobiles when they do not have a call in progress. FOCCs also serve as beacons which continually broadcast all of the traffic requests for all mobiles in the system.

1.6 Call transfer operation from one mobile phone to another

The operation of one phone placing a call to another mobile phone and the operation that takes place when a MS receives an incoming call is described in this section. Before describing the call transfer operation to/from one mobile to another mobile, the knowledge of the concept of duplex and control/voice channels must first be understood.

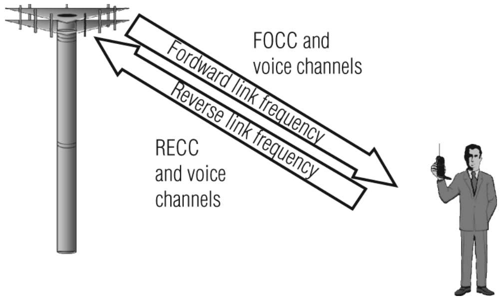


Figure 1.18(a) FDD

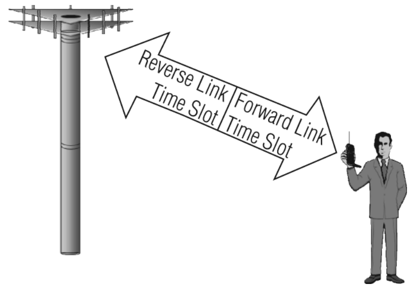


Figure 1.18(b) TDD

1.6.1 The duplex concept

One of the key elements of any radio communications system is the way in which radio communications are maintained in both directions. The various types of mobile radio transmission systems in use are simplex, half duplex, and full duplex.

Simplex: Communication is possible only in one direction (e.g. paging systems).

Half duplex: Two-way communication, but uses the same radio channel for both transmission and reception. User can only transmit or receive information (e.g. walkie-talkie).

Full duplex: Simultaneous two-way radio transmission and reception between the subscriber and the BS (e.g. FDD and time-division duplexing [TDD]).

Similar to landline phones, cellular phones must also be full duplex. For cellular systems, it is necessary to talk or to send data in both directions simultaneously and this places a number of constraints on the schemes that may be used to control the transmission flow. Using the FDD and TDD duplexing schemes, simultaneous two-way communication can be established.

***FDD** uses two separate frequencies for the uplink (from the mobile to the BS) and the downlink (from the BS to the mobile).*

***TDD** uses a single frequency to transmit signals in both the downlink and uplink directions.*

In FDD information from the mobile handset to the BS is carried on one frequency and information from the BS to the handset is carried on another (Fig. 1.18(a)). In TDD information from the handset to the BS is transmitted at one time on one frequency and information from the BS to the handset is transmitted at another time on the same frequency (Fig. 1.18(b)).

Example problem 1.1

The bandwidth allocated to a particular FDD cellular system is 33 MHz. It uses two 25 kHz simplex channels to provide full duplex voice and control channels, compute the number of channels available per cell if a system uses (a) 4 cell reuse and (b) 7 cell reuse (refer Section 1.6.1).

Solution

Total bandwidth allocated to cellular system = 33 MHz

Channel bandwidth required for duplex channel (uplink and downlink) = 25 kHz × 2 = 50 kHz

Total number of available channels = 33,000,000/50,000 = 660

a) For cluster size $N = 4$, number of channels available per cell = 660/4 = 165
b) For cluster size $N = 7$, number of channels available per cell = 660/7 = 95

1.6.2 Control and voice channels

There are two general types of channels in a cellular system: control channels and traffic (voice) channels.

Control channels: The control channels are referred to as setup channels and paging channels. Control channels are also sometimes called paging channels in the case of downlink, and access channels in the case of uplink. The paging channels are used to set up calls that originate from the BS, while the access channel is used to set up calls that originate from the mobile. Paging channel is designated as FOCC and the access channel as the RECC (Fig. 1.18(a)).

The FOCC and RECC establish the MS on the network (registration), to set up calls from the MS and to set up calls coming in for a particular MS (called a mobile page). After a call is established using the RECC and FOCC, the process switches to the voice channels.

Traffic channels: Traffic channels are active during voice conversations, but they also do contain the digital information needed to keep a call up. A MS thus tunes to and receives either a control channel or a traffic channel at any given moment.

1.6.3 Operation of one mobile phone placing a call to another mobile phone

The operation of one mobile phone placing a call to another involves two steps. One is the initialization of the mobile system and the second is the establishment of the call.

Initialization of mobile system: Five basic steps are involved in the mobile initialization procedure. They are power on, scanning, tuning, registering, and listening. When a mobile phone is turned on, it scans and selects the strongest and best bit-error rate (BER) (control channel) signal sent by adjacent BSs. Then a handshaking process takes place between the mobile phone and the MSC to identify the user and register its location. This procedure is repeated periodically as long as the mobile unit is on to monitor the location of the mobile.

Establishment of a call: If a user dials a number and presses START or TALK button, the mobile phone initializes a call by sending a call initiation request to its nearest BS. This request is sent on a special channel (RECC). The BS sends the request, which contains the telephone number of the called party, to the MSC (http://onlinelibrary.wiley.com/doi/10.1002/9780470050118.ecse055/full). The MSC validates the request and uses the number to make a connection to the called party via the PSTN. Then PSTN first connects itself to the MSC of the called party, and then the MSC instructs the BS and MS that placed the call to switch to voice channels. The MS that placed the call is then connected to the called station, using unused forward and backward voice channels.

1.6.4 Operation that takes place when a mobile station receives an incoming call

The following operation takes place when a MS receives an incoming call is described in this section.

MSs continually scan the FOCC for paging signals from BSs. Paging signal informs the mobile phone that it has a call coming in and should prepare the set up to receive it.

When a MSC receives a request for a connection to a MS in its area, it sends a broadcast message to all BSs under its control. The message contains the number of the MS that is being called. The BSs then broadcasts the message on all FOCCs. The correct MS acknowledges the page, by identifying itself over the RECC. The MSC receives the acknowledgment via the BS and instructs the BS and MS to switch to an unused voice channel. A data message is then transmitted over the FVC which instructs the mobile phone to ring.

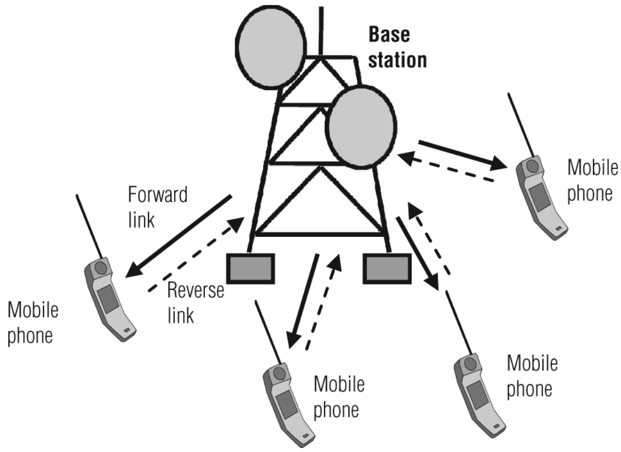


Figure 1.19 Multiple access control

1.7 Multiple access schemes

The radio spectrum is a scarce resource. Without access to radio spectrum, there can be no mobile communication. Multiple access refers to techniques that enable multiple users to share a finite portion of given frequency spectrum efficiently. The sharing of frequency spectrum is required to achieve high capacity by simultaneously allocating the bandwidth.

The multiple users will be assigned channels within that portion according to various techniques, known as multiple access schemes. A channel can be thought of as merely a portion of the limited radio resource (frequency slot or time slot or code), which is temporarily allocated for a specific purpose, such as someone's phone call. Multiple access control is shown in Figure 1.19.

The five most common schemes are given below and are discussed in more detail in Chapter 16.

- *Frequency division multiple access (FDMA)*, where the total spectrum assignment is divided into a number of discrete frequencies.
- *Time division multiple access (TDMA)*, where the total spectrum is divided in time between a number of users.
- *Code division multiple access (CDMA)*, where neither the frequencies nor the time are divided but users are distinguished through the use of a special code.
- *Orthogonal frequency division multiple access (OFDMA)*, where the spread spectrum technique spreads the data over a number of carriers that are spaced apart at precise frequencies.
- *Space division multiple access (SDMA)*, where different users will be served on same frequency channel at the same time.

1.8 Analogue and digital cellular mobile systems

The 1G cellular mobile systems were introduced in the beginning of the 1980s. Examples of these early analogue systems are AMPS and NMT. It used FDMA technology to achieve radio communications. In the 1990s the analogue systems were replaced by digital technology, which provided higher capacity, better quality, and new services. The most widely used 2G mobile system is **GSM**. The deployment of the 3G mobile networks is now ongoing. The 3G offers new data and multimedia services. The major 3G technology is called WCDMA. In this section, the analogue and digital cellular systems are briefly introduced.

1.8.1 Analogue cellular mobile radio systems (AMPS)

During the early 1980s, there were analogue technologies and the 1G cellular system was designed for analogue voice communications only. The following are examples of 1G cellular analogue radio system:

- Advanced mobile phone system (AMPS) in the United States
- Total access communication systems (TACS) in the United Kingdom
- Nippon advanced mobile telephone system (NAMTS) in Japan

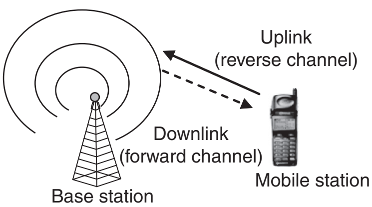


Figure 1.20 Forward and reverse communication channels in AMPS

AMPS are one of the leading analogue cellular phone systems in the United States. AMPS use the 800 MHz to 900 MHz band. It uses two separate analogue channels, one for forward (BS to MS) and the other for reverse (MS to BS) communication. The 824–849MHz band is used for reverse communication and the 869–894MHz band is for forward communication (Fig. 1.20). This spectrum is divided into 832 frequency channels consisting of 416 downlink and 416 uplink channels. The division of the spectrum into sub-band channels is achieved by using FDMA.

Salient features of analogue cellular mobile systems (AMPS)

- Frequency-division multiple access (FDMA) is the analogue cellular modulation standard.
- Cellular network uses a duplex mode of communication and two channels are required for each call: one channel for transmitting and one channel for receiving.
- User's mobile phone transmits on 824–849MHz band known as uplink or reverse channel.
- BS transmits on 869–894MHz band known as downlink or forward channel.
- In AMPS, the channel spacing is 30 kHz. Each uplink and downlink channel occupies 30 kHz of bandwidth. Every AMPS cellular call actually occupies a total of 60 kHz.

Limitations of AMPS include

- Low calling capacity
- Limited spectrum
- Poor data communication
- Minimal privacy
- Inadequate protection

Example problem 1.2

The American analogue technology standard, known as Advanced Mobile Phone Service (AMPS), employs frequency modulation and occupies a 30 kHz frequency slot for each voice channel. Suppose that a total of 30 MHz bandwidth is allocated to a particular cellular radio communication system with cluster size 7. How many channels per cell does the system provide? (refer Section 1.8.1).

Solution

Allocation of 15 MHz each for forward and reverse links provides a little more than 1,000 channels in each direction for the total system, and correspondingly a little less than 150 per cell.

IS-54/IS-136

This standard is based on the AMPS system and is commercially known as digital-AMPS (D-AMPS). It introduces TDM in the AMPS channels. It uses digital modulation (π/4 QPSK with speech coding in TDMA). It was standardized from the Telecommunications Industry Association (TIA) in 1990 and it uses the same frequencies with AMPS as are used, but in every time slot up to three full rate or six half rate users are multiplexed. The forward channel (downlink range) and reverse channel (uplink range) frequency bands are 869–894 MHz and 824–849 MHz, respectively. The final version of IS-54 Rev-C was called IS-136 and is still in use today.

1.8.2 Digital cellular mobile radio systems

While analogue cellular phone system (1G) was designed for analogue voice communication, the digital cellular mobile radio system (2G) was mainly designed for digitized voice. There are a number of different digital cellular technologies including the following:

- Global system for mobile communications (GSM)
- General-packet radio service (GPRS)
- Code-division multiple access (CDMA)
- Evolution-data optimized (EV-DO)
- Enhanced data rates for GSM evolution (EDGE)
- Digital enhanced cordless telecommunications (DECT)
- Digital AMPS (IS-136/TDMA)
- Integrated digital enhanced network (IDEN)

Two main groups have evolved in the digital cellular mobile radio system development. One group is from Europe and another is from America. The digital cellular mobile radio systems developed by the two groups are

- Global system for mobile communications (GSM) in Europe
- Code-division multiple access (CDMA)/Interim Standard (IS-95) in the United States

The above cellular systems (GSM and CDMA) are not compatible with each other.

1.8.2.1 Global system for mobile communications

It was developed in Europe in the year 1990. It provides a common 2G technology all over Europe. GSM uses TDMA and FDMA techniques as access mechanism. In GSM, the bandwidth is divided into time slots for better utilization of bandwidth. GSM operates in the 900 MHz band (890–915 MHz for forward link and 935–960 MHz for reverse link channels) in Europe and Asia and in the 1,900 MHz band in the United States. GSM uses two bands for duplex communication. Each band is 25 MHz in width, shifted towards 900 MHz. Each band is divided into 125 channels of 200 kHz separated by guard bands and each channel is subdivided into eight time slots or sub channels. One timeslot must be allocated for control channel purposes; therefore, up to seven subscribers can use a channel simultaneously. GSM users are almost eight times in number than CDMA users worldwide and in India, Bharti Telecom (Airtel) provides GSM standard. GSM mobile phone consists of two main components: handset and subscriber identity module (SIM).

SIM: The handset in a GSM system is different from analogue phones in that the identification information of the subscriber is programmed into a SIM module and not in the handset. The main functions of the handsets are receive/transmit and encoding and decoding of the voice transmission. The SIM is a microcontroller embedded into a small piece of plastic. The SIM card provides authentication, information storage, subscriber account information, and data encryption. SIM chips and handsets are swappable.

GSM: This is a digital-wireless standard which uses TDMA technology as its **air interface**. GSM has been deployed in the 900, 1,800, and 1,900 MHz bands. Each 25 MHz band provides a total of 125 forward channels and 125 reverse channels. Each channel has a bandwidth of 200 kHz (25 MHz bandwidth/125 channels). Where each channel is subdivided into eight time slots, or sub channels and the sub channel spacing is 25 kHz (200 kHz channel space divided by eight time slots). Each channel is shared by eight users giving a total of $125 \times 8 = 1{,}000$ users per cell.

Example problem 1.3

The global system for mobile (GSM) communications utilizes the frequency band 935–960 MHz for the forward link and frequency range 890–915 MHz for the reverse link. Each 25-MHz band is broken into radio channels of 200 kHz. Each radio channel consists of eight time slots. If no guard band is assumed,

a) Find the number of simultaneous users that can be accommodated in GSM.
b) How many users can be accommodated if a guard band of 100 kHz is provided at the upper and the lower end of the GSM spectrum? (Refer Section 1.8.2.)

Solution

a) The number of simultaneous users that can be accommodated in GSM in the first case is equal to
b) In the second case the number of simultaneous users = 992.

Example problem 1.4

The 2G cellular system GSM 900 operates its 125 forward channels in the uplink frequency band 890.2–915 MHz and 125 reverse channels in the frequency band 935.2–960 MHz. Each channel has a bandwidth of 200 kHz.

a) What is the bandwidth in the forward channels (uplink frequency band) and reverse channels (downlink frequency band)?
b) If each channel is subdivided into 16 time slots, what is the sub channel spacing?
c) If each channel is shared by 16 users then compute the total number of users per cell? (Refer Section 1.8.2.)

Solution

a) Number of forward channels = Number of reverse channels = 125
Total number of channels = 125 + 125 = 250
Bandwidth of each channel allocated = 200 kHz
Bandwidth of uplink = bandwidth of downlink =
Number of channels × bandwidth of each channel = 125 × 200 kHz = 25 MHz

b) Number of time slots in each channel = 16
Sub channel spacing = channel space/time slots in each channel
= 200 kHz/16 = 12.5 kHz

c) Number of users shared in each channel =16
Total number of users per cell = 125 channels × 16 = 2000 users

1.8.2.2 CDMA or (IS-95)

The use of CDMA technology started in the United States in the year 1990. IS-95 is a standard for CDMAone digital cellular where as CDMA2000 is a 3G specification (the North American version of wideband CDMA), and is backward compatible with IS-95 systems. In India, Reliance Communications (RCOM), BSNL, and Vodafone (formerly Hutch) serve more than 39.4 million subscribers using CDMA technology.

CDMA is a unique access technology that separates subscriber calls from one another by a pseudorandom noise (PRN) code instead of time or frequency. As a result, all available CDMA frequencies can be used in every cell, thereby increasing the total number of available voice channels and the overall system capacity.

CDMA is a wideband, spread-spectrum technology in which we allocate a unique code for every user separately and allocate bandwidth to the user. Today, each CDMA carrier can support around 22 voice calls. 3G CDMA2000 systems may deploy more carriers per BS, possibly six to eight carriers to accommodate the additional bandwidth requirements. CDMA networks have *pilot channels*, which carry no data but are used by the mobile phone to acquire the system and assist in the process of SHOs and synchronization. Table 1.2 summarizes the differences among the basic analogue (AMPS) and digital (GSM and CDMA) cellular systems that have been used. It gives an overview of the different mobile phone systems or cellular technologies that are in use today and those that have been used over the years. Although not every cellular technology is included, those that have been more widely used are included.

Channel Spacing in CDMA systems

Channel spacing refers to the actual bandwidth space that is allocated for every wireless channel out of the total amount of spectrum allocated to a wireless carrier. In AMPS, the channel spacing is 30 kHz. Each uplink and downlink channel occupies 30 kHz of bandwidth. Every AMPS cellular call actually occupies a total of 60 kHz. GSM systems allot their radio spectrum in 200 kHz carriers, where each carrier allocates 25 kHz to uplink or downlink. CDMA, by definition, is unique when it comes to channel spacing. In the most technical, literal sense, channel spacing in a CDMA system is 1.25 MHz because all calls that are carried on a 1.25 MHz CDMA carrier are spread out over the entire swath of that carrier. That is why CDMA is known as a spread spectrum technology.

Table 1.2 Comparison of AMPS, GSM, and CDMA standards

Name	AMPS	GSM	IS-95 CDMA
Generation	1	2	2
Year introduced and origin place	1983 US	1992/1994 Germany	1993 US
Frequency band (MHz) Uplink (BS receiving) Down link (BS transmission)	824–849 869–894	890–915/1850–1910 935–960/1930–1990	824–849/1850–1910 869–894/1930–1990
Multiple access scheme	FDMA	TDMA,FDMA	CDMA
Bandwidth/channel	30 kHz	200 kHz	1.25 MHz
Modulation type	FM	GMSK	QPSK & OQPSK
Number of channels/carrier	1	8	85
Total channels	832	1000	1700

Details of Forward and Reverse CDMA Channels

The CDMA standard details are illustrated in Table 1.2. It uses two frequency bands (800 and 1900 MHz bands) for duplex communication. These bands are designated as ISM 800 MHz band or the ISM 1900 MHz band. Each band (25 MHz) is divided into 20 channels of 1.228 MHz each separated by guard bands. Each service provider is allotted 10 channels. The forward and reverse CDMA channels used in the ISM 800 MHz frequency band are 824–849/1850–1910 MHz for uplink/downlink respectively. Similarly, the forward and reverse CDMA channels used in the ISM 1900MHz frequency band are 869–894/1930–1990 for uplink/downlink, respectively.

A more detailed description about the analogue and digital cellular systems is given in **Chapter 21**. All these methods rely on a distributed network of cell sites using frequency reuse concept.

1.9 Existing mobile communication technologies and current status

The following are some examples of mobile communication systems currently in use in addition to the cellular radio networks.

1.9.1 Paging

Radio paging is a low-cost service used for sending information to a person on move. Paging system consists of a BS that broadcasts the paging messages in the form of numeric digits (e.g. a telephone number), alphanumeric text messages or even voice messages to a service area. From the radio paging BS the transmission of data is one-way. The BS is interfaced to a paging controller (Fig. 1.21). The message is composed by an operator who receives the message from a user through a telephone. The message is broadcast through the BS using transmission medium. If the service area is very large, this system requires high-powered transmitters and low data rates for maximum coverage of each transmitter's designated area.

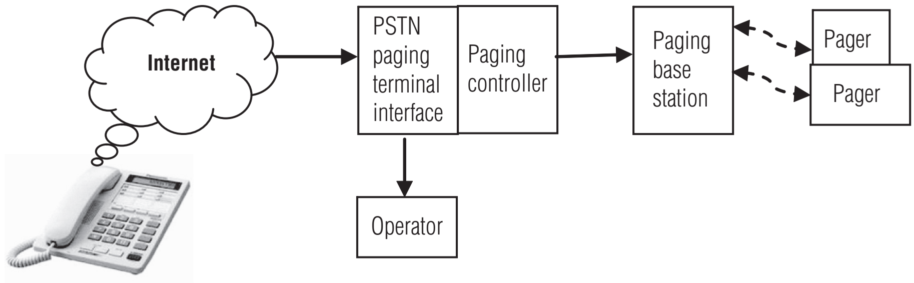


Figure 1.21 Paging system

The paging standard started in 1G and continued up to 2G systems.

Various pagers in use are *numeric pager* which displays only numeric data (such as a telephone number), *alphanumeric pager* which displays alphanumeric text, and *voice pager* that supports voice paging.

1.9.2 Communication satellites

A communication satellite provides long distance wireless communication and is similar to LOS microwave transmission in which one of the stations is a satellite orbiting the earth. The satellite acts like a tall antenna and repeater system (Fig. 1.22). Satellite system consists of number of transponders which allow radio, television, and telephone transmissions to be sent live to anywhere in the world. Satellite systems can make services available to airborne and sea-based users where cellular mobile communication provides service to world's land mass only.

Satellite systems are also playing a crucial role in *mobile communications* by providing coverage in zones where land-based infrastructures are unable or ineffective to supply mobile services. Mobile satellite communications began in 1976 with the launch by COMSAT of the MARISAT satellites to provide communications to ships at sea. Mobile satellite communication systems are used for transmitting point-to-point voice and data communications using a constellation of satellites. These systems include a number of user terminals, several terrestrial ground stations or gateways and a number of satellites for bi-directionally coupling the user terminals to terrestrial telecommunication networks and the Internet via the gateways.

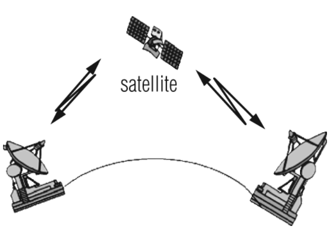


Figure 1.22 Point to point communication through satellite

Mobile satellite communication system, for example, Iridium low earth orbit (LEO) satellite system with 66 satellites, Globalstar satellites, and Ellipso mobile satellite system.

The main drawback of satellite communication systems is that they have quite a large propagation delay due to the distances travelled by radio waves.

1.9.3 Wireless local loop (WLL)

Wireless local loop (WLL) is a cellular-like phone without mobility. These are designed for fixed communications in situations where it is easier, cheaper, or more advantageous than wire line connections and are often based on cellular or cordless technologies. WLL employs a cellular like technology where the subscriber unit is fixed, like a wire line telephone, replacing the "local loop" between the exchange and the subscriber's home with a wireless link. They use the same basic architecture and principles of radio transmission. There are a number of problems with WLL when compared to the landline services. They are as follows:

- Voice quality is typically inferior.
- Data rates provided are typically lower.
- Call costs are typically more expensive.

1.9.4 Personal handy phone

Personal handy-phone system (PHS) has been in existence in Japan for years and operates in the 1880–1920 MHz spectrum. PHS is similar to cellular networks. However, PHS phones can also communicate directly with one another when in range. This is an advantage over cellular phones, which can only communicate with one another via BS transceivers. This system is very popular in heavily populated metropolitan areas.

Figure 1.23 Mobile radio

A PHS device functions as a cordless phone in the home and as a mobile phone elsewhere.

1.9.5 Mobile radio

Mobile radio is a half-duplex analogue communication system that uses single frequencies for sending and receiving signals. In mobile radio system, a button must be pressed to switch modes (Fig. 1.23). They are used for emergency services such as police control rooms and the security industry.

Figure 1.24(a) Cordless phone

1.9.6 Cordless phones

These are designed for indoor and office applications. The initial application of cordless phones (Fig. 1.24(a)) was as an extension to the fixed line in a home. The cordless BS is plugged into the home phone socket and provides a radio link to the handset and does not transfer well to outdoor, cellular-type systems. They are typically much simpler than cellular phones using lower power transmission and higher bit-rate speech coders.

1.9.7 DECT

Digital enhanced cordless telecommunications (DECT) (Fig. 1.24(b)) is a multicarrier/TDMA/TDD radio access system standard for cordless communications in residential, corporate, and public environments. DECT system consists of three components: (i) phone network, (ii) a BS, which is connected to the phone line socket and (iii) multiple number of mobile handsets. These mobile handsets are used as normal "cordless" phones. DECT system allows several cordless telephones to communicate with each other (internal) and to external network. The main advantage is that the additional handsets do not require additional telephone sockets or additional transceivers.

DECT frequency band: DECT operates in the 1880–1900 MHz band and defines ten channels from 1881.792 MHz to 1897.344 MHz with a band gap of 1728 kHz. Each BS frame provides 12 duplex speech channels with each time slot occupying any of channels.

Figure 1.24(b) A base station and multiple handsets in

The following are the advantages of DECT:

i) *Security*: All communications between base and handset are encrypted with a set of keys which are renewed for each call.
ii) *Long range without interference*: DECT range is at least 300 m in free space (typical 450 m), and reaches easily 50 m with an in-door base. It does not interfere with the 2.4 GHz WiFi band.
iii) *Low power consumption*: The DECT technology is simple and efficient. BSs use typically less than 3W of operating power, and handsets operate with two AAA rechargeable batteries.
iv) *High efficiency of spectrum use*: It can manage without any frequency planning a full array of bases every 10 m. No risk of interference, even if all neighbours are using DECT.
v) *GSM/DECT internetworking*: Part of the DECT standard describes how it can interact with the GSM standard so that users can be free to move with a telephone from the outdoors (and GSM signals) into an indoor environment (and a DECT system). It is expected that many GSM service providers may want to extend their service to support DECT signals inside buildings. A dual-mode phone would automatically search first for a DECT connection, then for a GSM connection if DECT is not available.
vi) *Home cordless phones*: A home owner could install a single-cell antenna within the home and use it for a number of cordless phones throughout the home and garden.

1.9.8 Bluetooth

Bluetooth wireless technology is an open specification for a low-cost, low-power, short-range radio technology for ad hoc wireless communication of voice and data anywhere in the world. It was developed as a new cable replacement technology, which provides a short-range (<10 m), low bit rate (<1 Mbps) access in the 2.4 GHz spectrum. It is also known as wireless personal area network (WPAN) for short-range and low mobility applications around a room in the office or at home. A Bluetooth WPAN involves up to eight devices, located within a 10 m radius personal operating space, that unite to exchange information or share services (Fig. 1.24(c)). Because connectivity can be done spontaneously according to immediate need, Bluetooth is also known as ad hoc networking. Because a WPAN involves directly networking between different points, without the use of network infrastructure, it is also referred to as a "point-to-point network".

Laptop to mobile phone connectivity: A mobile phone can communicate with a computer through a phone's vendor-specific cable, infrared, or Bluetooth. The latter two depend on both devices' hardware capabilities. Figure 1.24(d) presents a typical scheme for a laptop connecting to a server through the mobile phone and the cellular network. For laptop to mobile phone connectivity, the Bluetooth solution was selected for the following reasons: (i) compared to infrared, it does not have the restriction of having the phone being in the LOS of the laptop's sensor, (ii) compared to cable, it does not require the phone to be physically attached to the laptop and restricted to the cable's length and (iii) with Bluetooth, the phone only needs to be in a range of a few metres (<10) from the laptop.

Bluetooth uses a technique called spread-spectrum frequency hopping.

The important specifications of Bluetooth technology are illustrated in Table 1.3.

Advantages of Bluetooth technology are as follows:

- Wireless (no Cables)
- No setup needed
- Devices can be movable
- Industry wide support
- Easier synchronization due to omni-directional and no LOS requirement

Disadvantages of Bluetooth technology are as follows:

- Short range wireless radio technology operates in the range of 10 m
- Small throughput rates – data rate 1.0 Mbps
- mostly for personal use (PANs)
- fairly expensive

A more detailed description about Bluetooth technology is given in Chapter 27.

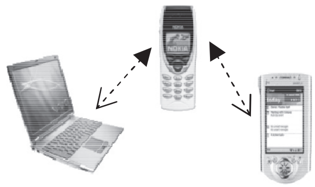


Figure 1.24(c) Bluetooth communication between portable devices

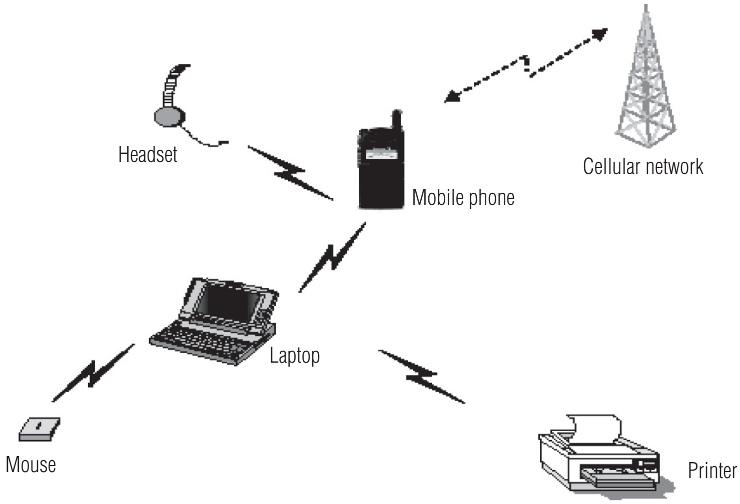


Figure 1.24(d) Typical scheme for a laptop connecting to a server through the mobile phone and the cellular network.

Table 1.3 Bluetooth technology specifications

Connection type	Spread spectrum (Frequency hopping)
Multiple access control scheme	FH-CDMA
Spectrum	2.4 GHz ISM
Modulation	Gaussian frequency shift keying
Transmission power	1 mw–100 mw
Aggregate data rate	1 Mbps
Range	30 ft
Supported stations	8 devices
Voice channels	3
Data security- Authentication key	128 bit key
Data security-encryption key	8–128 bits (configurable)

1.9.9 Current status of cellular radio

Tremendous changes are occurring in the area of wireless communications. With the rising demand for mobile communications, 3G systems have emerged, providing higher date rate to facilitate new multimedia applications such as video telephony and wireless Internet access. There are three primary standards that comprise 3G technology: W-CDMA, CDMA2000, and TD-CDMA. The mobile phone of yesterday is rapidly turning into a sophisticated mobile device capable of more applications than PCs. For example, the data rates provided by 3G networks enable a user to enjoy wireless access to the Internet at speeds up to 1.8 Mbps. Further enhancements in high speed downlink packet access (HSDPA) modulation schemes will soon increase this speed to greater than 10 Mbps.

Presently, we see 3G cell phones hitting the markets. In China, the 3G service is already in existence. The 3G has also reached India recently. The existence of several diverse 3G standards limits seamless *global roaming* between different cellular networks for a mobile user with a single handset. In addition, there is a fundamental difference between wireless cellular networks (1G, 2G, or 3G) and wireless data networks such as WLANs and PANs. The difference is that wireless cellular systems are *circuit-switched* while wireless data networks are *packet-switched*. Convergence issues for these differences between the wireless cellular systems and the wireless data networks will be addressed in the design of 4G cellular networks. It is projected that 4G networks will provide users with seamless wireless access to voice, data, and video services irrespective of which wireless network they belong to. A 4G system will be able to provide a comprehensive IP solution where voice, data, and streamed multimedia can be given to users on an “anytime, anywhere” basis, and at higher data rates than previous generations. 4G will be capable of providing data rates between 100 Mbps and 1 Gbps both indoors and outdoors.

As the 2G was a total replacement of the 1G networks and handsets and the 3G was a total replacement of 2G networks and handsets, in the same way the 4G is also a complete replacement of the current 3G networks and handsets. The ITU regulatory and standardization bodies are working for commercial deployment of 4G networks roughly in the 2012–2015 time scale.

1.10 Summary

The advent of the cellular concept is a crucial contribution in the development of mobile communication. A cell is viewed as the coverage area of a particular land site. Each cell within a cluster is allocated a distinct set of frequencies (channels).

These systems were called “cellular” because large coverage areas were split into smaller areas or “cells”. Each cell is served by a low power transmitter and receiver. The important points introduced in this chapter are summarized below.

- *Cellular standards and evolution of mobile from 1G–4G*: The service providers must use one of the approved cellular standards for developing the cellular network in that region. These standards are mutually agreed upon rules adopted by the industry on how the cell phone system operates. These standards describe the air interface, that is how cell phones and BSs must communicate with each other. These mutually agreed upon standards change over time, as technology progresses.
    - Before 1G, 0G refers to pre-cellular mobile phone technology.
    - ***1G*** was an analogue technology. In 1G, speech was converted to an FM signal and transmitted back and forth from user phones.
    - ***2G*** cellular networks used digital technology and provided enhanced services (e.g. messaging, caller-id, etc.). There were two 2G standards that service providers could choose between GSM based on TDMA and IS-95 based on CDMA.
    - Most of the present cellular systems are **2.5G**. They offer enhanced services over second generation systems (e.g. e-mailing, web-browsing, etc.).
    - Presently, service providers are setting up **3G** cellular systems. The idea behind 3G is to have a single network standard instead of the different types adopted in the United States, Europe, and Asia. 3G offers higher data rates than 2.5G. 3G allows users to send/receive pictures, video clips, and so on.
    - ***4G*** is the evolution based on 3G’s limitations and it will fulfil the idea of **WWWW** (World Wide Wireless Web) offering more services and smooth global roaming with inexpensive cost.
- *Frequency reuse*: This is the process in which the same set of frequencies (channels) can be allocated to more than one cell to increase the system capacity. Instead of deploying a powerful BS in a large coverage area, the area is divided into multiple smaller cells and a BS deployed in each cell can use smaller transmit power. Thus, two transmissions can employ the same frequency if they are far away enough such that the CCI level is below a desired threshold.
- *Handoff:* A crucial component of the cellular concept is the notion of handoffs. Handoff is the process of transferring an active call from one cell to another as the mobile unit moves from one cell to the other. Handoff operation involves identifying a new BS and allocation of voice and control signals associated.
- *Multiple access techniques:* In many wireless systems, multiple transmitters attempt to communicate with the same receiver. For example, in cellular systems, cell phone users in a local area typically communicate with the same cell tower. The limited spectrum has to be shared among these local transmitters. In such cases, the system adopts a multiple access policy. Three widely used policies are FDMA, TDMA, and CDMA.

Review questions

1. Why does the mobile phone cell – the basic geographic unit of cellular system – have a hexagonal shape?
2. Describe the principle of operation of cellular mobile system and explain the "cellular" concept with a neat diagram.
3. What is meant by 1G, 2G, 2.5G, 3G, and 4G cellular systems?
4. Name the wireless access techniques used in 1G, 2G, and 3G wireless systems.
5. Describe the analogue and digital cellular land mobile systems and the limitations of AMPS standard.
6. Compare the basic technological differences between the GSM and CDMA standards.
7. What is the fundamental difference between wireless cellular networks (1G, 2G, or 3G) and wireless data networks?
8. List the main features of 3G systems.
9. Explain neatly how the 4G technology is projected to provide users with seamless wireless access to voice, data, and video services irrespective of which wireless network they belong to.
10. What are the various mobile phone technologies that are existing in addition to the cellular networks?
11. Describe the functions of base station in a cellular network.
12. Mention the various components of a cellular network and describe them briefly.
13. What are the various multiple access and duplexing schemes used in cellular networks?
14. What are the applications of a satellite system?
15. What is a page and give the benefits of a page system?
16. What are the various types of mobile radio transmission systems in use?
17. State the reason why 800 MHz is selected for mobiles initially.
18. The 2G system GSM 900 has 125 channels in the uplink and 125 reverse channels in the downlink. Each channel has a bandwidth of 200 kHz.
    a) What is the total bandwidth occupied in both the uplink and downlink communications. (Ans: 50 MHz)
    b) If each channel is subdivided into 16 time slots, what is the sub-channel spacing? (Ans:16.66 kHz)
    c) If each channel is shared by 16 users then compute the total number of users per cell? (Ans: 1,500)
19. Describe the features of AMPS and compare with those of present-day systems. (Refer Section 1.8)
20. Briefly explain the evaluation of the analogue and digital cellular mobile systems. (Refer Section 1.8)
21. Differentiate between the generations in the cordless phones and cellular phones. (Refer Sections 1.9)
22. A total of 33 MHz of BW is allocated to a particular FDD which uses two 25 kHz simplex channels to provide full duplex voice and control channels. Compute the number of channels available per cell, if a system uses (a) 4-cell reuse, (b) 7-cell reuse, and (c) 12-cell reuse. If 1 MHz of the allocated spectrum is dedicated to control channels, determine an equitable distribution of control and voice channels in each cell for each of the three systems. (Refer Section 1.6) [Ans: (a) 165 (b) 95 (c) 55]
23. Explain the history of 800-MHz spectrum allocation to a cellular system. (Refer Section 1.8.2)

24. Differentiate between the analogue and digital cellular systems with their operating capacities. **(**Refer Section 1.8**)**
25. Explain the operation of a cellular system in detail (Refer Section 1.4).

Objective type questions and answers

1. The following concept refers to the use of radio channels on the same carrier frequency to cover different areas that are separated from one another by sufficient distances.
   (a) frequency reuse (b) handoff (c) cell splitting (d) cell geometry
2. The process of transferring an active call from one cell to another as the mobile unit moves from the first cell to the other cell without disconnecting the call.
   (a) frequency reuse (b) handoff (c) cell splitting (d) cell geometry
3. The interference received from co-channel cells is called
   (a) co-channel interference (b) frequency reuse
   (c) handoff (d) cell splitting
4. The actual radio coverage of a cell is known as ____________ and is determined from the field measurements
   (a) footprint (b) cluster (c) capacity (d) channel
5. The following method is used as the access technology for global system for mobile (GSM) communications
   (a) FDMA (b) TDMA (c) CDMA (d) SDMA
6. The mobile technology using general packet radio service (GPRS) standard has been termed as
   (a) 1G (b) 2G (c) 3G (d) 2.5G
7. The cellular radio system that was mainly designed for digitized voice.
   (a) 1G (b) 2G (c) 3G (d) 2.5G
8. The mobile phone system that was analogue and it only carried voice traffic
   (a) 1G (b) 2G (c) 3G (d) 2.5G
9. The capacity of a cellular system is directly proportional to
   (a) number of times a cluster is replicated in a fixed service area
   (b) number of cells in the cluster
   (c) Number of channels in the cell
   (d) none of the above
10. The systems were “cellular” because coverage areas were split into ____________ each of which is served by a low power transmitter and receiver.
    (a) cells (b) cluster (c) capacity (d) channels
11. Larger cells are more useful in ____________.
    (a) densely populated urban areas (b) rural areas
    (c) lightly populated urban areas (d) mountainous areas
12. The following cellular technology will bring almost perfect real world wireless or called “WWWW: World Wide Wireless Web”
    (a) 1G (b) 2G (c) 3G (d) 4G
13. The two 3G systems developed are
    (a) GSM, CDMA (b) AMPS, CDMA
    (c) UMTS, IMT-2000 (d) CDMAone, UMTS

Answers: 1. (a) 2. (b) 3. (a) 4. (a) 5. (b) 6. (d) 7. (b) 8. (b) 9. (a) 10. (a) 11. (b) 12. (d) 13. (c)

Open book questions

1. What are the limitations of conventional mobile telephone system and what are the classifications of wireless communication systems?
2. Distinguish between 1G and 2G cellular networks. Define a cell and cluster.
3. What are the channels used in mobile communication systems?
4. What are the basic units of a cellular system?
5. What is cell splitting?
6. Explain the cellular concept neatly in terms of frequency reuse (refer Section 1.4).
7. What are the main advantages and disadvantages of cell structures?
8. Describe various types of mobile radio transmission systems in use with examples.
9. Write the features of DECT.
10. What are the limitations of conventional mobile systems?
11. Explain inefficient spectrum utilization based on the existing mobile systems MTS and IMTS.
12. Draw the schematic and present the working of the cellular system.
13. Briefly explain cell shape and handoff.

Further reading

Behrouz A. Forouzan, *Data Communications and Networking*, 4th edition, New Delhi: Tata McGraw-Hill Publishing Company Ltd., 2006.

Couch, L.W. "Complex Envelope Representations for Modulated Signals", in Suthan S. Suthersan (ed.) *Mobile Communications Handbook*, Boca Raton: CRC Press LLC, 1999.

Dinan, E. H., and Jabbari, B. Spreading codes for direct sequence CDMA and wide band CDMA cellular networks. *IEEE Communications Magazine*, September 1998.

Garg, V. K., and Wilkes, J. E. *Wireless and Personal Communications Systems*. Upper Saddle River, NJ: Prentice Hall, 1996.

Garg, V. K., and Wilkes, J. E. *Principles and Applications of GSM*. Upper Saddle River, NJ: Prentice Hall, 1998.

Husni Hammuda, *Cellular Mobile Radio Systems*, England: John Wiley & Sons, 1997.

Negus, K., Stephens, A., and Lansford, J. Home RF: Wireless networking for the connected home. *IEEE Personal Communications*, February 2000, pp. 20–27.

Nikula, E., Toshala, A., Dahlman, E., Girard, L., and Klein, A. FRAMES multiple access for UMTS and IMT-2000. *IEEE Personal Communications Magazine*, April 1998, pp. 16–24.

Pirhonen, R., Rautava, T., and Pentinen, J. TDMA convergence for packet data services. *IEEE Personal Communications Magazine*, vol. 6, no. 3, June 1999, pp. 68–73.

Rapeli, J. UMTS: Targets, system concepts, and standardization in a global framework. *IEEE Personal Communications*, February 1995.

Report from UMTS forum, The European path towards UMTS, *IEEE Personal Communications Magazine*. Special Issue, Vol. 2, No. 1, February 1995.

Cellular Geometry, Frequency Reuse, Cell Splitting, and Sectoring

2

2.1 Introduction

Today, around the globe, billions of subscribers are using mobile phones and this number is increasing rapidly. Therefore, mobile communications need to offer efficiency in the use of the available frequency spectrum without any mutual interference. For example, each channel in the analogue cellular system needs 25 kHz bandwidth that includes guard bands between the adjacent channels to ensure no co-channel interference and also to offer sufficient voice quality. If the available frequency spectrum is 1 MHz wide, then we can accommodate only 40 users. Even if the frequency spectrum allocation to the system increases to 100 MHz, this would accommodate only 4,000 users to have access to the system. Therefore,

> *the main objective of cellular systems design is to handle as many calls as possible (called capacity in cellular terminology) in a given bandwidth in the most efficient way with reliability and quality of service in telephony.*

To achieve this objective, the cellular system employed two crucial features known as *frequency reuse* and *cell splitting*.

Frequency reuse refers to the usage of the same frequency carrier in different geographical locations that are distant enough so that the interference caused by using the same carrier is not a problem. The reason for the application of frequency reuse is to increase the number of simultaneous calls.

Cell splitting refers to the reconfiguration of a cell into smaller cells. This allows the system to adjust to an increase in the traffic demand in certain areas or in the whole network without any increase in the spectrum allocation.

In mobile communications, we talk in terms of cells that represent a small geographic area which has resulted in "cellular" technology that is popular nowadays. The users are called as *mobile stations* (MSs) to transmit/receive calls while moving in the cellular network. Each cell has a *base station* (BS) that supplies frequency channels to MSs. BSs are also referred to as cell sites. These cell sites are linked to a *mobile switching centre* (MSC) which is responsible for controlling the calls and acting as a gateway to other networks. An illustration of the general mobile system architecture is given in Figure 2.1.

In this chapter, the cellular concept is discussed in detail with emphasis on cellular geometry, frequency reuse, cell splitting, and sectoring. In the succeeding section, however, we discuss the different cell shapes and explain why the hexagonal shape is preferred over other geometries.

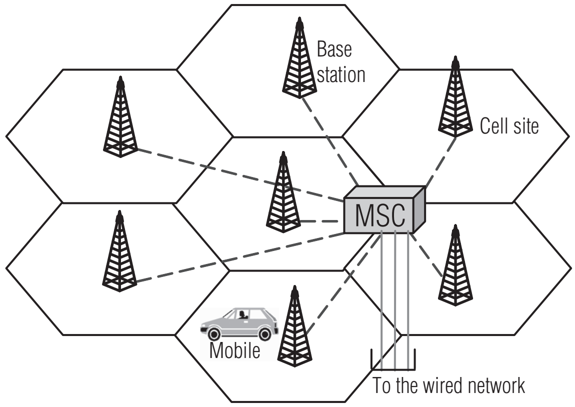


Figure 2.1 Mobile cellular system architecture

2.2 Cellular geometry

The geographical areas covered by cellular radio antennas are called *cells*. The cell site antenna is located at a point within the cell.

Cells are assumed to have a regular hexagonal shape.

Using this shape, let us picturize the cellular idea that approximates the covered area on a map. But why is a hexagon shape used to represent the cells? Why not a circle or a square or a triangle? Figure 2.2 is an illustration of the different cell shapes.

To justify the use of the hexagonal shape, we now analyse the different cellular geometries to approximate a given geographical area. When showing a cellular system, we want to depict an area totally covered by radio waves without any gaps. Any cellular system will have gaps in coverage, except the hexagonal shape. Let us now try to visualize, how the system is laid out. One can observe that the circular cells below leave gaps in the layout shown in Figure 2.3. On the other hand, the hexagonal geometry would not leave such gaps as far as the theoretical visualization of the layout is concerned.

Let us consider the simplest case of assuming that mobile subscribers are uniformly distributed throughout the concerned area. Our design objectives necessitate that the cellular system to be capable of growth in both coverage area and user capacity. To increase the user capacity within the coverage area, the system must have the ability to add cells. It must also have the ability to add cells to the peripheral area and to expand the service area. It is certainly possible to layout the cells and assign channels in a way that is suited for a given geographic area and customer distribution. But, while adding new cells for the growth would then require custom re-engineering. Such a design could easily result in a need to relocate BSs and redistribute

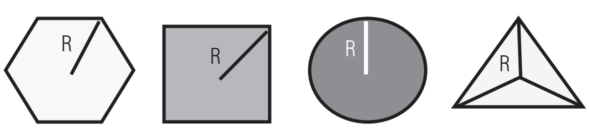


Figure 2.2 Different cell shapes

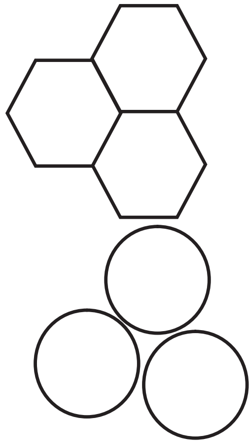


Figure 2.3 Gaps in the circular cells

cells. Construction of a BS is costly and in addition to the cost of erecting a tower, it may involve the purchase and approval of land from concerned authorities. Hence, the capability to expand must be incorporated while designing the cellular system right from the beginning. We will now investigate the extent to such a regular layout is possible and the principles on which the cell layout is based.

2.2.1 Circular geometry

To cover a specified service area, placement of cells require some assumption about the shape of the cells. *Based on the principle that equal-level signal contours surrounding a transmitting antenna are circles with the BS at the origin one can assume that these cells are circular in shape.*

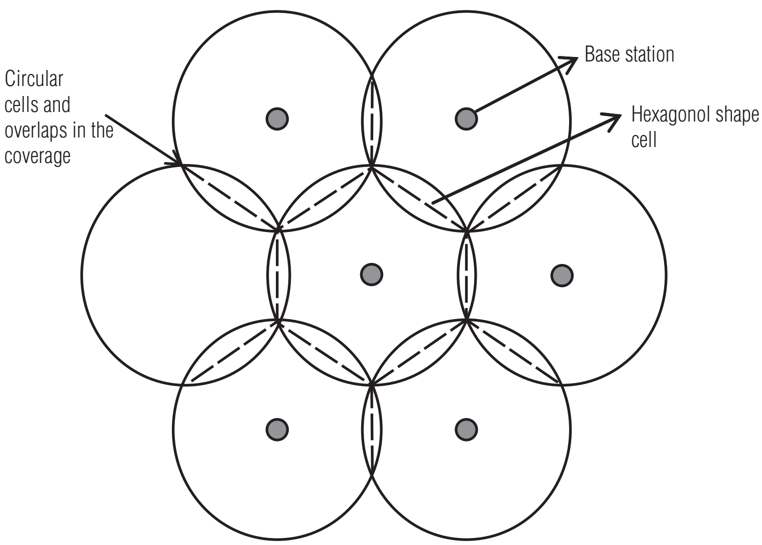


Figure 2.4 Ideal cells formation

Our approach is to initially assume that the BS coverage areas are circular and to focus on developing an efficient method for locating and growing cells that take care of the terrain and location of surrounding obstacles. However, in our real world it is unlikely that the coverage areas will be circular because with these circular areas it is not possible to uniformly cover a geographic area without having gaps or overlaps in the coverage (Fig. 2.4). Hence, it is required to determine a layout pattern that has no gaps and minimum overlaps.

2.2.2 Polygonal geometry

Figure 2.5 shows two overlapping circular coverage regions (cells). In the overlap region, there is a straight-line boundary that divides the locations for which the signal from BS_1 is stronger than the signal from BS_2 and vice versa. Consequently, if we cover our area with overlapping circular coverage regions arranged on a regular grid, the actual cells will be polygons.

There are three regular polygonal shapes (i.e. equilateral triangle, square, and hexagon) for which it is possible to completely cover an area without overlaps or gaps (Figure 2.6).

Figure 2.7 shows the coverage area pattern for each of these polygons and also shows how a polygon can be inscribed in a circle of radius *R*, the cell (coverage) radius. Now we can observe that within the circle of radius *R*, the area covered by the hexagonal pattern is the largest.

In Figure 2.8, different ways of circular overlapping areas are shown. The pattern in Figure 2.8(a) is resulting in a triangular grid while the pattern in Figure 2.8(b) is resulting in a rectangular

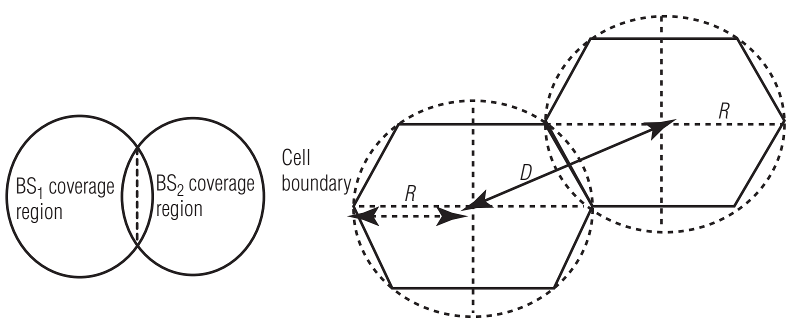


Figure 2.5 Boundary between two circular coverage regions (cells)

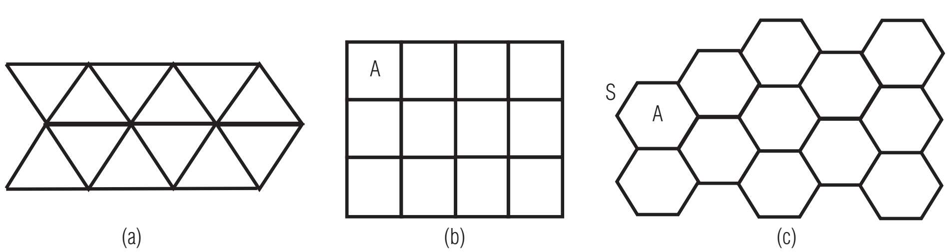


Figure 2.6 Covering a plane area with regular polygons (a) Equilateral triangles; (b) Squares; (c) Hexagons

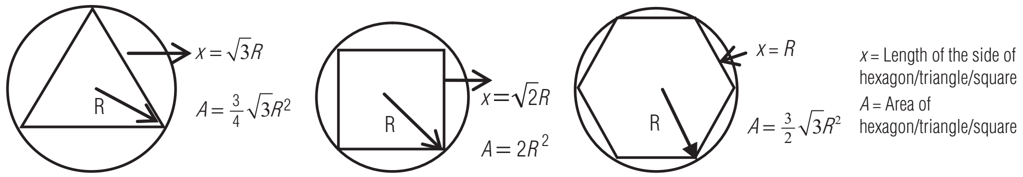


Figure 2.7 Coverage area within a regular polygon (a) Equilateral triangle; (b) Square; (c) Hexagon

grid. However, one can observe that the overlap is getting reduced when circular cells are deployed along a hexagonal grid as shown in Figure 2.8(c).

For these reasons, a hexagonal cell layout is chosen as the basis for designing cellular systems.

The hexagonal layout is the most economically efficient as it requires the least number of cells to cover a given area.

From now on, in every discussion that follows, we will treat cells as *having hexagonal shapes*.

Although this is never precisely true in fact as shown in Figure 2.9, the assumption provides a means for developing concepts about frequency reuse and cell size that may be applied in practice.

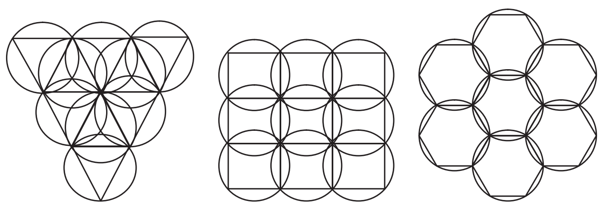


Figure 2.8 Overlap in circular cells using a (a) Triangular grid; (b) Rectangular grid; (c) Hexagonal grid

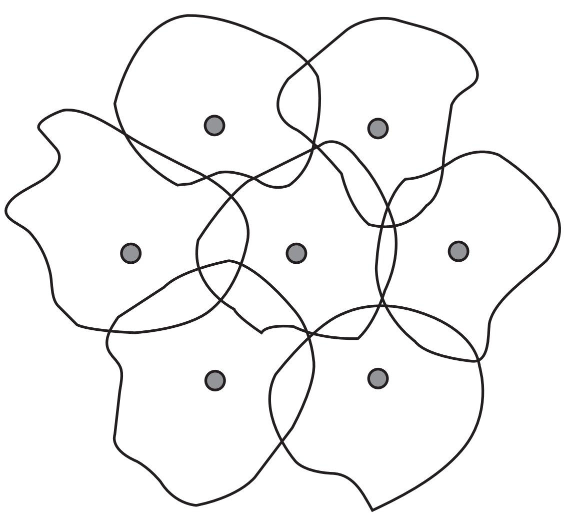


Figure 2.9 Real cells formation

Further, the hexagonal layout does provide a starting point for real-world design. We proceed to discuss the geometrical properties of a hexagonal grid as they relate to cell layout and frequency reuse.

2.2.3 Location of antenna to cover cellular region

Wherever the density of calls is more, the cells have to be split into individual areas to make them more efficient and to let them carry more calls. Sectorized antennas are used to cover these split areas by replacing the omni-directional antenna at the BS with several directional antennas. These antennas reduce the co-channel interference in a cellular mobile communication system. Antennas from other cell sites cover their own sectors. The covered area resembles a sort of *rhomboid*, as is evident from Figure 2.10. Specifically, Figure 2.11 shows the positioning of the antenna.

The cell site equipment provides each sector with its own set of channels. Each cell site transmits and receives on three different sets of channels, one for each part or sector of the three cells it covers.

One may visualize the cell as the hexagon bounded by solid lines, with the tower in the centre, with the antenna pointing in the directions indicated by the arrows. In reality, the grey-shaded hexagon is the cell with the towers at the corners, as depicted in Figure 2.10. In this book, the terms *cell* (the coverage area) and *cell site* (the BS location) are used interchangeably, although they are not the same.

In Figure 2.12, at the top of a cell phone tower, there are vertical arrays that constitute the antenna, usually with the form of long vertical rectangles or boxes, sometimes arranged in parallel groups. The vertical arrays are separated 120° from each other. This means that each array is pointing to the centre of its hexagonal cell. Therefore, each cell is covered by three arrays from three separate towers, one at the apex of each of its three angles.

2.3 Frequency reuse

The total number of frequencies/channels available in a cellular system are allocated to each cell by means of frequency reuse technique so as to minimize the co-channel and adjacent channel interference while meeting the performance requirements both in terms of received call quality

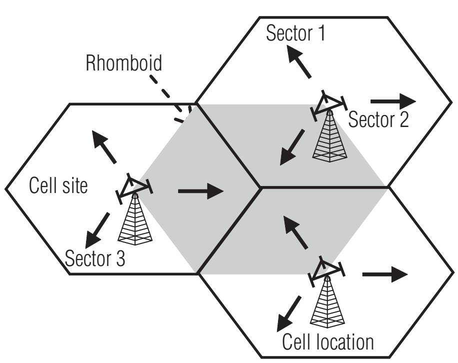


Figure 2.10 Cell shapes and coverage shape

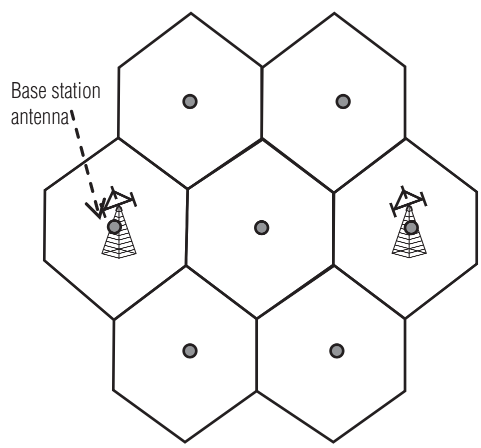


Figure 2.11 Antenna positioning

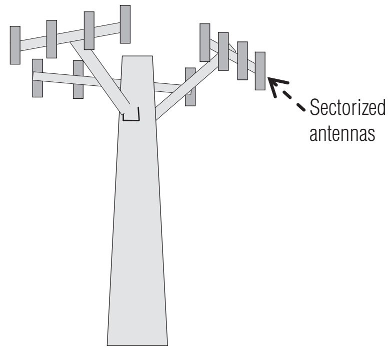


Figure 2.12 Cell tower

as well as traffic capacity in these cells. The distribution of the frequency channels in a cellular network is dependent on several parameters such as cellular geometry, signal interference, and signal propagation characteristics.

The assignment of frequency channels in the cellular concept is fixed, that is, a set of frequency channels is statically allotted to a cell. The same set is reused in another cell distant enough to allow the use of the frequency channels with acceptable signal interference.

Cells that use the same set of frequency channels are called ***co-channel cells*** and the distance between them is called ***co-channel reuse distance***. The total number of frequency carriers allotted to a network operator is divided into sets. Each set is assigned to a cell inside a cluster of cells. This cluster of cells forms a pattern. The pattern is reused according to the co-channel reuse distance. The choice of the number of cells per cluster is mainly governed by co-channel interference considerations. The frequency reuse concept is explained in the following section.

2.3.1 Cellular sysyem capacity and frequency reuse for a cluster size of "N" with each cell allocated a group of "K" channels

For a better understanding of the frequency reuse concept, consider a cellular system with a total of S duplex channels available for use in a cluster. If each cell is allocated a set of k channels ($k < S$) and the cluster size is N, then the total number of available radio channels can be expressed as

$$S = k \times N \tag{2.1}$$

Each cluster uses the same number of channels. If a cluster is replicated M times within the system, then the total number of duplex channels, C, which is a measure of capacity of the cellular system is given by

$$\textit{Cellular sysyem capacity } C = M \times k \times N = M \times S \tag{2.2}$$

The following observations can be made using Equation (2.2).

The capacity of a cellular system is directly proportional to the number of times a cluster is replicated in a fixed area.

The factor N is called the cluster size and it is typically equal to 4, 7, or 12.

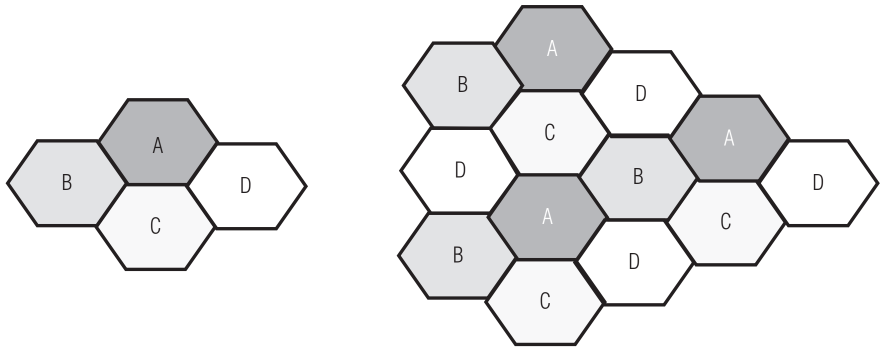


Figure 2.13 Reuse pattern of cluster size, $N = 4$

If the cluster size N is reduced keeping the cell size constant, it requires more clusters to cover a given area, hence more capacity is achieved. A larger cluster size N causes the ratio between the cell radius and the distance between co-channel cells to decrease, leading to weaker co-channel interference. Figure 2.13 shows the reuse pattern of a cluster of size 4.

For the same cell size at a given area, N decreases ⇒ M increases ⇒ C increases.

Example of frequency reuse for a cluster size of 7 in AMPS cellular system:

Cluster size of 7 means that the total voice channels (395 in AMPS) are divided into seven groups.

Thus, each cell has about 56 voice channels (approximately equal to 395/7). This is the number of users a cell can support, that is, roughly 10 square miles in normal environments.

This may or may not be sufficient based on the distribution of users.

To see how a system with cluster size 7 looks like, colour a cell with a particular colour.

This cell (if drawn as a hexagon) has six neighbours. Colour each of the seven neighbours using a different colour (different from each other). Now repeat this rule to get the overall "reuse pattern" as shown in Figure 2.14.

Example problem 2.1

We consider a cellular system in which the total available voice channels to handle the traffic are 1,200. The area of each cell is 9 km^2 and the total coverage area of the system is 3,600 km^2.

a. Calculate the *system capacity* if the cluster size, N is 4.
b. Calculate the system capacity if the cluster size is 7. Does decreasing the reuse factor N, increases the system capacity? Explain.
c. How many times should a cluster of size 7 be replicated to cover the entire cellular area?

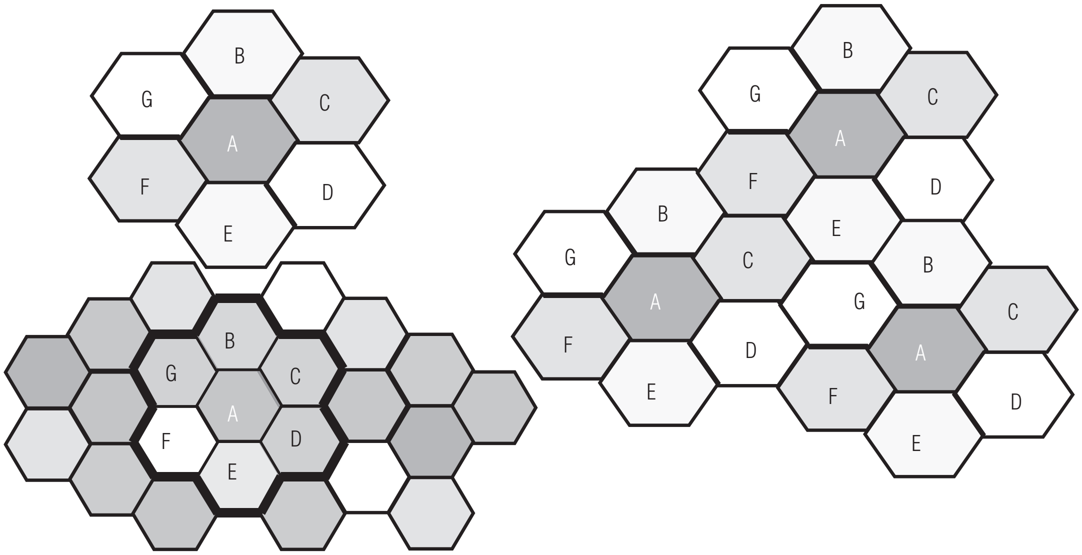


Figure 2.14 Reuse pattern of cluster size, $N = 7$

Solution

$$\text{Total available channels} = 1{,}200$$

$$\text{Cell area} = 9\ \text{km}^2$$

$$\text{Total coverage area} = 3{,}600\ \text{km}^2$$

a. $N = 4$

   Area of a cluster with cluster size $N = 4$ is $4 \times 9 = 36$ km²

   Number of clusters for covering total area with N equals $4 = 3{,}600/36 = 100$

   Number of channels per cell $= 1{,}200/4 = 300$

   System capacity $= 100 \times 1{,}200 = 120{,}000$ channels

b. $N = 7$

   Area of a cluster $N = 7$ is $7 \times 9 = 63$ km²

   Number of clusters for covering total area with N equals $7 = 3{,}600/63 = 57.14 \sim 57$

   Number of channels per cell $= 1{,}200/7 = 171.42 \sim 171$

   System capacity $= 57 \times 1{,}200 = 68{,}400$ channels

It is evident that when we decrease the value of N from 7 to 4, we increase the system capacity from 68,400 to 120,000 channels. Thus, decreasing the reuse factor (N) increases the system capacity.

c. To cover the entire circular area = Total coverage area/Area of a cluster with reuse $N = 7$ = 3,600 km²/63 km² = 57 times.

2.3.2 Spectrum efficiency and propagation path loss

Spectrum efficiency of a cellular system: The spectrum efficiency η_s expressed in Erlangs per square meter per hertz, yields a measure of how efficiently space, frequency, and time are used, and it is given by

$$\eta_s = \frac{\text{number of reuses}}{\text{coverage area}} \times \frac{\text{number of channels}}{\text{bandwidth available}} \times \frac{\text{time the channel is busy}}{\text{total time of the channel}} \tag{2.3}$$

The role of frequency reuse concept in improving the spectrum efficiency is illustrated in the following example.

Frequency reuse and spectrum efficiency

Figure 2.15(a) shows a single high power transmitter that can support 100 voice channels covering a given coverage area. The same coverage area is divided into seven smaller areas (cells) as shown in Figure 2.15(b) and each cell supported by low power transmitters. The available spectrum of 100 voice channels is divided into four distinct groups of 25 channels each. If the allocation of channel groups to cells is in such a way that the cells 1 and 7 uses group1 channels, cells 2 and 4 uses group 2 channels, cell 3 uses group 3 as well as 5.

Then the total number of channels available in the specified cellular system can be computed as:

Total number of channels allocated to all cells = number of channels per channel group × number of distinct cells.

= 25 × 7

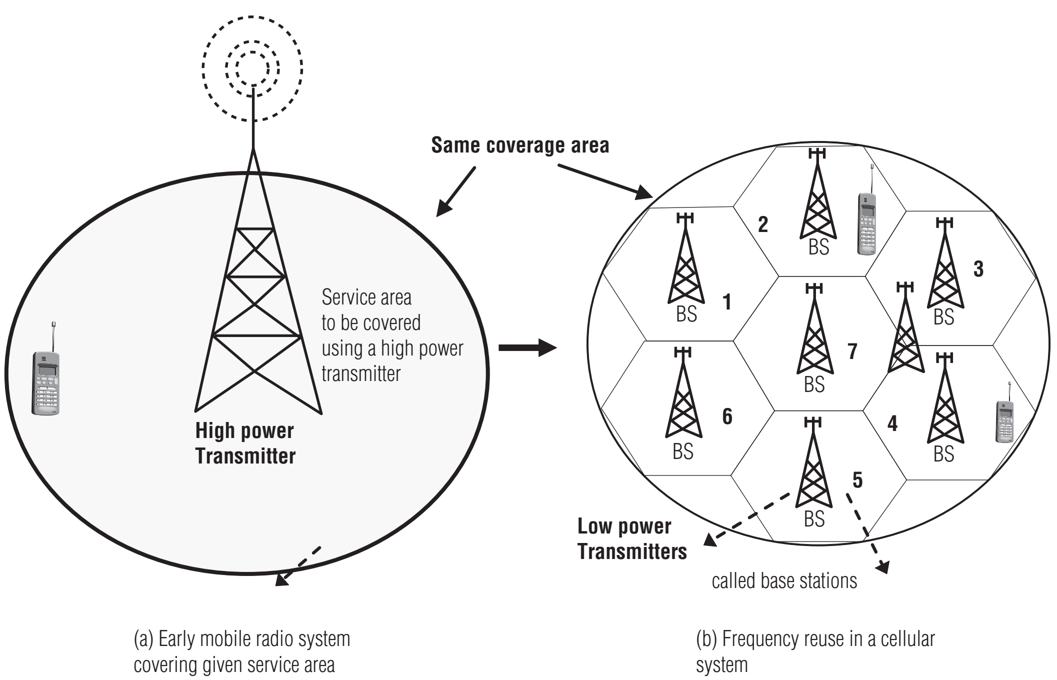


Figure 2.15 Illustration of frequency reuse concept and spectrum efficiency

Hence, *total number of channels available = 175 channels*.
From the above example, it can be concluded as follows:

- Total number of channels that can be supported by the given cellular system is increased to 175 from 100 to cover the same service area.
- Frequency reuse concept can significantly increase the *spectrum efficiency*, thereby increasing the *system capacity*

Propagation path loss: The propagation path loss of a signal is a function of several factors, such as environment, location, antenna type, antenna height, and so on. By considering omnidirectional antennas, the propagation path loss in a mobile radio environment is normally taken as 40 dB per decade, that is, the signal will suffer a 40 dB loss for each 10 km. The difference in power reception at two different distances d_1 and d_2 would be:

$$\frac{\mathrm{Pr}_2}{\mathrm{Pr}_1} = \left(\frac{d_2}{d_1}\right)^{-4} \tag{2.4}$$

where Pr_1 is the received carrier power at receiver 1, Pr_2 is the received carrier power at receiver 2, d_1 is the distance measured from the transmitter to receiver 1, and d_2 is the distance measured from the transmitter to receiver 2.

When expressed in decibels, Equation (2.4) becomes

$$\Delta\,\mathrm{Pr}\,(\text{in dB}) = \mathrm{Pr}_2 - \mathrm{Pr}_1\,(\text{in dB})$$

$$= 10\log\frac{\mathrm{Pr}_2}{\mathrm{Pr}_1} = 40\log\frac{d_1}{d_2}\,\mathrm{dB} \tag{2.5}$$

Example problem 2.2

Calculate the change in received signal powers (in decibels) in mobile radio propagation condition at two different distance points when the second distance point is twice the distance of the first point.

Solution

Let the received carrier signal power at a distance d_1 be Pr_1 and at a distance d_2 be Pr_2. The change in received signal strengths (in decibels), ΔPr in mobile radio propagation, between the distance points d_2 and d_1 is given by

$$\Delta\,\mathrm{Pr}\,(\text{in dB}) = 40\log\frac{d_1}{d_2}\,\mathrm{dB}$$

Here, $d_2 = 2d_1$ (given)

Therefore, $$\Delta\,\mathrm{Pr}\,(\text{in dB}) = 40\log\frac{d_1}{d_2} = 40\log(\frac{1}{2})$$

Hence, $\Delta Pr = 12dB$

From this result we observe that signal strength decays at the rate of 12dB/octave in the mobile radio-propagation environment condition.

Under the same conditions, but in free space, the propagation path loss would be of 20 dB/10 km. The propagation path loss will vary as:

$$Pr_{loss} = \alpha\ d^{-\gamma} \tag{2.6}$$

or in decibels

$$Pr_{loss}(\text{in dB}) = 10\log\alpha - 10\gamma\log d \tag{2.7}$$

where γ is the propagation path loss factor, α is a constant, and d is the distance from the transmitter to the receiver.

The γ parameter usually lies between 2 and 5; it cannot be lower than 2, the free-space condition.

2.3.3 Frequency reuse factor

One important characteristic of cellular networks is the *reuse of frequencies* in different cells. The cells using the same set of channels (or frequencies) are known as *co-channel cells*. For example, in Figure 2.14, the cells using channels A are co-channel cells. The distance between co-channel cells is known as *co-channel distance* or *frequency reuse distance* and the interference caused by the radiation from these cells is referred to as *co-channel interference*.

> *By reuse frequencies, a high capacity can be achieved. However, the reuse distance has to be high enough, so that the interference caused by subscribers using the same frequency (or an adjacent frequency) in another cells is sufficiently low.*

For proper functioning of any cellular system, the *co-channel interference* needs to be minimized. For example, to guarantee an appropriate speech quality, the *carrier-to-interference-power-ratio* (CIR) has to exceed a certain threshold CIR_{min} which is 9 dB for the GSM system.

Unlike thermal noise which can be overcome by increasing the signal-to-noise ratio (SNR), co-channel interference cannot be overcome by simply increasing the carrier power because an increase in carrier power increases the interference to neighbouring co-channel cells. To reduce co-channel interference, co-channel cells must be physically separated by a minimum distance.

> Therefore, *minimization of co-channel interference requires a minimum co-channel distance; that is, the distance cannot be smaller than this minimum distance.*

In a cellular system of equal cell size, the co-channel interference is a function of a *frequency reuse factor* or *co-channel reuse ratio* (q). The frequency reuse factor of a cellular system is defined by the ratio of distance between the frequency reusing cell sites (D) and the cell radius (R) of the serving cell sites and is known as D/R ratio (Fig. 2.16).

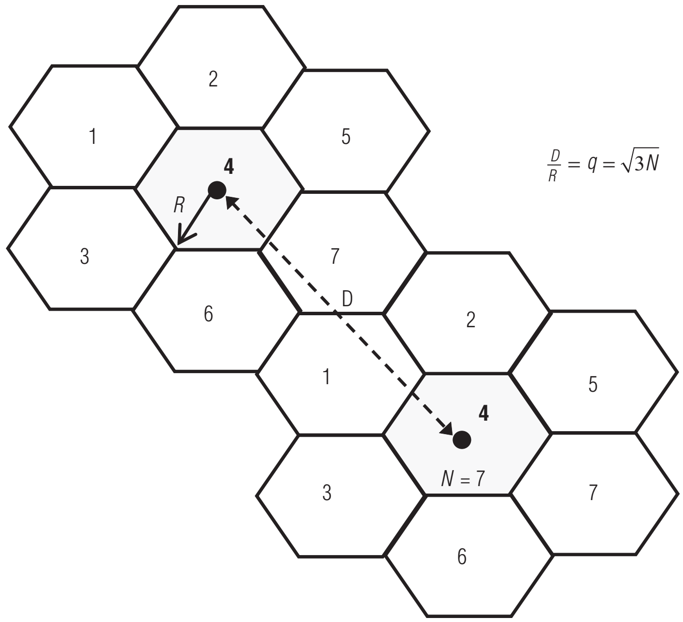


Figure 2.16 Frequency reuse factor or co-channel reuse ratio (q)

2.3.4 Relationship between frequency reuse factor (D/R) and cluster size (N)

Determination of relation between frequency reuse factor (D/R ratio) and the cluster size (N) involves mainly two steps: (i) finding the relation between the distance D (two co-channel cells) and R and (ii) locating the co-channel cells.

First let us derive the relation between distance (D) (two co-channel cells) and R in a hexagonal geometry (Fig. 2.17). The actual centre-to-centre distance between two adjacent hexagonal cells is

$$d = 2R\cos 30^{\circ} \quad \text{or} \quad \sqrt{3} \tag{2.8}$$

where R is the centre-to-vertex distance.

Proof of $d = \sqrt{3}R$

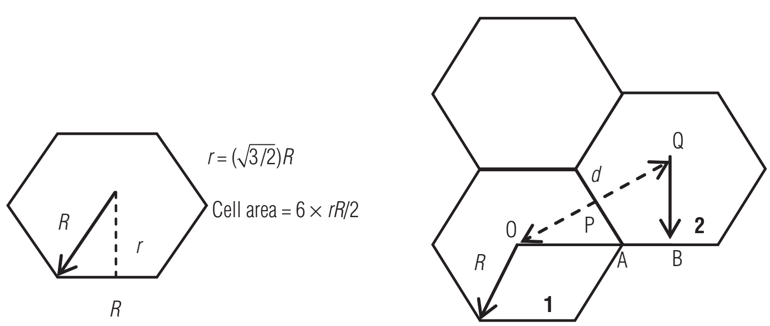


Figure 2.17 Centre-to-centre distance between two adjacent hexagonal cells 1 and 2

From Figure 2.17, we can observe that OA = R, $AB = R/2$ and OAP is a right-angled triangle. In the OAP triangle, $OP = OA$ Sin 60° = $(\sqrt{3}/2)\,R$.

Let d be the centre-to-centre (OQ) distance between two adjacent hexagonal cells 1 and 2, then

$$OQ = OP + PQ$$

$$d = (\sqrt{3}/2)R + (\sqrt{3}/2)R = \sqrt{3}R \tag{2.9}$$

$$\text{Area of a small hexagon with radius } R = 6 \times rR/2 = ((3\sqrt{3})/2)\cdot R^2 = 2\sqrt{3}r^2 \tag{2.10}$$

2.3.4.1 Method of locating co-channel cells

To locate the nearest co-channel cells, mark the centre of cell as (0,0) for which co-channel cells are required to be located. Define the unit distance as the distance of centres of two adjacent cells. The nearest co-channel cell in a hexagonal cellular structure to the cell under consideration can be located using shifting parameters (i, j). The two parameters i and j measure the number of nearest neighbouring cells between co-channel cells in a hexagonal geometry, where i and j are separated by 60° as shown in Figure 2.18. The shift parameters can have any value 0,1,2,3, and so on. The important steps to be followed now are as follows:

- Move i number of cells along any chain of hexagons.
- Turn 60° counter clockwise and move j number of cells along the chain that lies in this new heading.

The method of locating co-channel cells in a cellular system using the above procedure is shown in Figure 2.19 where co-channels are marked with letter A. Figure 2.19 illustrates the regular hexagonal geometry of one co-located cell. The parameters i and j measure the number of nearest neighbouring cells between co-channel cells. In this example, N = 19 for i = 3 and j = 2.

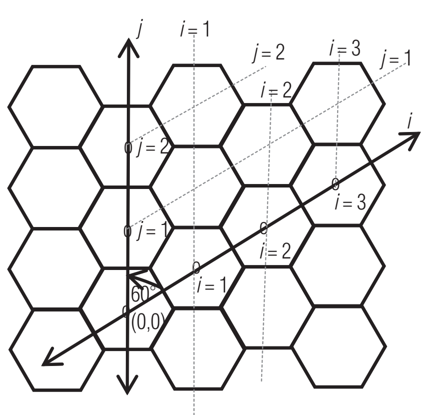


Figure 2.18 Shift parameters i and j in a hexagonal geometry

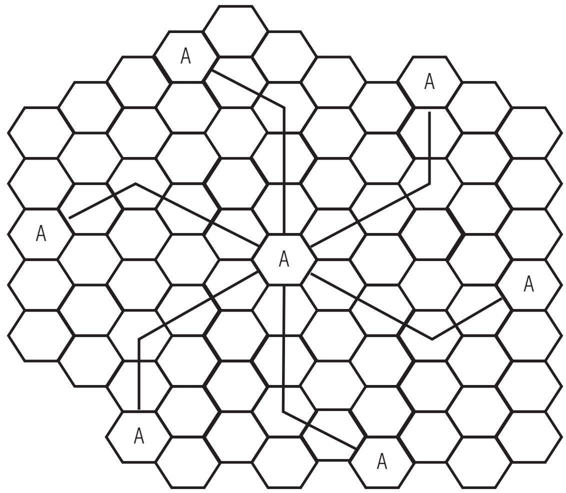


Figure 2.19 Locating co-channel cells in a ($N = 19$) hexagonal geometry for $i = 3$, $j = 2$

2.3.4.2 Establishment of relationship between D, d and shift parameters (i and j)

Since the hexagonal cell has six equidistant neighbouring cells, by following the two steps given under the method of locating co-channel cells corresponding to six sides of the hexagon. That is,

- firstly, move *i* number of cells along *j* axis
- secondly, turning 60° in counterclockwise
- Finally, move *j* number of cells along *j* axis, as shown in Figure 2.20.

Let *D* be the distance from the centre of the cell under consideration to the centre of a co-channel cell (XZ).

Apply cosine formula to triangle XYZ to derive the relation between *D*, *d*, and shift parameters (*i*, *j*)

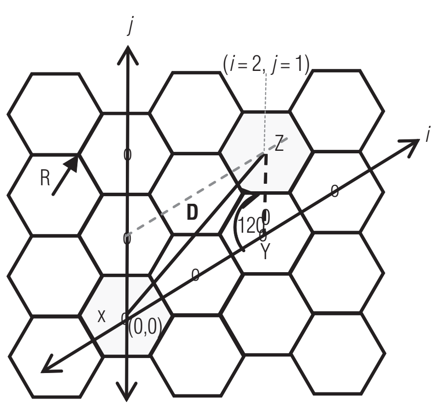


Figure 2.20 Relationship between *N* and shift parameters *i*, *j*

We have

$$XZ^2 = XY^2 - 2\ XY \times YZ \cos 120°$$

$$D^2 = (i \times d)^2 + (j \times d)^2 - 2 \times (i \times d) \times (j \times d) \cos 120°$$

or

$$D^2 = (i \times d)^2 + (j \times d)^2 - 2 \times (i \times d) \times (j \times d)\ (-1/2)$$

or

$$D^2 = (i \times d)^2 + (j \times d)^2 + (i \times d) \times (j \times d)$$

or

$$D^2 = d^2 \times (i^2 + j^2 + i\,j) \tag{2.11}$$

Using Equation (2.8), Equation (2.10) can be written as

$$D^2 = 3\ R^2\ (i^2 + j^2 + i\,j) \tag{2.12}$$

2.3.4.3 Establishment of relationship between D, R, and N

To establish the relationship between N and frequency reuse factor (D/R),

we know that the cluster size $$N = i^2 + j^2 + ij \tag{2.13}$$

Substituting Equation (2.13) in Equation (2.12), we get

$$D^2 = 3\ R^2 \times N$$

or

$$D^2/R^2 = 3 \times N$$

or

$$D/R = q = \sqrt{3N}$$

where q is the reuse ratio and N is the cluster size or reuse factor.

$$D/R = q = \sqrt{3N} \tag{2.14}$$

2.3.4.4 Relationship between area of a hexagon, number of cells in a large hexagon, and N

Figure 2.21 illustrates the geometry formed by seven clusters and each cluster contains seven cells. It can be observed that the larger hexagon is formed by joining the centres of co-channel cells in the first tier, which encloses seven cells of the middle cluster plus one-third of the number of seven cells of all surrounding six neighbouring clusters. In general, it can be computed that the larger hexagon encloses the centre cluster of N cells plus one-third the number of the cells associated with six other peripheral clusters in the first tier.

Area of large hexagon ($A_{larger\ hexagon}$):

By joining the six nearest neighbouring co-channel cells centres, a large hexagon is formed with radius equal to D, which is also the co-channel cell separation. The area of the large hexagon with radius D can be given as

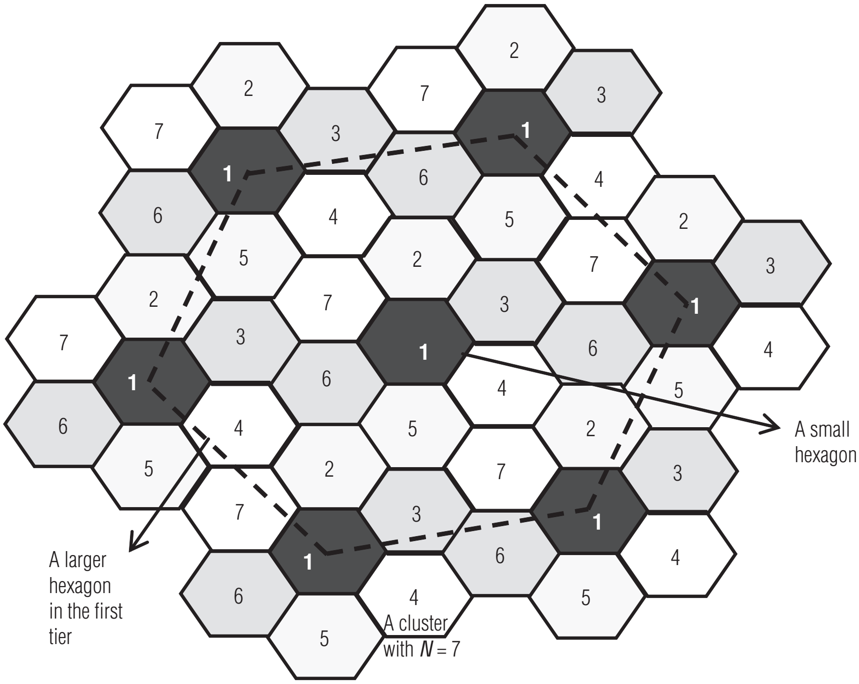


Figure 2.21 Number of clusters in the first tier for $N = 7$

$$(A_{\text{larger hexagon}}) = (3\sqrt{3}/2) \times D^2$$

By substituting Equation (2.12) in the above equation, we get

$$(A_{\text{larger hexagon}}) = (3\sqrt{3}/2) \times 3\, R^2\, (i^2 + j^2 + i\, j) \tag{2.15}$$

Area of small hexagon ($A_{small\ hexagon}$):

$$(A_{\text{small hexagon}}) = (3\sqrt{3}/2)\, R^2 \tag{2.16}$$

Number of cells in the $A_{\text{larger hexagon}}$ (L):
The number of cells in the larger hexagon ($A_{\text{larger hexagon}}$) shown in Figure 2.21 can be determined by comparing Equation (2.15) and Equation (2.16). We can write

$$L = \frac{A_{\text{larger hexagon}}}{A_{\text{small hexagon}}} = 3(i^2 + j^2 + ij) = \frac{D^2}{R^2} \tag{2.17}$$

In general, number of cells enclosed by the larger cell can be computed by adding number of cells contained in the centre cluster (N) plus one-third the number of cells associated with six other peripheral clusters in the first tier.

$$L = N + 6 \times (1/3) \times N = N + 2N = 3N \tag{2.18}$$

Relationship between *N* and shift parameters:

By comparing Equation (2.17) and Equation (2.18), we get

$$L = 3(i^2 + j^2 + ij) = 3N$$

$$3N = 3\ (i^2 + j^2 + ij)$$

$$N = i^2 + j^2 + ij \tag{2.19}$$

2.3.4.5 Important conclusions from Equation (2.14)

- Equation (2.14) is significant because it affects the traffic-carrying capacity of a cellular system and the co-channel interference. By reducing q, the number of cells per cluster is reduced.
- If the total number of RF channels are constant, then the number of channels per cell is increased, thereby increasing the system capacity. On the other hand, co-channel interference increases if q is small as discussed in Section 2.4. The reverse is true when q is increased. An increase in q reduces co-channel interference and also the traffic capacity of the cellular system.
- Table 2.1 shows the cluster size (N) determined from Equation (2.19) and the corresponding q values. It is observed that a 2 cell, 5 cell, 6 cell, etc., reuse pattern does not exist as they form asymmetrical reuse arrangements. In asymmetrical reuse arrangements, interferers are located at various distances. However, the basic assumption in Equation (2.14) is that all six first-tier interferers are located at the same distance from the desired cell.

Table 2.1 Frequency reuse ratio (q) vs. cluster size (N)

i	j	$N = i^2 + j^2 + ij$	$q = D/R = \sqrt{3N}$
1	0	1	1.73
1	1	3	3.00
2	0	4	3.46
2	1	7	4.58
3	0	9	5.20
2	2	12	6.00
3	1	13	6.24
4	0	16	6.93
3	2	19	7.55
4	1	21	7.94
4	2	28	9.17

- Due to the fact that the hexagonal geometry has exactly six equidistant neighbours and that the lines joining the centres of any cell and each of its neighbours are separated by multiples of 60°, there are only certain cluster sizes and cell layouts which are possible. The combination $i = j = 0$ cannot be used because $k = 0$ is meaningless. Similarly, $(i, j) = (0,1)$ or $(1, 0)$ results in $N = 1$ which is allowed only in CDMA systems in which all cells use the same frequency channels, but not in TDMA/FDMA cellular systems in which adjacent cells cannot be assigned the same frequency channels.

Example problem 2.3

As a total of 40 MHz of bandwidth is allocated to a particular frequency division duplex cellular telephone system which uses two 25 kHz simplex channels to provide full duplex voice and control channels. Compute the number of channels available per cell if a system uses (a) four-cell reuse, (b) seven-cell reuse, (c) 12-cell reuse. If 1 MHz of the allocated spectrum is dedicated to control channels, determine an equitable distribution of control channels and voice channels in each cell for each of the three systems.

Solution

Given bandwidth = 40 MHz
Channels bandwidth = 25 kHz × 2 simplex channels = 50 kHz/duplex channel
Total available channels = 40,000/50 = 800 channels

(a) For $N = 4$,total number of channels available per cell = 800/4 = 200 channels
(b) For $N = 7$, total number of channels available per cell = 800/7 ≈ 114 channels
(c) For $N = 12$, total number of channels available per cell = 800/12 ≈ 67 channels

For 1 MHz, the total available channels = 1 MHz/50 kHz = 20 channels.
For $N = 4$, total number of channels per cell = 20/4 = 5 channels.
For $N = 7$, total number of channels per cell = 20/7 ≈ 3 channels.
For $N = 12$, total number of channels per cell = 20/12 ≈ 2 channels.

2.3.5 The key trade-offs

The concept of frequency reuse leads to a cellular architecture that can allow for almost unlimited expansion in the geographic area and the number of subscribers that the system can serve.

In configuring a cellular layout, two parameters are of importance. These are the cell radius, R and the cluster size, N. This section begins by summarizing the design trade-offs in which these parameters are important. Then, we will focus on system expansion. Although, a cellular system can expand in area simply by adding cells at the periphery, we have not seen how a system can expand in user density. In this section, two methods will be discussed for dealing with increase in subscriber density. These are *sectoring* and *cell splitting*.

The cell radius governs both the geographic area covered by a cell and also the number of subscribers that the cell must serve for a given subscriber density.

Since every cell requires investment in a tower, land on which the tower is placed and radio transmission equipment, a large cell size reduces the cost per subscriber. Hence, the cell size should be as large as possible. The cell size is finally determined by the requirement that an adequate CIR is maintained over the coverage area. For example, *the system parameters that are involved in determining the SNR are transmitter power, antenna height, and receiver noise figure.*

Transmitter power is particularly limited in the reverse direction as the mobile units are small and battery powered. If a cell radius R and a cluster size N are given then the geographic area covered by the cluster is

$$A_{\text{cluster}} = N \times A_{\text{cell}} = N \times \frac{3\sqrt{3}}{2} R^2 \tag{2.20}$$

where a hexagon of radius R has an area A_{cell} given by

$$A_{\text{cell}} = \frac{3\sqrt{3}}{2} R^2 \tag{2.21}$$

If the market has a geographic area of A_{market}, then the number of clusters M is given by

$$M = \frac{A_{\text{market}}}{A_{\text{cluster}}} = \frac{A_{\text{market}}}{N \frac{3\sqrt{3}}{2} R^2} \tag{2.22}$$

Of the available channels, "n" number of channels (n_{channels}) are reused in every cluster. Hence, to make the maximum number of channels available to subscribers, the number of clusters M should be large. Equation (2.22) shows that the cell radius should be small. Ultimately, cell radius is determined by a ***trade-off***, that is,

R should be as large as possible to minimize the cost of installation per subscriber. At the same time, R should be as small as possible to maximize the number of customers that the system can accommodate.

Equation (2.22) also gives us an idea on the second key parameter. If we consider the cell radius as fixed, then the number of clusters can be maximized by minimizing the number of cells in the cluster. From Equation (2.14), we have

$$N = \frac{q^2}{3} \tag{2.23}$$

Equation (2.21) shows that the cluster size depends only on the frequency reuse ratio q. Now in selecting a value of N, we once again come across a trade-off.

When the AMPS cellular system was first deployed, the aim of the system designers was to guarantee the coverage. Initially, the number of users was not significant. Later, cells were configured with an eight-mile radius and a 12-cell cluster size was chosen.

The cell radius was chosen to provide a 17 dB SNR over 90 per cent of the coverage area.

In early 1980s this value of 17 dB SNR is chosen using the practical values for antenna height, mobile, and base power levels. However, the 12-cell cluster size did not provide adequate frequency reuse to service an explosively growing customer base. From the data available, a 7-cell cluster size should provide an adequate 18.7 dB C/I ratio. The margin, however, is slim, and the 17 dB C/I ratio requirement could not be met over 90 per cent of the coverage area.

In the next section, we will discuss techniques used for improving coverage and capacity in cellular systems.

2.4 Improving coverage and capacity in cellular systems

As demand increases, number of channels per cell become insufficient. Therefore, new cellular design techniques are needed to provide more channels per unit coverage area. Various techniques developed to expand the capacity of system are as follows:

- *Cell splitting:* Reduce radius of cell to increase frequency reuse.
- *Sectoring:* Uses directional antennas to control interference and frequency reuse.
- *Repeaters for range extension:* Use retransmitters to cover areas subjected to fading.
- *Zone microcells:* Distributes the coverage of a cell.

2.5 Cell splitting

Cell splitting is the process of subdividing a congested cell into smaller cells, each with its own BS and a corresponding reduction in antenna height and transmitter power.

> *Cell splitting increases the capacity of a cellular system since it increases the number of times that channels are reused.*

By making these the new cells to have smaller radius than the original cells and by installing these smaller cells between the existing cells, capacity increases due to the additional number of channels per unit area. Cell splitting achieves capacity improvement by essentially re-scaling the system.

> *By decreasing the radius R and keeping the co-channel reuse ratio D/R unchanged, cell splitting increases the number of channels per unit area.*

We have discussed that reducing the size of cells of a cellular system keeps the CIR constant but results in an expansion of the network capacity because the smaller cells cover less area and therefore more cells would be required to cover the whole region which directly reflects on the network capacity. If the network is already functioning, it may be found that the network needs expansion only in specific regions and not network-wide expansion. In such a case, we employed cell splitting concept.

> *Cell splitting is the attractive feature of cellular concept and is the process of dividing a larger congested cell into smaller cells, each with its own cell site with a corresponding reduction in transmitter power and antenna height.*

This is usually done to make more voice channels available to accommodate traffic growth in the area covered by the original cell. When the traffic in an area increases, larger cells are split into smaller cells so that frequency can be reused more frequently. By defining new smaller cells which have a smaller radius than the original larger cells and by installing these smaller cells (called microcells) between the existing cells, the system user capacity increases due to availability of additional number of channels per unit service area as well as the number of times that frequency channels are reused.

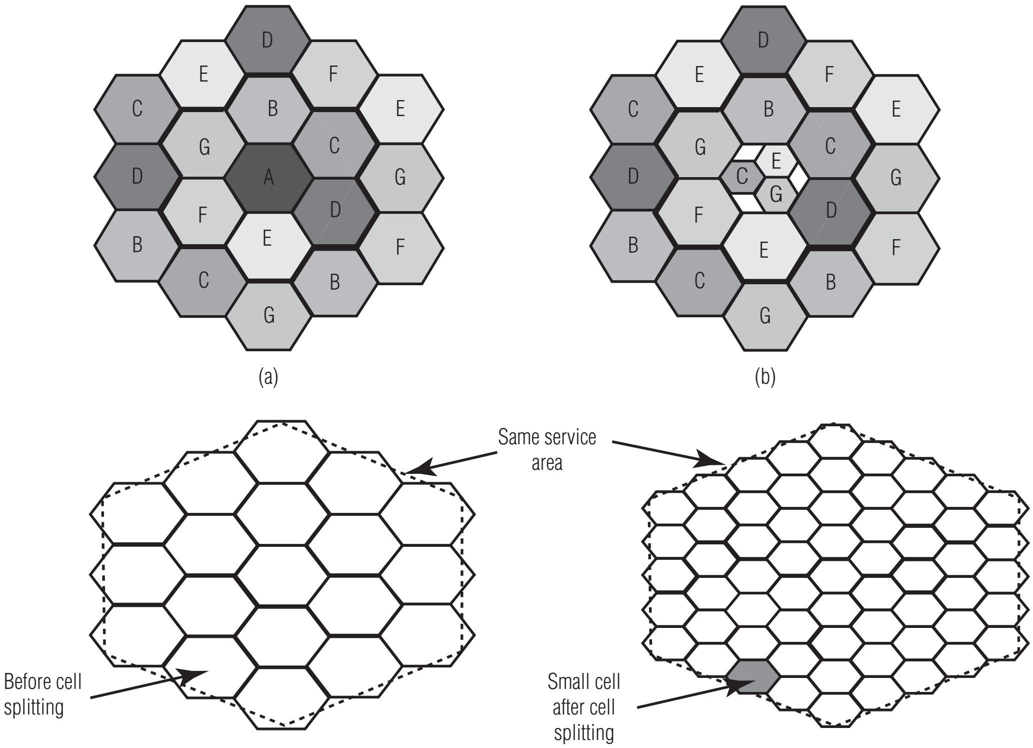


Figure 2.22 Original cell distribution and cell splitting

In other words, the increased number of cells would increase the number of channels as well as number of clusters over the coverage area, which in turn would increase the overall system capacity.

Figure 2.22(a) denotes that a cell that has reached its capacity and needs to be split. This cell is split into several cells. Since the area of a cell is proportional to R^2. So, reducing the new cell radius to one-half of its original value, for example, the area of the cell drops to one-quarter of its original value. Therefore, theoretically, four of the smaller cells can fit into one of the large cells. However, since it is not possible to fit four quarter-size hexagonal cells completely into one full-size hexagonal cell, some regions will have to be covered by adjacent cells. In order to overcome this, divide each cell (e.g. centre cell A) into three sub-cells (i.e. C, E, and G) as shown in Figure 2.22(b).

In the Figures 2.22(c) and 2.22(d), following cell splitting, the new small cells are reassigned new frequencies that do not cause co-channel interference with adjacent cells as shown in the above figure. In addition, the power transmitted in the small cells is reduced compared to the power transmitted in the large cells as it would require much less power to cover the cell compared to the large cells. In fact the power has to be reduced by a factor of

$$\frac{P_{\text{transmitted in small cell}}}{P_{\text{transmitted in large cell}}} = \left(\frac{R_{\text{small cell}}}{R_{\text{large cell}}}\right)^n \tag{2.24}$$

For example, if the cell radius of the small cells is half the radius of the large cell and the path loss exponent $n = 4$, the power transmitted by the tower of the small cell is only 1/16 that of the power transmitted by the tower of the large cell. In addition to the advantage of having a higher network capacity due to cell splitting, the reduced transmitted power, especially by the mobile phone, is another major advantage because it increases the battery life of these mobile phones. The main disadvantage of cell splitting is that it requires the construction of new towers, which is very costly.

2.6 Sectoring

Sectoring is another way to increase capacity. In sectoring, a cell has the same coverage space but instead of using a single omni-directional antenna that transmits in all directions, either three or six directional antennas are used and each with beamwidth of about 120° or 60° as shown in Figure 2.23.

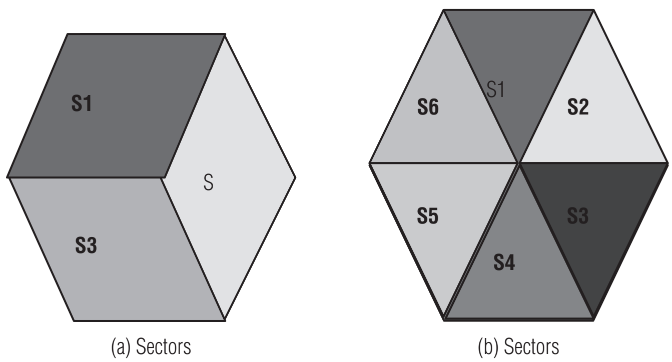


Figure 2.23 A cell divided into (a) 120° and (b) 60°

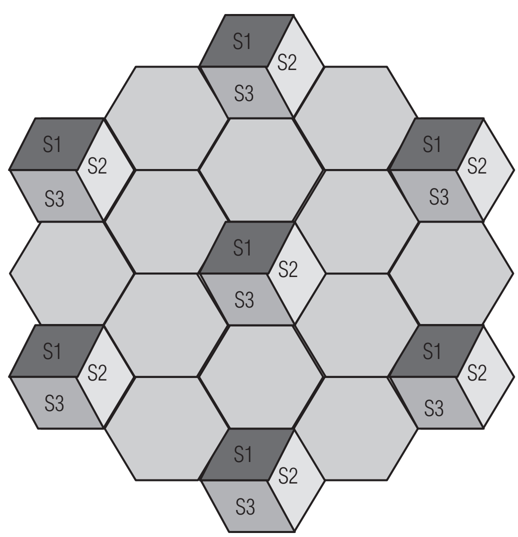


Figure 2.24 Sectoring for four-cell pattern

When sectoring is employed, the channels allocated to a particular cell are divided among the different sectors. It is done in such a way that channels assigned to a particular sector are always at the same direction in the different cells. For example, group A of channels assigned to the sector S2, group B of channels are assigned to the sector S1 at the top of all cells, and so on. Each sector causes interference to the cells that are in its transmission angle only. Unlike the case of no sectoring where six interfering co-channel cells from the first-tier co-channels cells cause interference, with 120° sectoring, two or three co-channel cells cause interference and with 60° sectoring, one or two co-channel cells cause interference. Figure 2.24 shows sectoring for a four-cell pattern.

Figure 2.25 shows the case of cluster size of $N = 4$ in which only two of the six co-channel cells cause interference to the middle cell for the sector labelled S2 in the case of 120° cell sectoring (the cells with radiation sectors coloured red and green).

The other four cells, although they are radiating at the same frequencies cause no interference because the middle cell is not in their radiation angles. For the case of 60° cell sectoring, only one cell causes interference (the cell with radiation sectors coloured green). The number of co-channel interfering cells depends on the cluster shape and size. By having less than six interfering first-tier co-channel cells causing interference, the CIR is increased for the same cluster size.

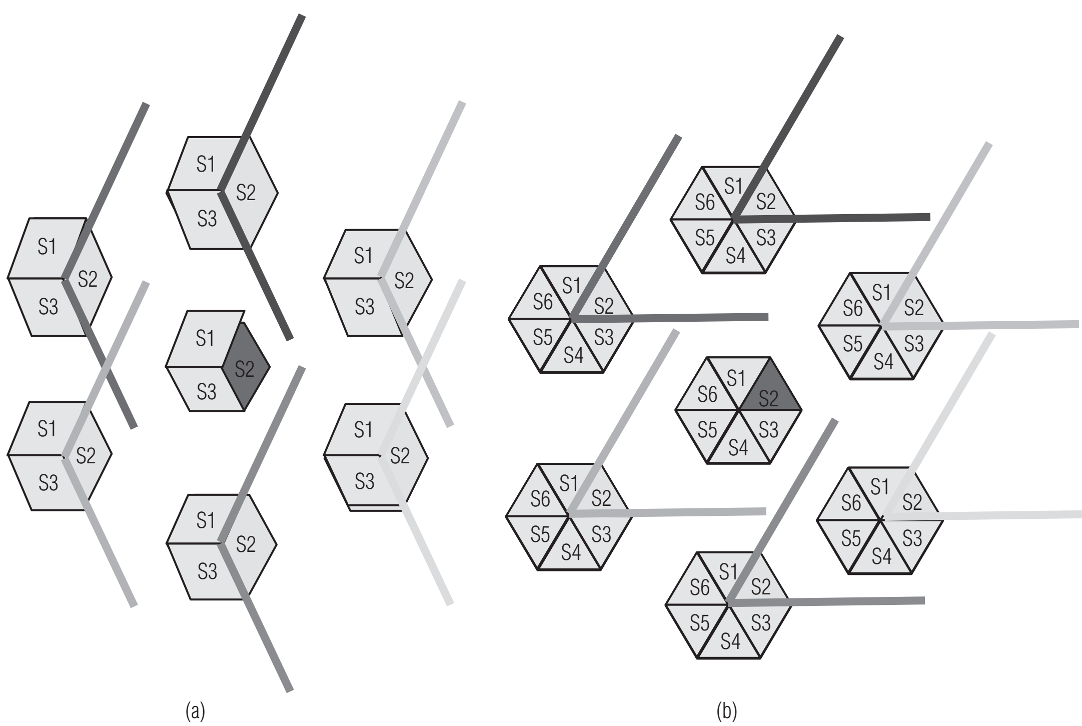


Figure 2.25 Interference in (a) 120˚ and (b) 60˚ cell sectoring

The resulting C/I for this case can be found using Equation (3.8) described in Section 3.3.2 of Chapter 3. The CIR must now be modified from

$$\frac{C}{I} = \frac{1}{6(q)^{-\gamma}} = \frac{q^{\gamma}}{6} \text{ (due to six interfering cells)}$$

to

$$\frac{C}{I} = \frac{1}{2(q)^{-\gamma}} = \frac{q^{\gamma}}{2} \text{ (due to two interfering cells)} \tag{2.25}$$

where q is the co-channel interference reduction factor and γ = is the path loss exponential constant.

The denominator has been reduced from 6 to 2 to account for the reduced number of interference sources.

Example problem 2.4

Find the signal-to-interference ratio (SIR) for a seven-cell-cluster layout with 120° sectors. Assume that the path loss exponent is $\gamma = 4$.

Solution

If $N = 7$, we know that $q = \sqrt{3N}$.

Then from Equation (2.23)

$$\frac{C}{I} = \frac{(\sqrt{3N})^{\gamma}}{2} = \frac{(\sqrt{3 \times 7})^4}{2} = 221 \quad \text{or} \quad \frac{C}{I}(\text{dB}) = 23.4\text{dB}$$

Some cellular systems divide their cells into six 60° sectors. The analysis is similar to the 120°-sector case.

In addition to the reduced number of interfering towers that sectoring produces, the SIR is increased for the same cluster size. This allows us to reduce the cluster size and achieve the same original CIR, which directly increases the network capacity. For example, if the cluster size is reduced from seven to four, the total number of channels is increased by a ratio 7/4 = 1.75, thereby increases capacity of cellular system up to 75 per cent. But there is a difficulty that the CIR is reduced due to another reason. Since interfering tower always fall behind the tower (i.e. if a sector is radiating to the right, e.g., the interfering cells must be to its left). Therefore, the worst case SIR occurs when the mobile phone being served by that sector is located at a relatively far corner with respect to the interfering cells. This means that among the six interfering co-channel cells in a non-sectored system, the sectored system gets rid of some of the worst interfering cells (the cells closest to the corner at which the mobile phone is located) as shown in Figure 2.26.

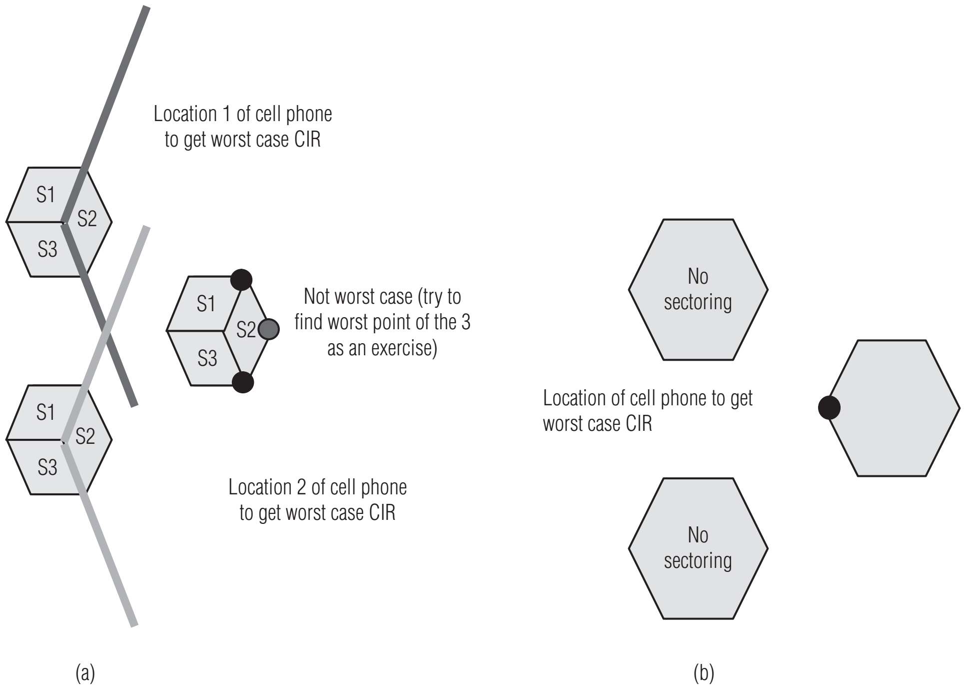


Figure 2.26 (a) With and without (b) Cell sectoring

And another problem with sectoring is that dividing a cell into sectors requires that a call in progress will have to be handed off when a mobile unit travels into a new sector. Although the need of handoffs between sectors as well as between cells does not directly reduce the number of customers that can be supported, it does increase the complexity of the system needed to support them.

2.7 Range extension by the use of repeaters

The use of repeater in cellular mobile communication system is for extending the range of the reception of the receiver. Especially, the repeater is used when it is hard for the transmitted signal to reach up to the receiver set. Repeaters are bidirectional in nature and simultaneously send signals to and receive signals from a serving BS. Upon receiving signals from BSs in forward link, the repeater amplifies and reradiates the BS signals to the specific coverage region. Repeaters are being widely used to provide coverage into and around buildings, where coverage has been traditionally weak. However, repeaters do not add any capacity to the system, they just increase the reach of a BS or MS into "shadowed" areas.

Example of the use of a repeater

Figure 2.27 show the scenario where a BS is unable to cover a difficult area because of a valley. The solution is to install a repeater attached to the same BS.

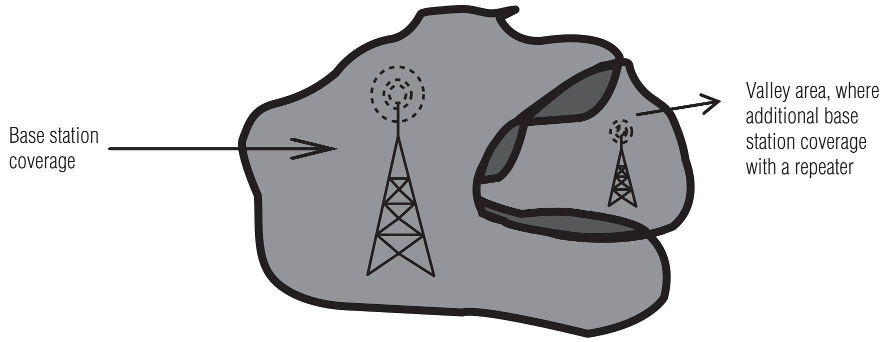


Figure 2.27 Range extension by the use of repeater

2.8 Microcell zone concept

By the use of sectorization technique, we can increase the system performance (i.e. quality of the signal) but side by side, there will be a large increment of handoffs which results in the increment of load on the switching and control link elements of the mobile system. So there must be some technique for the solution of this problem. So a *microcell zone concept* is introduced which leads to an increased capacity without any degradation in trunking efficiency caused by sectoring (Fig. 2.28).

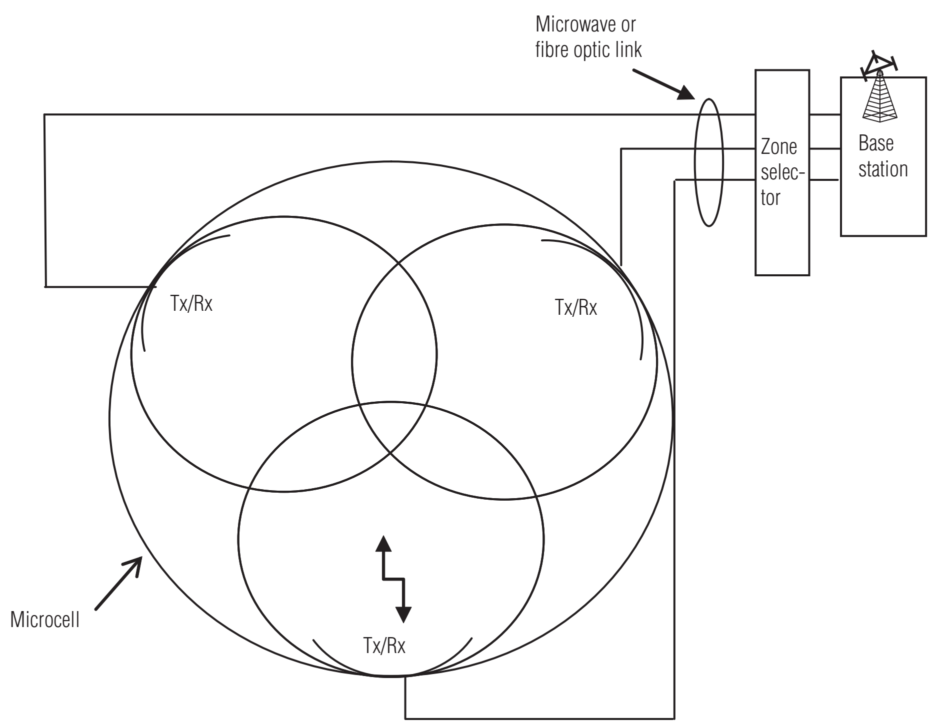


Figure 2.28 Microcell zone concept (for three microcells)

The microcell zone concept is related to sharing the same radio equipment by different microcells. This results in reduction of cluster size and hence increase in system capacity. The microcell coverage zone approach is used in practice to expand the capacity of cellular communication systems.

For example, each of the three or more zones (represented as Tx/Rx) in figure are connected to a single BS and share the same radio equipment (same frequency). Generally, the zones are connected by coaxial cable, fibre optic cable, or microwave link to the BS. All those multiple zones (three zones considered here) and a single BS make up a cell. As a mobile travels within a cell, it is served by the zone with the strongest signal. This approach (i.e. microcell zone concept) is superior to sectoring since antennas are placed at the outer edges of the cell, and any BS channel may be assigned to any zone by the BS. As the mobile travels from one zone to another within the cell, it retains the same channel. Thus, unlike in sectoring, a handoff is not required at the MSC when the mobile travels between zones within the cell since all the three zones have the same frequency. The BS simply switches the channel to a different zones' site. In this way, a given channel is active only in the particular zone in which the mobile is travelling, and hence the BS radiation is localized and interference is reduced.

2.8.1 Advantages of microcell zone concept

- A given channel is active only in a particular zone. Thus, interference is reduced and capacity is increased.
- Handoffs are reduced (also compared to decreasing the cell size) since the microcells within the cell operate at the same frequency; no handover occurs when the mobile unit moves between the microcells.
- Size of the zone apparatus is small. The zone site equipment being small can be mounted on the side of a building or on poles.

2.9 Picocell zone concept

A picocell is a small cellular BS. The use of picocell sites by wireless operators is driven by the desire to provide coverage and capacity to a given area or application. The picocell has a very small service area where several picocells in concept can cover the same area as a microcell. The picocell is a spot coverage and low-capacity site, as compared to a macrocell site. Picocell sites typically have a single omni-antenna, as do microcells. However, the power and thus the coverage of the picocell is less than a microcell.

Figure 2.29 explains the picocell zone concept in covering a small area, such as in buildings (offices, shopping malls, train stations, etc.), or more recently in aircrafts. In cellular networks, picocells are typically used to extend coverage to indoor areas where outdoor signals do not reach well, or to add network capacity in areas with very dense phone usage, such as train stations. Picocells provide coverage and capacity in areas difficult or expensive to reach using the macrocell concept.

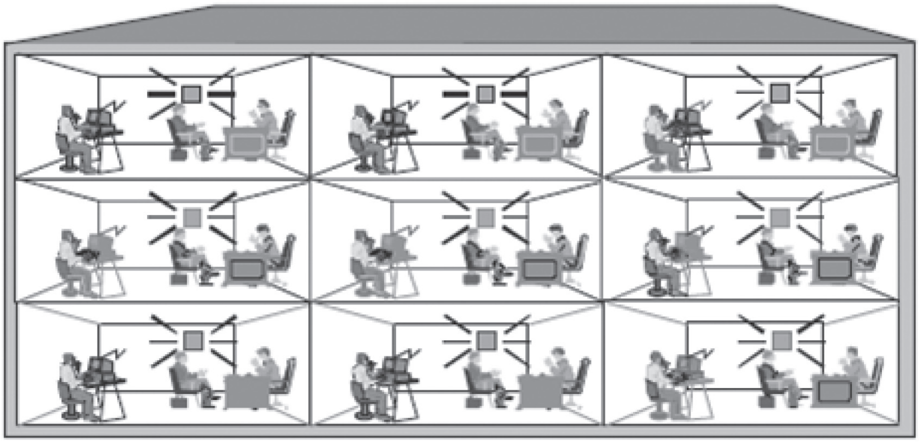


Figure 2.29 Picocell zone concept

2.10 Summary

- The concept of frequency reuse arises as a natural solution when the available radio spectrum is limited.
- Transmitter frequencies can be reused if the transmitters are separated by a sufficient distance. This leads to a system architecture which consists of an array of cells, each containing a BS that is assigned a specific set of channels.
- A hexagonal cell structure allows an unlimited geographical area to be covered by cells with minimum overlap and with no gaps in coverage.
- A cellular layout is characterized by integers i and j. Starting at any given cell, a co-channel cell that uses the same set of channels can be located by moving i cells in a direction parallel to one of the sides of the cell, then turning counter-clockwise 60° and moving j cells in the new direction.
- A set of cells that together are assigned all the channels is called a cluster. The size of a cluster is given by $N = i^2 + j^2 + ij$. Therefore, only certain cluster sizes are physically possible.
- The key design parameters for a cellular layout are the cell radius R and the cluster size N. The cluster size can be written in terms of the frequency reuse ratio $q = D/R$, where D is the distance between co-channel cells. We have shown that $N = q^2/3$.
- The cell radius is determined by a trade-off. The radius should be as large as possible to include as many subscribers as possible. This minimizes the cost per user of the BS tower and equipment. On the other hand, reducing the cell radius increases the number of cells that are needed to cover the market area, which in turn increases the number of times that each channel is reused. This increases the geographic density and total number of subscribers that a system can support.
- The cluster size is also determined by a trade-off. The cluster size should be made as small as possible to maximize the number of channels per cell and the number of clusters in the system. This maximizes the density of subscribers, and it also maximizes the number of times that each channel is reused.

- Cellular systems must be designed to allow for an expanding number of subscribers. Two methods are sectoring and cell splitting.
- Dividing each cell into sectors limits the number of sources of co-channel interference. This may allow the cluster size to be reduced, with a resulting increase in the subscriber density.
- Cell splitting is the process of subdividing a congested cell into smaller cells, each with its own BS and a corresponding reduction in antenna height and transmitter power. Cell splitting increases the capacity of a cellular system since it increases the number of times that channels are reused.

Example problem 2.5

A cellular system uses a frequency reuse factor $N = 4$ ($i = 0$, $j = 2$). If the path loss exponent $\gamma = 4$ and cell radius $R = 5$ km, find the following quantities in decibels:

(a) The SIR for the system with no cell sectoring
(b) The SIR for the system when 120° cell sectoring is used (note that worst occurs when mobile phone is at the furthest point from the interfering towers)
(c) The SIR for the system when 60° cell sectoring is used (note that worst occurs when mobile phone is at the furthest point from the interfering towers)

Solution

Given data:
Frequency reuse factor for $N = 4$ ($i = 0$, $j = 2$), $q = 3.46$
Path loss exponent $\gamma = 4$
Radius $R = 5$ km

(a) SIR for no cell sectoring

$$\text{We have formula } CIR = \frac{1}{(q)^{-\nu}}$$

$$= \sqrt{3(4)}^{4} = 143.31$$

$$= 10\log(48) = 21.56\,\text{dB}$$

(b) SIR when 120° cell sectoring is used

$$\text{We have formula } CIR = \frac{1}{3(q)^{-\nu}}$$

$$\text{We have } q = \sqrt{3N}$$

$$= \sqrt{3(4)} = 3.464$$

$$\text{Therefore } SIR = \frac{1}{3(3.464)^{-4}} = 47.99$$

$$= 10\log(47.99) = 16.81\,\text{dB}$$

(c) CIR when 60° cell sectoring is used

$$\text{We have formula } CIR = \frac{1}{6(q)^{-\gamma}}$$

$$= \frac{1}{6(3.464)^{-4}} = 23.85$$

$$= 10\log(23.85) = 13.80\,\text{dB}$$

Example problem 2.6

A region with area 10,000 km^2 has an evenly distributed population of 2.5 million people and is covered by a cellular system using a 12-cell reuse pattern and each cell has a radius of 3.06 km and the area allotted is 400 MHz of spectrum with a full duplex channel band width of 60 kHz. Assume grade of service (GOS) of 2 per cent for an Erlang B system is specified if the offered traffic per user is 0.03 Erlangs. Find the following:

(a) Number of channels per cell
(b) Traffic intensity of each cell
(c) Total number of cells
(d) Maximum carried traffic
(e) Total number of users served for 2 per cent of GOS

Solution

Given data: Area = 10,000 km^2

Population = 2.5 million

Cell frequency reuse = 12

Number of channels per cell [c]

$$= \frac{\text{Allocated spectrum}}{(\text{Channel width} \times \text{frequency reuse factor})}$$

$$= \frac{40000000}{60000 \times 12} = 55.5556$$

$$= 56 \text{ channels / cell}$$

(a) Traffic intensity of each cell

We have $C = 56$ and GOS = 0.02.

From Erlang B chart we have traffic intensity per cell A = 45 Erlangs/cell

(b) Total number of cells (N_c])

We have area of the cell (hexagon) to be 2.5981 R^2

Thus, each covers $2.5981 \times (3.06)^2 = 24.3$ km^2

Total number of cells are N_c =10,000/24.3 = 411 cells

(c) Maximum carried traffic

= Number of cells × traffic intensity per call

= 411 × 45

= 18,495 Erlangs

(d) Total number of users served for 2 per cent of GOS

Given traffic per user = 0.3 Erlang

Total number of users = total traffic time/traffic per user

=18,495/0.03

= 616,500 users

Example problem 2.7

Determine the distance from the nearest co-channel cell for a cell having a radius of 0.6 *km and a co-channel reuse factor of 12.*

Solution

The radius of a cell, $R = 0.64$ km (given)

The co-channel reuse factor, $q = 12$ (given)

To determine the distance from the nearest co-channel cell, D

We know that $q = D/R$,

or $D = q \times R$

Therefore, $D = 12 \times 0.6$ km = 7.2 km

Hence, the distance from the nearest co-channel cell, $D = 7.2$ km

---

Example problem 2.8

Determine the number of cells in clusters for the following values of the shift parameters i and j in a regular hexagonal geometry pattern:

(a) $i = 2$ *and* $j = 4$
(b) $i = 3$ *and* $j = 3$

Solution

(a) *To determine the number of cells in a cluster for* $i = 2$ *and* $j = 4$

Using the relationship, $K = i^2 + j^2 + i \times j$
The number of cells in a cluster, also called cluster size K, can be determined as
$K = 2^2 + 4^2 + 2 \times 4 = 4 + 16 + 8 = 28$

(b) *To determine the number of cells in a cluster for $i = 3$ and $j = 3$*

$K = 3^2 + 3^2 + 3 \times 3 = 9 + 9 + 9 = 27$

Example problem 2.9

Calculate the number of times the cluster of size 4 has to be replicated in order to approximately cover the entire service area of 1569 km² with the adequate number of uniform-sized cells of 7 km² each.

Solution

Size of the cluster, $K = 4$ (given)
Area of a cell, $A_{\text{cell}} = 7\ \text{km}^2$ (given)

Step 1. *To determine area of the cluster*

Area of a cluster, $A_{\text{cluster}} = K \times A_{\text{cell}}$

Therefore, $A_{\text{cluster}} = 4 \times 7\ \text{km}^2 = 28\ \text{km}^2$

Step 2. *To determine number of clusters in the service area*

Total service area, $A_{\text{system}} = 1{,}569\ \text{km}^2$ (given)

Number of clusters in service area = $A_{\text{system}}/A_{\text{cluster}}$

Number of clusters in service area = $1{,}569\ \text{km}^2/28\ \text{km}^2$

Number of clusters in service area = 56

Hence, the number of times the cluster of size 4 has to be replicated is 56.

Review questions

1. Describe the frequency reuse concept in cellular communication system. Using the cell geometry and co-ordinate system given in this chapter, derive the equation for the frequency reuse ratio.
2. What are the possible frequency reuse patterns?
3. What are spectrum efficiency and propagation path loss of a signal in a cellular communication system? Give appropriate equations for both.
4. Mention the various techniques used to expand the capacity of a cellular system.
5. Describe role of a use of a repeater in range extension of a cellular system.
6. Why do we divide the cell into various sectors? Show that in a seven-cell-cluster layout with 120° sectored cells the mobile units in sector **A** of the centre cell will receive co-channel interference from only two of the first-tier co-channel BSs, rather than from all six.
7. Show that in a 120° sectored cells, the CIR observed at the base station in the centre cell will be $\dfrac{C}{I} = \dfrac{(\sqrt{3N})^{\nu}}{2}$.
8. Describe the microzone and picozone concepts in a cellular system.
9. Draw the frequency reuse pattern for a cluster size of $N = 7$.

10. If the cluster size is four, the cluster is replicated seven times and each cell is allocated 20 channels. Find the total number of radio channels and the total number of duplex channels.
11. What is the need for frequency reuse? Prove that for a hexagonal geometry, the co-channel reuse ratio is given by $Q=\sqrt{3N}$, where $N = i^2 + ij + j^2$.
12. In a certain cellular system, the BSs radiate 15 W. Suppose the cells are split and the new cells have one-fourth the radius that the original cells had. Find the power that the BSs in the new layout must transmit to maintain the SNR at the cell boundaries. The path loss exponent is $v = 2$. (Ans. 0.936 W)
13. Calculate the change in received signal powers (in decibels) in mobile radio propagation condition at two different distance points. When the second distance point is ten times the distance of the first point. (Ans: –40 dB)
14. If a signal to interference ratio of 15 dB is required for satisfactory forward channel performance of a cellular system, what is the frequency reuse factor and cluster size that should be used for maximum capacity if the path loss exponent is (a) $v = 4$ and (b) $v = 3$? Assume that there are six co-channels cells in the first tier and all of them are at the same distance from the mobile. Use suitable approximations. (Ans. (a) 18.66 dB, (b) 12.05 dB)
15. What are the possible frequency reuse patterns?
16. Present the concepts of frequency reuse channels and frequency reuse distance. (Refer Sections 2.3, 2.3.1, and 2.3.3)
17. Explain the importance of $N = i^2 + ij + j^2$. (Refer Section 2.3.4 and Equation 2.19).

Objective type questions and answers

1. Compute the channel bandwidth if a total of 33 MHz of bandwidth is allocated to a cellular system which uses two 25 kHz simplex channels to provide full duplex voice and control channels.
   (a) 25 kHz × 1 simplex channels (b) 25 kHz × 2 simplex channels
   (c) 50 kHz × 2 simplex channels (d) 50 kHz × 2 simplex channels
2. Compute the total available channels if a total of 33 MHz of bandwidth is allocated to a cellular system which uses two 25 kHz simplex channels to provide full duplex voice and control channels.
   (a) 600 channels (b) 630 channels (c) 660 channels (d) 690 channels
3. The most appropriate measure of spectral efficiency for a cellular system is
   (a) users/MHz/km^2 (b) channels/MHz/km^2
   (c) Erlangs/MHz/km^2 (d)All the above
4. Cellular system capacity can be increased by
   (a) Frequency reuse (b) cell sectorization
   (c) a and b (d) bandwidth increase
5. Total number of available radio channels can be expressed for frequency reuse of a cluster size of "N" with each cell allocated a group of "k" channels
   (a) $2k \times N$ channels (b) $2k/N$ channels (c) k/N channels (d) $k \times N$ channels
6. The following cell layout is the most economically efficient as it requires least number of cells to cover a given area
   (a) triangular (b) circular (c) rectangular (d) hexagonal

7. Smaller cells in the cellular system have greater system capacity and the co-channel interference
   (a) increases (b) decreases
   (c) constant (d) none of the above
8. By dividing the cell geographically, usually into 120° sectors, its coverage area and the overall quality of the system
   (a) improves (b) degrades
   (c) no change (d) none of the above
9. In cellular design, the *D*/*R* ratio is used to characterize the
   (a) cell splitting (b) frequency reuse
   (c) cell sector (d) sectorized antenna

Answers: 1. (b), 2. (c), 3. (a), 4. (c), 5. (d), 6. (d), 7. (c), 8. (a), 9. (b).

True/False

1. Frequency reuse is maximized by increasing the size of cells. (b)
   (a) True (b) False
2. Cells are assumed to have a regular orthogonal shape. (b)
   (a) True (b) False
3. Cell splitting is based on the reduction of cell radius. (a)
   (a) True (b) False

Open book questions

1. What are co-channel cells and what is co-channel interference and describe its importance with respect to SNR.
2. State the different techniques used for improving coverage and capacity in cellular systems.
3. State the expression (a) that relates co-channel reuse ratio (*Q*) to radius (*R*) of a cell and (b) expression used to locate co-channel cells.
4. What is the purpose of cell sectoring and how co-channel interference in a cellular system may be decreased
5. Why hexagonal cell shape is used in cellular communication?
6. Write a short note on hexagonal cells.
7. What is a frequency reuse distance?
8. List some methods to reduce co-channel interference in cellular communication network.
9. What is the significance of cell size?
10. Distinguish between a cell and a cell site. What complications arise due to usage of smaller cells?
11. What is the core concept of the cellular communications? How is signal quality affected by employing this concept in cellular communication systems?
12. What could be the possible sources of interference which may limit the performance of cellular communication systems? Can the value of cluster size be increased more than 7 to minimize the effect of co-channel interference?

Key equations

1. The difference in power reception at two different distances d_1 and d_2 would be

$$\frac{\mathrm{Pr}_2}{\mathrm{Pr}_1} = \left(\frac{d_2}{d_1}\right)^{-4}$$

2. The propagation path loss

$$\mathrm{Pr}_{loss} = \alpha\ d^{-\gamma}$$

3. D/R ratio

$$\frac{D}{R} = q = \sqrt{3N}$$

4. The number of cells in the larger hexagon ($A_{\text{larger hexagon}}$) can be determined by comparing equation

$$L = \frac{A_{\text{larger hexagon}}}{A_{\text{small hexagon}}} = 3(i^2 + j^2 + ij) = \frac{D^2}{R^2}$$

5. If the market has a geographic area of A_{market}, then the number of clusters M is given by

$$M = \frac{A_{\text{market}}}{A_{\text{cluster}}} = \frac{A_{\text{market}}}{N\frac{3\sqrt{3}}{2}R^2}$$

Further reading

Beach M., *Propagation and System Aspects*, University of Bristol, Future Communication Systems course, April 1994.

Classen, F., Meyr, H., and Sehier, P., Maximum likelihood open loop carrier synchronizer for digital radio, Proceedings ICC'93, pp. 493–497, 1993.

Firmanto, W. T. and Gulliver.T. A., Code Combining of Reed–Muller Codes in An Indoor Wireless Environment, Wireless Personal Communications 6, 359–371, 1997, Kluwer Academic Publishers, the Netherlands.

Lee, W. C. Y., Elements of Cellular Radio System. IEEE Transactions on Vehicular Technology, 35, 48–56, May 1986.

Lee, W. C. Y., *Mobile Cellular Telecommunications System*. New York: McGraw-Hill, 1989.

Van Nee, R. and Prasad R., *OFDM for Wireless Multimedia Communications*, Artech House, Boston, pp. 80–81, 2000.

Proakis, J. G., *Digital Communication*, 3rd Ed. New York, NY, USA: McGraw-Hill. Inc. 1995

Rappaport, T. S., *Wireless Communications*, New Delhi: PHI Pvt. Ltd., 2007.

Tse, D. and Viswanath, P., *Fundamentals of Wireless Communication*, Cambridge: Cambridge University Press 2005.

Elements of Cellular Radio System Design 3

3.1 Introduction

The main requirement of mobile communication systems is to make services available regardless of time and place. This requirement enforces the reuse of limited radio frequency spectrum. The term co-channel interference (CCI) comes with the concept of frequency reuse.

In frequency reuse, several cells in the coverage area use the same set of frequencies.

The aim of any cellular system design is to efficiently utilize the available spectrum such that the CCI does not degrade the performance of the system below the guaranteed grade of service (GOS).

In this chapter, the various elements of cellular system design are discussed in detail. To this end, the chapter is divided into seven sections. Section 3.2 deals with the concept of frequency reuse channels. In Section 3.3, the CCI concept is introduced and the CCI reduction factor is derived. Sections 3.4 and 3.5 are concerned with the design of a cellular system in both the normal-case and worst-case scenarios using an omnidirectional antenna. The important concept of cell splitting is introduced in Section 3.6, and the challenges in cell splitting are discussed in Section 3.7. Finally, in Section 3.8 the components of a cellular system are considered for discussion as their choice affects the system design.

A challenge faced by the designer of any cellular system is that of serving the greatest number of customers with a specified system quality. In the process of meeting this challenge, a number of questions must be answered, which include listing out design specifications like the number of customers who can be served in a busy hour, the number of subscribers who can be taken into the system, and the number of frequency channels required.

3.2 Concept of frequency reuse channels

In cellular technology, a radio channel consists of a pair of frequencies used in a two-way radio link that is maintained between the user's mobile phone and the base station. The major drawback with earlier mobile communication systems was the inefficient use of allocated frequency spectrum. Repeated reuse of radio frequencies over a given geographic area provides a number of simultaneous conversations far in excess of the voice channels derived from simply parcelling out the available spectrum. As shown in Figure 3.1, a particular radio channel, say f_1, used in one

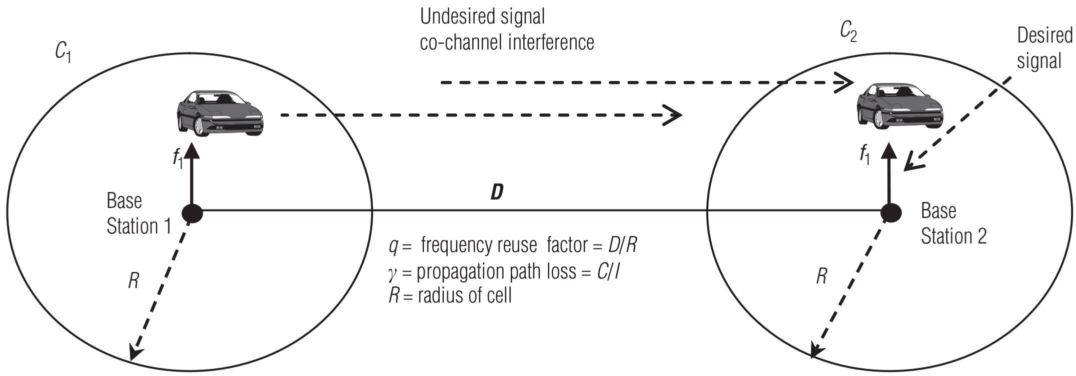


Figure 3.1 Ratio of D/R

cell, say C_1, with a coverage radius R can be used in another cell with the same coverage radius at a distance D. In a frequency reuse system, users in different cells may simultaneously use the same frequency channel. Though the frequency reuse system leads to increase in the spectrum efficiency, serious interference may occur if the system is not properly designed. CCI is a major concern in the concept of frequency reuse.

> *CCI occurs when the same carrier frequencies reach the mobile user from the transmitters of two base stations.*

3.2.1 Frequency reuse schemes

The frequency reuse concept can be used in both the time and space domains. In the time domain, frequency reuse results in different time slots being occupied by the same frequency. This is called *time-division multiplexing* (TDM). In the space domain, frequency reuse can be divided into two categories:

- Two different geographic areas being assigned the same frequency.
- Same frequency being repeatedly used in the same area in one system. There are many co-channel cells in the system. The total frequency spectrum allocation is divided into K frequency reuse patterns. This concept is explained in detail in Chapter 2.

3.2.2 Number of customers in the system

During a busy hour, the traffic conditions in the area help determine both the sizes of different cells and the number of channels in them when we design a system. At each particular cell, the traffic conditions drive the maximum number of calls per hour per cell. The maximum number of calls per hour can be taken care of in each cell after the maximum number of frequency channels per cell has been implemented in each cell.

3.3 Co-channel interference

The basic idea with cellular systems, as noted in the previous chapter, is to reuse frequency in different cells and to increase the capacity. But two problems arise. (When channels have been assigned to two co-channel cells that are not far enough apart geographically and their signals are strong enough to overlap with each other, they cause an interference called co-channel interference.) And the same frequency assignments cannot be made in adjacent cells because of *inter-channel interference*. The assignments must be spaced far enough apart geographically to keep interference to tolerable levels.

Unlike thermal noise, CCI can be overcome by increasing the carrier-to-interference ratio (CIR), defined to be the ratio of the desired average signal power at receiver to the total average interference power. This ratio is comparable with the signal-to-noise ratio and can be used as a performance measure in non-mobile communication systems. The signal-to-interference ratio (SIR) should then be greater than a specified threshold for proper signal operation.

For example, in the first-generation (1G) Advanced Mobile Phone System (AMPS), designed for voice calls, the desired performance threshold, in decibels (dB), is SIR_{dB} which is equal to 18 dB. ($SIR_{dB} = 10 \log_{10} SIR$.) This number has been found to be appropriate to provide acceptable voice quality in studies of mobile calls. For the second-generation (2G) digital AMPS (D-AMPS), a threshold of 14 dB is deemed suitable.

3.3.1 Co-channel interference reduction factor

CCI between the cells limits frequency reuse and hence is a major problem. Our aim is to find the minimum frequency reuse distance in order to reduce this CCI.

With the assumption that the size of all cells is the same, the cell size is determined by the coverage area of the signal strength in each cell. CCI is independent of the transmitted power of each cell as long as the size of the cell is fixed. It is a function of a parameter q and is defined as

$$q = \frac{D}{R} = \sqrt{3N} \tag{3.1}$$

where
D is the distance between the centres of cells
R is the radius of the cell
q is the reuse ratio
N is the cluster size

Distance to reuse ratio (D/R) defines the geographic distance required between the cells that are using same frequencies in order to avoid interference between these cells.

Here, q is the CCI reduction factor and N is the cluster size. CCI decreases when q increases. Further, the separation D in Equation (3.1) can be written as a function of K_1 and C/I,

$$D = f(K_1, C/I) \tag{3.2}$$

where
K_I is the number of co-channel interfering cells in the first tier
C/I is the received CIR at the desired receiver

$$\frac{C}{I} = \frac{C}{\sum_{k=1}^{K_1} I_k} \tag{3.3}$$

In a hexagonal-shaped cellular system, there are always six co-channel interfering cells in the first tier as shown in Figure 3.2; that is, $K_I = 6$.

Assume that the interference level is very high; the local noise can be neglected. Then, C/I can be expressed as

$$\frac{C}{I} = \frac{R^{-\gamma}}{\sum_{k=1}^{K_1} D_k^{-\gamma}} \tag{3.4}$$

where the actual terrain environment determines γ, the propagation path-loss slope. Usually, $\gamma = 4$ in a mobile radio medium. Weaker interference is caused by the six co-channel interfering cells in the second tier than those in the first tier and hence is negligible. Substituting Equation (3.1) into Equation (3.4),

$$\frac{C}{I} = \frac{1}{\sum_{k=1}^{K_1} \left(\frac{D_k}{R}\right)^{-\gamma}} = \frac{1}{\sum_{k=1}^{K_1} (q_k)^{-\gamma}} \tag{3.5}$$

where q_k is the CCI reduction factor with kth co-channel interfering cell, and

$$q_k = \frac{D_k}{R} \tag{3.6}$$

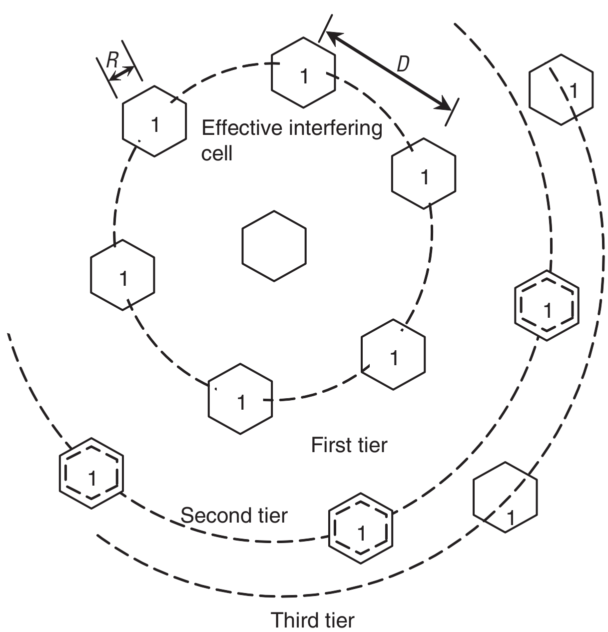


Figure 3.2 Six effective interfering cells of Tier 1

3.3.2 Relation between co-channel reduction factor and frequency reuse factor

From the above analysis, by substituting $K_I = 6$ in Equation (3.5), the C/I can be obtained as

$$\frac{C}{I} = \frac{1}{\sum_{k=1}^{6}\left(\frac{D_k}{R}\right)^{-\gamma}} \tag{3.7}$$

where
$2 \leq \gamma \leq 5$ is the propagation path loss, and depends upon the terrain environment
D_k is the distance between the mobile and the k^{th} interfering cell
R is the cell radius
If we assume D_k to be the same for the six interfering cells for simplification, or $D = D_k$, then Equation (3.7) becomes

$$\frac{C}{I} = \frac{1}{6(q)^{-\gamma}} = \frac{q^{\gamma}}{6} \tag{3.8}$$

where $q = D/R$ is the reuse factor.
Therefore,

$$q = \left[6\left(\frac{C}{I}\right)\right]^{\frac{1}{\gamma}} \tag{3.9}$$

Comparing Equations (3.9) and (3.1), that is, $q = \sqrt{3N}$, we have

$$N = \frac{1}{3}\left[6\left(\frac{C}{I}\right)\right]^{\frac{2}{\gamma}} \tag{3.10}$$

Example problem 3.1

Consider the AMPS in which the C/I of 18 dB is required for the accepted voice quality. What should be the cluster size for the system? Assume $\gamma = 4$. What will be the cluster size of the GSM system in which a C/I of 12 dB is required?

Solution

Using Equation (3.10), we get

$$N_{\text{AMP}} = \frac{1}{3}\left[6(10)^{1.8}\right]^{\frac{2}{4}} = 6.486 \sim 7$$

and

$$N_{\text{GSM}} = \frac{1}{3}\left[6(10)^{1.2}\right]^{\frac{2}{4}} = 3.251 \sim 4$$

Example problem 3.2

Consider a cellular system with 395 total allocated voice channel frequencies. Calculate the mean *C/I* for cell reuse factor equal to 4, 7, and 12. Assume omnidirectional antennas with six interferers in the first tier and a slope for path loss of 40 dB/decade ($\gamma = 4$).

Solution

For a reuse factor $q = 4$, the number of voice channels per cell site = 395/4 = 99.

Using Equation (3.10), then mean *C/I* for $N = 4$ is

$$4 = \frac{1}{3}\left[6\left(\frac{C}{I}\right)\right]^{\frac{2}{4}}$$

$\therefore$ *C/I* = 24(13.8 dB)
Similarly,
For $N = 7$
$\therefore$ *C/I* = 18.7 dB
And for $N = 12$
$\therefore$ *C/I* = 23.3 dB

Example problem 3.3

The base station transmitter antenna in a cell site radiates a +10 dBm RF signal and is connected using an RF coaxial cable that induces a 3 dB loss. The cable has two connectors at its either end that induce a loss of 2 dB each. What is the signal level at the input of the antenna system?

Solution

Given data:
The RF output signal level of cell-site transmitter antenna = +10 dBm
In terms of dB = +10 − 30 = −20 dB
Signal loss due to RF coaxial cable = 3 dB
Determination of total loss in cable connectors
Signal loss due to one connector of RF coaxial cable = 2 dB
Number of connectors on a coaxial cable = 2
Therefore, signal loss due to both connectors = 2 × 2 = 4 dB
Determination of signal loss due to cable and connectors
Total signal loss due to cable and connectors = 3 + 4 = 7 dB
Determination of signal level at the input of the antenna

Signal level at the input of the antenna = –20 – 7 = –27 dB
In terms of dBm = –27 + 30 = 3 dBm
Hence, signal level at the input of the antenna system = +3 dBm.

3.4 Desired *C/I* from normal case in an omnidirectional antenna system

The desired CIR (*C/I*) is related to the co-channel reuse factor q and is given in Equation (3.9). For an isotropic or omnidirectional antenna system, the desired *C/I* can be obtained with a simple analytical solution in which two cases are considered.

3.4.1 Analytic solution

In order to obtain the desired *C/I*, two cases are to be considered:

- The signal and CCI received by the mobile unit
- The signal and CCI received by the cell site

Both the cases are shown in Figure 3.3.

The local noises at the mobile unit and the cell site are small and can be neglected when compared with the interference level. The system is called a *balanced system* as long as the received CIRs at both the mobile unit and the cell site are the same. In order to analyse the system requirement, either of the cases can be chosen in a balanced system. In a simplified system,

$$\frac{C}{I} = \frac{R^{-\gamma}}{6D^{-\gamma}} = \frac{q^{\gamma}}{6} \tag{3.11}$$

Thus,

$$q^{\gamma} = 6\frac{C}{I} \tag{3.12}$$

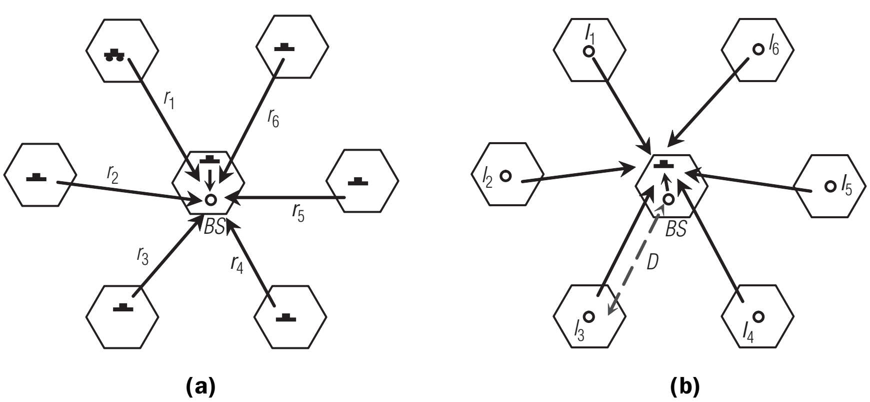


Figure 3.3 Co-channel interference from six interferers: (a) Receiving at the cell site and (b) Receiving at the mobile unit

and

$$q = \left(6\frac{C}{I}\right)^{\frac{1}{\gamma}} \tag{3.13}$$

In Equation (3.13), the values of C/I and γ are based on the required system performance and the terrain environment, respectively. On the basis of subjective tests, the normal cellular practice is to specify C/I to be 18 dB or higher. In a mobile radio environment, the path-loss slope γ is equal to about 4.

$$q = D/R = (6 \times 63.1)^{1/4} = 4.41 \tag{3.14}$$

Substituting q from Equation (3.14) in

$$q = \frac{D}{R} = \sqrt{3N} \tag{3.15}$$

We get, $N = 7$.

Equation (3.15) indicates that a seven-cell reuse pattern is needed for a C/I of 18 dB. The greater the value of q, the lower the CCI.

3.5 Cellular system design in worst-case scenario with an omnidirectional antenna

When the mobile unit receives the weakest signal from its own cell site but strong interferences from interfering cell sites, it is considered to be the worst-case scenario. Let us re-examine the seven-cell reuse pattern and consider the worst case in which the mobile unit is located at the cell boundary as shown in Figure 3.4. The distances from the six interfering cells are given in Figure 3.4.

The C/I can be given as

$$\frac{C}{I} = \frac{R^{-\gamma}}{2(D-R)^{-\gamma} + 2D^{-\gamma} + 2(D+R)^{-\gamma}} \tag{3.16}$$

Using $D/R = q$ and the path-loss exponent $\gamma = 4$, Equation (3.16) can be written as

$$\frac{C}{I} = \frac{1}{2(q-1)^{-\gamma} + 2q^{-\gamma} + 2(q+1)^{-\gamma}} \tag{3.17}$$

where $q = 4.6$ for a normal seven-cell reuse pattern ($N = 7$).

Substituting $q = 4.6$ in Equation (3.17), we get $C/I = 54.3$ or 17.3 dB. If we use the shortest distance ($D - R$), then

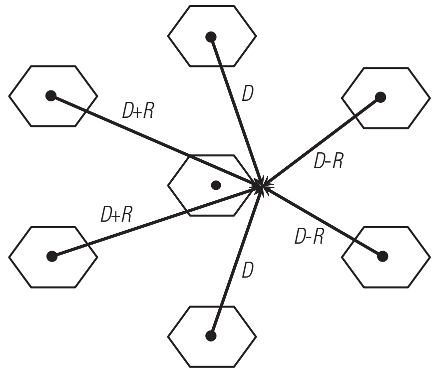


Figure 3.4 Worst-case scenarios for co-channel interference

$$\frac{C}{I} = \frac{1}{6(q-1)^{-4}} = \frac{1}{6(3.6)^{-4}} = 28(14.47\,\text{dB}) \tag{3.18}$$

The C/I could be 14 dB or lower in real situations. This can be due to imperfect cell-site locations, terrain configuration, and also heavy traffic. The cellular system should, hence, be designed around the C/I of worst case and for a seven-cell reuse (N = 7) pattern. We conclude that when considering the worst-case scenario, q = 4.6 is not enough in an omnidirectional cell system.

Therefore, in an omnidirectional cell system, $N = 9$ or $N = 12$ would be a correct choice. Then the values of q are

$$q = \begin{cases} \dfrac{D}{R} = \sqrt{3N} & \\ 5.2 & N = 9 \\ 6 & N = 12 \end{cases} \tag{3.19}$$

By substituting these values in Equation (3.17), we obtain

$$\frac{C}{I} = 84.5 \text{ or } 19.25 \text{ dB for } N = 9 \tag{3.20}$$

$$\frac{C}{I} = 179.33 \text{ or } 22.54 \text{ dB for } N = 12 \tag{3.21}$$

The N = 9 and N = 12 cell patterns, shown in Figures 3.5 and 3.6, are used when the traffic is light. Each cell covers an adequate area with adequate numbers of channels to handle the traffic.

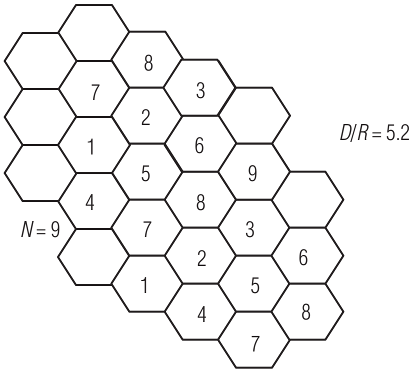


Figure 3.5 Interference with frequency reuse pattern, $N = 9$

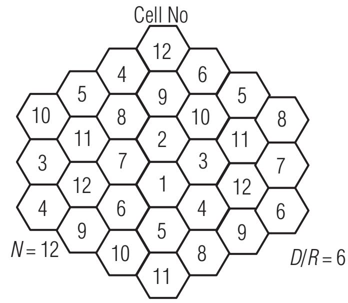


Figure 3.6 Interference with frequency reuse pattern, $N = 12$

Example problem 3.4

Determine the signal-to-co-channel interference ratio C/I at the mobile receiver located at the boundary of its omnidirectional operating cell, under the influence of interfering signals from six co-channel interfering cells in the first tier in a cellular system designed with (a) $N = 7$, (b) $N = 9$, and (c) $N = 12$.

Assume the path-loss exponent is 4. What would be the correct choice of N for a system requiring $C/I = 18$ dB?

Solution

Given data:

Configuration of cellular antenna design is omnidirectional.
Number of co-channel interferers = 6
The path-loss exponent $\gamma = 4$

(a) To determine the C/I at the mobile receiver for $N = 7$
To determine frequency reuse ratio, q
We know that:

$$q = 3N$$

$$\text{For } N = 7,\ q = 21 = 4.6.$$

To determine the C/I in the worst-case analysis.

In an omnidirectional antenna design of a cellular system based on regular hexagonal pattern, the worst case C/I, under the influence of interfering signals from six co-channel interfering cells in the first tier, is given by:

$$\frac{C}{I}(\text{Omni}) = \frac{1}{2(q-1)^{-\gamma} + 2q^{-\gamma} + 2(q+1)^{-\gamma}}$$

On substituting $q = 4.6$ and $\gamma = 4$, we get

$$\frac{C}{I}(\text{Omni}) = \frac{1}{2(4.6-1)^{-4} + 2(4.6)^{-\gamma} + 2(4.6+1)^{-4}} = 54$$

C/I (ratio) in decibels can be computed as

We know that C/I (dB) = 10 log10 C/I (ratio)
Or, C/I (dB) = 10 log 54
C/I (dB) = 17 dB

(b) Determine the C/I at the mobile receiver for $N = 9$
To determine frequency reuse ratio, q

$$q = \sqrt{3N} = 5.2$$

To determine the C/I in the worst-case analysis
On substituting $q = 5.2$ and $\gamma = 4$ in the relationship,
that is, on substituting these values in Equation (3.17), we get C/I =95
C/I in dB = 10 log 84.5 = 19.7 dB

(c) Determine the C/I at the mobile receiver for $N = 12$
To determine frequency reuse ratio, q

$$q = \sqrt{3N} = \sqrt{36} = 6$$

To determine the C/I in the worst-case analysis
On substituting q = 6 and $\gamma = 4$ in Equation (3.17), we get C/I =179.3
C/I in decibels = 10 log 84.5 = 22.5 dB
Correct choice of the value of N
Required value of C/I = 18 dB (given)
From the above results, C/I = 17 dB for $N = 7$, which is lower than the desired value of C/I = 18 dB. So the system design for $N = 7$ is imperfect.
It is observed that the values of C/I are 19.2 and 22.5 dB for $N = 9$ and for $N = 12$, respectively, which are higher than the desired value of C/I = 18 dB.
Therefore, in an omnidirectional cellular system, $N = 9$ or $N = 12$ cell patterns would be a correct choice.

3.6 Cell splitting

Theoretically, a cellular system can provide services for an unlimited number of users. However, once a system is installed, it can only provide services to a certain fixed number of users. As soon as the number of users increases and approaches the maximum that can be served, some technique must be developed to accommodate the increasing number of users. There are various techniques to enhance the capacity of a cellular system. One technique is *cell splitting*, a mechanism by which cells are split into smaller cells, each having the same number of channels as the original large cells, as shown in Figure 3.7.

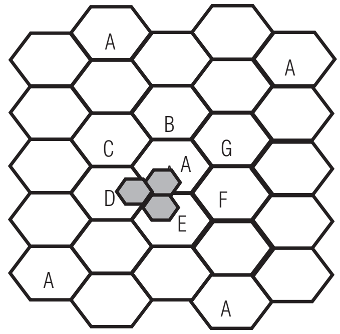


Figure 3.7 Seven-cell cluster layout and cell splitting

3.6.1 Types of cellular networks

On the basis of cell radius, there are four types of cellular networks:

- Macrocells
- Microcells
- Picocells
- Nanocells

3.6.1.1 *Macrocellular radio networks*

Macrocellular radio networks are mainly used to cover large areas with low traffic densities. Macrocells have radii between 1 and 10 km. Large macrocells, with radii between 5 and 10 km or even higher, are used for rural areas. Small cells, with radii between 1 and 5 km, are used if the traffic density in large cells is so high so as to cause blocking of calls. They, thus, provide large cells with extra capacity (cell splitting). Planning small cells is more difficult because traffic predictions for relatively small areas cannot be easily done, and these signals undergo multipath Rayleigh fading and lognormal shadowing.

3.6.1.2 *Microcellular radio networks*

Microcellular radio networks are used in areas with high traffic density, like urban areas. The cells have radii between 200 m and 1 km. For such small cells, it is hard to predict traffic densities and area coverage because the shape of the cell is time dynamic (i.e., the shape changes from time to time) due to propagation characteristics. One-dimensional microcells are placed in a chainlike manner along main highways with high traffic densities, whereas “two-dimensional” refers to the case where an antenna transmits the main ray and two additional rays are reflected off buildings on both sides of the street. One-dimensional microcells usually cover one or two house blocks. In these cells, antennas are placed at street lamp heights; therefore, the surrounding buildings block the signal propagation to adjacent co-channel cells. This improves the ability to reuse frequencies, as co-channel interference is reduced drastically by the shadowing effect caused by the infrastructure. The typical rms delay spread is 2 μs.

3.6.1.3 *Picocellular radio networks*

A *picocell* is the smallest of the cells in a wireless system, normally covering an office area to support an in-building wireless system. The term picocell is most commonly used in connection with third-generation (3G) personal communications systems, especially GSM. Picocells are almost always placed indoors. Picocells are also known as indoor cells, and they have cell radii between 10 and 200 m. For indoor applications, cells have three-dimensional structures. Fixed cluster sizes, fixed channel allocations, and prediction of traffic densities are difficult for indoor applications. Presently, picocellular radio systems are used for wireless communication in offices. The path-loss exponent varies from 1.2 to 6.8. Signals in picocells are always Rician faded and the Rician parameter lies between 6.8 and 11 dB. Typical values of rms delay spread lie between 50 and 300 ms.

3.6.1.4 Nanocellular radio networks

A *nanocell* is a miniature, portable radio unit that establishes cell phone service wherever coverage or capacity is needed. Installation is fast and easy. Nanocells require only a power source and a means for connecting the base station to a laptop PC, and then the minicell is ready to operate. Nanocells are the most cost-effective cell nodes available today. The design of a nanocell combines a small form factor with low power consumption to enable installation in a wide range of environments. Similar to picocells, a nanocell can be mounted on the wall, in a vehicle, in a briefcase, or in a weatherproof outdoor enclosure. While there is no limit to the number of users who can be allowed on a nanocell, up to seven handsets may access the network at one time in a typical installation. For greater coverage or capacity, multiple distributed nanocell base stations can be linked together using standard Ethernet cables. Up to four nanocell units can be combined and configured as a single network managed by one computer. The actual range of a nanocell or picocell will be determined by the antenna gain, terrestrial obstructions, and building materials, among other things.

The macrocells, microcells, picocells, and nanocells differ in the radius of coverage provided by each. This coverage is defined by the power output and antenna gain used in each system.

3.7 Challenges in cell splitting

When a new system is deployed, the user demand is less and users are assumed to be uniformly distributed over the area to be served. The initial system layout is designed to provide uniformly reliable coverage and uniform capacity over the entire service area. As the number of users increases, the demand for channels may begin to exceed the capacity of some base stations. This increased demand often first shows up in those areas of cities, where the population is dense during the working day. Once a system has been initially deployed, a system-wide reduction in the cluster size may not be needed, since user density does not grow at the same rate in all parts of the system.

Cell splitting is a technique that provides the capability to add smaller cells in specific areas of the system to support the increased demand in those areas, while minimizing the need to modify the existing cell parameters.

Cell splitting is based on the reduction of cell radius. It allows the system to grow gradually in response to an increasing demand of traffic. It takes place by reducing the cell radius by half and splitting the old cell into four new smaller cells. The reuse frequency can be used more often allowing the traffic to increase by four times in the same area where an old cell was placed. The ideal location for new cells is the mid-point between the neighbouring existing cells.

Smaller cells lead to increase in number of cells, which in turn requires additional base stations. The following are two challenges faced while increasing the system capacity by reducing the cell radius:

- Reconfiguring the system such that the existing base station towers do not have to be moved.
- Meeting an increasing demand that may vary between different geographic areas of a system.

For example, a city may have the highest density of users and, therefore, should be supported by cells with the smallest radius. The radii of cells in a typical system generally increase as one moves from urban to rural areas. This is because the user density typically decreases as one moves away from a city.

The key challenge, therefore, is to add the minimum number of smaller cells wherever increased demand dictates the need for increased capacity.

A gradual addition of new base stations and smaller cells means that the cellular system may have to operate with cells of more than one size. When two or more sizes of cells coexist in a mobile network, special attention needs to be paid to ensure that the minimum frequency reuse distance is being respected.

The use of channels in the new cell sites will not cause interference problems in the larger system. This is because the strength of the signal is smaller and is designed to comply with the constant D/R, but considering the radius of the small cell. The problem arises with the channels of the co-channel cell in the larger system because the small cell is within the frequency reuse distance of the larger cell. One way to deal with this problem is by making use of the *overlaid cell concept*.

3.7.1 Overlaid cell concept

In the overlaid cell concept, the cellular network is seen as a superposition of the smaller cell pattern on top of the complete larger cell pattern.

Each cell face will divide its channels between a larger and a smaller cell group. The selected channels in the larger group will be used in all the coverage areas of a larger cell. The selected channels installed in the smaller cell will make the smaller cell group. The formation and use of the channel groups is then governed by the presence or absence of real smaller cell neighbours.

For example, the use of any channel installed in a smaller cell must be restricted to the other smaller cell overlay areas in the nearest larger co-channels. If a mobile user using a smaller cell channel goes out of the perimeter of the small cell overlay area on the larger cell, it needs to handoff to a channel of the larger group or to a neighbouring cell. Therefore, the presence of two sizes of cells in a network reduces the capacity of some of the larger cells. This can force their cells to split even if they would not consider the growth of traffic in their areas.

Figure 3.8(a) shows a cellular layout with seven cell clusters. Assume that the cells in the centre of the diagram are congested and cell A in the centre is approaching user capacity. Figure 3.8(b) shows an overlay of smaller cells superimposed on the original layout. The new smaller cells have half the radius of the original cells. At half the radius, the new cells will have one-fourth of the area and consequently will need to support one-fourth the number of subscribers.

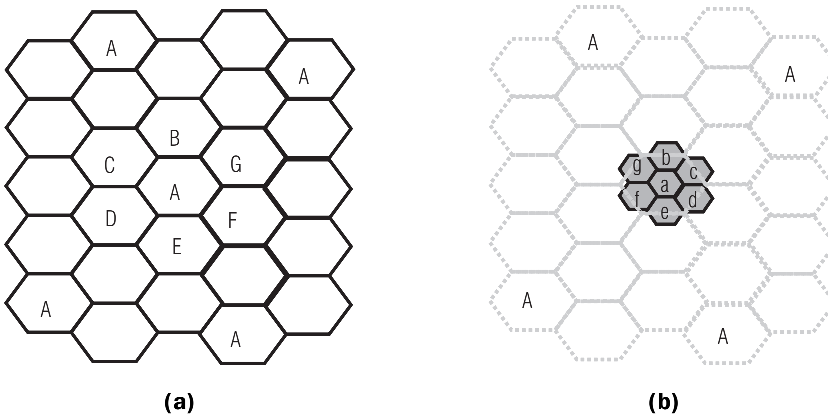


Figure 3.8 (a) Seven-cell cluster layout before splitting and (b) Cell splitting: overlay of half-radius cells

Notice that one of the new smaller cells lies in the centre of each of the larger cells. If we assume that base stations are located at the cell centres, this allows the original base stations to be maintained as the new cell pattern spreads outward from the centre. Of course, new base stations will have to be added for new cells that do not lie in the centre of the larger cells.

The organization of cells into clusters is independent of the cell radius. Owing to this, the cluster size can be the same in the small cell layout as it was in the large cell layout. Recall that the CIR is determined by the cluster size and not by cell radius.

Consequently, if the cluster size is maintained, the CIR will be the same after cell splitting as it was earlier. If the entire system is replaced with new half-radius cells and the cluster size is maintained, the number of channels per cell will be exactly as it was before and the number of subscribers per cell will have been reduced. In large cities, it is common for the cellular systems to be configured into "microcells" whose radii are measured in hundreds of meters, rather than in kilometres.

When the cell radius is reduced by a factor, it is also possible and desirable to reduce the power in transmitted signals.

The minimum required power level is determined by the need to maintain an adequate signal-to-noise ratio over a significant fraction of the cell area. This in turn requires a minimum signal-to-noise ratio at the cell radius.

Note that we are concerned with signal-to-noise ratio and not with CIR. The former is determined by the received power and noise level, whereas the latter depends only on the cluster size. However, the noise level is determined by the receiver noise figure, a parameter that is completely independent of the cell radius or layout. Therefore, we can focus on the received signal power. The received signal power at a distance d from the transmitting antenna is given by

$$P_{\mathrm{r}} = \frac{P_{\mathrm{r}0}}{(d/d_0)^{\gamma}} \tag{3.22}$$

where P_{r0} is the reference power received at some reference distance d_0. P_{r0} is directly proportional to the transmitted power and γ is the path-loss exponent. The following example illustrates the idea.

Example problem 3.5

In a certain cellular system, the base stations radiate 15 W. Suppose that the cells are split and the new cells have one-fourth the radius that the original cells had. Find the power that the base stations in the new layout must transmit to maintain the signal-to-noise ratio at the cell boundaries. The path-loss exponent is $\gamma = 4$.

Solution

Let P_{rOld} be the reference power measured at d_0 in the original cell configuration. For $\gamma = 4$, the received power at the cell boundary $d = R_{Old}$ is

$$P_r = \frac{P_{rOld}}{(R_{Old}/d_0)^4}$$

When the cells are split, the cell radius becomes $R_{New} = R_{Old}/4$. We want to maintain the received powers at the cell radius at P_r. We have

$$P_r = \frac{P_{rOld}}{(R_{Old}/d_0)^4} = \frac{P_{rNew}}{\left(\dfrac{R_{Old}/4}{d_0}\right)^4}$$

where P_{rNew} is the new reference power measured at d_0. On rearrangement,

$$P_{rNew} = P_{rOld}(1/4)^4$$

The transmitted powers change in the same proportion as the reference powers. Therefore,

$$P_{rNew} = 15(1/4)^4 = 0.0585 \text{ W}$$

In decibels, the power reduction is 10log (4^4) = 24 dB, from

$$P_{rOld}\,|_{dB} = 11.7 \text{ dB to } P_{rNew}\,|_{dB} = 12.3 \text{dB}.$$

If a cellular layout is replaced by a new layout with a smaller cell radius, the CIR will not change, provided the cluster size does not change. Special care must be taken, however, to avoid CCI when both large and small cell radii coexist. Therefore, the only way to avoid interference between the large cell and small cell systems is to assign entirely different sets of channels to the two systems.

For example, if a large cell system becomes congested in the city, channels can be taken away from the large cell system to make up channel sets for the small cell system. The capacity of the large cell system will be reduced, but the large cell system will now be used primarily in the suburban areas where the user density is low. The small cell system also does not have a full complement of channels. But as the cell area is small, there may not be enough users per cell to demand a full channel set. As the small cell system continues to spread, more and more channels can be reassigned from the large cell to the small cell system, until, ultimately, the large cell system is completely replaced.

3.8 Consideration of the components of the cellular system

In the previous sections, the elements of cellular mobile radio system design have been discussed. Now, we consider the various components of cellular systems, such as mobile radios, antennas, cell site, base-station controller (BSC), and mobile telephone switching centre (MTSO). Wrong choice of these components would affect our system design. A general view of the cellular system is shown in Figure 3.9. The various issues that affect the choice of antennas, switching equipment, and data links are described in this section.

3.8.1 Antennas

Various issues like antenna pattern, gain, tilting, and height affect the cellular system design. In both the vertical and the horizontal planes, the antenna pattern can be of any shape and the transmitted power is compensated by the antenna gain. The antenna patterns in cellular systems differ from the patterns seen in free space. Antenna tilting can enhance the weak spots in the cell by reducing the interference to the neighbouring cells. In addition, the area and shape of the coverage in the system are affected by the height of the cell-site antenna. For more details, the reader is referred to Chapter 8.

3.8.2 Switching equipment

In cellular systems, the capacity of the switching equipment is based on the capacity of the processor associated with the switches. The service life of the switching equipment is determined by the time it takes to reach its full capacity, not by the life cycle of the equipment. More modules can be added to increase the capacity of the equipment if the switching equipment is designed in modules, or as distributed switches. Digital switches may be more suitable for decentralized systems. The future trend seems to be the utilization of system handoff. This means that a call

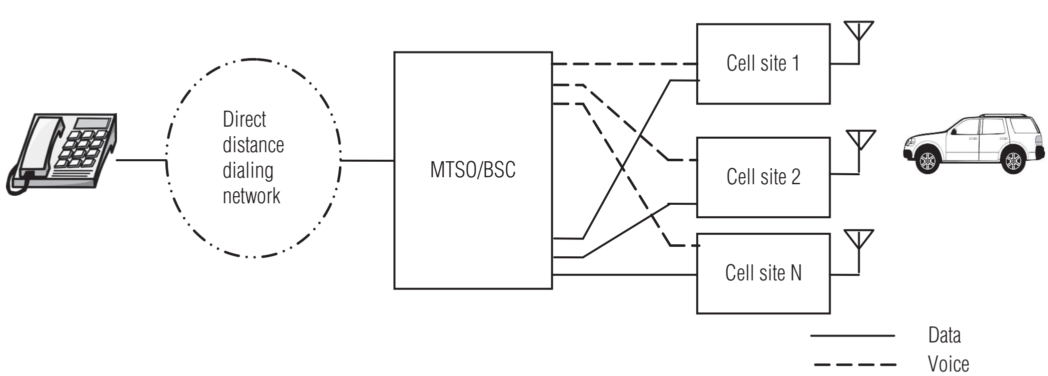


Figure 3.9 A general view of cellular telecommunication systems

can be carried from one system to another without the call being dropped by successive linking of switching equipment.

3.8.3 Data links

The data links are shown in Figure 3.9. Each data link can carry multiple channels of data from the cell site to the MTSO. This high-speed data transmission cannot be passed through a regular telephone line; data bank devices are needed for this purpose. They can be multiplexed with many data channels where the frequency is much higher than 850 MHz.

3.9 Summary

- The frequency reuse concept can be used in both the time and space domains.
- Frequency reuse: Several cells in coverage area use the same set of frequencies.
- Frequency reuse in the time domain takes the form of time division multiplexing. In the space domain, frequencies can be assigned in two different geographic areas or can be repeatedly used in the same area in one system.
- The value of C/I is based on the required system performance and the specified value of γ is based on the terrain environment. Normal cellular practice is to specify C/I to be 18 dB or higher based on subjective tests.
- CCI is a function of a parameter given by $q = D/R$, where q is the CCI reduction factor.
- Channel interference areas in a cellular system can be detected in two ways: To find the CCI area from a mobile receiver and to find the CCI area which affects a cell site.
- Cell splitting allows an array of cells to be replaced with a similar array of cells having a smaller radius. The cluster size and, therefore, the level of CCI are not changed. The increase in user density can be dramatic, but additional base stations must be added to the system. It is important to take advantage of the geometry so that the existing base stations do not have to be moved when cells are split.
- It is necessary to consider the various components of cellular systems, such as mobile radios, antennas, etc., while designing a cellular mobile radio system.

Review questions

1. Explain the co-channel interference reduction factor and derive the general formula for C/I.
2. Define co-channel interference and its reduction factor.
3. Derive the co-channel interference reduction factor for seven-cell reuse pattern.
4. Derive the C/I in an omnidirectional antenna system.
5. Derive the C/I in worst-case scenario with an omnidirectional antenna.
6. What are the components in a cellular system? Explain briefly.
7. Define cell splitting. How does cell splitting affect the system design?
8. Why is cell splitting needed? Define 4:1 and 3:1 cell splitting.
9. Explain cell splitting in detail and mention the various advantages of cell splitting concept.
10. Explain overlaid cell concept in detail.
11. Determine the minimum cluster size for a cellular system designed with an acceptable value of signal-to-co-channel interference ratio $C/I = 18$ dB. Assume the path-loss exponent as 4 and co-channel interference at the mobile unit from six equidistant co-channel cells in the first tier (Answer: $C/I = 63.1$, $q = (6 \times C/I)^{1/4} = 4.41$, $N = 7$).

12. In Problem 11, if the acceptable C/I is enhanced to 20 dB, will the cluster size determined be adequate? If not, then what should be the cluster size? (Answer: C/I for $N = 7$ is 74.2, C/I (18.73 dB), $N = 7$ cannot meet the desired C/I requirement, new frequency ratio $q = (6 \times C/I)^{1/4} = 4.41$, new N for $C/I = 20$ dB).
13. A cell-site transmitter generates a +15 dBm RF signal and is connected to an antenna using an RF coaxial cable that induces a 3 dB loss. The cable has two connectors at its either end that induce a loss of 2 dB each. What is the signal level at the input of the antenna? (Answer: 8 dBm)
14. Explain about the general description of the cellular radio system. (Refer Section 3.2)
15. Draw the general view of telecommunication and explain the function of the each unit. (Refer Section 3.8)
16. Mention the two frequency reuses schemes and explain N-cell reuse pattern in detail for four- and seven-cell reuse with illustrative diagrams. (Refer Section 3.2)
17. Derive the CCI reduction factor for seven-cell reuse pattern. (Refer Section 3.3)
18. Describe about desired C/I from a normal case in an omnidirectional antenna system. (Refer Section 3.4)
19. What do you mean by desired C/I? Explain. (Refer Section 3.4.1)

Objective type questions and answers

1. The desired performance threshold for proper signal operation in the first-generation AMPS system is
   (a) 4 dB (b) 12 dB (c) 18 dB (d) 24 dB
2. Reuse pattern needed for a C/I of 18 dB is
   (a) 7-cell (b) 4-cell (c) 12-cell (d) 2-cell
3. Reuse pattern needed in an omnidirectional antenna in the worst case is
   (a) 4-cell (b) 7-cell (c) 15-cell (d) 9-cell
4. The cellular network seen as a superposition of the smaller cell pattern on top of the complete larger cell pattern is
   (a) underlaid-cell concept (b) overlaid-cell concept
   (c) small cell superposition (d) none of the above
5. It is possible to reduce the power in transmitted signals if the cell radius is
   (a) increased (b) kept constant (c) made zero (d) reduced
6. The concept of repeating the same frequency is known as
   (a) cell splitting (b) frequency reuse
   (c) co-channel interference (d) none of the above
7. To increase the capacity of equipment while designing switching equipments the modules can be
   (a) distributed (b) reduced (c) added (d) altered
8. Which of the following design and concept sets cellular communications apart from all the preceding mobile radio systems?
   (a) TDMA (b) frequency reuse
   (c) half-duplex (d) none of the above
9. The cellular technical construct that requires the geographic distance that is required between the cells using identical frequencies to avoid interference between the radio transmission at these cells is
   (a) D/R ratio (b) call handoff
   (c) cell splitting (d) none of the above

10. The higher the frequency of a radio signal, the greater is the
    (a) absorption rate
    (b) free space
    (c) co-channel interfernce
    (d) none of the above
11. Co-channel interference describes which of the following:
    (a) interference between two base stations transmitting on the same frequency
    (b) interference between two mobile phones
    (c) inability of mobile to filter out signals of neighbouring cellular channels
    (d) none of the above

Answers: 1. (c), 2. (a), 3. (d), 4. (b), 5. (d), 6. (b), 7. (c), 8. (b), 9. (a), 10. (b), 11. (c).

Open book questions

1. How does the type and installation of cell site antenna system play a critical role in determining the impact of system interference?
2. Why do all cells not have uniform size in a practical cellular network?
3. What is meant by frequency reuse concept and what are the important points to be considered in frequency reuse concept?
4. Describe the changes that effect in the cellular architecture aspects due to cell splitting.
5. What are the elements of cellular mobile systems?
6. How do you find the differences between the simulation results and practical results in mobile environment?
7. During a busy hour the number of calls per hour Q_i for each of 10 calls is 2,000, 1,500, 3,000, 500, 1,000, 1,200, 1,800, 2,500, 2,800, and 900. Assume that 60 per cent of car phones will be used during the busy period and that one call is made per phone. Find out the total number of customers in the system. (Refer Section 3.2.2)
8. Explain the designing of the omnidirectional antenna under the practical case conditions for $k = 7$, $k = 9$, and $k = 12$ with all the suitable values and explaining each of them.
9. Explain the designing of the directional antenna for $K = 7$ with all suitable values explaining each of them, consider a noise margin of 6 dB.

Key equations

1. Carrier-to-interference ratio at the desired receiver.

$$\frac{C}{I} = \frac{C}{\sum_{k=1}^{K_1} I_k}$$

2. Assume that the interference level is very high; the local noise can be neglected. Then, C/I can be expressed as

$$\frac{C}{I} = \frac{R^{-\gamma}}{\sum_{k=1}^{K_1} D_k^{-\gamma}}$$

3. Co-channel interference reduction factor

$$q = \left[6\left(\frac{C}{I}\right)\right]^{\frac{1}{\gamma}}$$

4. Cluster size of the GSM system

$$N = \frac{1}{3}\left[6\left(\frac{C}{I}\right)\right]^{\frac{2}{\gamma}}$$

5. The received signal power at a distance *d* from the transmitting antenna is given by

$$P_r = \frac{P_{r0}}{(d/d_0)^{\gamma}}$$

Further reading

Mac Donald, V. H. "*Advanced Mobile Phone Service*," special issue, *Bell System Technical Journal*, Vol. 58, pp. 12–41, January 1979.

A. Mehrotra, *Cellular Radio Analog and Digital System*. Boston: Artech House, 1994.

AT&T Technical Center, "Cellular System Design and Performance Engineering I," CC 1400, version 1.12, 1993.

V. H. MacDonald, "The Cellular Concept," *Bell System Technical Journal*, Vol. 58, January 1979, pp. 15–42.

Lee W. C. Y., *Lee's Essentials of Wireless Communications*. New York: McGraw-Hill , 2001.

Interference 4

4.1 Introduction

In electronics and communications, especially in the field of telecommunications, interference is anything which alters, modifies, or disrupts a signal as it travels along a channel between a source and a receiver. The term typically refers to the addition of unwanted signals to a useful signal. In a communication environment, there exist both noise-limited and interference-limited environments. Point-to-point communication suffers from noise-limited situations, whereas mobile radio environment is interference limited as several transmitters and receivers are involved in the system.

Sources of interference include another mobile in the same cell, a call in progress in a neighbouring cell, another base station operating in the same frequency band, and any non-cellular system which leaks energy into the cellular frequency band. The two major types of system-generated cellular interferences are co-channel interference (CCI) and adjacent-channel interference (ACI). Although interfering signals are often generated within a cellular system, in practice they are difficult to control. The chapter introduces the two types of interferences and estimates the interference levels of either type. It also introduces the concept of a diversity receiver.

4.2 Types of interferences

Interference in mobile communications is of two types:

- Co-channel interference
- Adjacent-channel interference

The *co-channel interference* (CCI) is crosstalk from two different radio transmitters using the same frequency. In cellular mobile communications (GSM & LTE [Long Term Evolution] systems, for instance), frequency spectrum is a valuable resource which is divided into non-overlapping spectrum bands that are assigned to different cells. The CCI arises in the cellular mobile networks due to the phenomenon of frequency reuse. Thus, besides the intended signal from the cell, signals at the same frequencies (co-channel signals) arrive at the receiver from undesired transmitters located (far away) in some other cells and lead to a deterioration in the receiver performance.

The *adjacent-channel interference* (ACI), also known as inter-channel interference, is the interference caused by extraneous power from a signal in an adjacent channel. An ACI may be caused by inadequate filtering, such as incomplete filtering of unwanted modulation products in frequency modulation (FM) systems, improper tuning, or poor frequency control, in either the reference channel or the interfering channel, or in both. The problem can be particularly serious if an adjacent channel user is transmitting in a very close range to a subscriber's receiver, while the receiver attempts to receive a base station on the desired channel. This is referred to as the near-far effect, where a nearby transmitter (which may or may not be of the same type as that used by the cellular system) captures the receiver of the subscriber. Alternatively, the near-far effect occurs when a mobile close to a base station transmits on a channel close to one being used by a weak mobile. The base station may have difficulty in discriminating the desired mobile user from the "bleed over" caused by the close adjacent-channel mobile.

4.3 Co-channel interference areas in a system

The received voice quality is affected by the grade of coverage and the amount of CCI. In order to detect channel interference areas in a cellular system, we have to perform two tasks, discussed in the following.

4.3.1 To find the co-channel interference area from a mobile receiver

The CCI can be measured by selecting any one channel (as interference is equal in all the channels) and transmitting on that channel to all co-channel sites at night while the mobile receiver is moving in one of the co-channel cells. Co-channel or inter-channel interference is denoted as carrier-to-interference ratio (CIR) or signal-to-interference ratio (SIR).

We now look out for any change detected by a field-strength recorder in the mobile unit and compare the data with the condition of no co-channel sites being transmitted. This test must be repeated as the mobile unit moves in every co-channel cell. To facilitate this test, we can install a channel-scanning receiver in a car.

Suppose one channel (f_1) which receives the signal level (no co-channel condition), another channel (f_2) which receives the interference level (six-co-channel condition is the maximum), and a third channel receives f_3, which is not transmitting in the air. Therefore, the noise level is recorded only in f_3 (see Fig. 4.1).

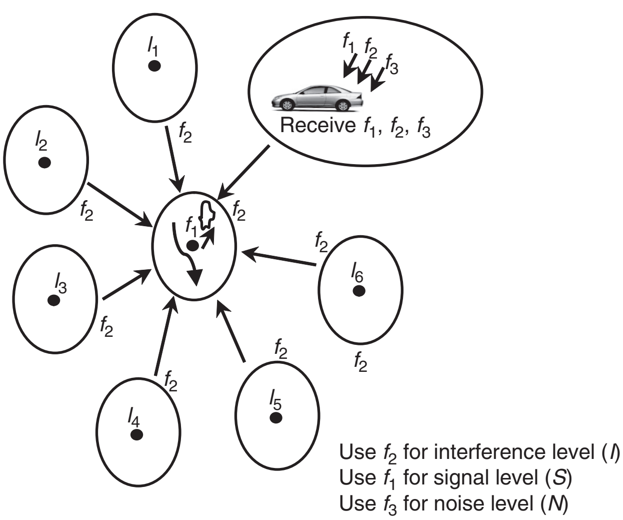


Figure 4.1 CCI at the mobile unit

Carrier-to-interference ratio, $C/I = f_1 - f_3$
Carrier-to-noise ratio, $C/N = f_2 - f_3$

The following four conditions are used to compare the results.

- If $C/I > 18$ dB throughout most of the cell, the system is properly designed for capacity.
- If $C/I < 18$ dB and $C/N > 18$ dB in some areas, it is an indication of the presence of CCI.
- If C/N and C/I are both less than 18 dB and $C/N = C/I$ in a given area, it is an indication of a coverage problem.
- If C/N and C/I are both less than 18 dB and $C/N > C/I$ in a given area, there is a coverage problem and CCI.

4.3.2 To find the co-channel interference area which affects a cell site

Reciprocity theorem is not applicable for CCI. Hence, the second task should be performed. In this task, we record the signal strength at every co-channel cell site while a mobile unit is travelling either in its own cell or in one of the co-channel cells shown in Figure 4.2.

First, we find the areas in an interfering cell in which the top 10 per cent level of the signal transmitted from the mobile unit in those areas is received at the desired site (J^{th} cell in Fig. 4.2). This top 10 per cent level can be distributed in different areas in a cell. The average value of the received top 10 per cent level signal strength is used as the interference level from that particular interfering cell. The mobile unit also travels in different interfering cells. Up to six interference levels are obtained from a mobile unit running in six interfering cells. We then calculate the average of the bottom 10 per cent level of the signal strength which is transmitted from a mobile unit in the desired cell (J^{th} cell) and received at the desired cell site as a carrier reception level.

Then we can re-establish the CIR received at a desired cell, say, the J^{th} cell, site as follows.

$$\frac{C_J}{I} = \frac{C_J}{\sum_{\substack{i=1 \\ i \neq j}}^{6} I_i} \tag{4.1}$$

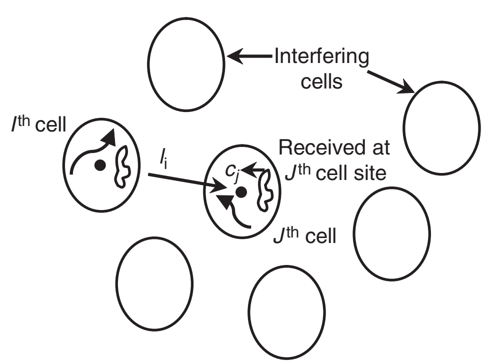


Figure 4.2 CCI at the cell site

where

C_J is the desired carrier power from J^{th} cell base station
I is the CCI
I_i is the interference power caused by the i^{th} interfering co-channel cell base station

The number of co-channel cells in the system can be less than six. We then compare

$$\frac{C_J}{I} \quad \text{and} \quad \frac{C_J}{N_J}$$

and determine the CCI condition, which will be the same as that in task 1. N_J is the noise level in the J^{th} cell assuming no interference exists.

4.4 Estimation of co-channel interference level

For a given cell size, the number of customers that a cellular system can support is maximized if the cluster size is minimum.

The factor that limits the extent to which cluster size can be reduced is the CCI. This is because reducing the cluster size has the effect of reducing the frequency reuse ratio,

$$q = D/R = \sqrt{3N}.$$

where

q is the frequency reuse factor
R is the radius of cell
D is the reuse distance
N is the cluster size or number of cells in the cluster

With increase in q, spatial separation of co-channel cells increases leading to a decrease in CCI. With decrease in q, N decreases leading to an increase in the number of replicas of the cluster (M), which results in an increase in channel capacity (C); however, the CCI increases.

In earlier days, in order to determine minimum signal-to-noise ratio (SNR) and SIR that would meet the quality-of-service objectives, simulators were used to conduct tests. The results of these early studies determined that these quality objectives could be met under the following conditions

- If the SNR is no less than 18 dB over 90 per cent of the coverage area for cells limited by receiver noise
- If the SIR is no less than 17 dB over 90 per cent of the coverage area for cells limited by interference.

Figure 4.3 begins our analysis by considering the interference from the nearest co-channel base stations. It is assumed that the receiver noise is negligible compared to the interference and also the reference base station is in the centre of the diagram.

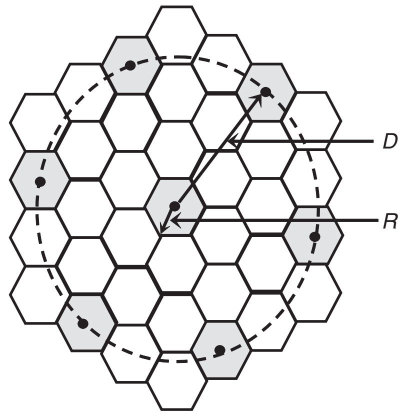


Figure 4.3 First-tier CCI sources

From the reference base station, it is considered that a mobile unit is located on the cell boundary at a distance equal to the cell radius *R*. This is the farthest point that a mobile unit should be from its serving base station. The nearest co-channel sources are mobile units in the co-channel cells and are all approximately at the reuse distance *D* from the reference base station.

The SIR C/I is given by Equation (4.2) with $j = 6$ interference sources

$$\frac{C}{I} = \frac{P_1}{\sum_{j=2}^{7} P_j} \tag{4.2}$$

where

P_1 is the desired signal power from desired base station

P_j is the interference power caused by J^{th} interfering co-channel cell base station.

Now if all of the mobile units have the same parameters and the environment is uniform in all directions, then

$$P_2 = P_3 = \ldots = P_7.$$

The SIR C/I at the desired mobile receiver is given by

$$\frac{C}{I} = \frac{C}{\sum_{i=1}^{N_i} I_i}$$

Let D_i be the distance between the i^{th} interferer and the mobile. The received interference, I_i, at a given mobile due to i^{th} interfering cell is proportional to $(D_i)^{-\gamma}$, where γ is the path-loss exponent and depends upon the terrain environment and $2 \le \gamma \le 5$.

The received signal power, *C* is proportional to $r^{-\gamma}$ where *r* is the distance between the mobile and serving base stations. The C/I at the desired mobile receiver is approximated by

$$\frac{C}{I} = \frac{r^{-\gamma}}{\sum_{i=1}^{N_i} (D_i)^{-\gamma}} \tag{4.3}$$

When the mobile is located at the cell boundary (i.e., $r = R$) and CCI from the second and other higher tiers is neglected, this means that $N_i = 6$ and using $(D_i) \equiv D$ for $i = 1, 2, \ldots, N_i$, we have

$$\frac{C}{I} = \frac{\frac{1}{R^{\gamma}}}{6\frac{1}{D^{\gamma}}} = \frac{1}{6}\left(\frac{D}{R}\right)^{\gamma} = \frac{1}{6}q^{\gamma} \tag{4.4}$$

and

$$q = \frac{D}{R} = \sqrt{3N} \tag{4.5}$$

Using Equations (4.4) and (4.5), we have

$$\frac{C}{I} = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma} \tag{4.6}$$

Example problem 4.1

Suppose, as in the AMPS system, that a SIR of 18 dB is required. The path-loss exponent is $\gamma = 4.0$. Considering only the nearest CCI sources, find the minimum cluster size.

Solution

Given $\gamma = 4.0$,
SIR (dB) = 18 dB which implies signal to interference = 63.1.

Using Equation (4.6), we have

$$63.1 = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma},$$

Therefore,

$$N = 6.49 \sim 7.$$

Table 4.1 includes a column of values of SIR assuming that the path-loss exponent $\gamma = 4$. Only first-tier interference is taken into account, as the first-tier interference predominates. In the presence of a deep fade, or when no first-tier sources are present, interference from second or further tiers may be noticeable.

Table 4.1 Approximate signal-to-interference ratio for several reuse ratios with $\gamma = 4$

i	j	D	N	q	C/I, dB
1	0	1	1	1.73	1.8
1	1	1.73	3	3	11.3
2	0	2	4	3.46	13.8
2	1	2.65	7	4.58	18.7
2	2	3.46	12	6	23.3
3	0	3	9	5.2	20.8
3	1	3.61	13	6.24	24.0
3	2	4.36	13	6.24	24.0
3	3	5.2	27	9	30.4
4	0	4	16	6.93	25.8

where

$$N = i^2 + j^2 + ij \quad \text{and} \quad q = \sqrt{3N}$$

Figure 4.4 shows the geometry of the second and third CCI tiers. The second-tier radius is the centre-to-centre distance between large hexagons, that is,

$$R_2 = \sqrt{3}R_1 = \sqrt{3}D \tag{4.7}$$

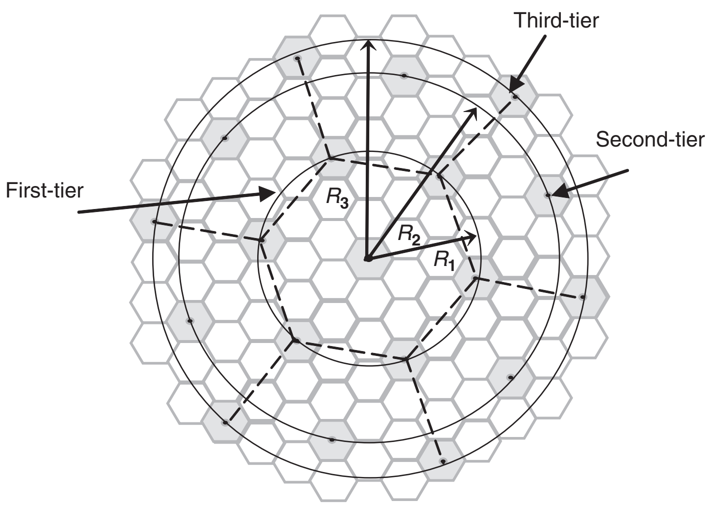


Figure 4.4 First-, second-, third-tier CCI

From the geometry, the third-tier radius is simply double the side of a large hexagon, that is, $R_3 = 2D$
For a path-loss exponent of $\gamma = 4$, interference from the second tier is

$$
\begin{aligned}
\left.\frac{C}{I}\right|_{2,\mathrm{dB}} &= 10\log\frac{1}{6}\left(\frac{\sqrt{3}D}{R}\right)^4 \\
&= 10\log\frac{1}{6}\left(\frac{D}{R}\right)^4 + 10\log\left(\sqrt{3}\right)^4 \quad \text{(From } \log(ab) = \log a + \log b\text{)} \\
&= \left.\frac{C}{I}\right|_{\mathrm{dB}} + 9.54\ \mathrm{dB}
\end{aligned}
\tag{4.8}
$$

where

$\left.\frac{C}{I}\right|_{\mathrm{dB}} = 10\log\frac{1}{6}\left(\frac{D}{R}\right)^4$ is the level of interference from the first tier.

Equation (4.8) indicates that the interference from the second tier is 9.54 dB below the level of interference from the first tier. Similarly, the interference from the third tier is 12 dB below the first-tier level. This method is applicable even when there is the need to consider additional tiers of interference.

4.5 Real-time co-channel interference measurement

The CCI is often the predominant factor that limits the capacity of a cellular mobile radio system. Power control is, thus, needed so that carrier-to-co-channel interference ratio can be maintained within a tolerable bound.

Many power control algorithms that require real-time CCI ratio measurement have been proposed and viable solutions for real-time CCI ratio estimations under various channel conditions were presented. These solutions are based on either the method of moments or the histogram matching concept.

The ultimate concern of a communication system designer is the quality of the demodulated baseband signals. When co-channel or inter-channel interference is present, CIR or SIR is an important receiver parameter that characterizes the degree of system performance deterioration. When a frequency reuse scheme is employed in a cellular mobile communication system to enhance the system's spectral utilization efficiency, a received signal inevitably suffers from interference from other co-channel users.

The signal quality requirement of such a cellular system is often specified by a CCI ratio threshold called the signal protection ratio. It is defined as "the minimum ratio of wanted to unwanted signal levels for satisfactory reception". For an FDMA system, the co-channel reuse distance, that is, the minimum distance between two co-channel users and therefore the cell size has to be such that the average CCI ratio is greater than or equal to this minimum required threshold value. Therefore, the system capacity is limited by the multiplicity of the simultaneous usage of the same channel. Furthermore, power control is necessary to reduce CCI and allow as many co-channel users as possible while each maintaining an acceptable CCI ratio. Besides power control, it is also seen that the criteria for CCI ratio in turn is based on the criteria for diversity reception, handoffs, and channel allocation. All of these algorithms assume that real-time CCI ratio measurement is available without mentioning how it is obtained.

Researchers used statistical properties of the received waveform's envelope and applied the method of moments to estimate the CCI ratio of a phase-modulated carrier that is corrupted by a single interferer. However, this method is based on a noiseless assumption. The estimated CCI ratio becomes less reliable whenever the operation scenario is different from the assumed one. It is found that the histogram matching method with a properly designed, non-uniform quantizer offers excellent CCI ratio estimation.

4.6 Diversity receiver

In communication systems, receiver specifications are set to accommodate a small received input power. Systems such as cellular base transceiver stations (BTS) receive signals from handsets that could be located in environments which greatly attenuate the signal, such as parking garage structures, multi-floor buildings, or crowded urban areas.

The signal transmitted from the handset will arrive at the BTS many times, having taken many different reflected paths. With only one antenna and receiver, many versions of the same signal will be present at the receive antenna, each with different phase and amplitude. The signals could add constructively or destructively due to the instantaneous phase relationship. In a mobile phone, for example, the mobile transmitter is not perfectly fixed in space, so the summation at the antenna is continually changing. This is referred to as fast fading and could cause lost reception.

Using diversity antennas increases the chances of finding a signal with sufficient receive strength, as the antennas are physically separated. While one antenna may be experiencing destructive interference, the other may not; this is diversity.

To demodulate the signal, the communication link is built with a minimum SNR required. Diversity allows for a higher probability that a signal will arrive at the BTS above the minimum SNR. To build a diversity receiver, at least one extra receive path is added for every channel. This may double the cost of the electronics and antenna. However, the cost is justified if it extends the range and quality of the BTS. It can reduce the number of base stations required, reducing the overall network capital cost.

The diversity scheme applied at the antenna receiver end is an effective technique for reducing interference because additional interference will not be caused by steps taken at the receiving end to improve the signal performance. We may use a selective combiner to combine a number of correlated signals as shown in Figure 4.5. The performance of other kinds of combiners can be at most 2 dB better than that of selective combiners. However, the selective combining technique is the easiest scheme to use.

In diversity receiver, the receiver needs to combine multiple streams from different antennas into a single stream. The challenge here is how to use "effectively" the information from all the antennas. In fact, it is just a matter of choosing the appropriate weight for each received signals as shown in Figure 4.5. In selective combining, the receiver selects the antenna with the highest received signal power and ignores observations from other antennas.

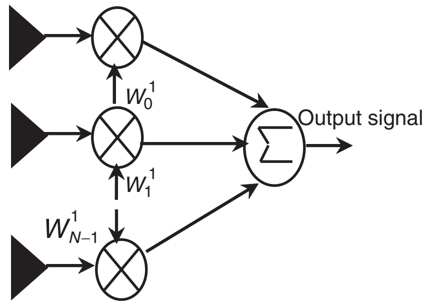


Figure 4.5 Diversity receiver

4.7 Non-co-channel interference

Two major types of non-CCI discussed in this section are as follows:
- Adjacent-channel interference
- Near-end to far-end ratio interference

4.7.1 Adjacent-channel interference

The ACI can be classified as either in-band or out-of-band interference. The term in-band is applied when the centre of the interfering signal bandwidth falls within the bandwidth of the desired signal. The term out-of-band is applied when the centre of the interfering signal bandwidth falls outside the bandwidth of the desired signal.

In the mobile radio environment, the desired signal and the adjacent-channel signal may be partially correlated with their fades. When ACI is compared with CCI at the same level of interfering power, the effects of the ACI are always less.

ACI can be eliminated on the basis of the channel assignment, the filter characteristics, reduction of near-end to far-end ratio interference, and also by keeping frequency separation between each channel as large as possible, avoiding the use of adjacent channels in neighbouring cell sites, and so on. ACI includes next-channel (the channel next to the operating channel) interference and neighbouring-channel (more than one channel away from the operating channel) interference. It can be reduced by frequency assignment.

4.7.1.1 Next-channel interference

For any particular mobile unit, the next-channel interference affecting it cannot be caused by transmitters in the common cell site but must originate at several other cell sites. This is due to the fact that any channel combiner at the cell site must combine the selected channels, normally 21 channels (630 kHz), or at least 10 channels away from the desired one. Without proper system design, next-channel interference will arrive at the mobile unit from other cell sites. In addition, interference can be caused by a mobile unit initiating a call on a control channel in a cell with the next control channel at another cell site. Next-channel interference reduction methods use the receiving end. Filters with a sharp falloff slope can help to reduce all the ACI, including the next-channel interference.

4.7.1.2 Neighbouring-channel interference

Neighbouring-channel interference is another type of ACI that is unique to the mobile radio system. It is caused by the channels that are several channels away from the desired channel.

In general, a fixed set of channels is assigned to each cell site. To reduce inter-modulation products, a sufficient amount of band isolation between channels is required for a multi-channel combiner if all the channels are simultaneously transmitted at one cell-site antenna. Evolving technologies are focusing on using multiple antennas instead of one antenna at the cell site with the assumption that band separation requirements can be resolved.

4.7.1.3 Transmitting and receiving channels interference

Transmitting and receiving channels interference is another type of ACI caused by the transmitting channels. This is because the transmitting channels are so strong that they can mask the weak signals received from the receiving channels. This effect can be reduced by the following means:

- In FDMA and TDMA systems, a guard band of 20 MHz is used to separate the transmitting and receiving channels.
- The duplexer can be used but it only provides an isolation of around 30–40 dB.
- By band isolation.

4.7.2 Near-end to far-end ratio interference

A type of interference which occurs only in mobile communication systems is the near-end to far-end type of interference. This kind of interference appears when the distance between a mobile unit and the base station transmitter becomes critical with respect to another mobile transmission that is close enough to override the desired base station signal. This phenomenon occurs when a mobile unit is relatively far from its desired base station transmitter at a distance d_0, but close enough to its undesired nearby mobile transmitter at a distance d_1, and $d_1 > d_0$ (refer Fig. 4.6).

The problem in such a situation is whether the two transmitters will transmit simultaneously at the same power and frequency, thus masking the signals received by the mobile unit from the desired source by the signals received from the undesired source. In addition, this type of interference can take place at the base station when signals are received simultaneously from two mobile units that are at unequal distances from the base station. The power difference due to the path loss between the receiving location and the two transmitters is called the near-end to far-end ratio interference and is expressed by the ratio of path loss at distance d_1 to the path loss at distance d_0.

This form of interference is unique to mobile radio systems. It may occur both within one cell and within cells of two systems.

4.7.2.1 In one cell

When mobile station A is located close to the base station, and at the same time mobile station B is located far away from the same base station (e.g., at the cell boundaries), mobile station A causes ACI to the base station and mobile station B (Fig. 4.6). The C/I at mobile station B is expressed by the following equation:

$$\frac{C}{I} = \left(\frac{d_0}{d_1}\right)^{-\gamma} \tag{4.9}$$

where

γ is the path-loss slope

d_0 is the distance from base station to desirable mobile station A

d_1 is the distance from base station to undesirable mobile station B

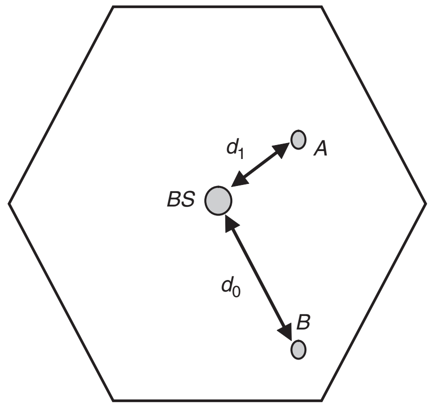


Figure 4.6 Near-far interference in one cells

Since $d_0 > d_1$, from Equation (4.9) we obtain $C/I < 1$. This means that the interfering signal is stronger than the desired signal. This problem can be rectified if the filters used for frequency separation have sharp cut-off slopes. The frequency separation can be expressed as follows:

Frequency band separation is 2^{G-1}B

where

$$G = \frac{\gamma \log_{10}\left(\dfrac{d_0}{d_1}\right)}{L} \tag{4.10}$$

B is the channel bandwidth
L is the filter cut-off slope.

4.7.2.2 *In cells of two systems*

If two different mobile operators cover an area, ACI may occur if the frequency channels of the two systems are not properly coordinated.

In Figure 4.7, two different mobile radio systems are depicted.

Mobile station A is located at the cell boundaries of system A, but very close to base station B. In addition, mobile station B is located at the cell boundaries of system B, but very close to base station A. Interference may occur at base station A from mobile station B and at mobile station B from base station A. The same interference will be introduced at base station B and at mobile station A.

This form of interference can be eliminated if the frequency channels of the two systems are properly coordinated, as mentioned earlier. If such a case occurs, two different systems operating in the same area may have co-located base stations.

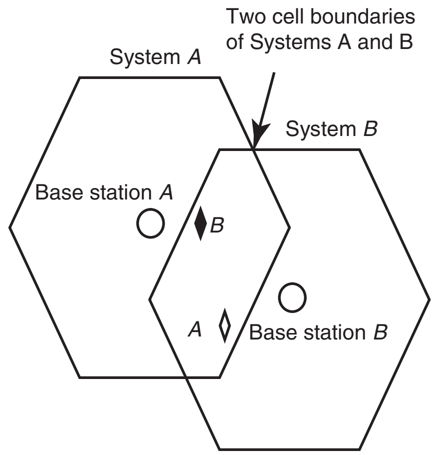


Figure 4.7 Near-far interference in cells of two systems

4.8 Estimation of adjacent-channel interference levels

A receiver in the cellular system must be designed to receive all channels in the cellular system band, as a telephone connection may be assigned to any of the possible channels. By using a highly selective filter, the receiver separates one channel from another. The pass band bandwidth of the filter is equal to the bandwidth of the channel.

To prevent signals in the adjacent channels from passing to the demodulator, the filter must cut off sharply at the pass band edges. A "brick wall" filter, which cuts off abruptly and completely at the pass band edges, is impossible to realize. In addition, sharp cut-off filters may be too expensive for mass consumer markets. For both analog and digital implementations, the performance of a filter is very sensitive to small errors in the component or coefficient values.

ACI can be a problem, even with highly selective channel filtering. There are a few strategies available for dealing with this problem. A common strategy is to avoid using adjacent channels in the same market area. This strategy is used in both AM and FM broadcasting and in television. In cellular systems, however, the number of channels available translates directly into the number of customers and in turn into revenue. Channels are too valuable to be set aside for interference avoidance.

With the dynamic control of the power in a mobile unit transmitter, less power can be transmitted when it is nearer the base station than it does when it is at a cell edge. In modern cellular systems, to maintain a constant received power level at a base station, a mobile unit's transmitted power is adjusted in 1 dB increments every few milliseconds, as the mobile unit moves over the cell's coverage area.

Finally, channel partition can be made so that adjacent channels are assigned to the same cell or to cells that are immediate neighbours. This will guarantee that an interference source cannot get physically close to a base station receiver. However, the available channels will be divided up among a relatively small number of cells when cluster sizes are small. In this case, it may be difficult to avoid assigning adjacent channels to the same or nearby cells, and ACI may significantly limit how small the clusters can be made.

4.9 Summary

- Interference is anything which alters, modifies, or disrupts a signal as it travels along a channel between a source and a receiver. The term typically refers to the addition of unwanted signals to a useful signal.
- CCI is the crosstalk from two different radio transmitters using the same frequency. It arises in the cellular mobile networks owing to the phenomenon of frequency reuse.
- ACI, also known as inter-channel interference, is the interference caused by extraneous power from a signal in an adjacent channel.
- CCI is a function of a parameter q defined as $q = D/R$ where q is the CCI reduction factor.
- Channel interference areas in a cellular system can be detected in two ways: by finding the CCI area from a mobile receiver and by finding the CCI area which affects a cell site.
- Adjacent-channel interference can be a problem, even with highly selective channel filtering. To counter this problem, there are two possible ways.
- The power level of a mobile unit transmitter can be controlled dynamically, so that it transmits less power when it is nearer the base station than it does when it is at a cell edge.
- Channels can be partitioned so that adjacent channels are not assigned to the same cell or to cells that are immediate neighbours.
- For a given cell size, if the cluster size is minimum, the number of customers that a cellular system can support is maximized.
- Power control is necessary to reduce CCI and allow as many co-channel users as possible while each maintaining an acceptable CCI ratio.
- Diversity allows for a higher probability that a signal will arrive at the BTS above the minimum SNR.

Example problem 4.2

If signal-to-noise interference ratio of 15 dB is required for a satisfactory forward channel performance of a cellular system, what is frequency reuse factor and cellular size that should be used for maximum capacity if the path-loss exponent is (a) $\gamma = 4$ and (b) $\gamma = 3$? Assume that there are six co-channel cells in the first tier and all of them are at the same distance from the mobile. Use suitable approximations.

Solution

(a) Path-loss exponent $\gamma = 4$
First, let cluster size $N = 7$

$$\text{Co-channel reuse ratio } (q) = \frac{D}{R} = \sqrt{3N} = 4.583$$

and

$$\frac{C}{I} = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma} = \frac{1}{6}(4.583)^4 = 75.3 = 18.66 \text{ dB}.$$

Since this is greater than minimum required $\frac{C}{I}$, $N = 7$ can be used.

(b) Path-loss exponent $\gamma = 3$
Let $N = 7$

$$\frac{C}{I} = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma} = \frac{1}{6}(4.583)^3 = 16.04 = 12.05 \text{ dB}$$

which is less than minimum required $\frac{C}{I}$, hence we need to use larger N.

$\therefore$ Next possible value of $N = 12$ for $i = j = 2$. $(N = i^2 + j^2 + ij)$

For $N = 12$,

$$q = \sqrt{3(N)} = \sqrt{3(12)} = 6.$$

$$\therefore \frac{C}{I} = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma} = \frac{1}{6}(6)^3 = 36 = 15.56 \text{ dB}$$

Since this is greater than minimum required $\frac{C}{I}$, $N = 12$ is used.

Example problem 4.3

If signal-to-noise interference ratio of 20 dB is required for a satisfactory forward channel performance of a cellular system, what is the frequency reuse factor and the cellular size that should be used for maximum capacity if the path-loss exponent is (a) $\gamma = 6$ and (b) $\gamma = 2$? Assume that there are six co-channel cells in the first tier and all of them at the same distance from the mobile. Use suitable approximations.

Solution

(a) Path-loss exponent $\gamma = 6$
First, let cluster size $N = 7$.

$$\text{Co-channel reuse ratio } (q) = \frac{D}{R} = \sqrt{3N} = 4.583$$

and

$$\frac{C}{I} = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma} = \frac{1}{6}(4.583)^6 = 1544.35 = 31.88 \text{ dB.}$$

Since this is greater than minimum required $\frac{C}{I}$, $N = 7$ can be used.

(b) Path-loss exponent $\gamma = 2$
Let $N = 7$

$$\frac{C}{I} = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma} = \frac{1}{6}(4.583)^2 = 3.500 = 5.44 \text{ dB}$$

which is less than minimum required $\frac{C}{I}$, hence we need to use larger N.

Next possible value of $N = 12$ for $i = j = 2$ $(N = i^2 + j^2 + ij)$

For $N = 12$,

$$q = \sqrt{3(N)} = \sqrt{3(12)} = 6.$$

$$\therefore \frac{C}{I} = \frac{1}{6}\left(\sqrt{3N}\right)^{\gamma} = \frac{1}{6}(6)^2 = 6 = 7.78 \text{ dB}$$

which is less than minimum required $\frac{C}{I}$ hence we need to use larger N. Continue this for lager values of N until $\frac{C}{I}$ is greater than the minimum required.

Review questions

1. What is adjacent-channel interference? How can it be minimized?
2. The C/I ratio was calculated by neglecting the interference from cells other than in the first tier. Calculate the amount of interference from the second tier of cells. Is it reasonable to neglect this interference?
3. Explain co-channel interference reduction factor and derive the general formula for C/I?
4. Explain how co-channel interference is measured in real-time mobile transreceiver.
5. Discuss the effect of near-end and far-end interference of mobile unit.
6. Explain CCI. (Refer Section 4.2)
7. Discuss in detail the various techniques to measure CCI. (Refer Section 4.3)
8. Distinguish between signal and co-channel interference received by the mobile unit and cell site. (Refer Sections 4.3.1 and 4.3.2)
9. Write short notes on power control. (Refer Section 4.5)
10. Explain the different types of non-CCI. (Refer Section 4.7)

Objective type questions and answers

1. Anything which alters, modifies, or disrupts a signal as it travels along a channel between a source and a receiver is called as
   (a) noise (b) interference
   (c) crosstalk (d) deterioration in receiver
2. For hexagonal cellular systems, the interference that results from the first tier is
   (a) next-channel interference (b) near-end to far-end interference
   (c) co-channel interference (d) adjacent-channel interference
3. Co-channel interference can be reduced by
   (a) decreasing D/R
   (b) increasing co-channel interference ratio, q
   (c) reducing number of channels
   (d) increasing cluster size (N)
4. Adjacent-channel interference is caused by signals from
   (a) same frequencies (b) same cell site
   (c) neighbouring frequencies (d) neighbouring cell site

5. Adjacent-channel interference is reduced if the separation between adjacent channels in a cell is
   (a) maximum (b) minimum (c) unaltered (d) doubled
6. Co-channel interference limits the extent to which
   (a) cluster size can be reduced
   (b) transmit power can be used
   (c) co-channel interference ratio can be increased
   (d) number of channels that can be used in a cell site
7. Interference caused by the channels that are several channels away from the desired channel is known as
   (a) next-channel interference
   (b) transmitting and receiving channel interference
   (c) neighbouring-channel interference
   (d) intra-channel interference

Answers: 1. (b), 2. (c), 3. (b), 4. (c), 5. (a), 6. (a), 7. (c).

Open book questions

1. How do you compute the $\frac{C}{I}$ ratio for cellular system? (S = signal; I = interference)
2. Write short notes on diversity.
3. What is adjacent-channel interference?
4. Distinguish between CCI and non-CCI.
5. Briefly explain different methods used for reducing near-end to far-end interference.
6. Discuss the diversity schemes for interference reductions at both mobile unit and cell site.

Key equations

1. Carrier-to-interference ratio received at a desired cell

$$\frac{C_J}{I} = \frac{C_J}{\sum_{\substack{i=1 \\ i \neq j}}^{6} I_i}$$

2. Frequency reuse ratio.

$$q = D/R = \sqrt{3N}$$

3. The signal-to-interference ratio C/I is given by equation with $J = 6$ interference sources

$$\frac{C}{I} = \frac{P_1}{\sum_{j=2}^{7} P_j}$$

4. The signal-to-interference ratio C/I at the desired mobile receiver is given by

$$\frac{C}{I} = \frac{C}{\sum_{i=1}^{N_i} I_i}$$

5. Frequency band separation

$$2^{G-1}B$$

Further reading

Bjelajac, B., "CIR Based Dynamic Channel Allocation Schemes and Handover Prioritisation for Mobile Satellite Systems," *IEEE VTC '96,* pp. 948–952, Vol. 2, Atlanta, GA, USA, April 28–May 1, 1996.

Buzzi, S., et al., ''Diversity Reception of Nonorthogonal Multipulse Signals in Multiuser Nakagami Fading Channels,'' *IEEE Communications Letters*, Vol. 5, May 2001, 188–190.

French, R. C., "The Effects of Fading and Shadowing on Channel Reuse in Mobile Radio," *IEEE Transactions on Vehicular Technology*, pp. 171–181, 28 August 1979.

MacDonald, V. H., "The Cellular Concept," *Bell System Technical Journal*, Vol. 58, No. 1, January 1979, 15–41.

Mehrotra, A., *Cellular Radio Analog and Digital System*. Boston, MA: Artech House, 1994.

Rappaport, T. S., *Wireless Communication*. Upper Saddle River, NJ: Prentice Hall, 2002.

Sapienza, F., M. Nilsson, and C. Beckmann, "Polarization Diversity in CDMA," *Proceedings of the 1998 IEEE Aerospace Conference*, Vol. 3, Atlanta, GA, USA, 1988, pp. 317–322.

Shin, E. and S. Safari-Naeini, "A Simple Theoretical Model for Polarization Diversity Reception in Wireless Mobile Environments," *IEEE International Symposium on Antennas and Propagation Society*, Vol. 2, August 1999, pp. 1332–1335, Orlando, FL, USA.

Stavroulakis, P., *Interference Analysis of Communications Systems*. New York: IEEE Press, 1980.

Whitehead, J. F., "Cellular System Design: An Emerging Engineering Discipline," *IEEE Communications Society Magazine*, Vol. 24, No. 2, February 1986, pp. 8–15.

Yardim, A., Laakso, T. I; Sabel, L. P; Cain, G. D, "Design of Efficient Receiver FIR Filters for Joint Minimization of Channel Noise, ISI and Adjacent Channel Interference," *IEEE Global Telecommunications Conference,* London, UK, pp. 948–952, Vol. 2, November 18–22, 1996.

Co-Channel Interference Models and Reduction

5

5.1 Introduction

In wireless communication networks, signal reception is often corrupted by interference from other sources, or users, that share the same propagation medium. Knowledge of the statistics of interference is important in achieving optimum signal detection and estimation. Even under ideal circumstances, insufficient spectrum is available for assigning a unique band of frequencies to each communication system. In less ideal circumstances, when a communication system experiences outages due to equipment failure or natural or man-made disasters, the demands on the existing spectrum become even greater. More efficient use of the available spectrum requires sharing between multiple systems. Co-channel interference models are used for evaluating the interference between communication systems either co-located or located in adjacent areas.

In this chapter, initially, the need for an interference model is discussed. The general features of the two kinds of models, that is geographical and statistical models, are studied. This is followed by a thorough comparative study of six different models belonging to the two categories.

5.2 Geographical and statistical models

In cellular land mobile systems, there are two major parameters which govern the spectral efficiency: the modulation technique used and the multiple access technique used to trunk the signals in the system. It is highly useful to relate the spectral efficiency of modulation techniques to speech quality experienced by the users in a cellular system. The speech quality experienced by the users is influenced by the signal-to-co-channel interference protection ratio determined by the modulation technique used. In general, the co-channel interference protection ratio can be defined as the system's capability to reject co-channel interference. In other words, it is defined as "the minimum ratio of wanted to unwanted signal levels for satisfactory reception". Yet in another way, it can be defined as "the level at which 75 per cent of the user's state that the voice quality is either good or excellent in 90 per cent of the service area".

In cellular mobile radio systems, the co-channel interference and not the total noise in the system is the limiting factor in their efficiency and performance. This is because the total noise power in the system is much lower than unwanted signal power. Mathematically,

$$\text{Protection ratio, } a = \frac{S}{(I + N_S)} \tag{5.1}$$

$$\therefore \text{ Protection ratio, } a \approx \frac{S}{I} \text{ for } I >> N_s$$

where S = the wanted signal power,

I = the unwanted co-channel interfering signal power

and N_s = the total noise power in the system.

The co-channel protection ratio can influence the spectral efficiency of cellular systems and is a good measure of the performance of a modulation technique. To relate the protection ratio of a modulation system and the co-channel reuse distance, it is necessary to model the cellular land mobile radio systems in order to include the propagation effects on the radio signal. Also, it is necessary to model the relative geographical locations of the transmitters and receivers in the system to predict the co-channel interference affecting the desired signal.

Co-channel interference models are primarily classified into two categories. The first is geographical, where the models are constructed by considering the relative geographical locations of the transmitters and receivers, considering different possible numbers of interferers in the cellular system. The second is statistical, in which the propagation effects (fading and shadowing) are included in a statistical fashion.

Both the categories are based on the following general assumptions:

- A cellular land mobile radio system with regular hexagonal cell shapes is considered.
- Base stations employ omni-directional antennas and are located at cell centres.
- The long-term median value of the signal power is inversely proportional to some power of the distance and decreases with radial distance from the base station.
- Only co-channel interference is considered in both categories.
- Intermodulation products will not be produced from a base station antenna with a large number of frequency channels. The channel combiner that is connected to the antenna is assumed to be well matched to each of the channel load impedance.
- Long-term and short-term frequency stability can be maintained for various types of modulation techniques.
- Techniques to improve signal quality such as diversity signal reception, pre-emphasis/de-emphasis, automatic frequency and gain control, and so on are assumed to be equally applicable to all systems and hence are not included as part of the models. However, the effect of employing such techniques will be reflected by the value of the protection ratio.
- Co-channel interference is independent of the actual amount of transmitted power of base stations. This follows from the assumption that the sizes of all cells in a given cellular system are roughly the same.
- The interference from base station/stations to a mobile station is likely considered as the worst case. This is because the power radiated from mobile stations is very much lower than that radiated by base stations and hence can be ignored.
- Provided that the propagation conditions are taken into account, the models can be adapted for other systems operating in frequency bands other than the 900/1,000 MHz frequency band to which they are suited.

5.2.1 General features of the geographical models

The models in this category have the following special features:

- The models are built depending on the relative geographical locations of the interfering and serving base stations with respect to the mobile station.

- The main difference between various models within this category is the number of active co-channel interferers in the systems which are taken into account.
- The models in this category account for the signal path loss due to free space and propagation loss over a "flat earth".
- The models in this category do not account for additional signal loss due to fading and/or shadowing. Nevertheless, these effects can be added by modifying the models once they have been fully developed.

5.2.2 General features of the statistical models

The statistical models have the following special features:

- All models are based on a one-interferer situation.
- The received signal has amplitude which follows Rayleigh distribution (fading) about a slowly varying mean.
- The variation in the signal mean follows log-normal distribution (shadowing).
- The signal mean is an inverse function of the distance between the transmitter and the receiver (the inverse power law).
- With fading and shadowing, co-channel interference can occur anywhere, even close to the serving base station.

Suppose that the signal E.M.F received at the mobile station from the serving base station T_s is y_s and from the interfering base station T_i is y_i, then for satisfactory reception it is required that

$$y_s \geq r y_i \tag{5.2}$$

where r is the amplitude protection ratio. With statistical models, we need to calculate probability P:

$$P[y_s \geq r y_i] \tag{5.3}$$

5.3 Interference models

There are three models which belong to the category of geographical models: geographical model with one interferer, with six interferers, and with many tiers of interferers. Similarly, there are three models which belong to the statistical category: fading only, shadowing only, and fading and shadowing statistical models. These are described below in detail.

5.3.1 Geographical model with one interferer

This is a basic model in which only one co-channel interferer is taken into account. It is depicted in Figure 5.1(a), where T_s is the serving base station and T_i is the interfering base station. The worst case is when the mobile station is nearest to the interfering base station and furthest from its own serving base station (i.e. at the edge of the cell towards the co-channel interferer).

Assume that the long-term median value of the signal power decreases with radial distance from the base station and is inversely proportional to some power a of the distance d. Then, the desired signal power S received from the serving base station T_s at the mobile station is inversely proportional to R^a, that is

$$S \propto \frac{1}{R^a} \tag{5.4}$$

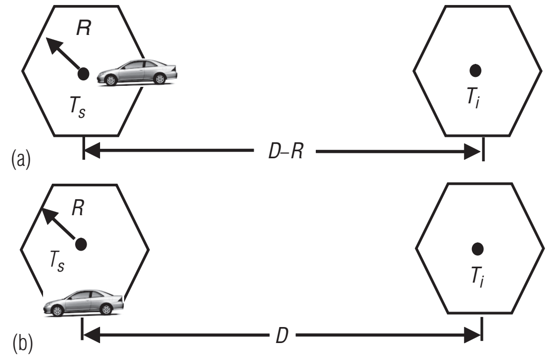


Figure 5.1 Geographical model with one interferer where T_s is serving base station and T_i is interfering base station: (a) worst case and (b) average case

Similarly, the undesired interfering signal power I received from the interfering base station T_s is inversely proportional to $(D - R)^a$, that is

$$I \propto \frac{1}{(D-R)^a} \tag{5.5}$$

Hence, combining the results in Equations (5.4) and (5.5) yields

$$\frac{S}{I} = \left[\frac{D-R}{R}\right]^a \tag{5.6}$$

$$\frac{D}{R} = \left(\left(\left(\frac{S}{I}\right)^{\frac{1}{a}}\right) + 1\right) \tag{5.7}$$

The frequency reuse ratio q is related to the number of cells in cluster (or frequency reuse factor) N by

$$q = \frac{D}{R} = \sqrt{3N}$$

$$N = \frac{\left[\left(\frac{S}{I}\right)^{\frac{1}{a}} + 1\right]^2}{3} \tag{5.8}$$

where, S/I is the protection ratio,

For higher values of a, it is possible to employ a smaller number of cells per cluster thereby having a higher spectral efficiency for a given modulation technique. However, a is dependent upon the nature of the terrain and urbanization degree and cannot be controlled by the system designer. In this model, as the mobile station is farthest from its own serving station, the height of the base station antenna can be considered to be small when compared with the distance of the mobile from the base station and a (path loss exponent) is well approximated by four.

Equation (5.8) gives the relationship between the protection ratio and the number of cells per cluster needed for a satisfactory signal reception. It shows that for lower values of protection ratios, the spectral efficiency is higher. Theoretically, N can only take particular integer values and, therefore, N is a discontinuous function of the protection ratio. Furthermore, since spectral efficiency is inversely proportional to N, it is a discontinuous function of the protection ratio.

Thus, the precise value of the protection ratio of a modulation technique might not be as crucial in assessing its spectral efficiency as was thought previously. Also, a slight advantage in the protection ratio of one modulation technique over another does not necessarily imply a higher spectral efficiency.

5.3.2 Geographical model with six interferers

In this model, the interference from the first tier co-channel cells (i.e. next nearest co-channel cells) is taken into account. In a fully developed hexagon-shaped cellular system, there are always six co-channel cells in the first tier regardless of the number of cells per cluster. It is assumed that all six co-channel interfering cells are active as in a busy hour situation. It is also assumed that interference from second and higher order tiers is negligible. This model is depicted in Figure 5.2, where T_{i_1}–T_{i_6} are the six closest interfering base stations.

Consider the average case, when the mobile station is furthest from its own serving base station and is at an average distance D from all six interfering base stations. Then, at the mobile station:

$$I \;\alpha\; 6\left(\frac{1}{D^{\alpha}}\right)$$

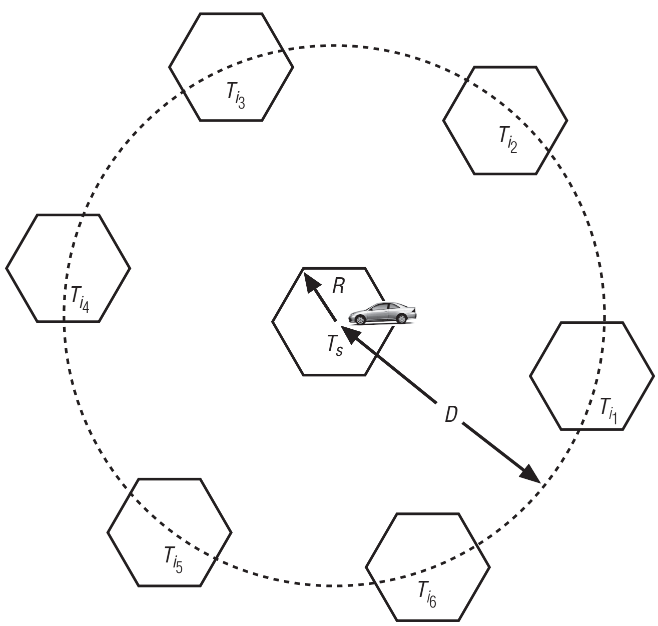


Figure 5.2 Geographical models with six interferers. T_s is serving base station and T_{i_1}–T_{i_6} is interfering base stations

Hence, the signal-to-interference ratio is given by

$$\frac{S}{I} = \frac{1}{6}\left(\frac{D}{R}\right)^{a} \tag{5.9}$$

$$\therefore \frac{D}{R} = \left[6\frac{S}{I}\right]^{\frac{1}{a}}$$

and

$$N = \left(\frac{1}{3}\right)\left[6\left(\frac{S}{I}\right)\right]^{\frac{2}{a}} \tag{5.10}$$

where S/I is the protection ratio and is given by

$$\frac{S}{I} = \frac{S}{\sum_{n=1}^{6} I_n} \tag{5.11}$$

where I_n is the interference from the n^{th} co-channel cell.

5.3.3 Geographical model with several tiers of interferers

In this model, the co-channel interference is considered from several tiers of co-channel cells. In a fully equipped hexagon-shaped cellular system, regardless of the number of cells per cluster, there are always 6 m co-channel cells in the m^{th} tier. It is assumed that all co-channel interfering base stations up to the m^{th} tier are active as in a busy hour situation. Also, it is assumed that the interference from cells in the higher order tiers (i.e. the $m + 1^{\text{th}}$ tier onwards) is negligible. This model is depicted in Figure 5.3, where R and D have their usual meanings. For a hexagonal cellular system, the average signal-to-interference ratio measured at a distance R from its own serving station can be found in and is given by the following relation:

$$\frac{S}{I} = \frac{(3N_c)^{a/2}}{6\sum_{t=1}^{T}\sum_{u=0}^{t-1}\frac{1}{(t^2+u^2-tu)^{\frac{a}{2}}}} \tag{5.12}$$

The following relations can be deduced from above:

$$\frac{S}{I} = \frac{\left(\frac{D}{R}\right)^{a}}{6\sum_{t=1}^{T}\sum_{u=0}^{t-1}\frac{1}{(t^2+u^2-tu)^{\frac{a}{2}}}} \tag{5.13}$$

$$\frac{D}{R} = \left\{6\frac{S}{I}\sum_{t=1}^{T}\sum_{u=0}^{t-1}\frac{1}{(t^2+u^2-tu)^{\frac{a}{2}}}\right\}^{\frac{1}{a}} \tag{5.14}$$

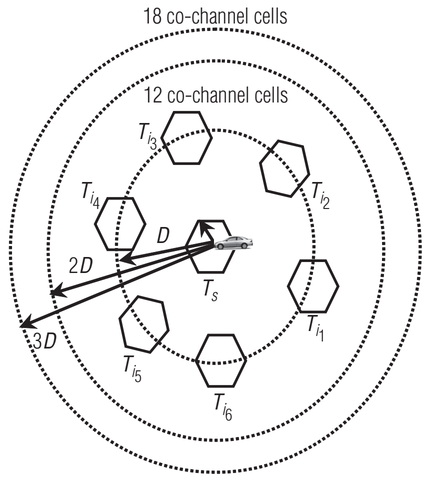


Figure 5.3 Geographical model with several tiers of interferers

and

$$N = \frac{\left\{ 6\frac{S}{I} \sum_{t=1}^{T} \sum_{u=0}^{t-1} \frac{1}{(t^2 + u^2 - tu)^{\frac{\alpha}{2}}} \right\}^{\frac{1}{\alpha}}}{3} \tag{5.15}$$

where T is the number of tiers of co-channel interfering cells. Also, (t, u) represents the normalized location of a unit cluster consisting of six base stations at equal distances from the serving base station. For example, the fourth tier can be expressed using four unit clusters:

$$(t, u) = (4, 0), (4, 1), (4, 2) \text{ and } (4, 3)$$

Using Equation (5.14), the relation between D/R and S/I can be established for several numbers of tiers of interference. In a similar fashion, the relation between N_c and S/I can be established for various numbers of tiers of interference using Equation (5.15). The summation is evaluated for $\alpha = 4$, without affecting the generality of the results.

i. Considering only the first tier of interference with six co-channel cells. In this case, $T = 1$ and $(t, u) = (1, 0)$.

$$\frac{D}{R} = \left\{ 6\frac{S}{I} \right\}^{\frac{1}{\alpha}} \tag{5.16}$$

and

$$N = \frac{\left\{ 6\frac{S}{I} \right\}^{\frac{2}{\alpha}}}{3} \tag{5.17}$$

The above result agrees with the geographical model with six interferers.

ii. Considering two tiers of interference with a total of 18 co-channel cells. In this case, $T = 2$ and $(t, u) = (2, 0)$ and $(2, 1)$.

$$\frac{D}{R} = \left\{7.04\frac{S}{I}\right\}^{\frac{1}{\alpha}} \tag{5.18}$$

and

$$N = \frac{\left\{7.04\frac{S}{I}\right\}^{\frac{2}{\alpha}}}{3} \tag{5.19}$$

iii. Considering three tiers of interference with a total of 36 co-channel cells. In this case, $T = 3$ and $(t, u) = (3, 0)$, $(3, 1)$, and $(3, 2)$.

$$\frac{D}{R} = \left\{7.36\frac{S}{I}\right\}^{\frac{1}{\alpha}} \tag{5.20}$$

and

$$N = \frac{\left\{7.36\frac{S}{I}\right\}^{\frac{2}{\alpha}}}{3} \tag{5.21}$$

5.3.4 Fading only statistical model

In this model, the received signals at the mobile station are assumed to have amplitude which follows Rayleigh distribution (see Fig. 5.4). The fading of the wanted and interfering signals is assumed to be uncorrelated and the shadowing effects on the signals are ignored. This model is based on one interferer. The mobile station is located between T_s and T_i at a distance xD from T_s, where $0 < x < 1$.

If the signal amplitude received at the mobile station from the serving base station T_s is y_s and from the interfering base station T_i is y_i, then for satisfactory reception, it is necessary that

$$y_s \geq r y_i \tag{5.22}$$

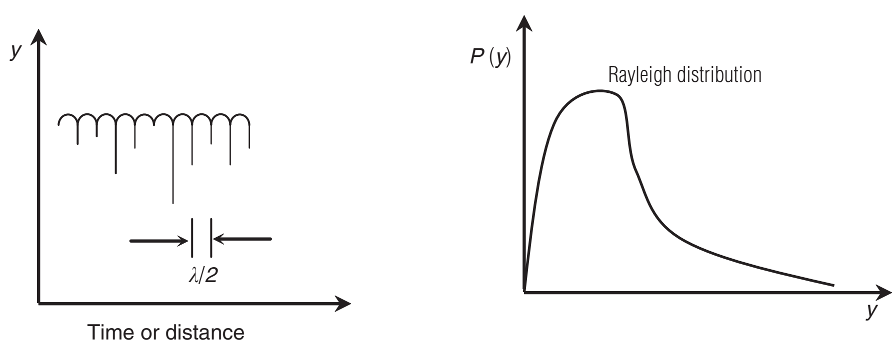


Figure 5.4 Fading only statistical model

where r is the "amplitude" protection ratio. We need to calculate the probability of y_s being greater than y_i by the amount of the amplitude protection ratio, r. Consider a variable y which is Rayleigh distributed with modal value σ. The probability density function (PDF) of the distribution is

$$p(y) = \left(\frac{y}{\sigma^2}\right)\exp\left(\frac{-y^2}{2\sigma^2}\right) \tag{5.23}$$

The probability P_1 such that $y_s \geq ry_i$ at a given y_i is

$$P_1 = \int_{ry_i}^{\infty} p(y_s)\mathrm{d}y_s \tag{5.24}$$

$$P_1 = \exp\left(\frac{-r^2y_i^2}{2\sigma_s^2}\right)$$

To obtain the probability P_2 such that $y_s > ry_i$ for all y_i, it is essential that we integrate this function over all y_i, thus

$$P_2 = \int_0^{\infty} P_1 p(y_i)\mathrm{d}y_i$$

$$= \int_0^{\infty} \exp\left(\frac{-r^2y_i^2}{2\sigma_s^2}\right)\left[\left(\frac{y_i}{\sigma_i^2}\right)\exp\left(\frac{-y_i^2}{2\sigma_i^2}\right)\right]\mathrm{d}y_i \tag{5.25}$$

$$= \int_0^{\infty} \left(\frac{y_i}{\sigma_i^2}\right)\exp\left\{\left(\frac{-y_i^2}{2}\right)\left[\left(\frac{r^2}{\sigma_s^2}\right)+\left(\frac{1}{\sigma_i^2}\right)\right]\right\}\mathrm{d}y_i \tag{5.26}$$

$$P_2 = \frac{\sigma_s^2}{r^2\sigma_i^2 + \sigma_s^2} \tag{5.27}$$

The mean signal power is an inverse function of the distance between the transmitter and the receiver (the inverse α^{th} power law). Hence, at the mobile station

$$\frac{\sigma_s^2}{\sigma_i^2} = \frac{W_s[(1-x)D]^{\alpha}}{W_i(xD)^{\alpha}} \tag{5.28}$$

or

$$\frac{\sigma_s^2}{\sigma_i^2} = \frac{W_s(1-x)^{\alpha}}{W_i x^{\alpha}} \tag{5.29}$$

where W_s and W_i are the omni-directionally radiated powers from T_s and T_i, respectively. Substituting Equation (5.29) in (5.27), P_2 becomes

$$P_2 = \frac{W_s(1-x)^a}{r^2 W_i x^a + W_s(1-x)^a} \tag{5.30}$$

P_2 is the probability that the wanted signal y_s is received at a level above the interfering signal y_i by the desired amplitude protection ratio, r.

Following the assumption that the co-channel interference is independent of the actual amount of power transmitted by the base stations and that the size of all cells are roughly the same $W_s = W_i$. Furthermore, considering the worst case of interference when the mobile station is nearest to the interfering base station T_i and furthest from its own serving base station T_s , $xD = R$ and Equation (5.30) can be rewritten as follows:

$$\frac{D}{R} = \left[\left(\frac{P_2}{1-P_2}\right)\left(\frac{S}{I}\right)\right]^{\frac{1}{a}} + 1 \tag{5.31}$$

and

$$N = \frac{1}{3} \times \left\{\left[\left(\frac{P_2}{1-P_2}\right)\left(\frac{S}{I}\right)\right]^{\frac{1}{a}} + 1\right\}^2 \tag{5.32}$$

where S/I is the "power" protection ratio a, where $a = r^2$.

5.3.5 Shadowing only statistical model

In this model, the received signals (both wanted and interfering) at the mobile station are assumed to suffer shadowing effects as a result of the signal being blocked by large structures such as hills and mountains (Fig. 5.5).

The shadowing of the wanted and interfering signals is assumed to be uncorrelated and the fading effects on the signals are ignored. For shadowing, the variation in the received signal level measured in decibels is best described by a normal (Gaussian) distribution.

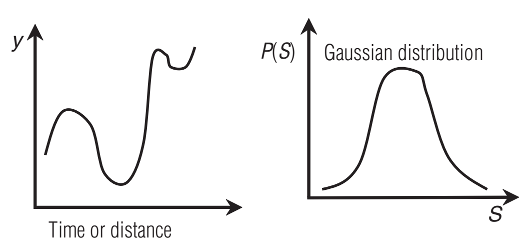


Figure 5.5 Shadowing only statistical model

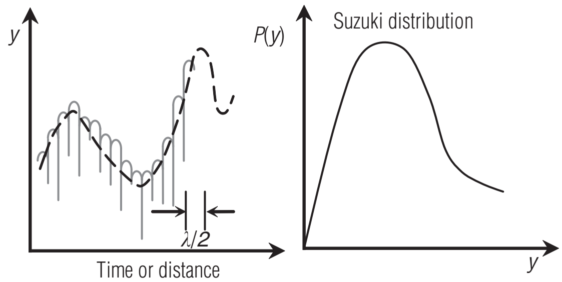


Figure 5.6 Fading and shadowing statistical model

5.3.6 Fading and shadowing statistical model

Fading and shadowing of the received signal are not separated from each other in land mobile radio. The fading and shadowing statistical model considers the general case when both the desired and interfering signals are undergoing fading and shadowing effects simultaneously in an uncorrelated manner. In this case, the signal local mean varies log-normally with a superimposed fading which follows a Rayleigh distribution (Fig. 5.6). The superimposition of the two types of variations, that is the Rayleigh and log-normal, is sometimes referred to as the Suzuki distribution.

5.4 Reduction of co-channel interference

One of the major challenges in cellular mobile systems is the reduction of co-channel interference. A few methods which can be considered are

- increasing the separation distance between the two co-channel cells
- use of directional antennas at the base station
- decreasing the antenna heights at the base station.

Method 1 is not advisable because the system efficiency decreases as the number of frequency-reuse cells increases. Method 3 is not recommended because the reception level at the mobile unit is weakened. Method 2 is good, when there is fixed number of frequency-reuse cells. The use of directional antennas results in further reduction of co-channel interference and increase in the channel capacity with increase in traffic. Few other methods are considered in the following section.

5.4.1 By using a notch in the tilted antenna pattern

Radiation from a co-channel serving site can easily interfere with another co-channel cell in normal conditions. Interference in the system can be reduced by installation of a 120° directional antenna, thereby eliminating the radiation to the remaining 240° sector. As the serving site can interfere with the co-channel cell that is directly ahead, co-channel interference can exist even when a directional antenna is used.

In general, the antenna pattern can be tilted down to reduce the signal strength. There are two ways to tilt down the antenna patterns: electronically and mechanically. The electronic down tilting involves changing the phases among the elements of a co-linear array antenna. The mechanical down tilting involves down tilting the antenna physically. To achieve a significant gain of C/I in the interference-receiving cell, we should use a notch in the centre of the antenna pattern at the interfering cell. A simple way to obtain a notch is to tilt the high-gain directional antenna mechanically downward.

5.4.2 Using an adaptive antenna

A mobile adaptive antenna system is used to improve the performance of the system in the presence of both noise and interference. This section describes a few useful applications of adaptive antennas. In the example shown below, the adaptive antennas at a base station minimize the impact of interference in a simple mobile communication case. Figure 5.7 shows a desired mobile and an interfering mobile transmitting co-channel signals s_1 and s_2, respectively, which are received at a base station having two independent antenna elements.

The channels from mobile *i* to element *j* are represented by a_{ij}, including all antenna, path loss, shadowing and fading effects. At the base station, the elements are then multiplied by complex weights *w* and summed resulting in the output *y*, which would then be demodulated by the base station. If the base station is able to estimate the channel coefficients a_{ij}, then it would be useful for it to set the weights w_j such that the output *y* minimizes the output due to the interferer while leaving the desired signal unaffected.

Weighting can be used to completely remove interfering co-channel signals. In order to do so, the base station must be able to estimate the channels between each of the mobiles and each of the antenna elements. It is also required that the number of elements is at least equal to the number of mobiles in order to solve the resulting system of simultaneous equations. In practical cases, the optimum weights would have to be estimated in the presence of noise, so an interferer is not removed completely. Instead, the weights are chosen to maximize the signal-to-interference-plus-noise ratio (SINR).

Several benefits are produced if a base station uses an adaptive array to direct its radiation pattern towards the mobile with which it is communicating. They are as follows:

- The transmit power for a given signal quality can be reduced in both uplink and downlink directions or the cell radius can be increased, thereby reducing the number of base stations required to cover a given area.
- As the mobile transmit power is reduced, its battery life can be extended.
- The channel delay spread is reduced because off-axis scatters are no longer illuminated.

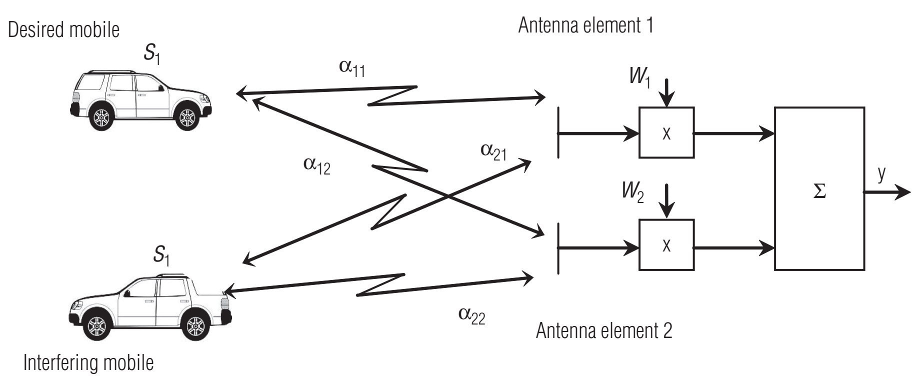


Figure 5.7 Interference reduction using adaptive antennas

- The probability of base stations causing interference to co-channel mobiles in surrounding cells is reduced depending on the direction of the mobile.
- The probability of mobiles causing interference to co-channel base stations is reduced.

5.4.3 Spatial filtering for interference reduction

This last point is known as spatial filtering for interference reduction (SFIR). SFIR is essentially an extreme form of sectorization, in which the sectors are much smaller and variable in direction compared with a non-adaptive system. This makes it relatively easy to integrate with existing systems, which already use sectorization. The capacity gain is, however, limited to the point at which all the available channels are reused in every cell (one cell reuse).

5.5 Summary

- Protection ratio is defined as the minimum ratio of wanted to unwanted signal levels for satisfactory reception.
- To relate the protection ratio of a modulation system and the co-channel reuse distance, it is necessary to model the cellular land mobile radio system in order to include the propagation effects on the radio signal.
- Co-channel interference models are primarily classified into two categories: geographical and statistical.
- Geographical model are of three types: with one interferer, with six interferers, and with many tiers of interferers.
- There are three models which belong to the statistical category: fading only, shadowing only, and fading and shadowing statistical models.
- The precise value of the protection ratio of a modulation technique might not be as crucial in assessing its spectral efficiency.
- A slight advantage in the protection ratio of one modulation technique over another does not necessarily imply a higher spectral efficiency.
- One of the major challenges in cellular mobile systems is the reduction of co-channel interference.
- Three methods which are discussed in this chapter for interference reduction are using a notch in the tilted antenna pattern, using an adaptive antenna, and spatial filtering.
- Weighting can be used to completely remove interfering co-channel signals.
- SFIR is essentially an extreme form of sectorization in which the sectors are much smaller and variable in direction compared with a non-adaptive system.

Example problem 5.1

If signal-to-interference ratio of 15 dB is required for satisfactory forward channel performance of a cellular system, what is the co-channel reuse factor and cluster size that should be used for maximum capacity if the path loss exponent is (a) $n = 4$ and (b) $n = 3$? (Assume there are six co-channel cells in the first tier and all of them are at the same distance from the mobile.)

Solution

(a) Consider seven-cell reuse pattern for $n = 4$

$$q = D/R = \sqrt{3N} = 4.583.$$

$$S/I = \left(\frac{(\sqrt{3N})^n}{6}\right) = (4.583)^4/6 = 73.5 = 18.66 \text{ dB}$$

$$\Rightarrow N = 7 \text{ can be used.}$$

(b) Consider seven-cell reuse pattern for $n = 3$

$S/I = (4.583)^3/6 = 16.04 = 12.05$ dB < 15 dB; therefore, a larger N should be used.

For $N = 12$

$q = D/R = 6$, $S/I = (6)^3/6 = 36 = 15.56$ dB > 15 dB; therefore, $N = 12$ should be used.

Example problem 5.2

Consider the advanced mobile phone system in which an S/I ratio of 18 dB is required for the accepted voice quality. What should be the reuse factor for the system? Assume *path loss exponent* $n = 4$. What will be the reuse factor of the global system of mobile (GSM) system in which an S/I of 12 dB is required?

Solution

Using the equations $q = \left(6\left(\frac{S}{I}\right)\right)^{\frac{1}{n}}$ and $q = \sqrt{3N}$, we get

$$N = \left(\frac{1}{3}\right)\left[6\left(\frac{S}{I}\right)\right]^{\frac{2}{n}}$$

From the above equation

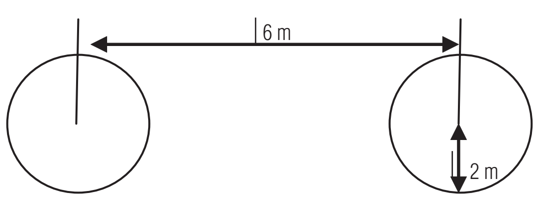
$$N_{\text{amp}} = \left(\frac{1}{3}\right)\left[6(10^{1.8})\right]^{\frac{2}{4}} = 6.486 \approx 7$$

and

$$N_{\text{GSM}} = \left(\frac{1}{3}\right)\left[6(10^{1.2})\right]^{\frac{2}{4}} = 3.250 \approx 4$$

Example problem 5.3

Find the co-channel interference reduction factor for a cellular communication system whose co-channels are separated by 6 m and radius of a cell is 2 m as shown in figure.



Solution

Given data:

Distance of separation between co-channel cells, $D = 6$ m

Radius of the cell, $R = 2$ m

Co-channel interference reduction factor $q = D/\mathrm{R} = 6/2 = 3$.

Review questions

1. Discuss the need for co-channel interference models.
2. Compare the characteristics of geographical and statistical models.
3. Establish the relation between D/R and S/I for several numbers of tiers of interference and deduce the result for the geographical model with six interferers.
4. What are the various methods of reducing co-channel interference?
5. What are the benefits obtained if a base station uses an adaptive array?
6. Explain how co-channel interference is measured in real-time mobile radio transceivers? (Refer Section 5.4)

Objective type questions and answers

1. The capability to reject co-channel interference is known as the
   (a) no-protection ratio (b) detection ratio
   (c) rejection ratio (d) protection ratio
2. The superimposition of the Rayleigh and log-normal types of variations is referred to as
   (a) Suzuki distribution (b) Kawasaki distribution
   (c) log-Rayleigh distribution (d) none of the above
3. Spatial filtering for interference reduction (SFIR) is an extreme form of
   (a) frequency reuse (b) cell splitting
   (c) sectorization (d) multiple access
4. To obtain a notch, the high-gain directional antenna must be tilted mechanically
   (a) upward (b) sideward (c) downward (d) forward
5. For shadowing, the variation in the received signal level is best described by a
   (a) Rayleigh distribution (b) Gaussian distribution
   (c) Suzuki distribution (d) all the above
6. In geographical models, propagation path loss is accounted over a
   (a) round earth (b) ellipsoidal earth
   (c) flat earth (d) none of the above
7. In statistical models, the variation in the signal mean follows
   (a) log-normal distribution (b) Rayleigh distribution
   (c) Gaussian distribution (d) some random distribution

Answers: 1. (d), 2. (a), 3. (c), 4. (c), 5. (b), 6. (c), 7. (a).

Open book questions

1. What are the various methods of reducing co-channel interference?
2. Compare the characteristics of various co-channel interference models?

Key equations

1. Protection ratio

$$a = \frac{S}{(I + N_S)}$$

2. The signal-to-interference ratio is given by

$$\frac{S}{I} = \frac{1}{6}\left(\frac{D}{R}\right)^{a}$$

3. For a hexagonal cellular system, the average signal-to-interference ratio measured at a distance R from its own serving station is given by the following relation:

$$\frac{S}{I} = \frac{\left(\frac{D}{R}\right)^{a}}{6\sum_{t=1}^{T}\sum_{u=0}^{t-1}\frac{1}{(t^2 + u^2 - tu)^{\frac{a}{2}}}}$$

4. The PDF of the distribution for Rayleigh distributed with modal value σ.

$$p(y) = \left(\frac{y}{\sigma^2}\right)\exp\left(\frac{-y^2}{2\sigma^2}\right)$$

Further reading

J. B. Andersen, T. Rappaport, and S. Yoshida, "Propagation measurements and models for wireless communications channels," IEEE Communication Magazine, Vol. 33, pp. 42–49, January 1995.

T. Aulin, "A modified model for the fading signal at a mobile radio channel," *IEEE Transactions on Vehicular Technology*, Vol. 28, pp. 182–203, August 1979.

M. D. Austin and G. L. Stuber, "Co-channel interference modeling for signal strength based handoff analysis," *Electronics Letters*, Vol. 30, pp. 1914–1915, November 1994.

Golden L. Stuber, *Principles of Mobile Communication*. New York: Kluwer Academic Publishers, 2002.

6 Teletraffic Engineering, Trunking, GoS, and Operational Techniques

6.1 Introduction

Teletraffic theory is defined as the application of probability theory to the solution of problems concerning planning, performance evaluation, operation, and maintenance of telecommunication systems. More generally, teletraffic theory can be viewed as a discipline of planning where the tools (stochastic processes, queuing theory, and numerical simulation) are taken from research operations.

The term teletraffic covers all kinds of data communication traffic and telecommunication traffic. The theory will primarily be illustrated by examples from telephone and data communication systems. The tools developed are, however, independent of the technology and applicable within other areas such as road traffic, air traffic, manufacturing and assembly belts, distribution, workshop and storage management, and all kinds of service systems.

Teletraffic engineers use their knowledge of statistics including queuing theory, the nature of traffic, their practical models, their measurements and simulations to make predictions and to plan telecommunication networks such as the Internet. These tools and basic knowledge help in providing reliable service at a lower cost.

In Chapters 2 and 3, we have introduced the central concept of frequency reuse and discussed in detail the hexagonal cellular structure that follows from it. We have investigated the key design trade-offs involving cell radius and cluster size. We have explored avenues for expansion, including sectoring and cell splitting. In this chapter, the major emphasis is laid on the concepts of trunking, blocking, call capacity, and grade of service (GoS).

6.2 Objectives of teletraffic engineering

To measure the traffic in well-defined units through mathematical models and to derive relationships between GoS and system capacity in such a way that the theory helps in planning investments. The task of teletraffic theory is to design cost-effective systems with a pre-defined GoS when we know the future traffic demand and the capacity of system elements.

Furthermore, it is the task of teletraffic engineering to specify methods for controlling, so that the GoS is fulfilling the requirements, and also to specify emergency actions when systems are overloaded or technical faults occur. This requires methods for forecasting the demand (based on traffic measurements), methods for calculating the capacity of the systems and specification of quantitative measures for the GoS. Applying the theory in practice results in a series of decision problems concerning both short-term as well as long-term arrangements.

Short-term decisions include the determination of the number of channels in a trunk group, the allocation of priorities to jobs in a computer system, and so on. Long-term decisions include decisions concerning the development and extension of data and telecommunication networks, transmission systems, extension of cables, establishing a new base station, and so on.

The work of A.K. Erlang led to creation of traffic engineering field. This was intended for circuit-switched networks but applied to packet-switched networks as well. According to the traffic engineering, the law of large numbers can be used for large systems. This makes the aggregate properties of a system over a long period of time much more predictable when compared to the behaviour of individual parts of the system. The application of the theory for design of new systems can help in comparing different solutions and thus eliminate non-optimal solutions at an early stage without having to go for implementation of prototypes.

The measurement of traffic in a public-switched telephone network (PSTN) allows network operators to determine and maintain the quality of service (QoS) and in particular, the GoS that is promised to the subscribers. The performance of a network depends on whether all origin–destination pairs are receiving a satisfactory service.

6.3 Concepts of trunking and blocking

Early in the development of the PSTN, it was recognized that having fixed connections between every possible pair of telephones is physically impractical and economically prohibitive. The number of connections required to provide fixed connections between every possible pair of N telephones is given by

$$c_N = \sum_{k=1}^{N-1} k = \frac{N(N-1)}{2} \tag{6.1}$$

where k is an integer.

For example, 100 telephones require 4,950 connections and 10,000 telephones require almost 50 million connections. Consider a limited geographic area served by a central office. Every telephone in the area is connected to the central office by a pair of wires. Within the central office, any telephone can be connected to another by joining the terminals of the wires. This is the function of the central office switch.

Assume that in a 10-digit telephone number, the first three digits are the "area code", the second three are the "central office code," and the last four represent the "line number" within the office. Clearly, a maximum of 10,000 phone numbers can be associated with a given central office code. If we consider connections among telephones only within a single central office, a worst-case maximum of only 5,000 connections would ever need to be made, compared with the 50 million identified in connection with the equation above. In fact, since the duration of a call is short, only a few hundred connections would be needed. The central office, however, needs lines that connect to other central offices and to long-distance switches that allow connections to other area codes. Here, resources are shared so that the number of lines is much smaller than the number of possible connections.

> *A line that connects switching offices and which is shared among users on need basis is called a trunk.*

As the number of trunks needed to make connections is much smaller than the maximum number that could be used, there might not be enough facilities to allow a call to be completed.

> *A call that cannot be completed due to lack of resources is said to be blocked.*

Then, the question arises as how to determine the quantity of equipment that is needed so that the event of a call being blocked is infrequent.

Though the frequency of service requests and the duration of service are not known in advance, we consider them predictable in a statistical sense. These and similar service systems are studied in a branch of applied statistics commonly known as queuing or traffic theory. In the jargon of telecommunications, the term trunking theory is often used when referring to the application of queuing theory to determine the number of trunks required for supporting connections between central offices.

> *The application of queuing theory to establish the quantity of resources necessary to provide a given level of service is referred to as traffic engineering.*

6.4 Call capacity

Call capacity is defined depending on a view of the mobile switching center (MSC). In general, there are two approaches – *global view* and *component view*.

6.4.1 Global view

The MSC is considered to be a single unit. Each single request to the MSC for service is counted as an attempt. This approach is applicable to central processors that are involved in call processing. In the case of the global view, the call volume of interest is expressed as the sum of the originating and incoming (OC + IC) calls.

Originating call (OC) includes the following call types:

- Partial dial calls–Calls with abandons and partial time-outs.
- Intra-office calls–All calls that originate from and terminate onto the same switch.
- Outgoing calls–All calls that originate from a line on the switch but terminate onto a line on a different switch.

Incoming call (IC) includes the following call types:

- Incoming-terminating calls–All calls that terminate onto a line on the switch but originate from a different switch.
- Tandem calls–Trunk to trunk calls that are routed through the switch.
- Direct inward dialling (DID)–Calls to a private branch exchange (PBX) system.

6.4.2 Component view

The component of interest is considered as a sub-system. Every request to the component for service is counted as an attempt. This view applies to the principal processors involved in call processing. In this view, the call volume of interest is expressed as the sum of the originating + terminating (OC + TR) half-calls.

1. *Originating half-calls:* One originating half-call is for each originating call as two peripheral equipment connections are required for a completed call. If the component serves both trunks and lines, incoming and outgoing half-calls are added to the total half-call volume.
2. *Terminating half-calls:* One terminating half-call is for each incoming terminating call and interoffice call.

False starts and permanent signals (ineffective attempts) are not counted as calls in both the cases. However, partial dial attempts, attempts receiving busy treatment and attempts not answered are all considered calls. The primary determinant of a stored program control system's call handling capacity is the maximum number of calls per hour that the processor can handle while still meeting service criteria. Central processors have high-day busy hour (OC + IC) call capacities, whereas peripheral processors have high-day busy hour (OC + TR) call capacities.

The following call types are defined (see Figure 6.1):

- Originating call (OC: Call placed by a subscriber of the office
- Terminating call (TR): Call received by a subscriber of the office
- Outgoing call: Call going out of the office
- Incoming call (IC): Call coming into the office from another MSC
- Intra-office call: Call that originates from and terminates to the same MSC
- Tandem calls: Calls that come in on a trunk from another MSC and go out over a trunk to a different MSC
- (OC + TR): The measure of traffic load on the line side of the MSC from within the office and other offices
- (OC + IC): The measure of incoming trunk-circuit traffic load and traffic load on the MSC

Bell Communication Research (Bellcore) suggests the use of standard call mixes (given in Table 6.1 above for central processors) for calculating the call capacity of central processors (in terms of (OC + IC) calls) and peripheral processors (in terms of (OC + TR) half-calls). These call mixes do not represent the best and worst case scenarios for call capacity.

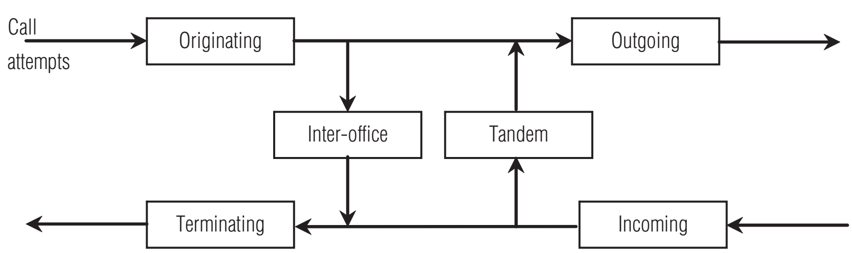


Figure 6.1 Call types in a telephone switch

Table 6.1 Standard call mixes for central processors

Type	Metro	City	Rural
Ineffective	11	14	20
Partial dial	4	5	6
Intra office	16	30	24
Outgoing	44	23	37
Incoming-Terminating	36	23	33
Tandem	0	19	0

However, they represent the typical or average case for each environment. These call mixes reflect the differences found in the following traffic environments:

- Metro: Office in a major metropolitan area with a high proportion of business office traffic. The typical traffic intensity of a metro is 3.5 to 4 calls/line.
- City: Office serving a medium-sized town.
- Rural: Office residing in a rural area with a low population density and almost no business traffic.

6.5 Grade of service (GoS)

In the context of a telephone system, the term GoS is used to mean the probability that a user's request for service will be blocked because a required facility such as a trunk or a cellular channel is not available. For example, a GoS of 5 per cent implies that on the average a user might not be successful in placing a call on 5 out of every 100 attempts. In practice, blocking frequency varies with time. One would expect more call attempts during business hours than during the middle of the night. Telephone operating companies maintain usage records and identify a "busy hour", that is, the hour of the day during which there is the greatest demand for service. Typically, telephone systems are engineered to provide a specified GoS during a specified busy hour. The busy hour chosen as a design target may be the busiest hour of the week, but is not normally the busiest hour of the year.

The two parameters that statistically characterize the user calling habits are the average number of call requests per unit time, λ_{user}, and the average holding time, H. The parameter λ_{user} is also called the average arrival rate, referring to the rate at which calls from a single user arrive at the switch. The average holding time (H) is the average duration of a call. The product

$$A_{user} = \lambda_{user} H \tag{6.2}$$

That is, the product of the average arrival rate and the average holding time is called the *offered traffic intensity or offered load* (A_{user}). This quantity represents the average traffic that is provided by a user to the system. The term offered refers to the fact that a call request may or may not be honoured, depending on the availability of a trunk or channel. Offered traffic intensity is a dimensionless quantity that is traditionally measured in Erlang.

One Erlang is equivalent to an arrival rate of one call per minute multiplied by a holding time of one minute.

If by accident, a new call arrived just as the previous one terminated, and this continued to happen, then one Erlang of traffic would continuously tie up one channel or trunk.

Example problem 6.1

Suppose during the busy hour, a user makes an average of three calls per hour and holds each call an average of 15 min. Find the offered traffic intensity.

Solution

The average arrival rate is λ_{user} = 3 calls/h = 3 calls/60 min = 1call/20min. Then, A_{user} = 1/20 calls/min × 15 min = 0.75 Erlang.

If the blocking probability or GoS is P_b, then 1– P_b represents the fraction of call requests that actually result in the assignment of a channel. If we use $(1 - P_b)\lambda_{user}$ instead of λ_{user} to calculate traffic intensity, we obtain the carried traffic intensity or carried load A_{user}. The carried traffic intensity is a measure of the traffic actually carried by the system.

Example problem 6.2

Suppose in the preceding example if the GoS during the busy hour is 10 per cent, find the carried load for an individual user.

Solution

The carried load $(1 - P_b)\, A_{user} = (1 - 0.1)0.75 = 0.675$ Erlang.

6.6 Blocking probability formulas

A call is said to be blocked or lost when it can't be completed because the connecting equipment is busy, even though the line that the caller intends to reach is idle. In traffic engineering, blocking probability is an important parameter. In conventional teletraffic engineering, three models are used for handling blocked calls.

- blocked call held (BCH)
- blocked call cleared (BCC)
- blocked call delayed (BCD)

The BCH concept assumes that the user will immediately re-attempt the call on receiving a congestion signal and will continue to redial. As and when the equipment is available, the user hopes to seize connection equipment or a trunk. The lost calls are held or waiting at the calling user's telephone. The principal traffic formula in North America is based on the BCH concept.

The BCC concept is primarily used in Asia, Europe, and Africa. The user hangs up and waits for some interval before re-attempting the call if the user hears the congestion tone on the first attempt. The *Erlang B* formula is based on this criterion.

In the BCD concept, it is assumed that the user is automatically put in a queue and is served when the connection equipment is available. The method by which a call is selected from the pool of waiting calls is based on the queue discipline. In the queuing system, the GoS is defined as the probability of delay. The *Erlang C* formula is based on this concept.

There are various blocking formulas to determine the number of circuits (or trunks) required on a route based on busy hour traffic load. The factors that are used include call arrivals and holding-time distributions, number of traffic sources, and availability and handling of lost calls. These factors help in determining which formulas to use given a particular set of circumstances.

Erlang B loss formula has been widely used outside the United States. Loss indicates the blockage probability at the switch due to either congestion or all trunks being busy. This is expressed as the GoS(G) or probability of finding channels busy.

6.6.1 Erlang B formula

The Erlang B formula is expressed as GoS or the probability of finding N channels busy.

The assumptions made in the Erlang B formula are as follows:

- Traffic originates independently from an infinite number of traffic sources.
- Lost calls are cleared assuming a zero holding time.
- Limited number of trunks or service channels.
- Full availability.
- Inter-arrival times of call requests are independent of each other.
- The service time (probability of a user occupying a channel) is based on an exponential distribution.
- Traffic requests are represented by a Poisson distribution implying exponentially distributed call inter-arrival times.

$B(N, A)$ = blocking probability

$$G_{\mathrm{B}} = \frac{A^N / A!}{\sum_{k=0}^{N-1} \frac{A^k}{k!}} \tag{6.3}$$

where N = number of serving channels and A = offered load.

6.6.2 Poisson's formula

The Poisson's formula is used for designing trunks on a route for a given GoS. It is used in the United States. The assumptions in Poisson's formula are as follows:

- Traffic originates from an infinite number of independent sources,.
- Traffic density per traffic source is equal,.
- Lost calls are held
- A limited number of trunks or service channels exist.

$$p_{\mathrm{b}} = e^{-A} \sum_{k=0}^{\infty} A^k / k! \tag{6.4}$$

where p_{b} = probability of blocking, A = offered load, and N = number of trunks or service channels.

6.6.3 Erlang C formula

The Erlang C formula assumes that a queue is formed to hold all requested calls that cannot be served immediately. Customers who find all N servers busy join a queue and wait as long as necessary to receive service. This means that the blocked customers are delayed. No server remains idle if a customer is waiting.

The assumptions in the Erlang C formula are as follows:

- Traffic originates from an infinite number of traffic sources independently.
- Lost calls are delayed.
- Number of trunks or service channels is limited.
- The probability of a user occupying a channel (called *service time*) is based on an exponential distribution.
- Calls are served in the order of arrival.

$$\text{Blocking probability } C(N,A) = \frac{A^N/[N!(1-A/N)]}{\sum_{k=0}^{N-1}\frac{A^k}{k!}+\frac{A^N}{N!(1-A/N)}} \tag{6.5}$$

where N = number of service channels and A = offered load.

The Erlang B formula holds even when the load is greater than number of servers ($A > N$) because, unlike the BCD model in which all calls are eventually served, the BCC model allows calls to be lost when all servers are busy. Therefore, the BCC system never becomes unstable.

6.6.4 Comparison of Erlang B and Poisson's formulas

A comparison between the Erlang B and Poisson's blocking formulas shows that Poisson's formula results in higher blocking than that obtained by the Erlang B formula for a given traffic load. For the Erlang B system, the offered load can be divided into the load lost and the carried traffic load A^* (i.e. amount of load serviced by system).

$$A^* = A[1 - B(N, A)] \tag{6.6}$$

where A = offered load and N = number of serving channels.

The carried traffic equals that portion of offered traffic load A that is not lost and $AB[N, A]$ which is the lost traffic. In a BCD system, no calls are lost. Thus, the carried load is equal to the offered load. The efficiency, ρ, of a BCC system is defined as the load carried per server (i.e. A^*/N).

6.6.5 Binomial formula

The assumptions used in the binomial formula are as follows:

- Traffic originates from a finite number of traffic sources independently.
- Traffic density per traffic source is equal.
- Lost calls are held in the system in a queue.

$$p_b = \left[\frac{s-D}{s}\right]^{s-1}\sum_{k=N}^{s-1}\binom{s-1}{N}\left(\frac{D}{s-D}\right)^k \tag{6.7}$$

where P_b = blocking probability, D = expected traffic density p, N = number of channels in the group of channels, and s = number of sources in group sources.

Example problem 6.3

The maximum calls per hour in a mobile cell equals 8,000 and the average call holding time is 80 s. If the GoS is 2 per cent, find the offered load A. How many service channels are required to handle the load?

Solution

$$A = \frac{8000 \times 80}{3600} = 166.67 \text{ Erlang (offered)}$$

Using the Erlang B table (Appendix A), N = 182 channels giving 168.3 Erlang at 2 per cent blocking.

Example problem 6.4

How many mobile subscribers can be supported with 36 service channels at 2 per cent GoS? Assume the average call holding time equals 120 s and the average busy hour call per subscriber is 1.2 calls/h.

Solution

From the Erlang B table in Appendix A, for 36 channels at 2 per cent blocking, the offered load = 27.34 Erlang. The carried load will be 27.34 × (1 – 0.02) = 26.79 Erlang.

$$\text{Average traffic per user} = \frac{1.2 \times 120}{3600} = 0.04 \text{ Erlang}$$

$$\text{Number of users} = \frac{26.79}{0.04} = 669$$

Call arrivals or requests for service are modelled as a Poisson random process. Holding times are predicted very well using an exponential probability distribution. This implies that calls of long duration are less frequent than short calls. If the traffic intensity (A) offered by a single user is A_{user}, then the intensity offered by N users is $A = NA_{user}$. This follows from the fact that a call arrival rate of λ_{user} for a single user implies an aggregate call arrival rate of $\lambda = N\lambda_{user}$ for N users. The statistical model is used to relate the offered traffic intensity, A, the GoS, $P_{b,}$ and the number of channels K needed to maintain the desired GoS.

Example problem 6.5

In a certain cellular system, an average subscriber places two calls per hour during the busy hour and holds for an average of 6 min. Each cell has 100 channels. If blocked calls are cleared, how many subscribers can be served by each cell at a 2 per cent GoS?

Solution

Using the Erlang B table in **Appendix A** with $K = 100$ and $P_b = 2$ per cent, we find that $A = 87.972$ Erlang. Now an individual subscriber offers a load of $A_{user} = 2$ calls/60 min × 6 min = 0.2 Erlang. Thus, the maximum number of subscribers is $N = \frac{A}{A_{user}} = 440$ subscribers.

Example problem 6.6

Suppose in the previous example, if the channels have been divided into two groups of 50 channels each and each subscriber is assigned to a group and can be served only by channels from that group. How many subscribers can be served by a two-group cell?

Solution

Using the Erlang B table with $K = 50$ and $P_b = 2$ per cent, we find that $A_{group} = 40.255$ Erlang. Then, the maximum number of subscribers per group is $N_{group} = \frac{A}{A_{user}} = 201$. Counting both groups, the maximum number of subscribers is $N = 2N_{group} = 402$ subscribers.

Example problem 6.7

Continuing the example, suppose that the set of channels has been divided into four groups of 25 channels each. How many subscribers can be served by a cell now?

Solution

With $K = 25$ and $P_b = 2$ per cent, the Erlang B table gives $A_{group} = 17.505$ Erlang. Then, $N_{group} = \frac{A}{A_{user}}$ $= 87$. Counting all four groups gives $N = 4N_{group} = 348$ subscribers maximum.

An important concept to learn from these examples is that the number of subscribers that can be supported by a given number of channels decreases as the pool of channels is subdivided. We can express this in terms of the trunking efficiency, ξ, defined as the carried load per channel, that is,

$$\xi = \frac{A_c}{K} = \frac{A(1-P_b)}{K}[0,1] \tag{6.8}$$

where, A_c = Carrier Load.

Example problem 6.8

If the GoS is 2 per cent, find the trunking efficiency for 100 channels offered as (a) a single group (b) two groups of 50 channels each, and (c) four groups of 25 channels each.

Table 6.2 Trunking efficiency for 100 channels

Number of groups	A_{group} (Erlang)	A (Erlang)	A_c (Erlang)	ξ
1	87.972	87.972	86.21	0.86
2	40.255	80.50	78.89	0.79
4	17.505	70.02	68.62	0.69

Solution

Using the offered load data from the previous series of examples, we can fill in Table 6.2. With a single group of 100 channels, each channel carries 86 per cent of the load it could carry if it were continuously in use. When the pool of channels is subdivided into four groups, then to maintain a 2 per cent GoS, each channel can carry only 69 per cent of a full load.

An AMPS system supports 395 voice channels per service block in each market area. For a seven cell cluster size, a cell can be assigned a maximum of 57 channels. As a complete pool, the 57 channels can support an offered load of A = 44,222 Erlang at a 1 per cent blocking probability. The carried load is A_c = (1 – 0.01)44,222 = 43.78 Erlang, and the trunking efficiency is ξ = 43.78/57 = 0.77 %per cent or 77 per cent.

If 120° sector antennas are used to improve the signal–-to-interference ratio, then there will be 19 channels available in each sector. The offered load is now $A = 3 \times 11.230 = 33.69$ Erlang (the offered load for one sector of 19 channels at 1 per cent GoS is 11.230 Erlang and there are three such sectors). The carried load is A_c = (1 – 0.01) 33.69 = 33.35 Erlang and the trunking efficiency is ξ = 33.35/57 = 0.59 per cent or 59 per cent. The carrier-to-interference ratio may be higher, but the amount of traffic that can be carried at the specified GoS is substantially decreased. The next example explores how system growth through sectoring is impacted by trunking efficiency considerations.

Example problem 6.9

A cellular system with an allocation of N_{chan} = 350 channels requires a 15 dB carrier-to-interference ratio. A seven cell cluster layout with omni-directional antennas has been performing adequately, but the system now needs additional channels to accommodate growth. An average subscriber places two calls per hour during the busy hours and holds each call an average of 3 min. Blocked calls are cleared and the system offers a GoS of 5 per cent. By what percentage can the number of subscribers be increased if 60° sectoring is introduced if the path loss exponent is $\nu = 4$.

Solution

Let us begin by estimating the number of subscribers that a cell can support before sectoring. Each subscriber produces an offered load of A_{user} = 2 calls/60 min × 3 min = 0.1 Erlang. For a seven cell cluster layout, the number of channels per cell is $N_{cell} = N_{chan}/N$ = 350/7 = 50. Using the Erlang B table in Appendix B with a GoS of 5 per cent and 50 channels, we find that each cell can support a carried load of $A = 44.53$ Erlang. The number N_s of subscribers is then $N_s = A/A_{user} = 44.53/0.1 = 445$ subscribers per cell.

We have encountered this sectoring problem previously in Example 2.4 of Chapter 2 in the Section 2.6. In that section, we showed that using 60° sectors, the 15 dB carrier to interference ratio could be maintained even if the cluster size is reduced to $N = 4$. Now, with the reduced cluster size, the number of channels per cell is increased to $N_{cell} = N_{chan}/N = 350/4 = 87$. Each cell is divided into six sectors, so the number of channels per sector is $N_{sector} = 87/6 = 14.5$. From the Erlang B table with 14 channels, we obtain $A_{sector} = 9.73$ Erlang. Then for the entire cell, $A = 6 \times 9.73 = 58.38$ Erlang. This gives the number of subscribers per cell of $N_s = A/A_{user} = 58.38/0.1 = 583$.

We now see that the number of subscribers has increased by 583/445 = 1.31 per cent or 31 per cent. This is a substantial increase, but recall that the example in the Section 2.6 showed an increase in the number N_{market} of channels of 75 per cent. Counting the channels does not give the true picture of the increase in the capacity, since dividing the cells into sectors causes a loss of trunking efficiency.

6.7 Operational techniques and technologies

Operational techniques are important because they are closer to implementations. These techniques are used to improve the overall efficiency of the system.

6.7.1 Adjusting system parameters

In this section, we deal with the adjustment of the parameters of the system. This is done in three ways which includes increasing the coverage area of a noise-limited system, reducing the interference levels and increasing the traffic capacity.

6.7.1.1 Increasing the coverage for a noise-limited system

There is no co-channel interference or adjacent-channel interference in a noise-limited system. This implies that either there are no co-channels and adjacent channels in the system or there would be negligible interference as the channel reuse distance is so large. To increase the coverage, the following techniques are used at the cell site.

- Increasing the transmitted power of each channel results in large coverage area. Here, two cases can be considered. In the first case, the transmitted power remains unchanged but the received power changes. If the received power is to be strong, the cell radius should be smaller. In the second case, the transmitted power changes but the received power doesn't.
- In a flat terrain, doubling the antenna height causes a gain increase of 6 dB. If the terrain contour is hilly, then an effective antenna height should be used, depending on the location of the mobile unit. Sometimes, doubling the actual antenna height results in a gain increase of less than 6 dB and sometimes more.
- Use of a high-gain or a directional antenna at the cell site also results in large coverage area.
- Lowering the threshold level of a received signal lowers the acceptable received power while increases the radius of the cell.
- In a noise-limited environment with a constant received power level, the carrier-to-noise ratio is large in a receiver with low front-end noise compared to that with a high front-end noise. This low-noise receiver can receive a signal from a farther distance than can a high-noise receiver.

- A diversity receiver is very useful in reducing multi-path fading. When the fading reduces, the reception level can be increased.
- Another approach is the selection of a proper site. For coverage purposes, select a high site if there is no risk of interference. However, when we need to cover an important area within the coverage area, it is necessary to move around the site location.
- Repeaters and enhancers are also used to enlarge the coverage area or to fill in holes.
- The technique of engineering the antenna patterns can be used to cover a desired service area.

6.7.1.2 Reducing the interference

If co-channels or adjacent channels were used in the system, the approaches discussed in Section 6.7.1.1 would cause interference. The methods for reducing the interference are as follows:

- Use of a good frequency-management chart.
- Following an intelligent frequency assignment strategy.
- Depending on the current situation, some channels may be noisy, some may be quiet, and some may be vulnerable to channel interference. These factors should be considered in assigning of frequency channels to a particular mobile unit.
- On the basis of antenna direction, a proper design of an antenna pattern is required for reducing the interference. This design tool should include the findings of signal requirements.
- Omni-directional antennas that use umbrella-pattern or downward tilting directional antennas are used in order to confine the energy within a small area.
- Antenna height can be reduced for reducing interference. This method can be used because reducing interference is more important than the radio coverage.
- In some cases, reducing transmitted power can be more effective in eliminating interference than reducing the height of the antenna.

6.7.1.3 Increasing the traffic capacity

Increase in traffic capacity is accomplished in the following manner:

- Reducing the size of the cell by controlling the radiation pattern increases the traffic capacity.
- Increasing the number of radio channels in each cell.
- Enhancing the frequency spectrum.
- Queuing of hand-off calls.
- Fixed channel assignment schemes can reduce the interference among channels.
- Dynamic, rather than fixed, channel assignment is another means of increasing traffic capacity. The external environmental factors, such as traffic volume, are considered in dynamic channel assignment.

6.7.2 Coverage-hole filler

Weak spots are created in a general area during antenna radiation similar to the formation of water puddles during a rainstorm. There are several methods for filling these holes and these are discussed in this section.

6.7.2.1 Enhancers (Repeaters)

An enhancer is used in an area that is a hole (weak spot) in the serving cell site. There are two types of enhancer: wideband and channelized enhancers.

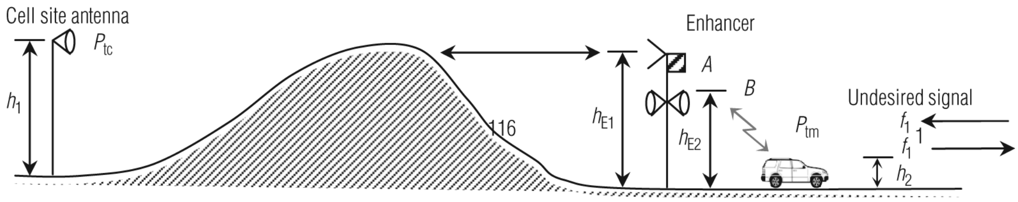


Figure 6.2 Enhancer

The wideband enhancer is a repeater that amplifies all the received signals. Sometimes, it can create inter-modulation products; therefore, implementation of an enhancer in an appropriate place is a challenging job. The amplifier in the application shown in Figure 6.2 requires only low amplification. At the enhancer site, a high directional antenna which is mounted at a high altitude receives the signal transmitted from the cell site which is radiated by the lower antenna.

The mobile units which receive the signal use the reverse channel to respond to calls through the enhancer to the cell site. The enhancer cannot improve the signal-to-noise (S/N) ratio as it amplifies both the signal and the noise.

The channelized enhancer amplifies only those channels that it selected previously with a good design. Therefore, it is a useful apparatus for filling the holes.

6.7.2.2 Passive reflector

The reflector system should be installed in a field far from both the transmitting antenna and the receiving antenna so as to redirect the incident energy. The approximate separation between the antenna and the reflector is

$$d_1 > \frac{2A_\mathrm{T}}{\lambda} + \frac{2A_1}{\lambda} \qquad \text{and} \qquad d_2 > \frac{2A_1}{\lambda} + \frac{2A_\mathrm{R}}{\lambda} \tag{6.9}$$

where,

A_T, A_R = effective aperture of transmitting and receiving antennas, respectively,
A_1 = effective aperture of the reflector,
d_1, d_2 = distance from reflector to transmitting and receiving antennas, respectively,
λ = wavelength.

6.7.2.3 Diversity

One method that is used to fill the holes is to make use of diversity receiver. A diversity receiver has an ability to receive signals that has lower signal levels. Therefore, the hole that existed in a normal receiver reception now becomes a no-hole or lesser hole situation. The diversity schemes can be classified into different categories. They are

- Polarization diversity
- Field component energy density
- Space diversity
- Frequency diversity
- Time diversity and
- Angle diversity

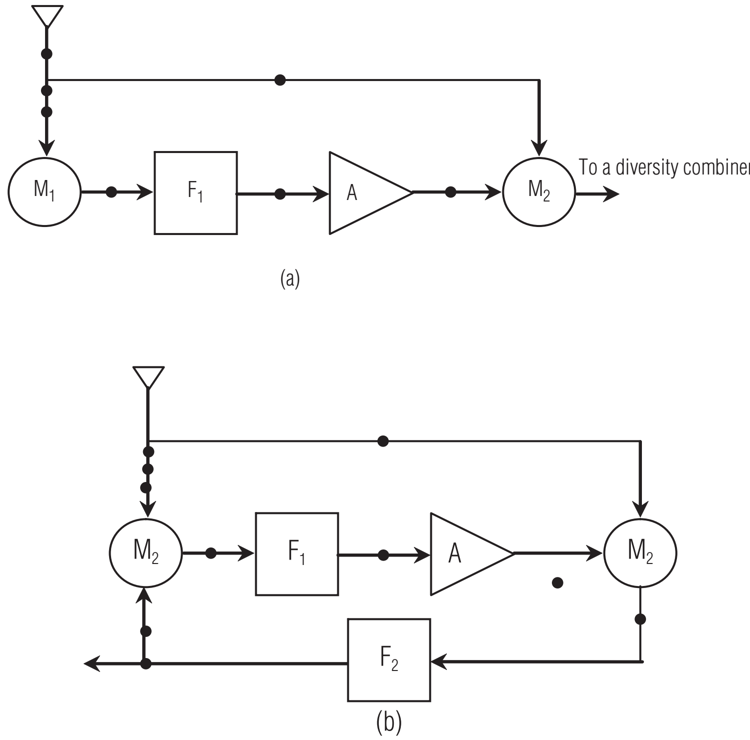


Figure 6.3 Two co-phase techniques: (a) Feed-forward combiner and (b) Feedback combiner (F = filter, M = mixer and A = limiting amplifier)

From any of the above diversity schemes, the performance obtained for any two independent branches is same.

This is because the correlation coefficient of the two received signals becomes zero. If the two signals obtained are dependent on the correlation coefficient, then the performance can be degraded. The performance can also vary with different diversity-combiner techniques. The maximal ratio combiner is the best performance combiner.

6.7.2.4 Co-phase technique

The co-phase technique is used to bring all signal phases from different branches to a common phase point. The point at which the random phase in each branch is reduced is known as *common phase point*. There are two kinds of co-phase techniques: feed-forward and feedback. These circuits are shown in Figure 6.3(a) and (b), respectively.

An important application of the feed-forward co-phase technique is in satellite communication. It is simpler than the phase-locked loop.

The feedback co-phase technique is also called as the *Granlund combiner.* A comparative study reveals that the outcome of the feedback technique is always better than that of the feed forward technique provided the two filters in the circuit have been properly designed to avoid any significant time delay.

6.7.3 Leaky feeder

A leaky feeder is a form of transmission line that enables radio communication to take place with or between mobile sets in its vicinity through its leakage fields, while substantially providing the linear range of the system through its internal propagation properties. Such systems are used in confined spaces, such as tunnels, mines, or large buildings, where natural propagation is inadequate. Leaky feeders may also be used on the surface, to confine radio coverage and thus improve spectrum efficiency.

6.7.4 Cell splitting

As mentioned in Chapter 2, when the call traffic in an area increases, cell splitting is used so that we can reuse the frequencies more often. This technique involves reducing the radius of a cell by half and splitting an old cell into four new small cells. Therefore, the traffic is increased fourfold.

Basically, there are two techniques of cell splitting. They are (a) permanent splitting and (b) dynamic splitting, which are briefly described below.

6.7.4.1 Permanent splitting

Permanent splitting is easy to handle as long as the cutover from large cells to small cells takes place during a low traffic period. Here, the frequency assignment should follow the rules based on the frequency-reuse distance ratio (q) with the power adjusted. However, in this technique selecting small cell sites is a tough job.

6.7.4.2 Dynamic splitting

Dynamic splitting is also called as real-time splitting. This scheme is based on utilizing the allocated spectrum efficiency in real time. Cell splitting proceeds gradually over a cellular operating system to prevent dropped calls. Suppose there exist a communication line between two stations 1A and 2A. If the area between two 2A sectors requires increased traffic capacity, then mid-point between two old 2A sectors is named as "new 2A". The new 1A is found by rotating the old 1A-2A line clockwise by 120°. Then the orientation of new set of seven split cells is determined.

To maintain service for the ongoing calls while doing the cell splitting, the channels assigned to the old 2A sector is separated into two groups: 2A = (2A) + (2A), where (2A) represents the frequency channels used in both new and old cells in small sectors, and (2A) represents the frequency channels used only in the old cells. At the early splitting stage only a few channels are in (2A). Gradually more channels will be transferred from (2A) to (2A). When no channel remains in (2A) the cell-splitting procedure will be completed.

6.7.5 Small cells (microcells)

> *A microcell is a cell in a mobile phone network served by a low power cellular base station (tower), covering a limited area, such as a mall, a hotel, or a transportation hub.*

A microcell uses power control to limit the radius of its coverage area. Typically, the range of a microcell is less than 5 km wide, a picocell is 200 m wide or less, and the width of a femto cell is of the order of 10 m.

A microcell is usually larger than a picocell.

A micro cellular network is a radio network composed of microcells.

Microcells are usually used to add network capacity in areas with very dense phone usage such as train stations. Microcells are often deployed temporarily during sporting events and other occasions in which extra capacity is known to be needed at a specific location ahead of time.

Cell size flexibility is a feature of 2G (and later) networks and is a significant part of how such networks have been able to improve capacity. Power controls implemented on digital networks make it easier to prevent interference from nearby cells using the same frequencies. By subdividing cells and creating more cells to help serve high-density areas, a cellular network operator can optimize the use of spectrum and ensure capacity can increase. By comparison, older analog systems have fixed limits beyond which attempts to subdivide cells simply would result in an unacceptable level of interference.

6.7.6 Narrow beam concept

The narrow-beam sector concept is another method for increasing the traffic capacity. This technique is shown in Figure 6.4. As an example, if we consider the total number of voice channels to be 310, then for $N = 7$ frequency-reuse pattern with 120° sectors as a conventional configuration, each sector will contain approximately 15 voice channels.

$$\frac{310}{3\times 7} \approx 15 \text{ channels per } 120° \text{ sector}$$

Similarly, for $N = 4$ frequency-reuse pattern with 60° sectors, the number of channels in each 60° sector is 13 channels.

$$\frac{310}{4\times 6} = 13 \text{ channels per } 60° \text{ sector}$$

In the $N = 7$ pattern, there are 21 sectors with 15 channels in each sector. In the $N = 4$ pattern, there are a total of 24 sectors with 13 channels in each sector. The spectrum efficiency of using these two patterns can be calculated using the Erlang B table in Appendix A. With a blocking

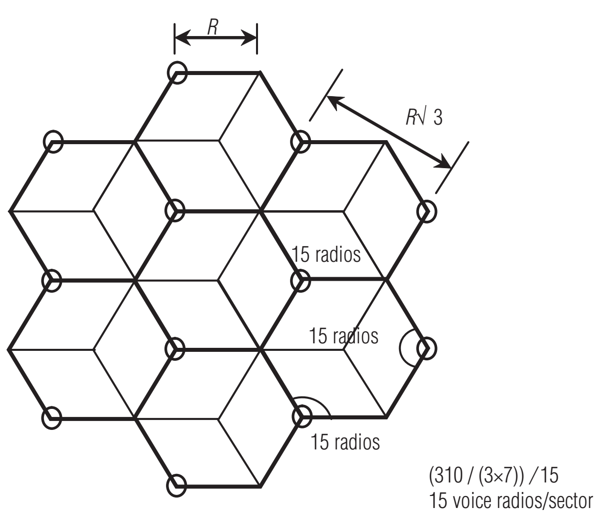


Figure 6.4 Ideally located cell sites over a flat terrain ($N = 7$)

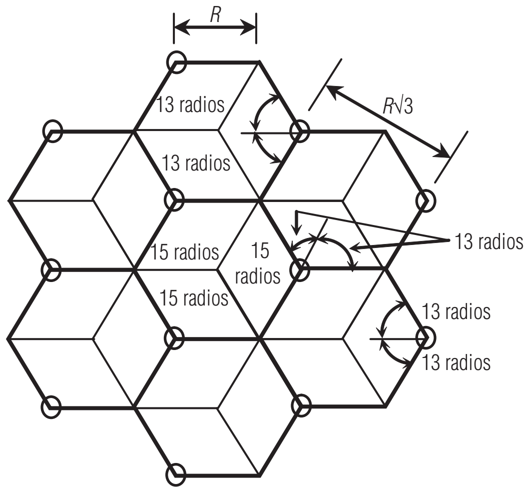


Figure 6.5 Mixed 120⁰ and 60⁰ sectors in ideally located cell sites (N = 7)

probability of 2 per cent, the results are an offer load of 189 Erlang for $N = 7$ and 177 Erlang for $N = 4$. This means that the $N = 7$ pattern offers a 7 per cent higher spectrum efficiency than the $N = 4$ pattern does.

However, the $N = 4$ arrangement results in increased handoff processing. Also, the antennas erected in each site with $N = 4$ pattern should be relatively higher than those with $N = 7$ pattern. Otherwise, channel interference among channels will be increased because the wrong frequency channels will be assigned to the mobile units due to the low antenna height in the system. As a result, the actual location of the mobile units in smaller sectors may be incorrect.

Therefore, for customizing the channel distribution, we could use the scheme shown in Figure 6.5; that is, usage of the 120° and 60° sectors can be mixed. In the $N = 7$ pattern, some 120° sectors can be replaced by two 60° sectors, thereby increasing the number of channels from 15 to 26.

This scheme would be suitable for small-cell systems. Besides, the 24 subgroups (each containing 13 channels) are used as needed in certain areas. These sector-mixed systems follow the $N = 7$ frequency-reuse pattern and the traffic capacity is dramatically increased as a result of customizing the channel distribution according to the real traffic condition.

6.8 Summary

- Traffic engineering uses a statistical model to determine how many users can be supported by a given number of channels.
- Congestion is defined as the situation when exchanges are inundated with calls and are unable to serve all the subscribers.
- A line that connects switching offices and which is shared among users on basis of requirement is called a trunk.
- As the number of trunks needed to make connections is much smaller than the maximum number that could be used, there might not be enough facilities to allow a call to be completed.

- A call that cannot be completed due to a lack of resources is said to be blocked.
- The application of queuing theory to establish the quantity of resources necessary to provide a given level of service is referred to as **traffic engineering**.
- An important parameter is the GoS which is the probability that a call will be blocked because no channel is available for it.
- Call capacity is defined depending on a view of the mobile switching center (MSC). In general, there are two approaches – global view and component view.
- One Erlang is equivalent to an arrival rate of one call per minute multiplied by a holding time of one minute.
- If by accident, a new call arrived just as the previous one terminated, and this continued to happen, then one Erlang of traffic would continuously tie up one channel or trunk.
- The Erlang formulas relate the number of channels, the GoS, and the number of subscribers.
- If a set of channels is subdivided into groups, the trunking efficiency goes down and fewer subscribers can be supported at a given GoS. This is what happens precisely when cells are divided into sectors, and some of the expected capacity gain is not realized owing to a loss in trunking efficiency.
- In conventional teletraffic engineering, three models are used for handling blocked calls.
  - blocked call held (BCH)
  - blocked call cleared (BCC)
  - blocked call delayed (BCD)
- The BCH concept assumes that the user will immediately re-attempt the call on receiving a congestion signal and will continue to redial.
- In the BCC concept, the user hangs up and waits for some interval before re-attempting the call if the user hears the congestion tone on the first attempt.
- In the BCD concept, it is assumed that the user is automatically put in a queue and is served when the connection equipment is available.
- The Erlang B formula is expressed as GoS or the probability of finding N channels busy.
- The Poisson formula is used for designing trunks on a route for a given GoS.
- The Erlang C formula assumes that a queue is formed to hold all requested calls that cannot be served immediately.
- A comparison between the Erlang B and Poisson's blocking formulas shows that Poisson's formula results in higher blocking than that obtained by the Erlang B formula for a given traffic load.
- Operational techniques are used to improve the overall efficiency of the system. The following are some of the technique discussed in this chapter
  - Adjusting system parameters
  - Coverage-hole filler
  - Leaky feeder
  - Cell splitting
  - Microcells
  - Narrow-beam concept
- The feedback co-phase technique is also called as the *Granlund combiner.*
- A microcellular network is a radio network composed of microcells.
- A leaky feeder is a form of transmission line that enables radio communication to take place with or between mobile sets in its vicinity through its leakage fields.

Review questions

1. A hexagonal cell within a four-cell system has a radius of 1.387 km. A total of 60 channels are used within the entire system. If the load per user is 0.029 Erlang and λ = 1call/h, compute the following for an Erlang C system that has a 5 per cent probability of a delayed call:
   (a) How many users per square kilometre will this system support?
   (b) What is the probability that a delayed call will have to wait for more than l0 s?
   (c) What is the probability that a call will be delayed for more than 10 s?
   (Ans: (a) 62 users/ km^2; (b) 56.29 per cent; (c) 2.81 per cent)
2. In a certain cellular system, an average subscriber places two calls per hour during the busy hour and holds for an average of 5 min. Each cell has 100 channels. If blocked calls are cleared, how many subscribers can be served by each cell at a 2 per cent GoS?
3. Assume each user of a single base station mobile radio system averages three calls per hour, each call lasting an average of 5 min.
   (a) What is the traffic intensity for each user?
   (b) Find the number of users that could use the system with 1 per cent blocking if only one channel is available.
   (c) Find the number of users that could use the system with 1 per cent blocking if five-trunked channels are available.
   (d) If the number of users you found in (c) is suddenly doubled, what is the new blocking probability of the five channel trunked mobile radio system? Would this be acceptable performance? Justify why or why not.
4. The maximum calls per hour in a mobile cell equals 4,000 and the average call holding time is 160 s. If the GoS is 2 per cent, find the offered load A. How many service channels are required to handle the load? (Ans: A = 166.67 Erlang, N = 182 channels)
5. How many mobile subscribers can be supported with 50 service channels at 2 per cent GoS? Assume the average call holding time equals 120 s and the average busy hour call per subscriber is 1.2 calls/h. (Ans: 986 subscribers)
6. How many subscribers can be supported if the total number of voice channels in a network is 504? Assume that the network uses a mixed 120° and 60° sectoring for a cluster size of (a) 4, (b) 7, and (c) 9.
7. Explain the concept of cell-splitting technique. (Refer Section 6.7.4)
8. Explain about trunking efficiency. (Refer Section 6.3)
9. Discus and explain the blocking probability of the cellular system. (Refer Section 6.6)
10. Let the maximum number of calls per hour Q_i in one cell be 5,000 and an average calling time T be 1.76 min. The blocking probability is 2 per cent. Find the offered load. If Q_i is 30,000, find the offered load and compare this with number of channels by using the Erlang B model charts. (Refer Section 6.6) (Ans: A = 146.67 and 880 Erlang, N = 160 and 900)

Objective type questions and answers

1. A call that cannot be completed due to a lack of resources is said to be
   (a) locked (b) blocked (c) incomplete (d) jailed
2. The probability that a call will be blocked because no channel is available for it is
   (a) blocking service (b) quality of service (c) service error (d) GoS

3. The field of traffic engineering was created by the work of
   (a) A. K. Erlang (b) T. S. Rappaport (c) Martin Cooper (d) Marconi
4. The situation that arises when exchanges are inundated with calls and are unable to serve all the subscribers is called
   (a) Flooding (b) Inundation (c) Congestion (d) Call tsunami
5. Having fixed connections between every possible pair of telephones is
   (a) physically impractical (b) economically prohibitive
   (c) very practical (d) both a and b
6. Call capacity is defined depending on a view of the
   (a) BS (b) MS (c) MSC (d) MTSO
7. The Erlang B formula is based on
   (a) BCC concept (b) BCD concept
   (c) BCH concept (d) none of the above
8. If a set of channels is subdivided into groups, the trunking efficiency
   (a) goes down (b) rises
   (c) remains constant (d) oscillates about a fixed value
9. The user will immediately reattempt the call on receiving a congestion signal and will continue to redial. This is assumed in the
   (a) BCC concept (b) BCD concept
   (c) trunking concept (d) call forwarding
10. If we lower the threshold level of a received signal at the cell site, then the radius of the cell
    (a) Increases (b) decreases (c) remains constant (d) becomes zero
11. To limit the radius of its coverage area, a microcell uses
    (a) diversity technique (b) high gain antenna
    (c) power control (d) Granlund combiner

Answer: 1. (b), 2. (d), 3. (a), 4. (c), 5. (d), 6. (c), 7. (b), 8. (a), 9. (b), 10. (a), 11. (c).

Open book questions

1. What are the different possible ways to adjust the system parameters in operational techniques?
2. What is the main advantage of cell splitting?
3. What does the Erlang formula relate to?
4. What is need for cell splitting and explain cell splitting? Explain the types of cell splitting.
5. Explain the factors used to measure the service quality of a cellular system.
6. Let the maximum number of calls per hour Q_i in one cell be 5,000 and an average calling time T be 1.76 min. The blocking probability is 2%. Find the offered load. If Q_i is 35,000, find the offered load and compare this with number of channels by using the Erlang B model.
7. For a cellular system, if blocking probability is 40 per cent and assuming each subscriber generates 0.1 Erlang traffic load, find the number of subscribers if number of channels are
   (i) 1
   (ii) 20
   (iii) 100

Key equations

1. The number of connections required to provide fixed connections between every possible pair of N telephones is given by

$$c_N = \sum_{k=1}^{N-1} k = \frac{N(N-1)}{2}$$

2. Blocking probability for Erlang B formula $B\,(N, A)$

$$G_B = \frac{A^N / A!}{\sum_{k=0}^{N-1} \frac{A^k}{k!}}$$

3. Blocking probability for Erlang C formula $C\,(N, A)$

$$\frac{A^N / [N!(1 - A/N)]}{\sum_{k=0}^{N-1} \frac{A^k}{k!} + \frac{A^N}{N!(1 - A/N)}}$$

4. Blocking probability for binomial formula

$$p_b = \left[\frac{s-D}{s}\right]^{s-1} \sum_{k=N}^{s-1} \binom{s-1}{N} \left(\frac{D}{s-D}\right)^k$$

5. Trunking efficiency is given by

$$\xi = \frac{A_c}{K} = \frac{A(1-P_b)}{K}$$

Further reading

Bellcore. "LATA Switching Systems Generic Requirements — Teletraffic Capacity and Environment," Technical Reference, TR-TSY-000517, Issue 3, March 1989.
Freeman, R. L. Telecommunication System Engineering. New York: John Wiley & Sons, 1989.
Lee, W. C. Y. *Mobile Cellular Telecommunications System*. New York: McGraw-Hill,1989.
Rappaport, T. S. "*Wireless Communications*." Upper Saddle River, NJ: Prentice Hall, 1996.
Sharma, R. L. et al. *Network Systems*. New York: Van Nostrand Reinhold Co., 1982.

7 Basic Antenna Theory

7.1 Introduction

An antenna converts electromagnetic (EM) energy to radio frequency (RF) waves in the case of transmission or RF waves into electrical energy in the case of reception. The physical dimensions (especially length) of an antenna are directly related to the frequency at which the antenna can propagate or receive waves. The physical structure of an antenna is directly related to the shape of the area in which it concentrates most of its radiated RF energy. Same antenna can be used for both transmitting and receiving the signals and this property of interchangeability is known as antenna *reciprocity*.

Any current flow in a conductor will set up a magnetic field. An alternating current flow will generate a varying magnetic field that in turn establishes a varying electric field (Maxwell's 3rd law known as Faraday's law of electromagnetic induction). These two alternating fields interact with one another and if the length of the conductor is sufficiently long, the conductor will *radiate*. In other words,

> *The coupling between time varying electric and magnetic fields produces EM waves capable of travelling through free space and other media.*

The shape and size of the current carrying structure determines the direction and the amount of energy to be radiated. We also know that an EM field will induce current in a wire. The shape and size of the structure determines how efficiently the field is converted into current, or in other words, determines how well the radiation is captured. The shape and size also determines from which direction the radiation is preferentially captured. The generic categories of RF antennas are omni-directional, semi directional, highly directional. Each category has multiple types of antennas, each having different RF characteristics and appropriate uses. The most common wireless LAN antenna is a dipole. Dipole simply means it is in two parts and radiates energy equally in all directions about its axis.

The basic antenna parameters are radiation pattern, beam area and beam efficiency (BE), effective aperture and aperture efficiency, directivity and gain, and radiation resistance. In this chapter, we will review the basics of antenna theory, types of antennas, properties of antennas used in wireless and microwave communication. These aspects are necessary for better understanding of the mobile antennas and cell site antennas presented in Chapters 8 and 9, respectively.

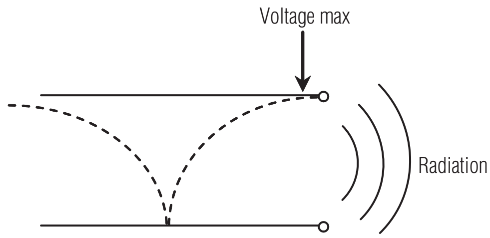


Figure 7.1 Open-ended transmission line

7.2 Basics of antennas

Before getting into a discussion on the various characteristics of antennas, we will first describe the principle of working of an antenna.

Antenna principle

The easiest way to explain antennas is to start with a transmission line with its output end left open. Try to recall from the theory pertaining to the open-circuited transmission lines that the voltage at this point is maximum, and there will be radiation of the energy from the open-ended line (Fig. 7.1). Any open-ended transmission line including microstrip transmission lines will radiate and act as an antenna. Although a certain amount of EM energy radiates from all open-circuit transmission lines, the distance for which the radiation of any consequence is minimal.

Therefore, we need some method to get the open-ended line to radiate power over a much longer range. Figure 7.2 illustrates a good start in getting more range out of the transmission-line antenna. The open transmission line now has a flared end on it and allows more energy to be radiated out into the air. The flared structure is the beginning of a familiar structure called a dipole. When we talk about a dipole antenna, we mean that the antenna has two poles associated with it.

One is on the top conductor and the other is on the bottom conductor. The combination of the two poles makes an economical and efficient antenna. Figure 7.2 is the beginning of the dipole antenna because the antenna usually does not look like the flared transmission line in the figure. It is presented like this to lead you into the actual construction of the common dipole antenna. The more recognizable representation of a dipole antenna is shown in Figure 7.3(a) with the conductors expanded until the distance between them is a quarter-wavelength ($\lambda/4$). This type of antenna is called a *quarter-wave dipole* or a *Marconi antenna*.

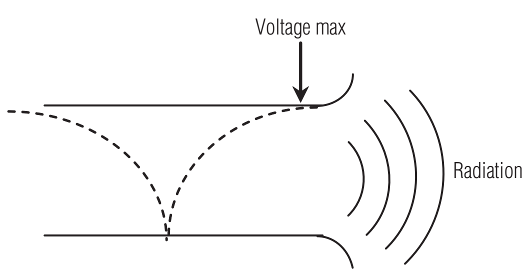


Figure 7.2 Transmission line with flared-end

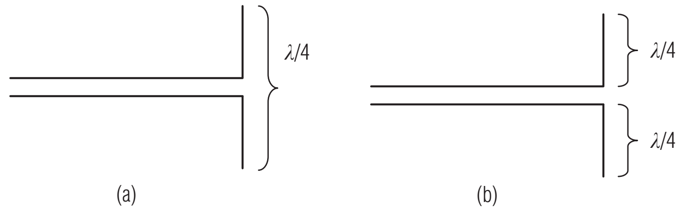


Figure 7.3 (a) λ/4 dipole antenna and (b) λ/2 dipole antenna

Figure 7.3(b) shows the most popular type of dipole antenna, which is the same type of antenna just described, except that we have continued to spread out the transmission line until each side is a $\lambda/4$. It now becomes a $\lambda/2$ dipole antenna commonly called a hertz antenna (Fig. 7.4(a)).

As the name suggests, the dipole antenna consists of two terminals or "poles" into which RF current flows. This current and the associated voltage causes an EM or radio signal to be radiated. Being more specific, a dipole is generally taken to be an antenna that consists of a resonant length of conductor cut to enable it to be connected to the feeder. For resonance, the conductor is an odd number of half wavelengths long. In most cases, a single $\lambda/2$ is used, although three, five,..., $(2n + 1)$ wavelength antennas are equally valid (where n is a positive integer).

The current distribution along a dipole is roughly sinusoidal. It falls to zero at the end and is at a maximum in the middle. Conversely, the voltage is low at the middle and rises to a maximum at the ends. It is generally fed at the centre, at the point where the current is at a maximum and the voltage is at a minimum. This provides a low impedance feed point, which is convenient to handle. High voltage feed points are far less convenient and more difficult to use.

To further understand antennas, you must know the terminology used to describe them.

Directivity (D): This parameter tells us how well the antenna radiates in a given direction, compared to the total radiated power.

Efficiency (η): An antenna is made usually of metals and dielectrics that have certain losses. The efficiency is defined as the total power radiated by the antenna divided by the input power to the antenna.

Gain (G): The gain is the product of the directivity and the antenna efficiency $(G = \eta D)$.

Impedance: An antenna is fed from some circuit and it looks like a load to that circuit. It is important to match the impedance of the antenna properly if we want it to radiate efficiently.

Bandwidth: The bandwidth of an antenna tells us in what frequency range the antenna has its specified characteristics. Usually, the impedance of the antenna changes with frequency, so the match to the circuit will change, resulting in degraded performance.

Polarization: The polarization of an antenna is defined with respect to the direction of an electric field vector that is radiated.

Radiation pattern: A graphical description of the radiation is the radiation pattern.

Feed point: An antenna is generally connected to a waveguide or transmission line in order to carry power to or from the antenna. The port at which this connection is made is called the feed point of the antenna.

A detailed description of the important antenna characteristics such as radiation pattern, directivity, polarization, and impedance is given in the following subsections.

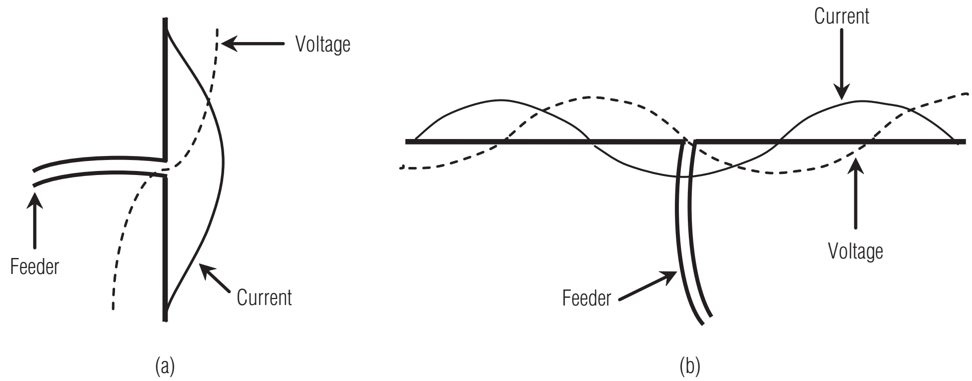


Figure 7.4 The basic (a) $\lambda/2$ and (b) $3\lambda/2$ wave dipole antennas

7.2.1 Radiation pattern

An antenna will radiate better in some directions, and worse in others. The *radiation pattern of an antenna* is a representation (pictorial or mathematical) of the distribution of the power flowing out (radiated) from the antenna (in the case of transmitting antenna), or flowing in (received) to the antenna (in the case of receiving antenna) as a function of direction angles from the antenna.

A radiation pattern is a polar diagram showing field strengths or the power densities at various angular positions relative to the antenna.

Distance between antenna's location and any point on the radiation pattern is a measure of radiation strength at that point. Non-spherical radiation patterns of antennas indicate that their performance is different in different directions.

The simplest type of antenna normally radiates most of its energy in one direction and this is called the "main lobe". The angular width of the main lobe is determined by the size and design of the antenna. It is usually described by a parameter known as *beamwidth*.

Beamwidth is the angle between the two directions in which the radiated power is half the maximum value of the beam.

The weaker secondary maxima in other directions are called *side lobes*. Although the pattern is a function of both elevation and azimuth angle, it is often specified only as a function of elevation angle in two special orthogonal planes, called the E-plane and the H-plane. The radiation pattern plot of a generic directional antenna is shown in Figure 7.5.

Some common parameters used to compare radiation patterns are defined as follows:

- *HPBW*: The half power beamwidth (HPBW) can be defined as the angle subtended by the half power points of the main lobe.
- *Main lobe*: This is the radiation lobe containing the direction of maximum radiation.
- *Minor lobe*: All the lobes other than the main lobe are called the minor lobes. These lobes represent the radiation in undesired directions. The level of minor lobes is usually expressed as a ratio of the power density in the lobe in question to that of the major lobe. This ratio is called as the side lobe level (expressed in decibels).
- *Back lobe*: This is the minor lobe diametrically opposite to the main lobe.
- *Side lobes*: These are the minor lobes adjacent to the main lobe and are separated by various nulls. Side lobes are generally the largest among the minor lobes.

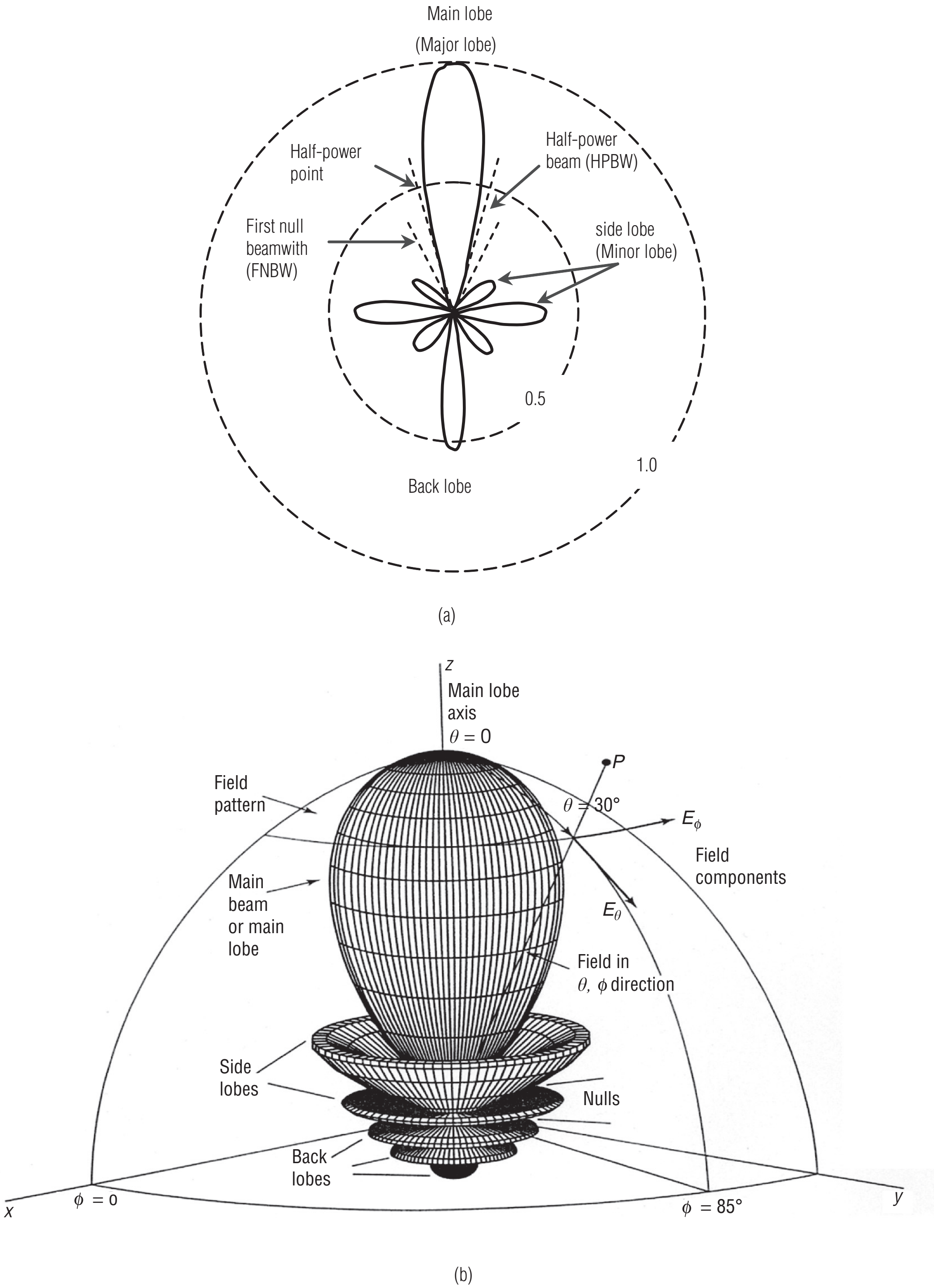


Figure 7.5 (a) Radition and (b) 3D radiation pattern of a generic antenna

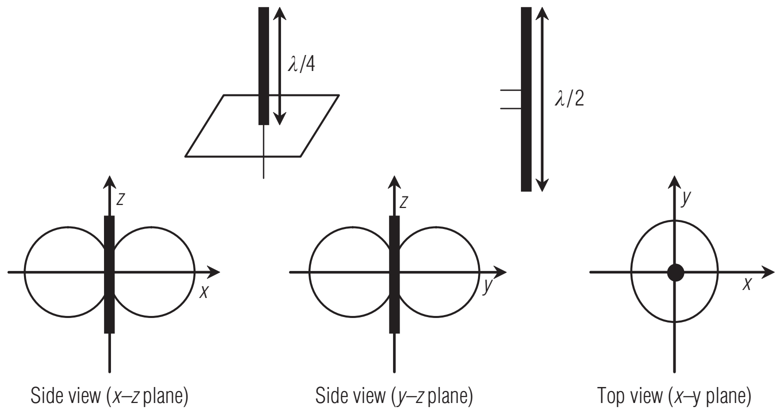


Figure 7.6 Radiation pattern of a simple hertzian dipole

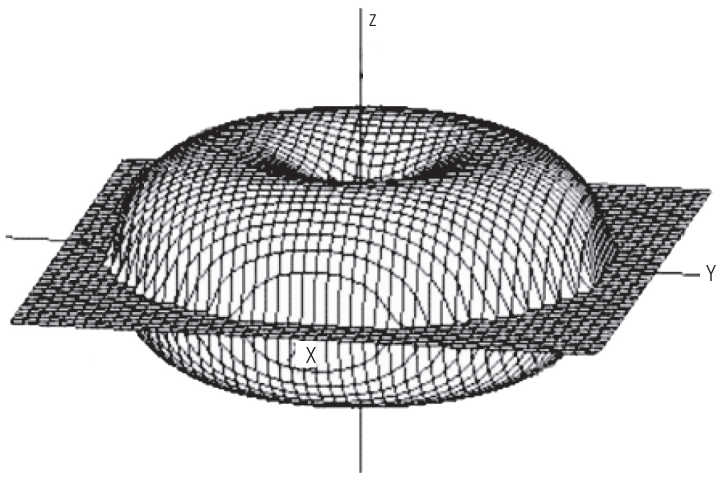


Figure 7.7 Hertzian dipole radiation pattern in x–y plane

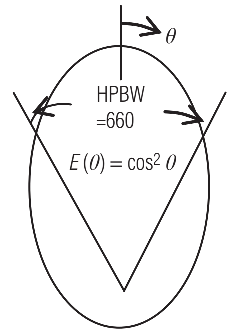
In most wireless systems, minor lobes are undesirable. Hence, a good antenna design should minimize the minor lobes.

Two special cases of radiation patterns are often referred to. The first is the *isotropic antenna*, a hypothetical antenna, which radiates power equally in all directions. This cannot be achieved in practice, but acts as a useful point of comparison. More practical is the *omnidirectional antenna* whose radiation pattern is constant in, say, the horizontal plane but may vary vertically. Real antennas are not isotropic radiators. Dipoles with lengths $\lambda/2$ are known as Hertzian dipole or $\lambda/4$ monopole (Fig. 7.6).

An antenna having a physical length of ½ the wavelength λ is called a hertz antenna or a half wave dipole antenna.

The Hertzian dipole is clearly omni-directional in the x–y plane as illustrated in Figures 7.6 and 7.7.

Example problem 7.1



Find the HPBW of an antenna having $E(\theta) = \cos^2\theta$ for $0° < \theta < 90°$

Solution

$E(\theta)$ at half power
$0.707 = \cos^2 \theta$ ($\because$ field pattern magnitude $= 1/\sqrt{2}$)
$\theta = 33°$ HPBW = 66°

7.2.2 Directivity

The directivity of a transmitting antenna is defined as the ratio of the radiation intensity flowing in a given direction to the average radiation intensity of all directions. Directivity is sometimes referred to as *antenna directive gain*.

In other words, the directivity of a non-isotropic source is equal to the ratio of its radiation intensity in a given direction, over that of an isotropic source.

$$D = \frac{U}{U_i} = \frac{4\pi U}{P} \tag{7.1}$$

where

D is the directivity of the antenna
U is the radiation intensity of the antenna
U_i is the radiation intensity of an isotropic source
P is the total power radiated

Sometimes, the direction of radiation is not specified. In this case, the direction of the maximum radiation intensity is implied and the maximum directivity is given by

$$D_{\max} = \frac{U_{\max}}{U_i} = \frac{4\pi U_{\max}}{P} \tag{7.2}$$

where

$D_{\max}$ is the maximum directivity
$U_{\max}$ is the maximum radiation intensity

Directivity is a dimensionless quantity since it is the ratio of two radiation intensities. Hence, it is generally expressed in decibels over isotropic (dBi).

$$\text{Directivity } (D) \text{ in dBi} = 10 \log D$$

The directivity of an antenna can be easily estimated from the radiation pattern of the antenna. An antenna that has a narrow main lobe would have better directivity than the one which has a broad main lobe. Based on directivity, antennas can be categorized into the following types:

- Omni-directional antennas
- Semi-directional antennas
- Highly directional antennas
- Sectorized and phased-array antennas

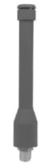


Figure 7.8 Dipole antenna

Omni-directional antennas: The omni-directional antenna radiates or receives equally well in all directions. The most popular omni-directional antenna is the dipole antenna (Fig. 7.8). They have 360° horizontal propagation pattern. They provide coverage on a horizontal plane with some coverage vertically and outward from the antenna. This means they may provide some coverage to floors above and below.

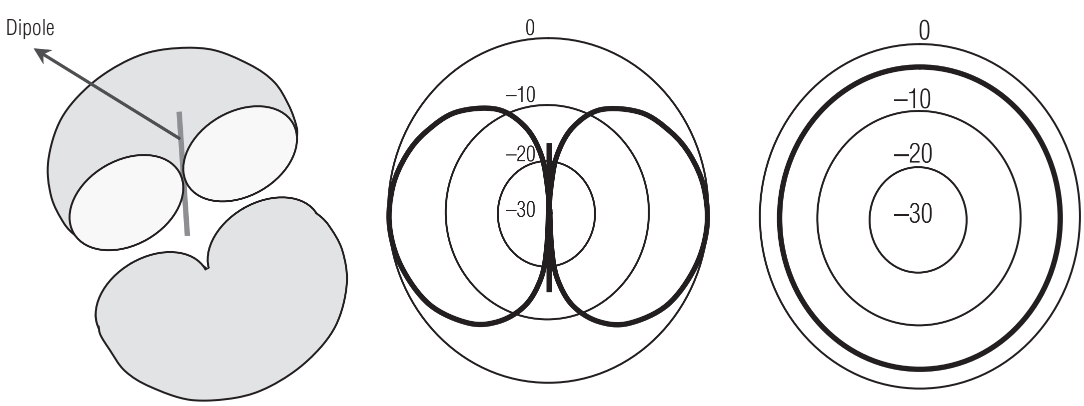
Figure 7.9(a) shows the radiation from an ideal vertical dipole. In the vertical plane, the cross-section of the radiation is in the shape of *figure of eight* on its side as shown in Figure 7.9(b), and radiation is uniform in the horizontal plane as shown in Figure 7.9(c). A directional antenna is one that radiates most of its power in one particular direction.

Examples of directional antennas are horns, reflector systems, log-periodics, and Yagis.

Semi-directional antennas are antennas that focus most of their energy in a particular direction. Examples are patch, panel, and Yagi antennas. Patch and panel antennas usually focus their energy in a horizontal arc of 180° or less, whereas Yagi antennas usually have a coverage pattern of 90° or less (Fig. 7.10).

Note: The dBi (the i stands for isotropic) represents a measurement of power gain used for RF antennas. It is a comparison of the gain of the antenna and the output of a theoretical isotropic radiator.

Highly directional antennas are antennas that transmit with a very narrow beam. These types of antennas often look like the satellite dish. They are generally called *parabolic dish* or *grid* antennas (Fig. 7.11(a)). Figure 7.12 illustrates the r*adiation pattern of highly directional antennas*.



(a) Perspective view

(b) Radiation pattern of a ½ λ dipole-elevation cut

(c) Radiation pattern of a ½ λ dipole-Azimuth cut

Figure 7.9 The radiation obtained from an idealized omni-directional antenna

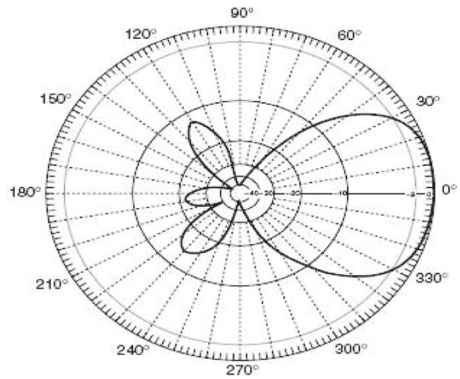


Figure 7.10 9 dBi Yagi antenna coverage pattern

Figure 7.11 Parabolic dish and grid antennas

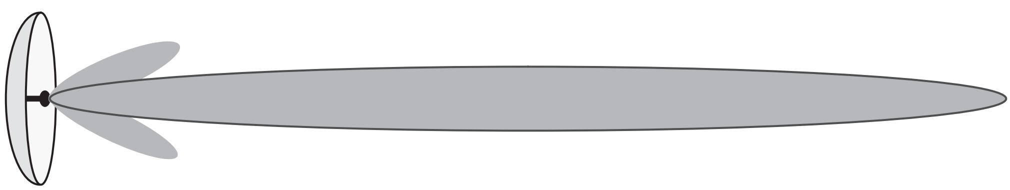


Figure 7.12 Radiation pattern of highly directional antennas

Advantages of highly directional antennas: A highly directional antenna, when used as a reciever improves the signal reception because it is pointed towards the origin of the signal. Multi- and omni-directional antennas pick up all signals from all directions, resulting in too many incoming signals and a weaker signal from the direction of choice. Highly directional antennas also "ignore" signals coming from places other than the one they are directed to, which cuts down the interference with a chosen signal. Another advantage these antennas offer is the ability to change the focus of the receiver to another direction. For example, when switching channels on a TV set, the grid aerial or antenna is moved around to get the best reception or signal.

Sectorized antenna is a high-gain antenna that works back-to-back with other sectorized antennas (Fig. 7.13). It is a directional antenna with a sector shaped radiation pattern (Fig. 7.14).

Phased-array antenna is a special antenna system that is actually composed of multiple antennas connected to a single processor. The antennas are used to transmit different phases that result in a directed beam of RF energy aimed at client devices.

Figure 7.13 Sectorized antenna

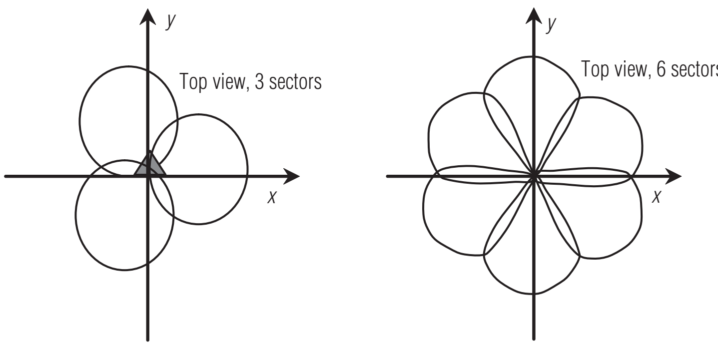


Figure 7.14 Sectorized antenna radiation pattern

7.2.3 Polarization

An antenna is a transducer that converts RF electric current to EM waves that are then radiated into space. Antenna polarization is an important consideration when selecting and installing antennas. The polarization of an antenna refers to the polarization of the electric field vector of the radiated wave. In other words, the position and direction of the electric field with reference to the earth's surface or ground determines the wave polarization.

The polarization of an EM wave is, by definition, the direction of the electric field.

The following are the most commonly encountered polarizations assuming that the wave is approaching.

- Linear
- Circular
- Elliptical

Linear polarization

Linearly polarized wave is a transverse EM wave whose electric field vector lies along a *straight line* at all times. In linear polarization, the path of the electric field vector is back and forth along a line (Fig. 7.15).

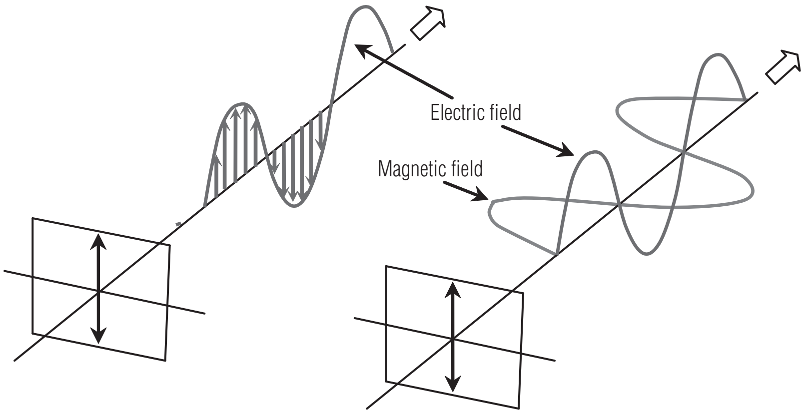


Figure 7.15 A linearly (vertically) polarized wave

Most wireless communication systems use either linear (vertical, horizontal) or circular polarization. Linear polarization can be of two types:

- Horizontal polarization
- Vertical polarization

Horizontal polarization: A linearly polarized wave is said to be horizontally polarized if its electric field is aligned in parallel with the horizontal axis.
Vertical polarization: A linearly polarized wave is said to be vertically polarized if its electric field is aligned in parallel with the vertical axis. The orientation of the electric field vector in case of vertical and horizontal polarizations is illustrated in Figure 7.16.

Circular polarization

In a circularly polarized wave, the electric field vector remains constant in length but rotates around in a circular path. Figure 7.17(a) shows an EM wave propagating in z-direction, where E_y is in y–z plane and E_x is in x–z plane. The resultant E-field will trace the spiral path as shown. When the resultant field is viewed from z-direction, it appears as a circle in x–y plane.

A left hand circular polarized wave is one in which the wave rotates counter clockwise as shown in Figure 7.17(b), whereas right hand circular polarized wave exhibits clockwise motion as shown in Figure 7.17(c).

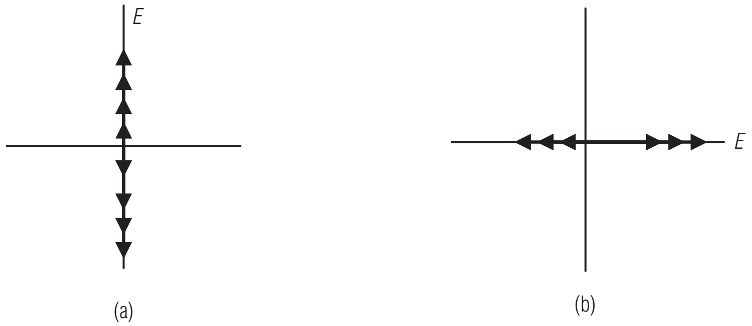


Figure 7.16 Orientation of electric field vector in vertical (a) and horizontal (b) polarizations

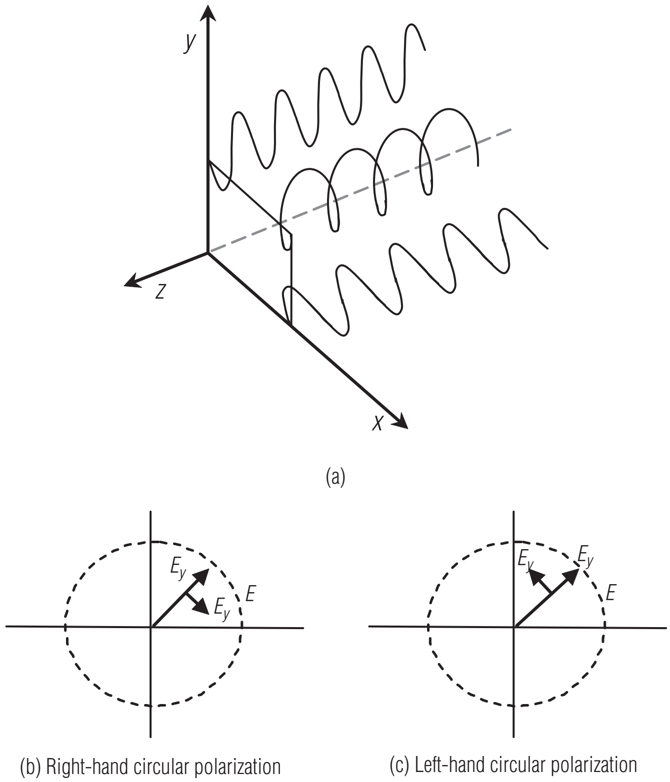


Figure 7.17 Circular polarization

Elliptical polarization

A wave is said to be elliptically polarized if either or both of the following conditions are satisfied:

- If E_x and E_y are not in phase, but have constant phase difference other than 90°.
- If the ratio of amplitudes of E_x and E_y is constant but not equal to one, then the wave is said to be elliptically polarized.

Here, the resultant electric field vector traces out an ellipse as shown in Figure 7.18. To understand clearly, let us consider a wave travelling in *z*-direction, where the *x* and *y* components have different amplitudes and a phase difference of 90°. The field satisfying such conditions is

$$E_0 = A\cos\omega t + B\sin\omega t \tag{7.3}$$

When the field in *x*-direction is *A cos ωt* and that in *y*-direction is *B* sin *ωt*,

$$\tilde{E}_x = A\cos\omega t \quad \text{and} \quad \tilde{E}_y = B\sin\omega t$$

When $A \neq B$

$$\frac{\tilde{E}_x^2}{A^2} + \frac{\tilde{E}_y^2}{B^2} = 1 \qquad (\because \cos^2\omega t + \sin\ \omega t = 1) \tag{7.4}$$

Thus, the wave is elliptically polarized. Elliptical polarization is a general case, whereas if the amplitudes of both the field components is equal and difference is 90°, then the polarization is said to be circular polarization. Such a field can be represented by

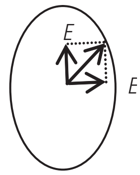


Figure 7.18 Elliptical polarization

$$E_0 = A\cos\omega t + A\sin\omega t$$

$$\tilde{E}_x = A\cos\omega t \quad \text{and} \quad \tilde{E}_y = A\sin\omega t$$

$$\tilde{E}_x^2 + \tilde{E}_y^2 = A^2$$

Example problem 7.2

A travelling wave has two linearly polarized components $E_x = 2\cos\omega t$ and $E_y = 3\cos\left(\omega t + \frac{\pi}{2}\right)$. Calculate (i) the axial ratio and (ii) the tilt angle of major axis of polarization ellipse.

Solution

Given that $E_x = 2\cos\omega t$ and $E_y = 3\cos\left(\omega t + \frac{\pi}{2}\right)$

(i) Axial ratio for $E_{mx}\cos(\omega t)$ and $E_{my}\cos(\omega t + \theta)$ is given by

$$\frac{E_{my}}{E_{mx}} = \frac{3}{2} = 1.5$$

(ii) The tilt angle of the major axis of polarization ellipse is the phase difference between

$$E_x \text{ and } E_y = \theta = 90°$$

7.2.4 Impedance

One major feature of an antenna that does change with frequency is its impedance. Antenna *impedance relates the voltage to the current at the input to the antenna.*

Case 1. Let's say an antenna has an impedance of $Z = 50\ \Omega$. This means that if a sinusoidal voltage is applied at the antenna terminals with an amplitude of 1 V, then the current will have an amplitude of $1/50 = 0.02$ amps. If the impedance is a real number, the voltage is in phase with the current.

Case 2. On the other hand, suppose if the impedance is a complex number, say

$$Z = 50 + j \times 50\ \Omega,$$

then the impedance has a magnitude $= \sqrt{50^2 + 50^2} = 70.71$ and

$$\text{phase} = \tan^{-1}\left(\frac{\text{Im}(Z)}{\text{Re}(Z)}\right) = 45^0$$

From the above equation, we can say that the current waveform is delayed relative to the voltage waveform by 45°.
The voltage (with frequency f) at the antenna terminals ($V(t)$) and current ($I(t)$) is given

$$\text{by } V(t) = \cos 2\pi ft,\ \ I(t) = \frac{1}{70.71}\cos(2\pi ft - \frac{\pi}{180}.45)$$

Hence, we can conclude that the real part of the antenna impedance represents power that is either radiated away or absorbed within the antenna. The imaginary part of the impedance indicates non-radiated power that is stored in the near field of the antenna. An antenna is said to be resonant, if the imaginary part of the impedance is zero.

From the point of view of the microwave circuit behind, which drives the antenna, the antenna can be represented as complex load impedance. The characteristics of this load depends on the radiation patterns of the antenna and hence the design of the antenna. The goal of a good design is to match the impedance of the antenna to the impedance of the transmission line connecting the antenna to the receiver. The impedance match can be characterized by any one of the following parameters:

- The voltage reflection coefficient, ρ_v
- The return loss (in decibels), $R_L = -20\log|\rho_v|$
- The voltage standing-wave ratio, $VSWR = \dfrac{1+|\rho_v|}{1-|\rho_v|}$

7.3 Other important antenna parameters

There are several other critical parameters that affect an antenna's performance and can be adjusted during the design process. These parameters are the resonant frequency, gain, bandwidth, reciprocity, effective area, beamwidth, and efficiency. These are dealt with in the following subsections.

7.3.1 Resonant frequency

An RF antenna is a form of tuned circuit consisting of inductance and capacitance, and as a result, it has a resonant frequency. This is the frequency where the capacitive and inductive reactances cancel out each other. At this point, the RF antenna appears purely resistive, the resistance being a combination of the loss resistance and the radiation resistance. The capacitance and inductance of an RF antenna are determined by its physical properties and the environment where it is located. The major feature or parameter of the RF antenna design is its dimensions. It is found that the larger the antenna or more strictly the antenna elements, the lower the resonant frequency.

The resonant frequency and electrical resonance is related to the electrical length of the antenna. The electrical length is usually the physical length of the wire multiplied by the ratio of the speed of wave propagation in the wire to the speed of wave in free space. Typically, an antenna is tuned for a specific frequency and is effective for a range of frequencies usually centred on that resonant frequency. However, the other properties of the antenna (especially radiation pattern and impedance) change with frequency.Antennas can be made resonant on harmonic frequencies with lengths that are fractions of the target wavelength. Some antenna designs have multiple resonant frequencies and some are relatively effective over a very broad range of frequencies. The most commonly known type of wide band aerial is the logarithmic or log periodic, but its gain is usually much lower than that of a specific or narrower band aerial.

7.3.2 Gain

Gain of an antenna relates the power radiated by the antenna to that radiated by an isotropic antenna.

In antenna design, "gain" is the logarithm of the ratio of the intensity of an antenna's radiation pattern in the direction of strongest radiation to that of a reference antenna. If the reference antenna is an isotropic antenna, the gain is often expressed in units of decibels over isotropic (dBi).

For example, when we say that the gain of an antenna is, for instance, 20 dBi (100 in linear terms) we mean that an isotropic antenna would have to radiate 100 times more power to give the same intensity at the same distance as that of the directional antenna.

Often, the dipole antenna is used as the reference since a perfect isotropic reference is impossible to build, in which case, the gain of the antenna in question is measured in decibels over dipole (dBd). The two gain measurements can also be converted using the following formula

$$\text{dBi} = \text{dBd} + 2.1 \tag{7.5}$$

The gain of an antenna is a passive phenomena – power is not produced by the antenna, but simply redistributed to provide more radiated power in a certain direction than would be transmitted by an isotropic antenna. If an antenna has a positive gain in some directions, it must have a negative gain in other directions as energy is conserved by the antenna. Therefore, there is a trade-off between the range of directions that must be covered by an antenna and the gain of the antenna. For example, a dish antenna on a spacecraft has a very large gain, but only over a very small range of directions. It must be accurately pointed at earth but a radio transmitter has a very small gain, as it is required to radiate in all directions.

If the directivity of the transmitting and receiving antennas is known then it is possible to compute the power received by the receiving antenna using either of the following formulas.

When using gain ratios and powers in "W" (watts):

$$P_{\text{received}} = \frac{P_{\text{transmitted}} G_T G_R \lambda^2}{16\pi^2 d^2} \tag{7.6}$$

where G_T is the transmitter gain,

G_R is the receiver gain,

D is the distance between transmitting and receiving antennas

Antenna gains should be expressed as a number, distances and wavelengths in m, and powers in W.

In decibels,

$$P_{\text{received}} = P_{\text{transmitted}} + G_T + G_R + 20\log(\lambda) - 20\log(d) - 21.98 \tag{7.7}$$

Path loss (L_p) or attenuation is the loss occurring in a transmission path due to spreading of the signal and is given by

$$L_p = \frac{P_T}{P_R} = \frac{(4\pi d)^2}{G_T G_R \lambda^2}$$

L_p in decibels = 32.5 + 20 log f + 20 log d – 10 log G_T – 10 log G_R

where f is in MHz and d in km.

Antenna gain should be expressed in dBi, wavelength and distances in 'm' (meters) and powers in dBm or dBW.

7.3.3 Bandwidth

While designing efficient antennas, care is to be taken to keep imaginary value of the impedance as small as possible, that is they are to be designed around the resonant point. Therefore, the bandwidth of antenna is very limited for efficient operations. If the antennas are operated away from resonant point, then not only levels of reactance rise, but also other characteristics of antenna may get impaired. This may result in the damage of the transmitter, if adequate safety measures are not taken. This constraint is relaxed in case of antennas used for receiving purpose.

In general, bandwidth is the range of frequencies over which the antenna system's SWR remains below a maximum value, typically 2.0. It can also be defined as the range of frequencies, on either side of a centre frequency, where the antenna characteristics such as input impedance, radiation pattern, beamwidth, polarization, side lobe level, gain, beam direction, and radiation efficiency are within the acceptable limits. For broadband antennas, the bandwidth is usually expressed as the ratio of the upper-to-lower frequencies for acceptable operation. The bandwidth percentage is expressed as a percentage of the frequency difference over the centre frequency of the bandwidth.

When we refer to the bandwidth, we are usually referring to the frequency bandwidth. The frequency band of 3–30 MHz refers to frequencies above 3 and up to 30 MHz. An antenna has its maximum gain at its tuned frequency, but can still receive other frequencies, albeit with reduced gain. Another parameter of importance is bandwidth percentage given as

$$\text{Bandwidth\%} = \frac{\text{Bandwidth of antenna}}{\text{Centre frequency}} \times 100$$

Example problem 7.3

An antenna operates with a bandwidth of 6 MHz and the minimum and maximum operating frequencies are 97 MHz and 103 MHz, respectively. Find the bandwidth percentage?

Solution

Given data: Bandwidth of antenna = 6 MHz
Maximum and minimum operating frequencies = 97 MHz and 103 MHz, respectively

$$\text{Bandwidth\%} = \frac{6\text{ MHz}}{100\text{ MHz}} \times 100 = 6\%$$

7.3.4 Reciprocity

- Antenna reciprocity means that the characteristics and performance of an antenna are the same whether the antenna is radiating or intercepting an EM signal.
- A transmitting antenna takes a voltage from the transmitter and converts it into an EM signal.
- A receiving antenna has a voltage induced into it by the EM signal that passes across it.
- *Reciprocal antennas have identical radiation patterns as transmitter and receiver antennas.*

7.3.5 Effective area

Antennas capture power from passing waves and deliver some of it to the terminals. Given the power density of the incident wave (S) and the effective area (A_e) of the antenna, the power delivered (P_d) to the terminals is the product of the both.

$$P_d \text{ in Watts} = S \quad \text{in Watts} / m^2 \times A_e \text{ in square metres}$$

For an aperture antenna such as a horn, parabolic reflector, or flat-plate array, effective area is physical area multiplied by aperture efficiency. In general, losses due to material, distribution, and mismatch reduce the ratio of the effective area to the physical area. Typical estimated aperture efficiency for a parabolic reflector is 55%. Even antennas with infinitesimal physical areas, such as dipoles, have effective areas because they remove power from passing waves.

The relationship between antenna gain (G) and effective area is given by

$$G = \frac{4\pi A_e}{\lambda^2} = \frac{4\pi f^2 A_e}{c^2} \tag{7.8}$$

where,

G is the antenna gain
A_e is the effective area
f is the carrier frequency
c is the speed of light
λ is the carrier wavelength

7.3.6 Beamwidth

When we refer to the beamwidth of an antenna, we are referring only to the width of the main lobe or beam of the antenna and not the side lobes. In general, the larger the antenna the smaller is its beamwidth for the same frequency so that the beamwidth of an antenna is inversely proportional to its physical size. If the antenna does not have the same dimensions in all planes, the plane containing the largest dimension will have the narrowest beamwidth. The beamwidth of an antenna defined in three ways are as follows:

- Half power beamwidth
- 10 dB beamwidth
- First null beamwidth

When the beamwidth is quoted, it is usually assumed that it is the half power or 3 dB beamwidth, that is the width in degrees (or sometimes in radians) of the main beam across which the gain drops to 3 dB below that of its level at bore-sight. In the case of highly directional antennas that have a narrow main lobe, the 10 dB beamwidth is often used, that is the width at the points on either side of main beam where the radiated power is one-tenth of the maximum value. Large reflector antennas may have gains as high as 60 dBi, that is linear gains one million times greater than that of an isotropic antenna. For wider beamwidth antennas, the beamwidth is quoted to the first nulls, that is the width across which the main beam drops to the first nulls. A tuned half wave dipole has a beamwidth of about 78° whereas a Hertzian dipole has a beamwidth of about 90°.

The directivity of an antenna increases as its beamwidth is made smaller ($\because$ D= 4π/beam area). For large antennas, with a single major lobe, the half-power beamwidths of the antenna in the $\theta°$ and $\phi°$ directions may be related to its directivity by the following approximate formula:

$$D \approx \frac{41{,}000}{\theta^0_{HP}\phi^0_{HP}} \tag{7.9}$$

where θ^0_{HP} and ϕ^0_{HP} are the half-power beamwidths in degrees in elevation and azimuthal planes, respectively.

7.3.7 Efficiency

Efficiency (η) is the ratio of radiation resistance to total antenna input resistance and is given by

$$\eta = \frac{P_{\text{radiated}}}{P_{\text{input}}} = \frac{R_{\text{radiated}}}{R_{\text{loss}} + R_{\text{radiation}}} \tag{7.10}$$

Where

R_{radiated} is radiation resistance. It is antenna's effective resistance that is caused by radiation of EM waves from antenna. It depends on geometry of antenna. If I is the current fed to antenna and P is power in radiated EM field, then

$$R_{\text{radiated}} = \frac{P}{I2}$$

Radiation resistance *does not dissipate power in the form of heat; the power is dissipated as radiated EM energy.*

The efficiency of practical antennas varies from less than 1% for certain types of low frequency antennas to 99% for some types of wire antennas.

Half-wave dipole radiation resistance

A half-wave dipole acts as a resonant circuit. At its resonant frequency, the antenna appears to be a pure resistance of 73 Ω (Fig. 7.19).

Example problem 7.4

Find the efficiency of an antenna whose radiation resistance and loss resistance are 100 Ω and 20 Ω respectively.

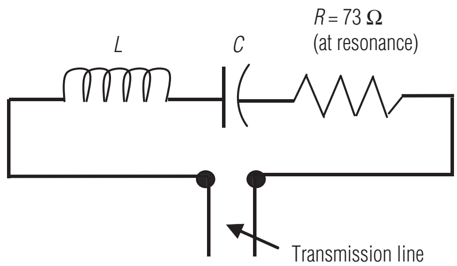


Figure 7.19 A half-wave dipole radiation resistance

Solution

The efficiency of an antenna, $\eta = \dfrac{R_{\text{radiated}}}{R_{\text{loss}} + R_{\text{radiation}}}$

$$\eta = \frac{100 \text{ ohms}}{(100 + 20) \text{ ohms}} = 83.33\%$$

Antenna efficiency in terms of directivity and gain

The efficiency η of an antenna is the ratio of the power delivered at the terminals of the antenna to the radiated power, expressed as a percentage:

$$\eta = G/D \times 100 \tag{7.11}$$

where

D is the directivity in linear terms

G is the gain in linear terms.

The efficiencies of antennas vary between about 50% and 100%.

If the gain and directivity are in decibels, then the efficiency would also be in decibels and is not usually expressed as a percentage in this case:

$$\eta_{\text{dB}} = G_{\text{dB}} - D_{\text{dB}}$$

Efficiencies of 50% and 100% are 0.5 and 1.0, respectivley, and since 10 log(0.5) is –3 and 10 log(1) is zero, the values in decibels would be –3 and 0 dB, respectively.

Beam efficiency

Another parameter that is frequently used to judge the quality of transmitting and receiving antennas is the BE. For an antenna with its major lobe directed along the z-axis ($\theta = 0$), the BEis defined by

$$\text{BE} = \frac{\text{Power transmitted (received) within cone angle } \theta_1}{\text{Power transmitted (received) by the antenna}} \tag{7.12}$$

where θ_1 is the half-angle of the cone within which the percentage of the total power is to be found.

7.4 Antenna arrays

In some applications, single-element antennas are unable to meet the gain or radiation pattern requirements of a particular application, such as satellite communications or point-to-point telecommunications. One possible solution to this problem is to combine several single antenna elements into an antenna array.

An important characteristic of an array is the change of its radiation pattern in response to different excitations of its antenna elements. Unlike a single antenna whose radiation pattern is fixed, an antenna array's radiation pattern, called the array pattern, can be changed upon exciting its elements with different currents (both current magnitudes and current phases). This gives us a freedom to choose (or design) a certain desired array pattern from an array, without changing its physical dimensions. Furthermore, by manipulating the received signals from the individual antenna elements in different ways, we can achieve many signal processing functions such as spatial filtering, interference suppression, gain enhancement, target tracking, and so on.

Advantages of using antenna arrays are as follows:

- They can provide the capability of a steerable beam (radiation direction change) as in smart antennas.
- They can provide a high gain (array gain) by using simple antenna elements.
- They provide a diversity gain in multipath signal reception.
- They enable array signal processing.

Arrays can be designed in the following configurations:

- *Broadside* or *end-fire*
- Parasitic
- Driven arrays

7.4.1 Broadside versus end-fire arrays

Arrays can be designed in either *broadside* or *end-fire* configurations. Broadside arrays radiate perpendicular to the array orientation (the *z*-axis in Fig. 7.20), while end-fire arrays radiate in the same direction as the array orientation (the *y*-axis in Fig. 7.20).The broadside array results in increased directivity in both the horizontal and vertical plane. The broadside array is a stacked collinear antenna (Fig. 7.21(a)).

7.4.2 End-fire antenna

- The *end-fire array* uses two half-wave dipoles spaced one-half wavelength apart (Fig. 7.22).
- The end-fire array has a bidirectional radiation pattern, but with narrower beamwidths and lower gain.
- The radiation is in the plane of the driven elements.
- A highly unidirectional antenna can be created by careful selection of the optimal number of elements with the appropriately related spacing.

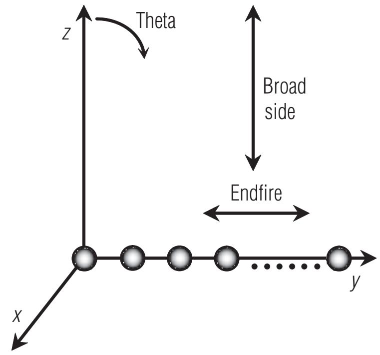


Figure 7.20 Topology of a linear array

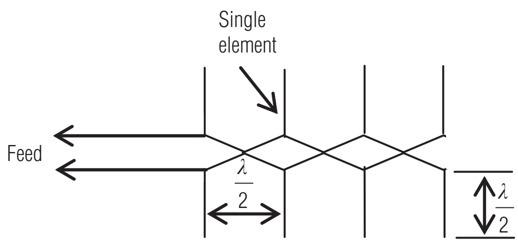


Figure 7.21 Broadside array

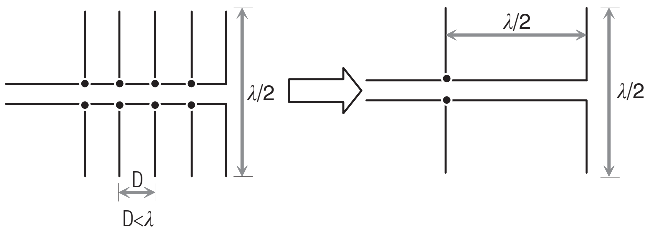


Figure 7.22 End-fire antennas: (a) bidirectional and (b) unidirectional

7.4.3 Parasitic arrays

In this array, only one element is connected to the transmitter. The other elements are coupled to the driven element through the electric fields and magnetic fields that exist in the near field region of the driven element.

Consider the half-wave dipole with a single half-wave parasitic element below. Figure 7.23 shows the radiation pattern with and without the reflector.

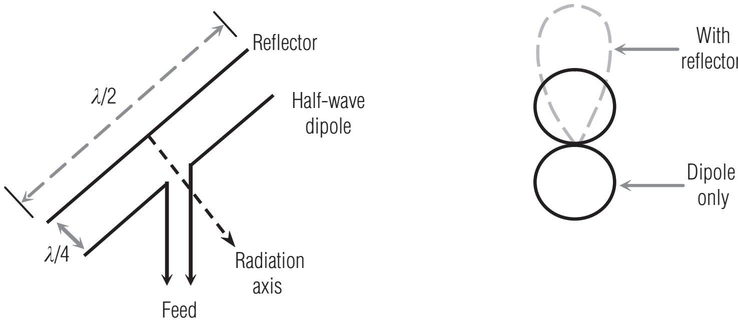


Figure 7. 23 Single half-wave parasitic element radiation pattern

Parasitic element is also called a reflector because it reflects the energy of the driven element.

Operation of parasitic element (Fig. 7.23): The driven element radiates as normal. This induces voltages and currents in the parasitic element causing it to radiate also. Reflection introduces a 180° phase shift. Radiation arriving back at the dipole is in phase, while the radiation going in the reverse direction is out of phase and causes cancellation.

7.4.4 Driven arrays

A *driven array* is a multi-element antenna in which all of the elements are excited through a transmission line. There are many types of driven arrays. The four most common types are

- collinear array
- broadside array
- log periodic array
- Yagi–Uda array

Collinear antenna

Collinear antennas usually consist of two or more half-wave dipoles mounted end to end.

They typically use half-wave sections separated by shorted quarter-wave matching stubs, which ensure that the signals radiated by each half-wave section are in phase.

Collinear antennas are generally used only on VHF and UHF bands because their length becomes prohibited at the lower frequencies. Figure 7.24 illustrates the radiation pattern of four element collinear antenna.

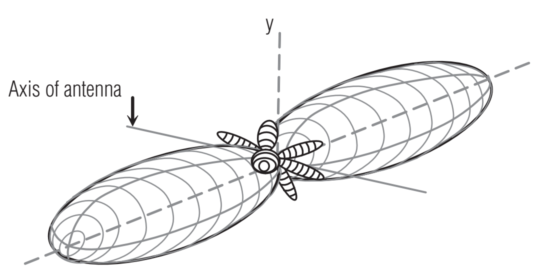


Figure 7.24 Radiation pattern of a four-element collinear antenna

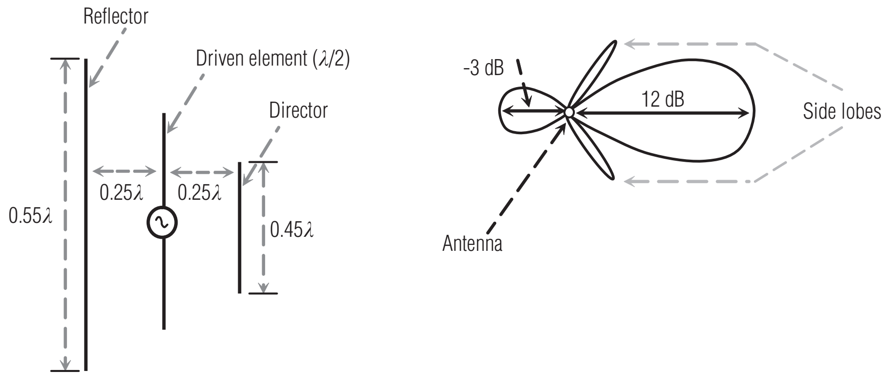


Figure 7.25 Yagi–Uda array (Yagi)

Yagi–Uda array (Yagi)

This antenna, named after its inventors, Yagi and Uda, consists of a driven element and one or more parasitic elements called reflectors and directors. The director is a parasitic element that "directs" EM energy in the desired direction (Fig. 7.25). It is used extensively for the reception of TV signals and can be seen on the roofs of most houses. The driven element is connected to the transmitter and the remaining elements are coupled to the driven element through its EM field.

The driven element is usually a folded dipole since it offers an improved bandwidth over a standard half-wave dipole antenna and moreover allows the antenna to be fed from a 300 Ω balanced line. The directors guide the EM waves either away from or towards the driven element depending on whether it is transmitting or receiving. Since the gain is closely related to the directivity of an antenna, the addition of directors in front of this driven element will improve the gain of an antenna. Each director improves the gain by approximately 1 dB. Typical beamwidths are 20–40°.

The front-to-back ratio (F/B ratio) is the ratio of the power radiated in the forward direction to the power radiated in the backward direction.

$$\text{F/B} = 10 \log P_f/P_b \text{ dB}$$

where

Pf is the forward power and *Pb* is the backward power.

If the radiation patterns are plotted in decibels, the F/B ratio is simply the difference between the forward value and the backward value, in decibels.

More complicated Yagi–Uda antennas consist of a reflector and many directors to improve gain. This type antenna design is common for HF transmitting antennas and VHF/UHF television receiving antennas (Fig. 7.26).

Log-periodic dipole array

> *The main advantage of an* log-periodic dipole array *(LPDA) is that it is frequency independent. Its input impedance and gain remain more or less constant over its operating bandwidth, which can be very large.*

The LPDA consists of a system of driven elements, but not all elements in the system are active on a single frequency of operation (Fig. 7.27). The EM fields produced by these active elements

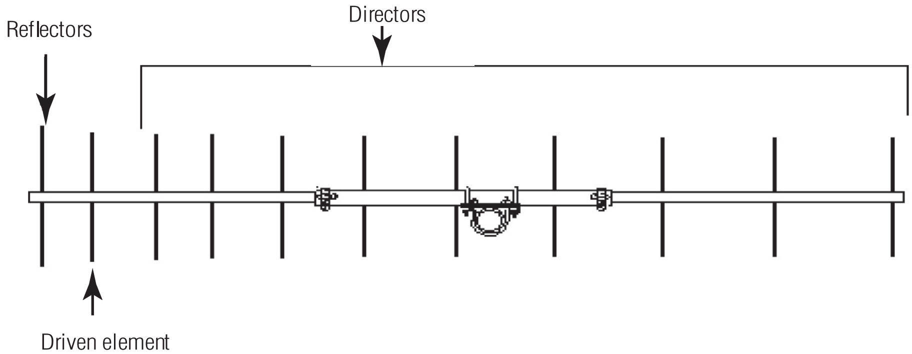


Figure 7.26 Element Yagi–Uda array antenna

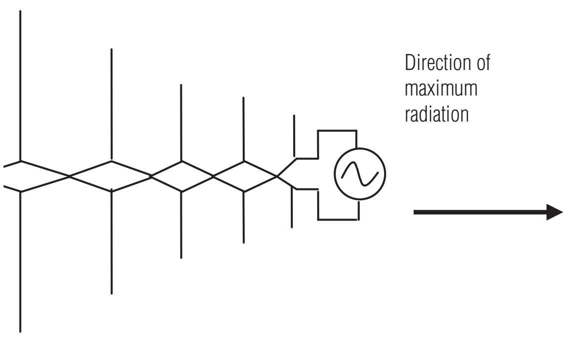


Figure 7.27 Log-periodic dipole array

add up to produce a unidirectional radiation pattern, in which maximum radiation is from the small end of the array. The radiation in the opposite direction is typically 15–20 dB below the maximum. The ratio of maximum forward to minimum rearward radiation is called the FB ratio and is normally measured in decibels.

The length ratio between adjacent dipoles is a constant (τ) and the ratio of element spacing to twice the next larger element length is a constant (σ). The dipoles are connected to the source using a twin transmission line in such a way that the phase is reversed at each connection relative to the adjacent elements.

Figure 7.28 shows a simplified way of connecting the dipoles to a transmission line. Each dipole is effective over a narrow band of frequencies determined by its length. When they are all connected to the twin transmission line, their narrow bandwidths add up to give a wider bandwidth. The length ratio (τ) is chosen in such a way that the antenna's performance will be uniform over the whole bandwidth. The shortest dipole corresponds to the highest frequency band and the longest dipole to the lowest frequency band of an antenna. A number of the dipoles whose frequency bands are in the vicinity of the selected resonant frequency will also be active. The bandwidth ratio (B) is given approximately by

$$B = B_s/B_{ar}$$

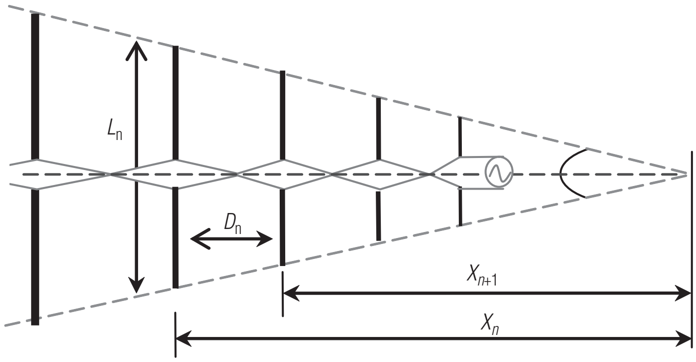


Figure 7.28 Log-periodic array antenna

where

B_s is the structure bandwidth and B_{ar} is the bandwidth of the active region.

$$\tau = \frac{X_{n+1}}{X_n} \qquad \text{and} \qquad \sigma = \frac{d_n}{2L_n}$$

where

$n = 1,2,3....N$

The angle α is used for the calculation of B_{ar} and is given by

$$\alpha = 2\tan^{-1}\left(\frac{1-\tau}{4\sigma}\right)$$

The length (L) between the longest and shortest element is

$$L = \frac{\lambda_{\max}}{4}\left(1-\frac{1}{B_s}\right) \cot(\alpha)$$

The bandwidth of the active region (B_{ar}) is

$$B_{ar} = 1.1 + 7.7\,(1 - \tau)^2 \cot(\alpha)$$

The number of dipoles required is obtained from the formula

$$N = 1 + \frac{\log B_S}{\log(1/\tau)} \tag{7.13}$$

The analysis of log periodic dipole arrays shows that the characteristic impedance of the elements varies with frequency and thus an empirical formula (7.14) for calculating the average characteristic impedance has been devised.

$$Z_a = 120(\ln \frac{h}{a} - 2.25) \tag{7.14}$$

where

h is the half length of the longest dipole

a is the radius of the dipoles

The dipoles are mounted on two boom structures, which in essence are transmission lines since the dipoles mounted on them are uninsulated. As a consequence, the booms have to be isolated from each other by using dielectric spacers. The impedance of this parallel rod feeder can be obtained by using Equation (7.15).

$$Z_0 = \frac{R_0^2}{8\sigma' Z_a} + R_0\sqrt{\frac{R_0}{8\sigma' Z_a} + 1} \tag{7.15}$$

where

R_0 is the input impedance of the LPDA and can be found by using Equation (7.16).

$$R_0 = \frac{Z_0}{\sqrt{1 + \frac{Z_0}{4\sigma' Z_a}}} \tag{7.16}$$

where σ is the mean spacing factor and is given by $\sigma/\sqrt{\tau}$.

7.5 Summary

- An antenna is a device used to efficiently transmit and/or receive EM waves. The essential part of a wireless communication system is characterized by its radiation pattern, directivity, polarization, and impedance. To design a particular antenna, its parameters need to be considered.
- **Antenna performance parameters**
  Radiation pattern – angular plot of the radiation
  Omni-directional pattern – uniform radiation in one plane
  Directive patterns – narrow beam(s) of high radiation
  Directivity – ratio of antenna power density at a distant point relative to that of an isotropic radiator (*isotropic radiator* – an antenna that radiates uniformly in all directions (point source radiator))
  Gain – directivity reduced by losses
  Polarization – trace of the radiated electric field vector (linear, circular and elliptical)
  Impedance – antenna input impedance at its terminals
  Bandwidth – range of frequencies over which performance is acceptable (resonant antennas, broadband antennas)
- **Antenna pattern** – a graphical representation of the antenna radiation properties as a function of position (spherical coordinates) Common types of antenna patterns
  Power pattern – normalized power versus spherical coordinate position
  Field pattern – normalized *E* or *H* versus spherical coordinate position
- **Antenna field types**
  Reactive field – the portion of the antenna field characterized by standing (stationary) waves, which represent stored energy

Radiation field – the portion of the antenna field characterized by radiating (propagating) waves, which represent transmitted energy

- **Antenna field regions**
  Reactive near field region – the region immediately surrounding the antenna where the reactive field (stored energy–standing waves) is dominant
  Near-field (Fresnel) region – the region between the reactive near-field and the far-field where the radiation fields are dominant and the field distribution is dependent on the distance from the antenna
  Far-field region – the region farthest away from the antenna where the field distribution is essentially independent of the distance from the antenna (propagating waves)
- **Antenna pattern definitions**
  Isotropic pattern – an antenna pattern defined by uniform radiation in all directions, produced by an isotropic radiator (point source, a non-physical antenna, which is the only non-directional antenna)
  Directional pattern – a pattern characterized by more efficient radiation in one direction than another (all physically realizable antennas are directional antennas)
  Omnidirectional pattern – a pattern which is uniform in a given plane
  Principal plane patterns – the E-plane and H-plane patterns of a linearly polarized antenna
  E-plane – the plane containing the electric field vector and the direction of maximum radiation
  H-plane – the plane containing the magnetic field vector and the direction of maximum radiation
- **Polarization** The polarization of a plane wave is defined by the figure traced by the instantaneous electric field at a fixed observation point. The polarization of the antenna in a given direction is defined as the polarization of the wave radiated in that direction by the antenna.
- A linear array allows beam steering in one dimension permitting directivity to be obtained in a single plane. Hence, an omni-directional pattern can be synthesized.
- The collinear array consists of $\lambda/2$ dipoles oriented end-to-end. The centre dipole is fed by the transmitter and sections of shorted transmission line known as phasing lines connect the ends of the dipoles.
- A broadside array consists of an array of dipoles mounted one above another. Each dipole has its own feed line and the lengths of all feed lines are equal so that the currents in all the dipoles are in phase.
- The main advantage of a log-periodic dipole array (LPDA) is that it is frequency-independent. Its input impedance and gain remain more or less constant over its operating bandwidth, which can be very large.

Review questions

1. Define an antenna.
2. Define directivity.
3. What is meant by polarization? Mention various types of polarizations.
4. Write short notes on antenna arrays.

5. What are the different antenna parameters? Write in brief about each parameter.
6. What are the characteristics of a simple mobile antenna?
7. Write short notes on Yagi–Uda array.
8. Obtain the relation between directivity and beamwidth of an antenna.
9. Define antenna efficiency. What is radiation resistance?

Exercise problems

1. An antenna is having a gain of 8 db over a isotropic reference antenna and is radiating 800 watts. Calculate the power of the reference antenna. (Ans: 5047.65watts)
2. An Antenna is having a loss resistance of 10 Ω, power gain and directivity are in the ratio10:11. Find the radiation resistance of the antenna. (Ans: 100 Ω)
3. What are the lengths of elements for a Yagi–Uda antenna, which is operating at a frequency of 120 MHz? (Ans: Length of driven element = 3.988 ft, length of reflector = 2.86 ft, length of director = 3.845 ft)
4. Suppose an antenna has $D = 4$, $R_{\text{rad}} = 40\ \Omega$, and $R_{\text{diss}} = 10\ \Omega$. Find antenna efficiency and maximum power gain. (Ans: $e = 0.80$, $G_{\text{max}} = 3.2$).
5. A power of 100 W is supplied to an isotropic radiator. What is the power density (W/m^2) at a distance of 10 km? (Hint: power density (S) is given by $S(R,\theta) = P_t/4\pi R^2$, where P_t is the transmitter power.

Objective type questions and answers

1. The process of interchangeability of receiving and transmitting operations of antennas is known as
   (a) polarization (b) efficiency (c) reciprocity (d) directivity
2. A half-wave dipole antenna is also known as
   (a) Marconi antenna (b) vertical antenna (c) hertz antenna (d) phased array
3. An antenna that is a quarter-wavelength long connected such that the ground acts as a reflecting quarter-wavelength section is called a
   (a) Marconi antenna (b) vertical antenna (c) hertz antenna (d) phased array
4. The folded dipole antenna has
   (a) greater bandwidth than a half-wave dipole (b) 288 Ω
   (c) a and b (d) None
5. As the height of a half-wavelength antenna is reduced below a quarter-wavelength, the radiation resistance
   (a) increases (b) decreases (c) constant (d) zero
6. The advantage of an LPDA is
   (a) frequency dependent (b) frequency independent
   (c) constant frequency (d) none of the above
7. The minimum number of elements that a Yagi consists of
   (a) a single element (b) two elements
   (c) more (d) none of the above

8. Impedance match can be characterized by
   (a) voltage reflection coefficient (b) the return loss
   (c) VSWR (d) all
9. ____ antenna is defined as a hypothetical loss-less having equal radiation in all directions.
   (a) Monopole (b) Quarter-wave dipole
   (c) Half-wave dipole (d) Isotropic
10. ______ antenna is one which has the property of radiating or receiving EM waves more effectively in some directions than in others.
    (a) Omni-directional (b) Directional (c) Isotropic (d) Smart
11. The minimum separation between a transmitting antenna and a receiving antenna is necessary to avoid the ______________ problem.
    (a) co-channel interference (b) receiver desensitization
    (c) a & b (d) none

Answers: 1. (c) 2. (c) 3. (a) 4. (c) 5. (b) 6. (b) 7. (b) 8. (d) 9. (d) 10. (b) 11. (b)

Open book questions

1. Explain the various characteristics of antennas.
2. Define an antenna. Mention different types of antennas.
3. What is meant by radiation pattern?
4. Differentiate between the absolute gain and the relative gain of a transmitting antenna in a given direction.
5. What is meant by a passive antenna?
6. Describe the effects of antenna parameters on the cell interferers.

Key equations

1. The directivity of a non-isotropic source is equal to the ratio of its radiation intensity in a given direction, over that of an isotropic source.

$$D = \frac{U_{\max}}{U_{\mathrm{i}}} = \frac{4\pi U_{\max}}{P}$$

2. Efficiency (η) is the ratio of radiation resistance to total antenna input resistance and is given by

$$\eta = \frac{P_{\text{radiated}}}{P_{\text{input}}} = \frac{R_{\text{radiated}}}{R_{\text{loss}} + R_{\text{radiation}}}$$

3. The voltage standing-wave ratio

$$VSWR = \frac{1+|\rho_{\mathrm{v}}|}{1-|\rho_{\mathrm{v}}|}$$

4. Power received is

$$P_{\text{received}} = \frac{P_{\text{transmitted}} G_{\mathrm{T}} G_{\mathrm{R}} \lambda^2}{16\pi^2 d^2}$$

In decibel

$$P_{\text{received}} = P_{\text{transmitted}} + G_{\text{T}} + G_{\text{R}} + 20\log(\lambda) - 20\log(d) - 21.98$$

5. Bandwidth percentage

$$\text{Bandwidth \%} = \frac{\text{Bandwidth of antenna}}{\text{Centre frequency}} \times 100$$

6. The power delivered (P_{d}) to the terminals is the product of power density (S) of incident wave and effective area of antenna.

$$P_{\text{d}} = SA_{\text{e}}$$

7. The relationship between antenna gain (G) and effective area

$$G = \frac{4\pi A_{\text{e}}}{\lambda^2} = \frac{4\pi f^2 A_{\text{e}}}{c^2}$$

8. Directivity of antenna in terms of the half-power beamwidths of the antenna in the $\theta°$ and $\phi°$ directions is

$$D \approx \frac{41{,}000}{\theta^0_{\text{HP}}\phi^0_{\text{HP}}}$$

9. Beam efficiency is

$$\text{BE} = \frac{\text{Power transmitted (received) within cone angle } \theta_1}{\text{Power transmitted (received) by the antenna}}$$

10. The number of dipoles required in a log periodic array is

$$N = 1 + \frac{\log B_{\text{S}}}{\log(1/\tau)}$$

11. The average characteristic impedance of elements in log periodic array is

$$Z_{\text{a}} = 120(\ln \frac{h}{a} - 2.25)$$

Further reading

C. A. Balanis, *Antenna Theory: Analysis and Design*, 2nd edition. Chichester: John Wiley & Sons Ltd, ISBN 0-471-59268-4, 1997.

R. J. Coe and G. Held, "A parasitic slot array," *IEEE Trans. Antennas Propagation*, Vol. AP-12, No. 1, pp. 10–16, January 1964.

F. R. Connor, *Noise,* 2nd edition. London: Edward Arnold, ISBN 0-7131-3459-3, 1982.

D. DeMaw (ed.), *The Radio Amateur's Handbook*, 56th edition. Newington, CT: American Radio Relay League, 1979.

J. D. Kraus and R. Marfehka, *Antennas*, 2nd edition. US: McGraw Hill Education, ISBN 0-07-232103-2, 2001.

E. Ledinegg, W. Paponsek, and H. L. Brueckmann, "Low-frequency loop antenna arrays: Ground reaction and mutual interaction," *IEEE Trans. Antennas Propagation.*, Vol. AP-21, No. 1, pp. 1–8, January 1973.

J. E. Lindsay, Jr., "A parasitic end-fire array of circular loop elements," *IEEE Trans. Antennas Propagation.*, Vol. AP-15, No. 5, pp. 697–698, September 1967.

A. Shoamanesh and L. Shafai, "Design data for coaxial Yagi array of circular loops," *IEEE Trans. Antennas Propagation*, Vol. AP-27, No. 5, pp. 711–713, September1979.

A. Shoamanesh and L. Shafai, "Properties of Coaxial Yagi Loop Arrays", *IEEE Trans. Antennas Propagation,* Vol. AP-26, No. 4, pp. 547–550, July 1978.

W. L. Stutzman and G. A. Thiele, *Antenna Theory and Design*. Chichester: John Wiley & Sons Ltd, ISBN 0-471-04458, 1981.

L. C. Shen and G. W. Raffoul, "Optimum design of Yagi Arrays of Loops," *IEEE Trans. Antennas Propagation*, Vol. AP-22, pp. 829–830, November 1974.

8 Mobile Antennas

8.1 Introduction

Personal wireless communication is a true success story and has become a part of people's everyday lives around the world. In the early days of mobile communications quality of service (Qos) was often very poor. Nowadays, it is assumed the service will be ubiquitous, high speech quality, and the ability to watch and share streaming video or even broadcast television programmes. Even rural areas are obtaining good coverage in many countries. However, there are still geographical remote and isolated areas without good coverage and several countries do not yet have coverage in towns and cities. In other words, satellite mobile communication offers the benefits of true global coverage reaching into remote areas as well as populated areas. In this chapter, we will see what antennas are actually used for mobile communication and how it works, its radiation patterns, and functionality.

A mobile antenna is a device that is used to transfer guided electromagnetic (EM) waves (signals) to radiating waves in an unbounded medium, usually free space, and vice versa (i.e. in either the transmitting or receiving mode of operation). Mobile antennas are frequency dependent, that is an antenna is designed for a certain frequency band. Beyond the operating band, the antenna rejects the signal. Therefore, the antenna functions as a band-pass filter and a transducer. Mobile antennas are essential parts in communication systems. Therefore, understanding their principles is important. In this chapter, we introduce basic antenna fundamentals and their types.

There are many different types of antenna. The isotropic point source radiator (one of the basic theoretical radiator) is useful, because it can be considered as a reference to other antennas. The isotropic point source radiator radiates equally in all directions in free space. Physically, such an isotropic point source cannot exist. Most antennas gains are measured with reference to an isotropic radiator and are rated in decibels with respect to an isotropic radiator (dBi).

8.1.1 Principle of basic mobile antenna

Antennas transform wire propagated waves into space propagated waves. They receive EM waves and pass them onto a receiver or they transmit EM waves which have been produced by a transmitter. As a matter of principle, all the features of passive antennas can be applied for reception and transmission alike. From a connection point of view, the antenna appears to be a dual gate, although in reality it is a quad gate. The connection which is not made to a RF-cable is connected to the environment; therefore, one must always note that the surroundings of the antenna have a strong influence on the antennas electrical features (Fig. 8.1).

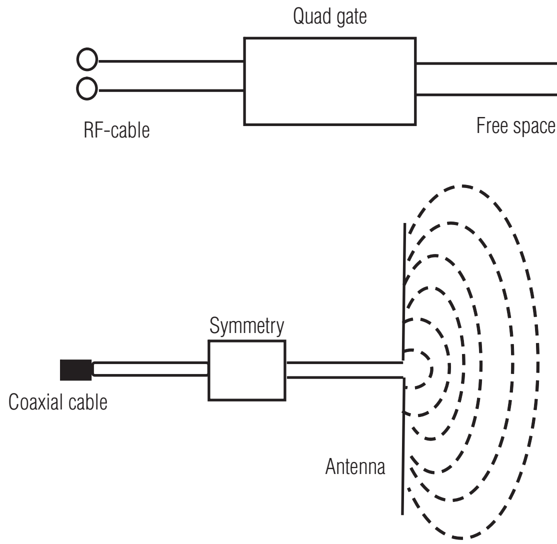


Figure 8.1 The antenna as quad gate

The principle of an antenna can be shown by bending a co-axial cable open (Fig. 8.2):

- A transmitter sends a high frequency wave into a co-axial cable. A pulsing electrical field is created between the wires, which cannot free itself from the cable.
- The end of the cable is bent open. The field lines become longer and are orthogonal to the wires.
- The cable is bent open at right angles. The field lines have reached a certain length, allows the wave to free itself from the cable.

The apparatus radiates an EM wave, whereby the length of the two bent pieces of wire corresponds to half of the wavelength.

This simplified explanation describes the basic principle of almost every antenna – the λ/2-dipole. Not only it is an electrical field (E) created due to the voltage potential (U) but also a magnetic field (H) which is based on the current (I) (Fig. 8.3). The amplitude distribution of both fields corresponds to the voltage and current distribution on the dipole.

The free propagation of the wave from the dipole is achieved by the permanent transformation from electrical into magnetic energy and vice versa. Thereby resulting electrical and magnetic fields are at right angles to the direction of propagation (Fig. 8.4).

8.1.2 Performance requirements

Mobile terminal antennas include those used in cellular phones, walkie-talkies for private and emergency service applications, and data terminals such as laptops and personal digital assistants.

Some of the important performance requirements of the mobile terminal antennas are mentioned below:

- These antennas are subjected to a wide range of variations in the environment, which they encounter.
- These antennas should work under various propagation conditions and they vary from very wide multipath arrivals to a strong line of sight component.

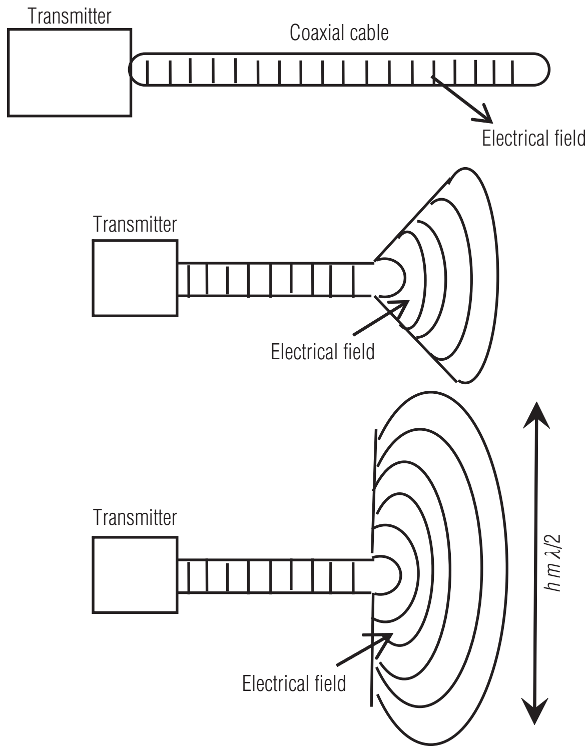


Figure 8.2 Evolving an antenna from a coaxial

- Particularly, the orientation of the terminal is often random, when a phone is in standby mode.
- They must be able to operate in close proximity of the user's head and hand.
- They must also be suitable for manufacturing in very large volumes at an acceptable cost.
- Increasingly, also the users prefer that the antenna be fully integrated with the casing of the terminal rather than being separately identifiable.
- *Radiation pattern*: Approximately, omnidirectional in azimuth and wide beamwidth in the vertical direction, although the precise pattern is usually uncritical and given the random orientation, the large degree of multipath and the pattern disturbance which is inevitable, given the close proximity of the user.
- *Input impedance*: Should be stable and well matched to the source impedance over the whole bandwidth of operation, even in the presence of detuning from the proximity of the user and other objects. Many user terminals now operate over a wide variety of standards. So multi-band, multi-mode operation via several resonances is increasingly a requirement.
- *Efficiency*: Given the low gain of the antenna, it is important to achieve a high translation of input RF power into radiation over the whole range of conditions of use.
- *Manufacturability*: It should be possible to manufacture the antenna in large volumes efficiently, without the need for tuning of individual elements, while being robust enough against mechanical and environmental hazards encountered while moving.
- *Size*: Generally as small as possible, consistent with meeting the performance requirements. Increasingly, the ability to adapt the shape to fit into casing acceptable to a consumer product is important.

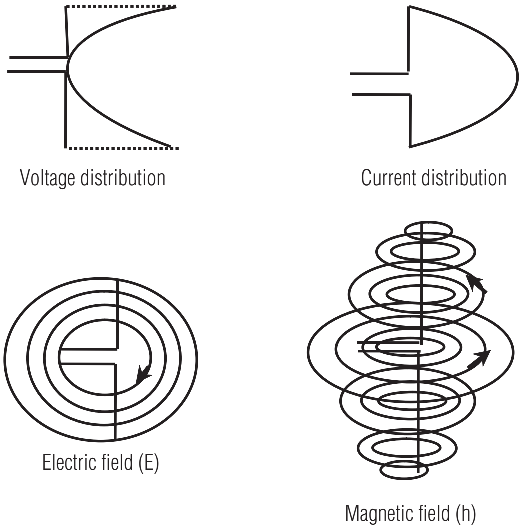


Figure 8.3 Field distribution on a λ/2 dipole

8.2 Antenna fundamentals

The mobile antennas are characterized by the following parameters:

8.2.1 Polarization

Polarization can be defined as the direction of oscillation of the electrical field vector. For the following communication systems respective polarization are mentioned:

Mobile communications: vertical polarization
Broadcast systems: horizontal polarization
See Section 7.2.3 for detailed description on polarization.

8.2.2 Propagation pattern

In most cases, the propagation characteristics of an antenna can be described via elevations through the horizontal and vertical radiation diagrams. In mobile communications, this is defined by the magnetic field components (H-plane) and the electrical field components (E-plane).

8.2.3 Half-power-beam-width

This term defines the aperture of the antenna. The half-power-beam-width (HPBW) is defined by the points in the horizontal and vertical diagram, which shows where the radiated power has reached half the amplitude of the main radiation direction. These points are also called 3 dB points.

8.2.4 Gain

In reality, one does not achieve an increment in energy via antenna gain. An antenna without gain radiates energy in every direction. An antenna with gain concentrates the energy in a defined angle segment of three-dimensional space. The λ/2-dipole is used as a reference for defining gain. At higher frequencies, the gain is often defined with reference to the isotropic radiator.

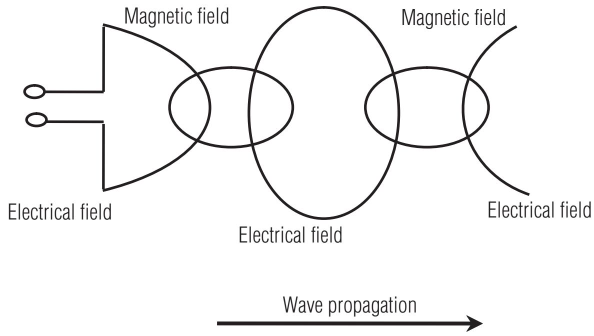


Figure 8.4 Wave propagation

The isotropic radiator is a non-existent ideal antenna, which has also an omni-directional radiation characteristic in the E-plane and H-plane.

Calculation:

Gain (with reference to the isotropic radiator dBi) = Gain (with reference to λ/2-dipole dBd) + 2.15 dB

The gain of an antenna is linked to the radiation characteristics of the antenna. The gain can be roughly calculated by checking the HPBWs in the horizontal and vertical planes.

8.2.5 Impedance

The frequency-dependent impedance of a dipole or an antenna is often adjusted via a symmetry or transformation circuit to meet the 50 Ω criterion. Adjustment across a wider frequency range is achieved using compensation circuits.

8.2.6 Voltage standing wave ratio/return loss

An impedance of exactly 50 Ω can only be practically achieved at one frequency. The voltage standing wave ratio (VSWR) defines how far the impedance differs from 50 Ω with a wide-band antenna. The power delivered from the transmitter can no longer be radiated without loss because of this incorrect compensation. Part of this power is reflected at the antenna and is returned to the transmitter (Fig. 8.5). The forward and return power form a standing wave with corresponding voltage maxima and minima (U_{min}/U_{max}). This VSWR defines the level of compensation of the antenna and was previously measured by interval sensor measurements.

A VSWR of 1.5 is standard in mobile communications. In this case, the real component of the complex impedance may vary between the following values:

Maximum value: 50 Ω × 1.5 = 75 Ω

Minimum value: 50 Ω /1.5 = 33 Ω

The term return loss attenuation is being used more often in recent times. The reason for this is that the voltage ratio of the return to the forward-wave UR/UV can be measured via a directional coupler. This factor is defined as the co-efficient of reflection.

8.2.7 Mechanical features

Antennas are always mounted at exposed sites. As a result, the antenna must be designed to withstand the required mechanical loading. For example, vehicle antennas must withstand a

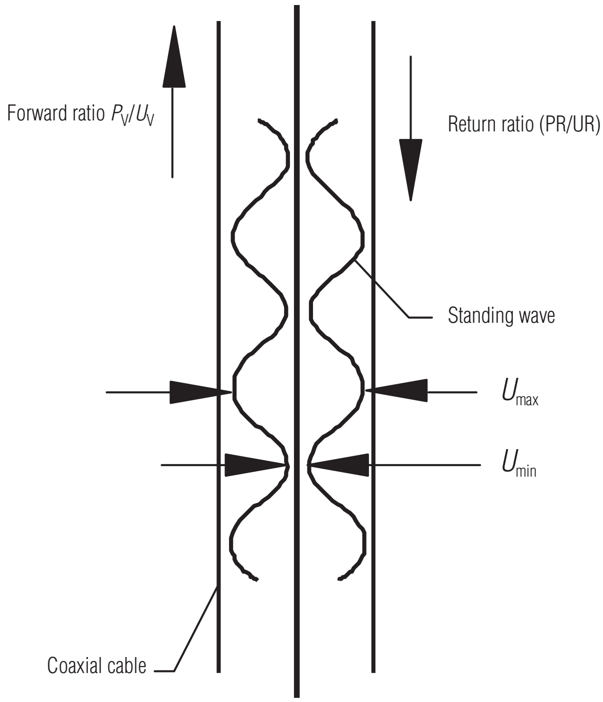


Figure 8.5 VSWR, return loss attenuation, and factor of reflection

high wind velocity, vibrations and still fulfil a limited wind noise requirement. Antennas for portable radio equipment are often exposed to ill-handling and sometimes even played with by the user. Base station antennas are exposed to high wind speed, vibrations, ice, snow, a corrosive environment, and also extreme electrostatic loading via lightening.

8.3 Types of antennas

Here we introduce several antennas that were recently developed and could be considered as relatively new. These antennas are conventional microstrip antennas as narrow band planar printed antennas, suspended planar antennas as wideband antennas, and planar monopole as an ultra-wideband antenna (UWB).

8.3.1 Monopole antenna

The monopole antenna shown in Figure 8.6 results from applying the image theory to the dipole. According to this theory, if a conducting plane is placed below a single element of length $L/2$ carrying a current, then the combination of the element and its image acts identically to a dipole of length L except that the radiation occurs only in the space above the plane.

For this type of antenna, the directivity is doubled and the radiation resistance is halved when compared to the dipole. Thus, a half-wavelength dipole can be approximated by a quarter-wave monopole ($L/2 = \lambda/4$). The monopole is very useful in mobile antennas where the conducting plane can be the car body or the handset case. The typical gain for the quarter-wavelength monopole is 2–6 dB and it has a bandwidth of about 10 per cent. Its radiation resistance is 36.5 Ω and its directivity is 3.28 (5.16 dB). The radiation pattern for the monopole is shown in Figure 8.7.

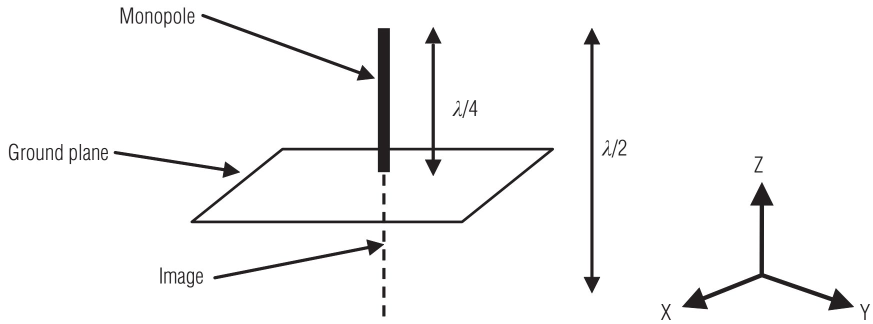


Figure 8.6 Monopole antenna

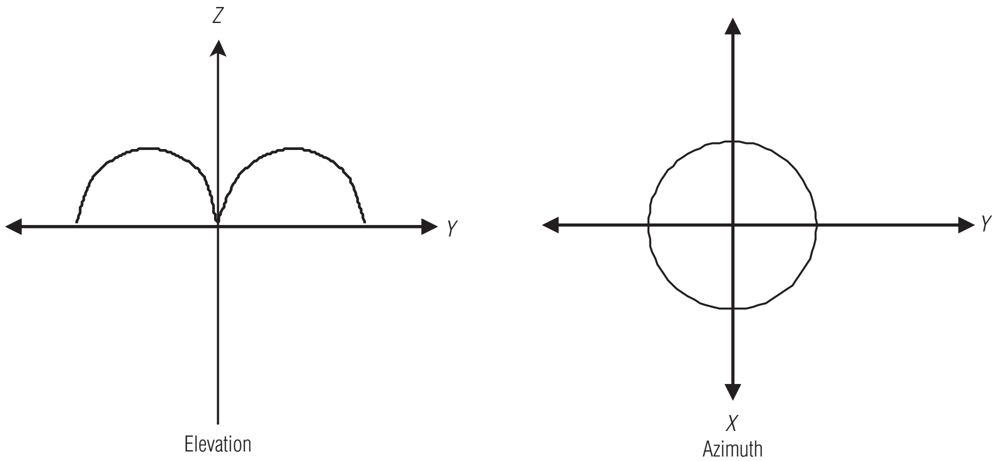


Figure 8.7 Radiation pattern for the monopole antenna

8.3.2 Dipole

A dipole is a conductive rod usually split in the centre and fed from a balanced transmission line that carries equal and oppositely flowing currents. Not all dipoles are split and fed in the centre because currents can be excited on it electromagnetically or it can be shunt fed. The dipole length determines possible current distributions in modes, and when we place a continuous rod near an antenna, radiating a linear polarization component directed along the rod, it excites a standing-wave current on the rod. The amount excited on the rod depends on how close its length is to resonance and the antenna spacing. The continuous rod loads the fed antenna through mutual coupling. The continuous rod can be fed from a coax line by attaching the outer conductor to the centre and then connecting the centre conductor away from the centre in a shunt feed. Feeding a dipole or loop requires a balun to prevent current flow either along the outside of a coaxial feeder or excitation of unbalanced currents along a two-wired line.

The current flowing along the outside of the coax or unbalanced currents on the two-wired line radiates in unwanted directions or radiates undesired polarization. When we design an antenna without considering or knowing its final mounting, we produce an uncontrolled situation without a balun. Our initial configuration may work without a balun, but the antenna may fail to produce the desired pattern in the final location. If you control the installation completely, you can reduce your design effort and may be able to eliminate the balun.

8.3.2.1 *Ideal half-wavelength dipole*

This type of antenna is a special case where each wire is exactly one-quarter of the wavelength, for a total of a half wavelength as shown in Figure 8.8. The radiation resistance is about 73 Ω if wire diameter is ignored, making it easily matched to a coaxial transmission line (Fig. 8.9). The directivity is a constant 1.64 or 2.15 dB. Actual gain will be a little less due to ohmic losses.

If the dipole is not driven at the centre then the feed point resistance will be higher. If the feed point is distance x from one end of a half-wave ($\lambda/2$) dipole, the resistance R_r will be described by the following equation.

$$R_r = \frac{75}{\sin^2\left(\frac{2\pi x}{\lambda}\right)}\Omega \tag{8.1}$$

If taken to the extreme then the feed point resistance of a $\lambda/2$ long rod is infinite, but it is possible to use a $\lambda/2$ pole as an aerial; the right way to drive it is to connect one terminal of a parallel LC resonant circuit. The other side of the circuit must be connected to the braid of a coaxial cable lead and the core of the coaxial cable can be connected part way up the coil from the RF ground side. An alternative means of feeding this system is to use a second coil, which is magnetically coupled to the coil attached to the aerial.

The most common form of dipole has an electrical length of half wavelength. As a result, this antenna is called a half-wavelength dipole. As before, the lengths of the wires are both the same. As the total length of the dipole is a half wavelength, this makes each section or leg of the dipole a quarter-wavelength long.

8.3.2.2 *Folded dipole*

A folded dipole is a half-wavelength dipole with an additional wire connecting its two ends as shown in Figure 8.10. If the additional wire has the same diameter and cross-section as the dipole,

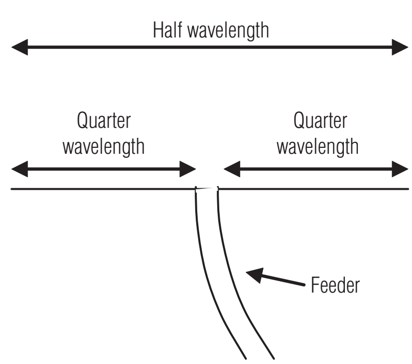


Figure 8.8 Half-wavelength dipole

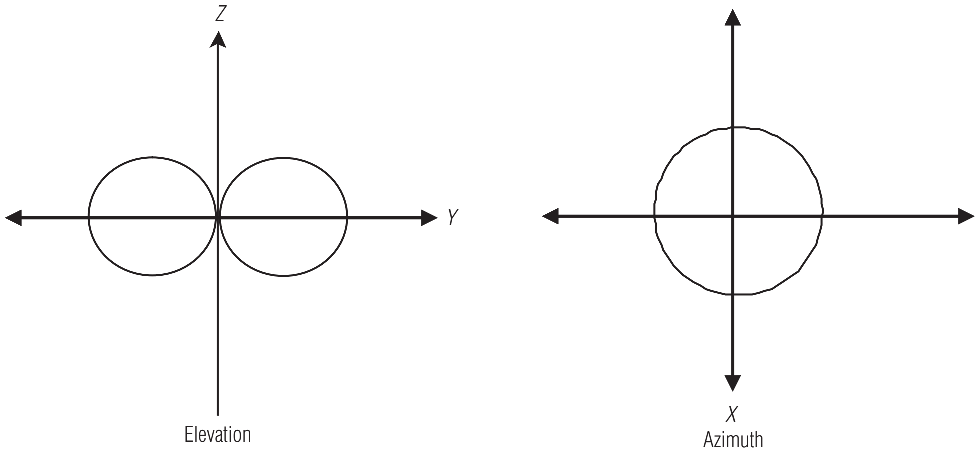


Figure 8.9 Radiation pattern for half-wave dipole

nearly two identical radiating currents are generated. The resulting far-field emission pattern is nearly identical to the one for the single-wire dipole described above; however, at resonance its input (feed-point) impedance R_{fd} is four times the radiation resistance of a single-wire dipole. This is because for a fixed amount of power, the total radiating current I_0 is equal to twice the current in each wire and thus equal to twice the current at the feed point. Equating the average radiated power to the average power delivered at the feed-point, we may write

$$\frac{1}{2} R_{\frac{\lambda}{2}} I_0^2 = \frac{1}{2} R_{fd} \left(I_0 / 2 \right)^2 \tag{8.2}$$

It follows that

$$R_{fd} = 4 R_{\frac{\lambda}{2}} = 292.52 \ \Omega \tag{8.3}$$

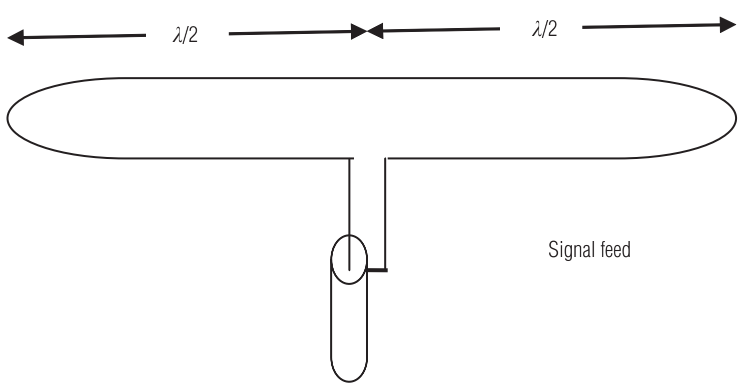


Figure 8.10 Folded dipole

Therefore, the folded dipole is well matched to 300 Ω balanced transmission lines.

If the radii of the two conductors are equal, then current flows in both the conductors in same direction, that is current is equal in magnitude and phase in the two dipoles.

The total power developed in folded dipole is equal to power developed in conventional dipole.

The following are the advantages of folded dipole over standard dipole:

- The basic form of a dipole consists of a single wire or conductor cut in the middle to accommodate the feeder and this feed impedance is altered by the other objects, especially other parasitic elements that may be used in other forms of antenna.
- Many antennas have to operate over large bandwidths and a standard dipole may be unable to fulfil the requirements adequately.
- This antenna offers a wider bandwidth and a considerable increase in feed impedance.
- There is four-fold increase in the feed impedance.
- In a standard dipole, the currents flowing along the conductors are in phase and as a result there is no cancellation of the fields and radiation occurs as a result.
- But in the case of folded dipole due to the addition of second conductor and this second conductor is a extension to the standard dipole with the ends folded back to meet each other. As a result, the currents in the new section flow on the same direction as those in the original dipole. The currents along both the half waves are therefore in phase and the antenna will radiate with the same radiation patterns as a simple half-wavelength dipole.
- The current consumption in the folded dipole is reduced to half of that in a standard dipole because power is shared evenly between the two sections.

8.3.2.3 Hertzian dipole (Current element)

The Hertzian dipole is a theoretical short dipole (significantly smaller than the wavelength) with a uniform current along its length. A true Hertzian dipole cannot physically exist, since the assumed current distribution implies an infinite accumulation of charge at its ends.

The radiation resistance R_r is given by:

$$R_r = \frac{2\pi}{3} Z_0 \left(\frac{l}{\lambda}\right)^2 \Omega \tag{8.4}$$

where Z_0 is the impedance of free space (120π). This is precisely four times the radiation resistance of the real short dipole with the linearly tapered current distribution.

The radiation resistance is typically a fraction of an ohm, making the infinitesimal dipole to an inefficient radiator. The directivity D is the theoretical gain of the antenna assuming no ohmic losses (not real-world), is a constant of 1.5, which corresponds to 1.76 dB. Actual gain will be much less due to the ohmic losses and the loss inherent in connecting a transmission line to the antenna, which is very hard to do efficiently considering the incredibly low radiation resistance. The maximum effective aperture A_e is:

$$A_e = \frac{3\lambda^2}{8\pi} (\text{metre})^2 \tag{8.5}$$

Even though the Hertzian dipole is minute, its effective aperture is comparable to antennas many times its size.

8.3.3 Horn antennas

The horn antenna is a natural evolution of the idea that any antenna represents a region of transition between guided and propagating waves. Horn antennas are highly suitable for frequencies (typically several gigahertz and above) where waveguides are the standard feed method, as they consist essentially of a waveguide whose end walls are flared outwards to form a megaphone-like structure (Fig. 8.11). The aperture is maintained as a rectangle, but circular and elliptical versions are also possible. The dimensions of the aperture are chosen to select an appropriate resonant mode, giving rise to a controlled field distribution over the aperture. The best patterns (narrow main lobe and low side lobes) are produced by making the length of the horn large compared to the aperture width, but this must be chosen as a compromise with the overall volume occupied. A common application of horn antennas is as the feed element for parabolic dish antennas in satellite systems.

The horn can have square, rectangular, or circular open mouths. For airborne applications, they are usually used as a feed for a reflector system or as "suppressed" antennas since they are not suitable aerodynamically unless they are enclosed in a random. Suppressed antennas are not like conformal antennas that are shaped to the airframe skin. These antennas do not protrude from the airframe because a hole is cut in the airframe and the antenna is installed inside the profile of the airframe, and a dielectric/fibre glass covering/ random is installed (over the antenna) flush with the aircraft's outer skin. Horns are usually fed by waveguides which are in turn fed by coaxial cables using a waveguide to co-axial adapter.

The feed horns used in weather radar reflectors are usually circular waveguides since circular polarization is used in heavy rain/sea clutter and horizontal polarization is used in clear weather. Search radar usually employs linear polarization, which may be vertical or horizontal. Older antenna systems have a pair of suppressed rectangular pyramidal horns.

Horns provide high gain, low VSWR, relatively wide bandwidth, low weight, and are easy to construct. The aperture of the horn can be rectangular, circular, or elliptical. However, rectangular horns are widely used. The three basic types of horn antennas that utilize a rectangular geometry are shown in Figure 8.12. These horns are fed by a rectangular waveguide, which have a broad horizontal wall as shown in the figure. For dominant waveguide mode excitation, the E-plane is vertical and H-plane horizontal. If the broad wall dimension of the horn is flared with the narrow wall of the waveguide being left as it is, then it is called an H-plane sectoral horn antenna as shown in the figure. If the flaring occurs only in the E-plane dimension, it is called an E-plane sectoral horn antenna. A pyramidal horn antenna is obtained when flaring occurs along both the

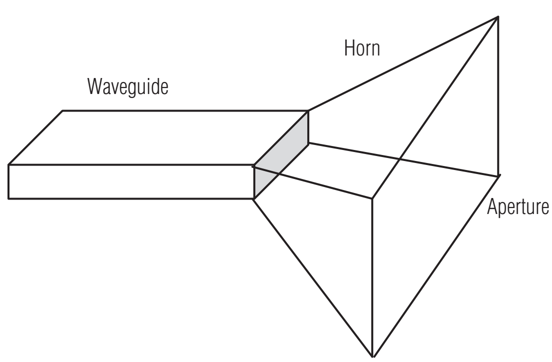


Figure 8.11 The rectangular horn antenna

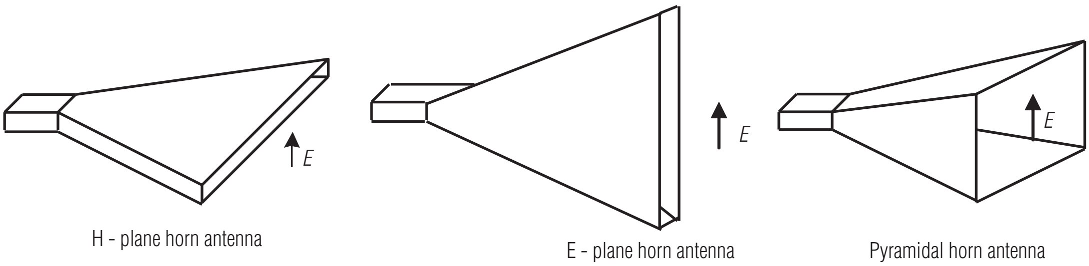


Figure 8.12 Types of horn antenna

dimensions. The horn acts as a transition from the waveguide mode to the free-space mode and this transition reduces the reflected waves and emphasizes the travelling waves which lead to low VSWR and wide bandwidth. The horn is widely used as a feed element for large radio astronomy, satellite tracking, and communication dishes.

8.3.4 Loop antennas

The loop antenna is a conductor bent into the shape of a closed curve such as a circle or a square with a gap in the conductor to form the terminals as shown in Figure 8.13. There are two types of loop antennas – electrically small loop antennas and electrically large loop antennas. If the total loop circumference is very small as compared to the wavelength ($L <<< \lambda$), then the loop antenna is said to be electrically small. An electrically large loop antenna typically has its circumference close to a wavelength. The far-field radiation patterns of the small loop antenna are insensitive to shape.

As shown in Figure 8.14, the radiation patterns are identical to that of a dipole despite the fact that the dipole is vertically polarized whereas the small circular loop is horizontally polarized.

Small loop antenna's radiation pattern is different from that of large loop antenna. A one-wavelength square loop antenna is a large loop antenna and its maximum radiation is perpendicular to the plane of loop. There is a lobe in the direction normal to side containing feed and in the direction parallel to the feed there is a null. Small loop antennas are generally used for receiving purpose. In pagers single-turn loop antennas are used and in AM broadcast receivers multi-turn loop antennas are used.

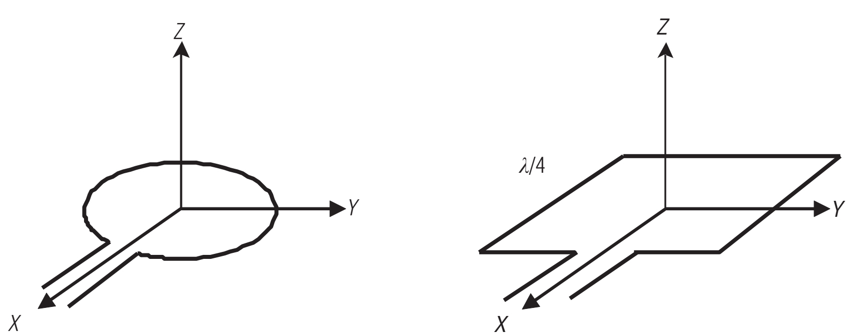


Figure 8.13 Loop antenna

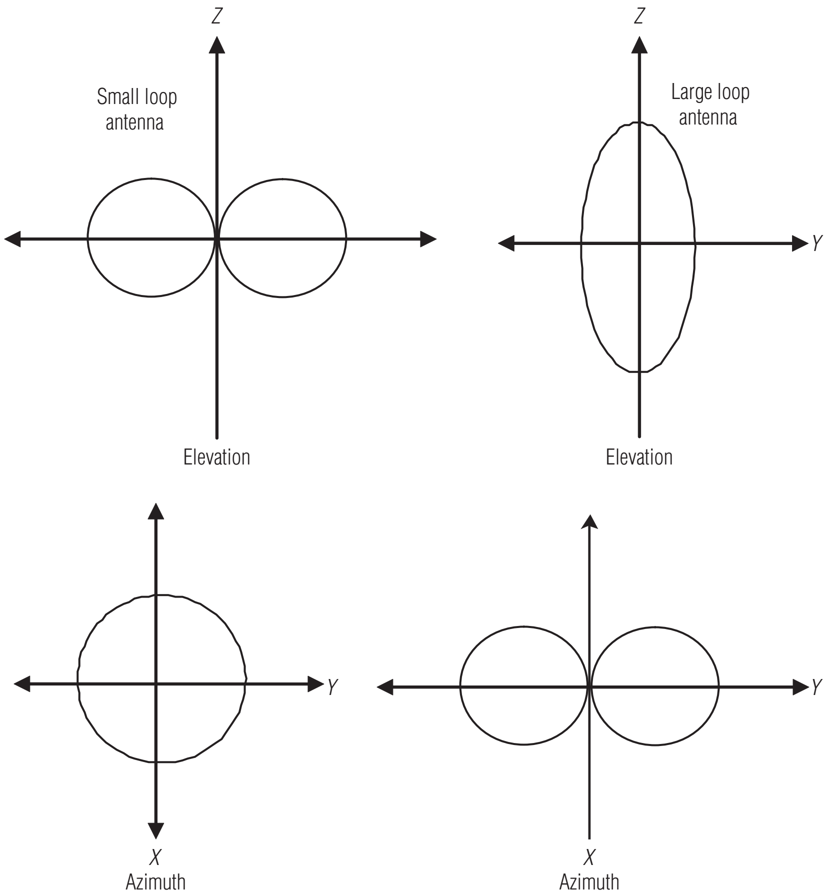


Figure 8.14 Radiation pattern of small and large loop antenna

8.3.5 Helical antennas

The helical antenna (Fig. 8.15) can be considered as a vertical array of loops, at least for the case when the diameter of the helix is small compared to a wavelength. The result is normal mode radiation with higher gain than a single loop, providing an omnidirectional antenna with compact size and reasonable efficiency, but rather narrow bandwidth. It is commonly used for hand-portable mobile applications where it is more desirable to reduce the length of the antenna below that of a quarter-wave monopole.

In case the diameter is around one wavelength or greater, the mode of radiation changes completely to the axial mode, where the operation of the antenna is similar to that of a Yagi, but with circular polarization. This mode is commonly used for satellite communications, particularly at lower frequencies where a dish would be impractically large.

Figure 8.16 shows the radiation patterns for the normal mode as well as the axial mode of operations. In the normal mode of operation, the antenna field is maximum in a plane normal to the helix axis and minimum along its axis. This mode provides low bandwidth and is generally used for hand-portable mobile applications.

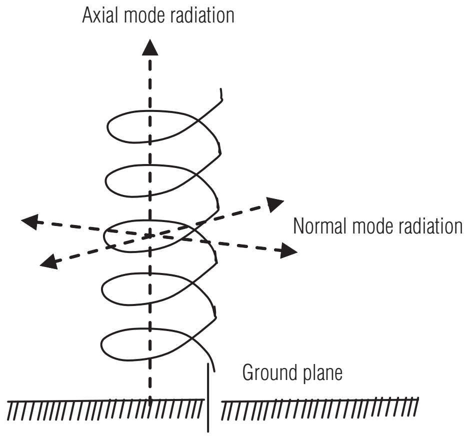


Figure 8.15 The helical antenna

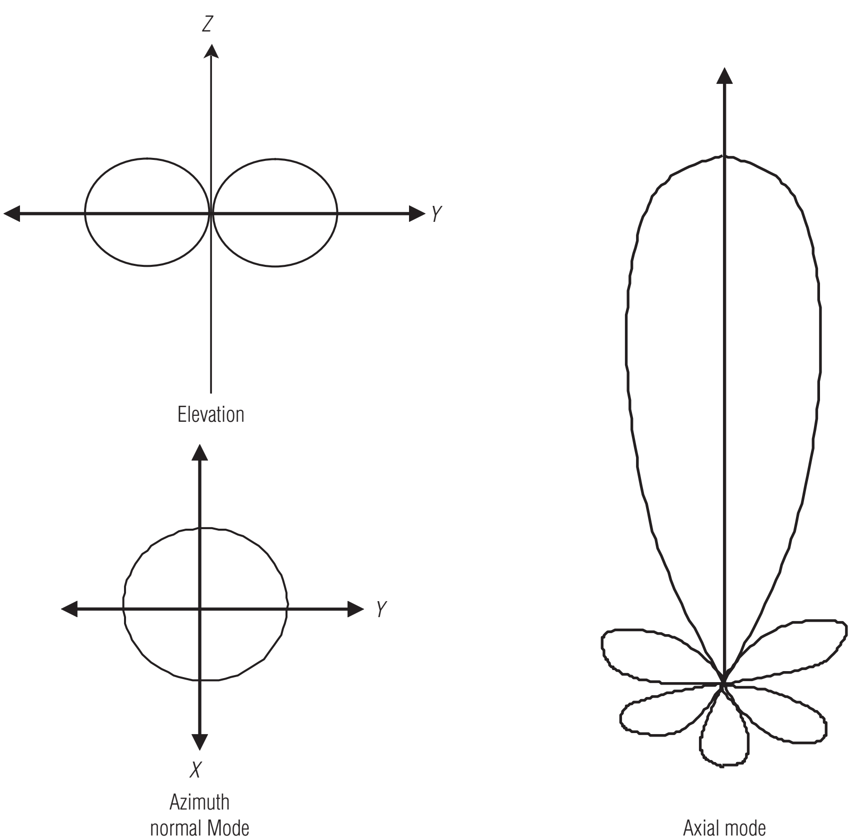


Figure 8.16 Radiation pattern of helix antenna

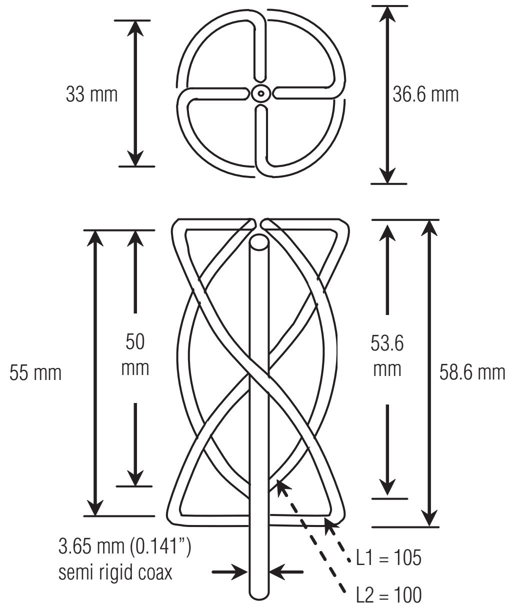


Figure 8.17 Top and side view of GPS helical antenna

In the axial mode of operation, the antenna radiates as an end-fire radiator with a single beam along the helix axis. This mode provides better gain (up to 15 dB) and high bandwidth ratio (1.78:1) as compared to the normal mode of operation. For this mode of operation, the beam becomes narrower as the number of turns on the helix is increased. Due to its broadband nature of operation, the antenna in the axial mode is used mainly for satellite communications.

8.3.5.1 Helical antenna for global positioning system

Global positioning system's (GPS) Helical antenna is constructed by winding a stiff copper wire around a wooden dowel. After drilling the centre hole, the drill is left in the hole. Four copper wires of 1.8 mm thickness are taken. Two wires are of length L1 and remaining two are of length L2. Each wire is stuck into one of the side holes such that the wire touches the centre drill. Wire is put into corresponding hole after it is bent half of a turn around the dowel. Ones required shape is attained, they are soldered to a semi rigid coax as shown in Figure 8.17.

The measurements shown in the figure are inside measurements to the left and outside measurements to the right.

The wires at "9" and "12" hours are connected to the outside (the braid), the wires at "3" and "6" connected to the inner conductor. Most Garmin receivers can supply a voltage of 5 V for an active antenna. This voltage must be blocked in this type of helical antenna with a small condenser.

8.3.6 Patch antennas

Patch antenna, as seen in Figure 8.18, are based upon printed circuit technology to create flat radiating structures on top of dielectric, ground-plane-backed substrates. The appeal of such structures is in allowing compact antennas with low manufacturing cost and high

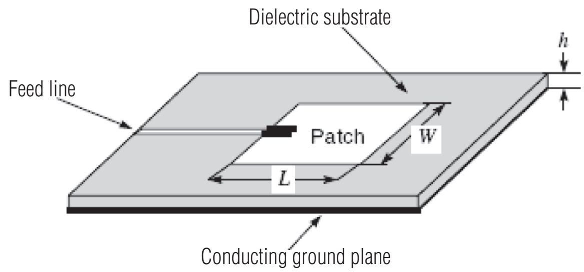


Figure 8.18 Microstrip patch antenna

reliability. It is in practice difficult to achieve this at the same time as acceptably high bandwidth and efficiency. The improvements in the properties of the dielectric materials and in design techniques have led to enormous growth in their popularity and there are now a large number of commercial applications. Many shapes of patch antennas are possible, with varying applications, but the most popular are rectangular (pictured), circular, and thin strips (i.e. printed dipoles).

In the rectangular patch, the length L is typically up to half of the free space wavelength. The incident wave fed into the feed line sets up a strong resonance within the patch, leading to a specific distribution of fields in the region of the dielectric immediately beneath the patch, in which the electric fields are approximately perpendicular to the patch surface and the magnetic fields are parallel to it. The fields around the edges of the patch create the radiation, with contributions from the edges adding as if they constituted a four-element array. Thus, the resultant radiation pattern can be varied over a wide range by altering the length L and width W. In this case, the polarization is approximately linear, but patches can be created with circular polarization by altering the patch shape and/or the feed arrangements. A major application of patch antennas is in arrays, where all of the elements, plus the feed and matching networks, can be created in a single printed structure. The necessary dimensions can be calculated approximately by assuming that the fields encounter a relative dielectric constant of 2 due to the combination of fields in the air and in the dielectric substrate.
where

L is the length of the patch
W is the width of the patch
h is the thickness or height of dielectric slab

Advantages:

- Planer (and can be made conformal to shaped surface)
- Low profile
- Ease of integration with microstrip technology
- Can be integrated with circuit elements
- Ability to have polarization diversity (vertical, horizontal, right-hand circular polarization (RHCP), or left-hand circular polarization (LHCP))
- Lightweight and inexpensive

Disadvantages:

- Narrow bandwidth (typically less than 5 per cent), requiring bandwidth widening techniques
- Can handle low RF power
- Large ohmic loss

8.3.7 Aperture antennas

The most common aperture antennas used on an aircraft are reflectors, phased arrays, and horns. These antennas are not affected to any significant extent by the ground plane, as long as the direct radiation from the antenna does not strike the ground plane. These antennas do not require a ground plane, as is the case for monopole antennas. In the case of aperture antennas, the beamwidth is inversely proportional to the dimensions so that the larger the antenna the smaller is its beamwidth, assuming that the frequency is unchanged. The beamwidth depends on the electrical dimensions of the antenna, that is its dimensions in terms of wavelength. Thus, if the frequency of operation is doubled the wavelength is halved, and so an antenna of only half the physical dimensions would be required to obtain the same beamwidth. On the other hand, the bore-sight gain is proportional to the dimensions of the antenna, thus the larger the antenna the higher is its gain, assuming that its frequency is not changed. The gain depends on the electrical dimensions of the antenna, that is its dimensions in terms of wavelength. Thus, if the frequency of operation is doubled the wavelength is halved, and so an antenna of only half the physical dimensions would be required to obtain the same gain.

8.3.8 Planar inverted-L/F antennas

A planar inverted-L/F antenna (PIFA) is an improved version of the monopole antenna. The straight wire monopole is the antenna with the most basic form. Its dominant resonance appears at around one-quarter of the operating wavelength. The height of quarter-wavelength has restricted their application to instances where a low-profile design is necessary.

Figures 8.19 and 8.20 show the geometry of a narrow-strip monopole with a horizontal bent portion, and the PIFA is shown in Figure 8.21.

A PIFA can be considered as a kind of linear inverted-F antenna (LIFA) with the wire radiator element replaced by a plate to expand bandwidth.

Advantages:

- Reduced height
- Reduced backward radiation
- Moderate to high gain in both vertical and horizontal polarizations

Disadvantages:

- Narrow bandwidth

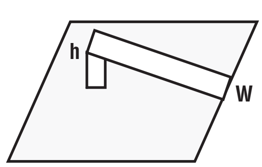


Figure 8.19 Geometry of a narrow strip monopole with a horizontal bent portion

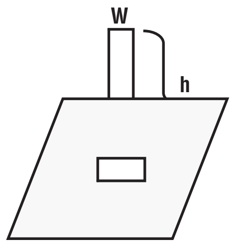


Figure 8.20 Geometry of a monopole

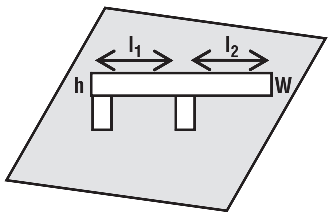


Figure 8.21 Geometry of the PIFA

8.3.9 Reflectors

Reflector antennas are based on optical systems, which usually have one or two reflectors, fed by a horn, planar spiral, conical log spiral, or a dipole. For broadband operation, the feed could be a log-periodic array.

A log-periodic array is an array of dipoles or monopoles that are used instead of single elements, to give broadband frequency operation. In the case of the single reflector, the feed is placed at the focus of the reflector, which is usually parabolic as shown in Figure 8.22(a). However, the feed may be offset, so that it is not on the principal axis of the reflector as shown in Figure 8.22(b). In the Cassegrain system, two reflectors are used. The feed is placed at one focus of the hyperbolic sub-reflector, and the second focus of the sub-reflector is coincident with the focus of the main parabolic reflector, as shown in Figure 8.22(c).

8.4 Mean effective gain

The performance of a practical mobile antenna in its realistic operating environment may be very different than would be expected from measurements of the gain of the antenna in isolation. This arises because the mobile is usually operated surrounded by scattering objects which spread the signal over a wide range of angles around the mobile. The concept of a mean effective gain (MEG), which combines the radiation performance of the antenna itself with the propagation characteristics of the surrounding environment, was explained with an example.

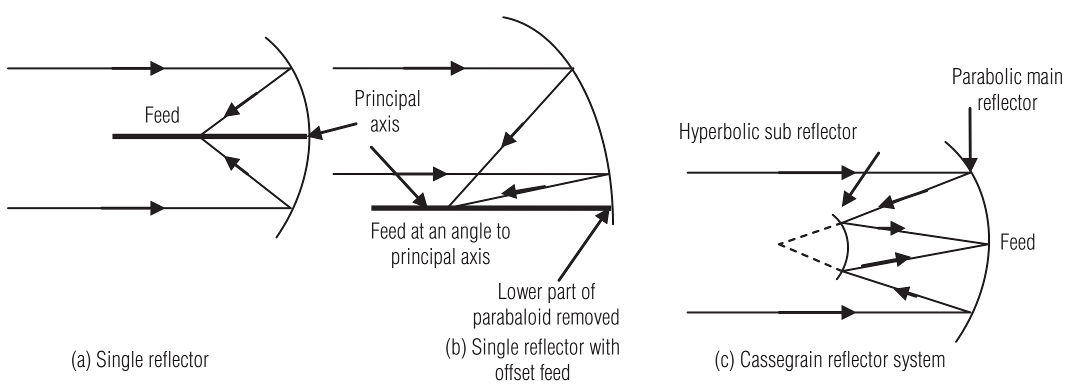


Figure 8.22 Reflector antenna systems

Consider a mobile antenna which receives power from a base station after scattering has occurred through a combination of buildings, trees, and other clutter in the environment. The total average power incident on the mobile is composed of both horizontally and vertically polarized component. All powers are considered as averages, taken after the mobile has moved along a route of several wavelengths. The MEG of the antenna, G_e, is then defined as the ratio between the power which the mobile actually receives and the total power which is available.

8.5 Human body interactions and specific absorption rate

In the frequency range of interest for practical radio communications, EM radiation is referred to as "non-ionizing radiation" as distinct from the ionizing radiation produced by radioactive sources. The energy associated with the quantum packets or photons at these frequencies is insufficient to dissociate electrons from atoms, whatever is the power density, so the main source of interactions between non-ionizing radiation and surrounding human tissue is simple heating. We are all continually exposed to EM radiation from a variety of sources, including mobile phone systems, common sources of radio waves include radio, and television broadcasts; there is understandable concern to ensure that human health is not adversely affected.

The potential health impact of EM fields has been studied for many years by both civil and military organizations, as well as the effects and interactions of handheld antennas with the human head. A number of bodies have commissioned research into such effects and the World Health Organization has produced guidelines to ensure that this research is conducted according to appropriate standards. The conclusions from these investigations have been used to set regulatory limits on exposure which reflects a precautionary principle based on the current state of knowledge. Many administrations require equipment manufacturers to ensure that the fields absorbed are below given limits and to quote the values produced by individual equipment under suitable reference conditions. Therefore, it is essential at this stage to establish procedures and metrics to assess the impact of antennas on absorption within the body.

8.6 Mobile satellite antennas

The key requirements for the antenna on a mobile handset for non-geostationary satellite systems are summarized in the following.

8.6.1 Omni-directional, near-hemispherical radiation pattern

This allows the handset to communicate with satellites received from any elevation and azimuth angle, without any special cooperation by the user. The elevation pattern should extend down to at least the minimum elevation angle of the satellite, but should not provide too much illumination of angles below the horizon, since this would lead to pick-up of radiated noise from the ground and degradation of the receiver noise figure. The pattern need not necessarily be uniform within the beamwidth; indeed, it may be an advantage in some systems to emphasize lower elevation angles at the expense of higher ones in order to overcome the extra free space and shadowing losses associated with lower satellites.

8.6.2 Circular polarization with axial ratio close to unity

This limits the polarization mismatch. A typical specification is that the axial ratio should not be more than 5 dB elevation angles down to the minimum elevation angle of the satellite constellation.

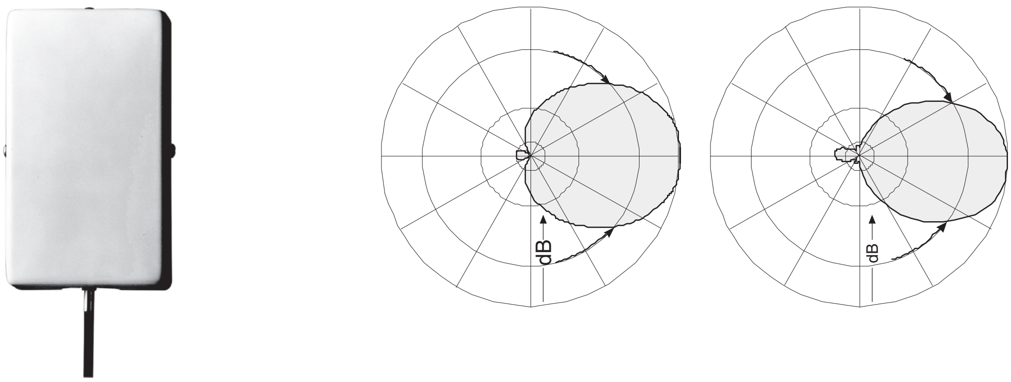


Figure 8.23 Indoor patch antenna and radiation pattern with relative field strength

8.6.3 Patch antennas

Patch antennas are also used for satellite mobile terminals, and became very attractive due to their low cost and easy manufacturing, as well as the reduction in size because of the technology used in their construction. Circular polarization may be achieved by dual feeding of a square or a circular patch at right angles with quadrature phasing or by perturbing the patch shape (e.g. cutting off one corner) so as to create anti-phase currents which produce the same result. A good example of the use of such antennas is in the GPS, where patch antennas are the most common configuration, although the quadrifilar helix antenna is still often applied for high-performance requirements (Fig. 8.23).

8.7 Summary

- A mobile antenna is a device that is used to transfer guided EM waves (signals) to radiating waves in an unbounded medium.
- A dipole is a conductive rod usually split in the centre and fed from a balanced transmission line that carries equal and oppositely flowing currents.
- Horn antennas are highly suitable for frequencies (typically several gigahertz and above) where waveguides are the standard feed method, as they consist essentially of a waveguide whose end walls are flared outwards to form a megaphone-like structure.
- Loop antenna pattern has exactly the shape of the Hertzian dipole pattern, except that the electric and magnetic fields are reversed in their roles.
- The most common aperture antennas used on an aircraft are reflectors, phased arrays, and horns. These antennas are not affected to any significant extent by the ground plane, as long as the direct radiation from the antenna does not strike the ground plane.
- A planar inverted-L/F antenna is an improved version of the monopole antenna. The straight wire monopole is the antenna with the most basic form.
- Reflector antennas are based on optical systems which usually have one or two reflectors, fed by a horn, planar spiral, conical log spiral, or a dipole.
- Patch antennas are also used for satellite mobile terminals, and become very attractive due to their low cost and easy manufacturing, as well as the reduction in size as a result of the technology used in their construction.

Review questions

1. Write short notes on high gain antenna.
2. Draw the structure of horn antennas.
3. Draw the structure of loop antennas.
4. Describe the application of helical antennas.
5. Draw the structure of patch antennas and explain its operation.
6. Write short notes on reflector antenna system.
7. Write short notes on aperture and PIF antennas.

Objective type questions and answers

1. Mobile antennas are ____ dependent.
   (a) frequency (b) power (c) height (d) all of the above
2. Polarization can be defined as
   (a) direction of oscillation of the magnetic field vector
   (b) direction of oscillation of the electrical field vector
   (c) direction of oscillation of the both EM
   (d) none of the above
3. Standard value of VSWR in mobile communication is
   (a) 1 (b) 0 (c) 1.5 (d) 2
4. If a mobile antenna has again more than 3 dB what will be its receiving capacity?
   (a) receives only limited portion of total multipath signals
   (b) receives only portion of total multipath signals
   (c) cannot receive at all
   (d) receive large signal with noise
5. Compare the position of glass mounted antenna with roof mounted antenna.
   (a) it should be lower than roof mounted antenna
   (b) it should be larger than roof mounted antenna
   (c) they are equal positions
   (d) cannot be compared
6. Comment an antenna beam pattern of directional antenna.
   (a) suppressed horizontally (b) suppressed vertically
   (c) elevated vertically (d) elevated horizontally
7. Comment an antenna beam pattern of high gain antenna.
   (a) suppressed horizontally (b) suppressed vertically
   (c) elevated vertically (d) elevated horizontally
8. The motor vehicle mounted antenna is the following antenna
   (a) directional antenna (b) omni-directional antenna
   (c) smart antenna (d) microstrip antenna
9. What will be the radiation pattern of roof mounted antenna?
   (a) uniformly distributed (b) not uniformly distributed
   (c) (a) or (b) (d) cannot be predicted
10. Several vertically stacked umbrella pattern can form a
    (a) abnormal antenna (b) high gain antenna
    (c) interference reduction antenna (d) mobile antenna

Answers: 1. (a), 2. (b), 3. (c), 4. (a), 5. (a), 6. (a), 7. (b), 8. (b), 9. (a), 10. (b).

Open book questions

1. What is HPBW and when do we use the vertical and horizontal polarization?
2. Write short notes on folded dipole antenna.
3. Define the terms: antenna radiation pattern, antenna gain, antenna polarization, and the front-to-back ratio.
4. Why are high-gain omni-directional or directional antennas used at the cell site?

Key equations

1. The resistance R_r of a half-wave dipole is given by

$$R_r = \frac{75}{\sin^2\left(\frac{2\pi x}{\lambda}\right)}\Omega$$

2. The radiation resistance R_r of a Hertzian dipole is given by

$$R_r = \frac{2\pi}{3} Z_0 \left(\frac{l}{\lambda}\right)^2 \Omega$$

3. The maximum effective aperture A_e of a Hertzian dipole is given by

$$A_e = \frac{3\lambda^2}{8\pi}(\text{metre})^2$$

Further reading

"Bandwidth enhancement by flared microstrip dipole antenna," IEEE Antennas and Propagation Society International Symposium, 1991, vol. 1, AP-S. Digest, 24–28 (June 1991): 342–345.

Dey, S., Venugopalan, P., Jose, K. A., Aanandan, C. K., Mohanan, P., and Nair, K. G., "Bandwidth enhancement by flared microstrip dipole antenna", IEEE Antennas and Propagation Society International Symposium, Vol. 1, AP-S Digest, 24–28 (June 1991): 342–345.

Evtioushkine, G. A., Kim, J. W., and Han, K. S. "Very wideband printed dipole antenna array," Electronics Letters. vol. 34 (1998): 2292–2293.

Kim, J. I., Lee, B. M., and Yoon, Y. J., "Wideband printed dipole antenna for multiple wireless services," IEEE Radio and Wireless Conference (August 19–22, 2001): 153–156.

Kiminami, K., Hirata, A., and Shiozawa, T., "Double-sided printed bow-tie antenna for UWB communications," IEEE Antenna and Wireless Propagation Letters. vol. 3 (2004): 152–153.

Levine, E., Shtrikman, S., and Treves, D. "Double-sided printed arrays with large bandwidth," IEE Proceedings – Part H, vol. 135, no. 1 (1988): 54–59.

Lin, Y. D., and Tsai, S. N. "Coplanar waveguide-fed uniplanar bow-tie antenna," Electronics Letters. vol. 45 (1997): 305–306.

Cell-Site Antennas for Mobile Communication

9

9.1 Introduction

A base station (BS) is the interface between wireless phones and traditional wired phones. It allows you to use your cell phone to call your home phone. The base station, a wireless system, uses microwave radio communication. It is composed of several antennas mounted on a tower and a building housed with electronics at the base. When you make a call with your cell phone, the cell phone and base station communicate back and forth by radio, and the radio waves they use are in the microwave region of the electromagnetic spectrum.

The base station antenna is mounted on tall towers because from high point it is easier to stay in communication with cell phone users, who are often near the ground. The actual antenna elements of a base station are usually less than 10 cm (about 4 in), but may be grouped into clusters or "arrays" with heights of about a metre (about 3 ft). They need to be mounted on a tower to overcome obstacles, such as trees, hills, or tall buildings, which stand between the base station and the cell phone. This chapter deals with the base station antenna and their types, specific adoption, and antennas at cell sites and their operations.

9.1.1 Base station antennas

Most cells are divided into three sectors, and hence each antenna rack will have a triangular shape. Each face of the rack will usually have three antennas installed, of which two are for receiving and one is for transmitting. Two antennas are used on the receive side so that the base station can compare signals and select the best antenna for each user within the cell. This is known as "diversity" reception which manages the power differences between the cellular base station transmitter and the small battery powered cell phone transmitter. The cellular tower transmitting antenna is usually placed between the two receiving antennas. Even though the received frequency is different from the transmitted frequency, the antennas are separated by several metres because of the large difference in power levels. The physical size of the antennas is generally related to the frequency of their operation.

To transmit calls, the base station requires a powerful transmitting amplifier to generate strong signals. This "power amplifier" is linked to the transmitting antenna by a length of coaxial cable. Connected by cable to the receiving antennas, low-noise amplifiers can detect the weak incoming signals and separate them from any background noise, if present.

A bank of electronic circuits, called the transceiver rack, connects to the low-noise and power amplifiers and converts their radio signals into digital signals and vice versa. The transceiver is connected to the electronic switching device that routes calls between the base station and the main telephone system.

Table 9.1 Frequency bands and services

Frequency band (MHz)	Short reference (MHz)	Service
450–470	450	Phone+data
824–890	850	Phone+data
880–960	900	Phone+data
824–960	Low Bands	Phone+data
1710–1880	1800	Phone+data
1850–1990	1900	Phone+data
1900–2170	2100	Phone+data
1710–2170	High Bands	Phone+data

The electronic switching function is known as the "Mobile Switching Centre" (MSC) for GSM, AMPTS, NMT, and CDMA systems, and for 3G/4G mobile systems it is known as the "Media Gateway" (MGW)/"Gateway MSC Server" (GMSC).

In addition to providing a link between a base station and a mobile station, base station antennas provide an essential and increasingly important tool for the control of frequency re-use and the optimization of channel capacity in a mobile radio network.

The descriptions are of general applications to many frequency bands; hence to avoid tedious repetition, the following bands will be defined and are usually referred by their abbreviations. In the context of multiband arrays, bands of 850–900 MHz will be referred to as the low bands and those of 1,710–2,170 MHz as the high bands. The nomenclature for base stations and mobile stations differs between the air interface specifications, but for convenience the terms base station (BS) and mobile station (MS) will be used here without implying reference to any particular air interface.

9.1.1.1 Reciprocity

The principle of reciprocity applies to most of the performance parameters of an antenna. Gain, radiation pattern, efficiency, and many other characteristics have the same value irrespective of whether an antenna is transmitting or receiving a signal; hence in the following discussion the direction of transmission is chosen to suit the simplest understanding of the phenomena described.

9.1.1.2 Frequency bands

Table 9.1 includes most major worldwide assignments, but other frequency bands are allocated to mobile radio services in some countries. Future bands for UMTS or additional 3G services are not included. Following the transfer of broadcast TV services to digital format, a significant amount of the present analog TV spectrum will be reassigned to mobile services.

9.2 Cell-site antennas

Cellular, PCS, Data, ISM, and WLL are technologies migrating to wireless systems completely. Base station antennas provide the critical link between the user and the system provider. They also provide connectivity within the system without being directly accessed by the user. Base station antennas, as their name implies, are usually fixed in a specific location in the network and provide connectivity over a geographic area or from point to point. There are two types of base station antennas used, omnidirectional and directional antennas.

9.2.1 Omnidirectional antennas

Omnidirectional antennas have radiation patterns which cover the horizon uniformly. Gain greater than unity is achieved by forming a collinear vertical array, which reduces the elevation beamwidth but leaves the azimuth (horizon) pattern unaffected. Omnidirectional antennas are generally used in low traffic volume areas or for very small or indoor "cells."

9.2.1.1 Ground plane and λ/4 skirt antennas in comparison

The classical omni directional λ/2 antennas are of a ground plane or λ/4-skirt nature (Figure 9.1). The names indicate how the antenna is decoupled from the mast. In the first case, a conductive plane is achieved via three counterweighted poles, and in the other case the decoupling is achieved by using a λ/4-skirt. The second type, however, only works across a very limited bandwidth, so that, for example, three versions are needed to cover the 2-m band. On the other hand, being a wideband antenna, the ground plane antenna, can cover the complete frequency range.

9.2.1.2 Design of omnidirectional antenna cellular system

The carrier-to-interference (C/I) ratio requirement is determined by the transmission and modulation scheme adopted in a system and is specified for an acceptable voice quality. A voice quality is termed as acceptable if 10–15 per cent of the mobile subscribers rate the received signal quality as good or excellent. The parameter C/I is generally determined on the basis of subjective voice quality tests.

Let N_I be the number of co-channel interfering cells and I_i be the interference power caused by transmissions from the i^{th} interfering co-channel cell. The signal-to-co-channel interference ratio (C/I) at the desired mobile receiver is given by

$$\frac{C}{I} \approx \frac{C}{\sum_{i=1}^{N_I} I_i} \tag{9.1}$$

(a) (b)

Figure 9.1 (a) Ground plane antenna and (b) λ/4-skirt antenna

In addition to the co-channel interference, there is always an inherent background noise. However, in an interference-dominated mobile environment, we may neglect the background noise. It may be noted that the desired received signal power C is proportional to $r^{-\gamma}$, where r is the distance between the mobile and the serving base stations and γ is a propagation path-loss exponent determined by the actual terrain environment and varies in the range $2 \le \gamma \le 5$. When the transmit powers from all base stations are equal and the path-loss exponent is the same throughout the geographical coverage area, the co-channel interference from the i^{th} co-channel cell, I_i, for all i, depends on D_i and γ only, D_i being the distance between the i^{th} interfering co-channel cell and the mobile unit.

When the mobile unit is located at the cell boundary having radius R (i.e., $r = R$), the worst-case co-channel interference occurs as the power of the desired signal is minimum. Since the average received signal strength at any point decays as a power law of the distance between the transmitter and the receiver (C varies as $R^{-\gamma}$ and I varies as $D^{-\gamma}$), the C/I at a typical mobile receiver can be approximated by the following:

Assuming as a simple case,

$D_1 \approx D$ for $I = 1, 2, \ldots, N_I$, in Equation (9.1), we get

$$\frac{C}{I} \approx \frac{R^{-\gamma}}{N_1 D^{-\gamma}} \tag{9.2}$$

$$\frac{C}{I} \approx \frac{D^{\gamma}}{N_1 R^{\gamma}} \tag{9.3}$$

Thus, for a minimum reuse distance D and the cell radius R, the carrier-to-interference ratio due to co-channel interference is given as

$$\frac{C}{I} \approx \frac{1}{N_I}\left(\frac{D}{R}\right)^{\gamma} \tag{9.4}$$

The separation between co-channel cells, D, is mainly determined by C/I that is required to achieve the desired transmission quality and the fade margin that is necessary to take care of statistical fluctuations in the desired received signal level induced by the mobile environment. Since by definition, frequency reuse ratio, $q = D/R$, therefore,

$$\frac{C}{I} \approx \frac{1}{N_I}(q)^{\gamma} \tag{9.5}$$

In a mobile radio environment, γ is usually taken as 4. In a fully developed hexagonal cellular system, the number of co-channel interfering cells is six in the first tier, ignoring co-channel interfering cells in the second tier, which cause weaker interference than those in the first tier. Therefore, putting $N_I = 6$ and $\gamma = 4$ in Equation (9.5), we get

$$\frac{C}{I} \approx \frac{1}{6}(q)^{4} \tag{9.6}$$

$$q = \left(6 \times \frac{C}{I}\right)^{1/4} \tag{9.7}$$

9.2.2 Directional antennas

Directional antennas are used in communication systems where higher gains are required. Directional antennas are useful in remote locations where high gain is required and the direction to a desired transmitter/receiver is known. They are used in cell and microcell applications to divide a geographical region into sectors. This reduces interference in the network, allowing a greater number of users to be served. Antenna gain and directivity are increased by increasing the effective aperture of the antenna. In a Yagi, this implies lengthening the boom and adding more elements (directors) to the antenna.

9.2.3 Base station antenna series omnis

Base station antenna (BSA) series omniantennas are engineered to provide lasting performance in the most demanding field conditions. A practical BSA series omins and a field tunable low band omni base station kit of 45–50 MHz are shown in Figure 9.2(a) and (b).

The design features of BSA series omnis include the following:

- Heavy, nickel-plated brass square nut radial collar (square nut allows easy removal for extra portability and convenience)
- 150-MHz and 220-MHz models are DC grounded
- Wind load rating of 100 mph

(a) (b)

Figure 9.2 (a) BSA series omnis and (b) Field-tunable low-band omni base station kit, 45–50 MHz

9.2.3.1 Yagis

A Yagi is a parasitic linear array of parallel dipoles. The typical construction of Yagis is shown in Figure 9.3. It uses a single-driven dipole with a reflector and director elements excited by near field compiling to the driven element. Owing to their rugged construction and relatively high gain, Yagis are popular. YA series Yagis are built to precise specifications to perform in extreme weather conditions. Yagis are welded for high strength and low noise performance.

The features of Yagis include the following:

- Fully welded design
- Solid aluminium elements
- Aluminium tube boom
- Wind load rating of 100 mph
- Optional heavy-duty bracket for up to 2.5 pipe

9.2.3.2 FB series omnis

The Front-to-Back ratio (FB) series omni antenna shown in Figure 9.4 is designed for optimum performance in extreme weather conditions.

The design features include the following:

- Compact, easily transportable design
- All-weather construction
- Wind load rating of 100 mph

9.2.4 Antennas for wireless LAN

The vast majority of wireless local area networks (WLAN) can be found operating at two frequency bands – 2.4-GHz ISM band (IEEE 802.11b and 802.11g standards) and the 5.4-GHz band (802.11a standard) – with maximum data rates from 11 Mbps (802.11b) up through to 54 Mbps (802.11g/a) and up to over 100 Mbps (IEEE 802.11n, operating in either frequency band).

Figure 9.3 Yagis

Figure 9.4 FB series omnis

Spatial diversity is often employed in WLAN access points to overcome multipath fading effects and combine the various replicas of the received signal coherently, achieving substantial spatial diversity gain. Indeed, in 802.11n standard, multiple antennas are an absolute requirement to achieve high data rates. Omnidirectional antennas are preferred for some applications, but this depends on whether uniform coverage is required, that is, if the access point and antennas are located in the middle of a room. Some WLAN access points have integrated antennas, which are often microstrip elements, designed to provide coverage underneath the access point, in an "umbrella" fashion. Floor penetration is sometimes difficult to achieve, especially at the relatively low transmit powers used in access points from 50 to 200 mW EIRP.

When coverage enhancement is required, especially for corridors, tunnels, or to connect two buildings, directional antennas with narrow beamwidth are employed. In this case, parabolic reflectors, Yagi-Uda antennas, and phased-array panels are often used. As the number of channels which can be used is very limited (only three non-overlapping channels in the 2.4 GHz band in the many countries where 11 or 12 channels are available), interference management and sectorization (also known as zoning for indoor systems) are also important, and hence stringent directional requirements must be enforced to maximize system performance. Such high gain antennas will usually increase effective transmit power beyond regulatory limits. Hence, transmit power from the access point should be reduced pro-rata so that the gain is still effective in increasing the range at the receiver.

9.2.5 Antennas at cell site

The antennas used for cellular radio systems have to fulfil the requirement of not only providing adequate radio coverage but also reducing co-channel interference arising due to frequency reuse in cellular architecture. Hence, there is a need for high-gain (usually 6 dB or 9 dB) omnidirectional antennas and antennas with beamwidths of 60° or 120° for sectorized cells. Narrower beamwidths are sometimes needed for filling weak or dead signal spots. Typical cellular antennas use variations of the collinear antennas backed by reflectors for omnidirectional patterns or log-periodic antennas for directional patterns.

9.2.5.1 High-gain antennas

There are standard 6-dB and 9-dB gain omnidirectional antennas. The antenna patterns for 6-dB gain and 9-dB gain are shown in Figure 9.5.

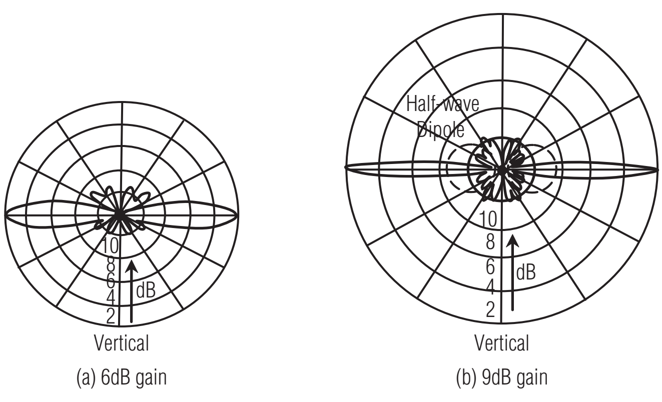


Figure 9.5 High-gain antenna pattern

9.2.5.2 Start-up system configuration

The omni-cell consists of three omnidirectional transmitting antennas. Each antenna can transmit signals from 16 radio transmitters, and have three transmitting antennas which serve 45 voice radio transmitters simultaneously. Each transmitting signal is amplified by its own channel amplifier in each radio transmitter, and then 16 channels pass through a 16-channel combiner and transmit signals through a transmitting antenna. Two receiving antennas normally can receive all 45 voice radio signals simultaneously. Then in each channel, two identical signals received by two receiving antennas pass through a diversity receiver of that channel. The receiving antenna configuration is shown in Figure 9.6(a). In abnormal antenna configuration, where the call traffic in each cell increases as the number of customers increase, some cells require a great number of radios to handle the increasing traffic. In this configuration an omni-site cell can be equipped with up to 90 voice radios. In such cases, six transmitting antennas should be used as shown in Figure 9.6(b). In the meantime, receiving antennas are still two. In order to reduce the number of transmitting antennas, a hybrid ring combiner to combine two 16 channels is found.

For interference reduction use directional antennas: When the frequency reuse scheme is used, co-channel interference will occur. The co-channel interference reduction factor, $q = D/R = 4.6$, is based on the assumption that the terrain is flat. A 120° corner reflector or 120° place reflector can be used in a 120° sector cell. A 60° corner reflector can be used in a 60° sector cell. A typical pattern for a directional antenna of 120° is shown in Figure 9.7.

9.2.5.3 Location antennas

In each cell site, a location receiver connects to the respective location antenna. This antenna can be either omnidirection or shared directional. The location receiver can tune a channel to one of 333 channels either upon demand or periodically.

Setup channel antenna: It is used to page a called mobile unit or to access a call from a mobile unit. It transmits only data. The setup channel antenna can be an omnidirectional antenna or consists of several directional antennas at one cell site. In general, in both omni-cell and sector-cell systems, one omnidirectional antenna is used for transmitting signals and the other for receiving signals in each cell site.

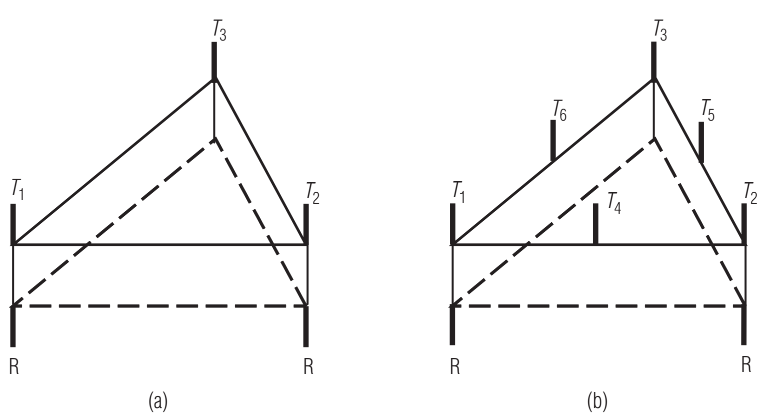


Figure 9.6 Cell-site antennas for omni-cells (a) for 45 channels (b) for 90 channels

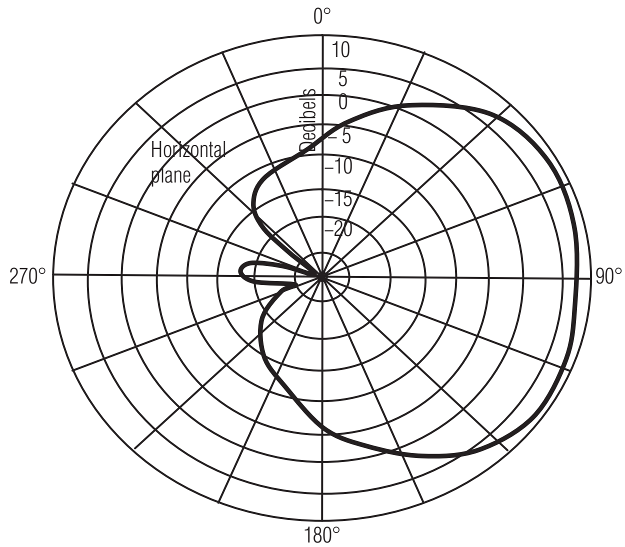


Figure 9.7 Azimuthal pattern of 8-dB directional antenna

Figure 9.8 (a) Diversity antenna spacing at the cell site and (b) Proper arrangement (*h*, antenna height; *d*, antenna separation)

Space diversity antennas used at cell site: Two branch space diversity antennas are used at the cell site to receive the same signal with different fading envelopes, one at each antenna. The degree of correlation between two fading envelopes is determined by the degree of separation between two receiving antennas. When the two fading envelopes are combined, the degree of fading is reduced.

It should be aligned as shown in Figure 9.8. The use of the space diversity antennas is at the base station.

9.2.5.4 Umbrella-pattern antennas

In certain situations, umbrella-pattern antennas should be used for the cell-site antennas.

Normal umbrella-pattern antenna: For controlling the energy in a confined area, the umbrella-pattern antenna can be developed by using a monopole with a top disc as shown in Figure 9.9.

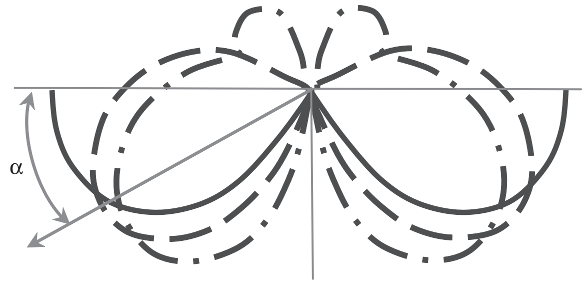


Figure 9.9 Vertical-plane patterns of quarter-wavelength stub antenna on infinite ground plane (solid) and on finite ground planes several wavelengths in diameter (dashed line) and about one wavelength in diameter (dotted line)

Broadband umbrella-pattern antenna: The parameters of discone antenna in which one of the cones is extended to 180° to form a disc are shown in Figure 9.10.

High-gain broadband umbrella-pattern antennas: A high-gain antenna can be constructed by vertically stacking a number of umbrella-pattern antennas

$$E_0 = \sin\left[(Nd/2\lambda)\cos\varphi\right]/\sin\left[(d/2\lambda)\cos\varphi\right] \cdot \text{(individual umbrella pattern)}$$

where φ is the direction of wave travel and N is the number of elements.

9.3 Design of directional antennas

In antennas, gain and directivity are interrelated. The directivity is a measure of how the RF energy is focused in any direction. As total RF energy is distributed over less area, the signal strength appears to be higher. This increase in signal strength is termed antenna gain.

Gain is measured in decibels with respect to an isotropic radiator or a dipole. The isotropic radiator is a theoretical construct. It is spherical in shape and is capable of radiating equally in all directions. The radiation pattern of Directional antenna is shown in Figure 9.11.

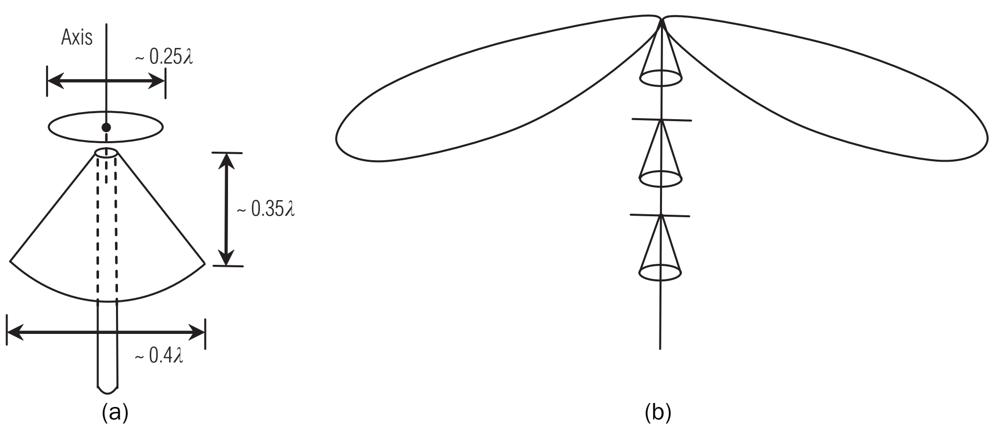


Figure 9.10 Discone antennas: (a) Single antenna and (b) An array of antennas

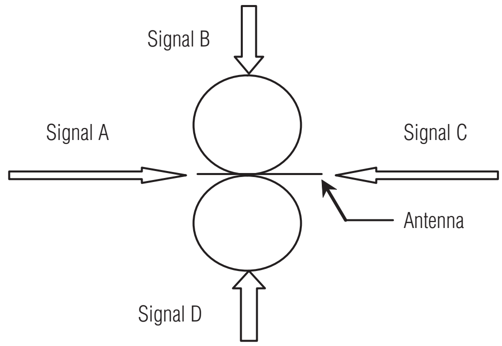


Figure 9.11 Directional antenna pattern

Solving Equations (2.14) $(q = \sqrt{3N})$ and (9.7)

$$q = \sqrt{3N} = \left(6 \times \frac{C}{I}\right)^{1/4} \tag{9.8}$$

$$\frac{C}{I} = \frac{3}{2}N^2 \tag{9.9}$$

As seen from Equation (9.9), increasing the value of N can increase C/I. When N increases, the number of frequency channels available in a cell decreases for a given frequency spectrum. This results in overall decrease of spectrum efficiency that is contrary to the advantage of frequency-reuse scheme in a cellular system. Hence, it is needed that when the call traffic increases, the frequency spectrum efficiency should be enhanced without increasing the cluster size N. Instead of increasing the cluster size K beyond 7, a suitable arrangement of directional antennas in place of an omnidirectional antenna at each cell site can be deployed to reduce co-channel interference, and thereby increasing C/I. This simply means that each omni-cell is divided into three or six sectors radially, while retaining the same number of channels per cell in the 7-cell cluster and accordingly uses three or six directional antennas at a cell site. This technique reduces co-channel interference.

9.3.1 Interference reduction antennas

A design for an antenna configuration that reduces interference in two critical directions is shown in Figure 9.12. The parasitic element is about 1.05 times longer than the active element.

Mobile antennas: The requirement of a mobile antenna is an omnidirectional antenna which can be located as high as possible from the point of reception. However, the physical limitation of antenna height on the vehicle restricts this requirement. Generally, the antenna should at least clear the top of the vehicle.

Roof-mounted antennas: The pattern of a roof-mounted antenna is more or less uniformly distributed around the mobile unit when measured at an antenna range in the free space as shown in Figure 9.13. The 3-dB antenna shows 3-dB gain over the quarter-wave antenna. However, the gain of the antenna used at the mobile unit must be limited to 3 dB because the cell-site antenna is rarely as high as the broadcasting antenna and out-of-sight conditions often prevail.

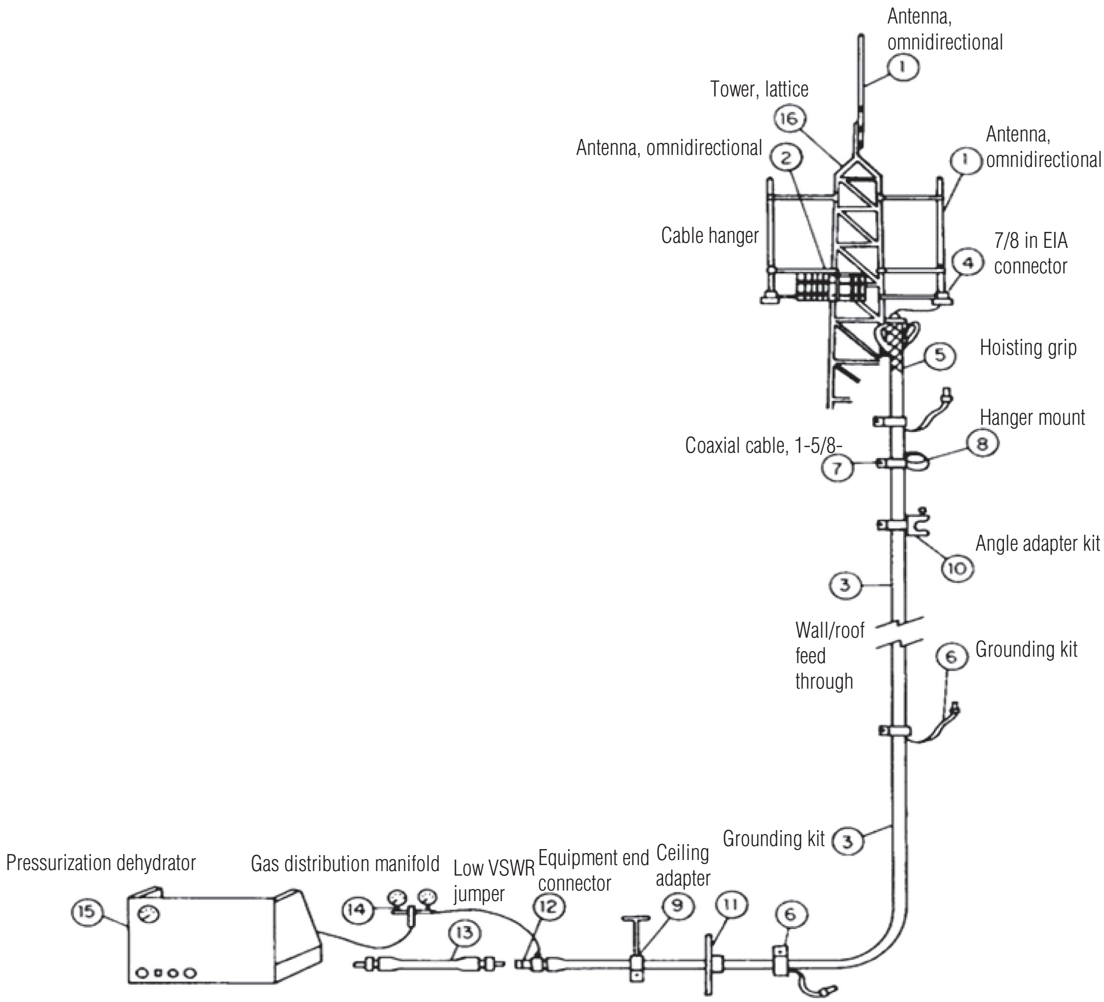


Figure 9.12 Antenna configuration for interference reduction

The mobile antenna with a gain of more than 3 dB can receive only a limited portion of the total multipath signal in the elevation as measured under the out-of-sight condition.

Glass-mounted antennas: In this type of antenna, energy is coupled through the glass; hence, drilling a hole in not required. However, some energy is dissipated on passage through the glass. Its gain range is 1–3 dB depending on frequency. Its position is lower than roof-mounted antenna. Also, it cannot be installed on shaded glass found in motor vehicles.

Mobile high-gain antennas: A high-gain antenna used on a mobile unit should be distinguished from a directional antenna. In directional antenna, the pattern is suppressed horizontally, and in high-gain antenna, the pattern is suppressed vertically. In a mobile radio environment, the scattered signals arrive at the mobile unit from every direction with every probability. Hence, an omnidirectional must be used. Moreover, measurements reveal that elevation angle for scattered signals received in urban areas is greater than in suburban areas.

Horizontally oriented space diversity antennas: The two branch space diversity receiver mounted on a motor vehicle has the advantage of reducing fading and thus can operate at a lower reception level. The discussion here concerns a space diversity scheme in which two vehicle-mounted antennas separated horizontally by 0.5 wavelength can achieve this diversity.

9.4 Macrocell antenna

A macrocell provides the largest area of coverage within a mobile network. The antennas for macrocells can be mounted on ground-based masts, rooftops, or other existing structures. They must be positioned at a height that is not obstructed by terrain or buildings. Macrocells provide radio coverage over varying distances depending on the frequency used, the number of calls made, and the physical terrain. Macrocell base stations have a typical power output in tens of watts

9.4.1 Performance requirements in macrocells

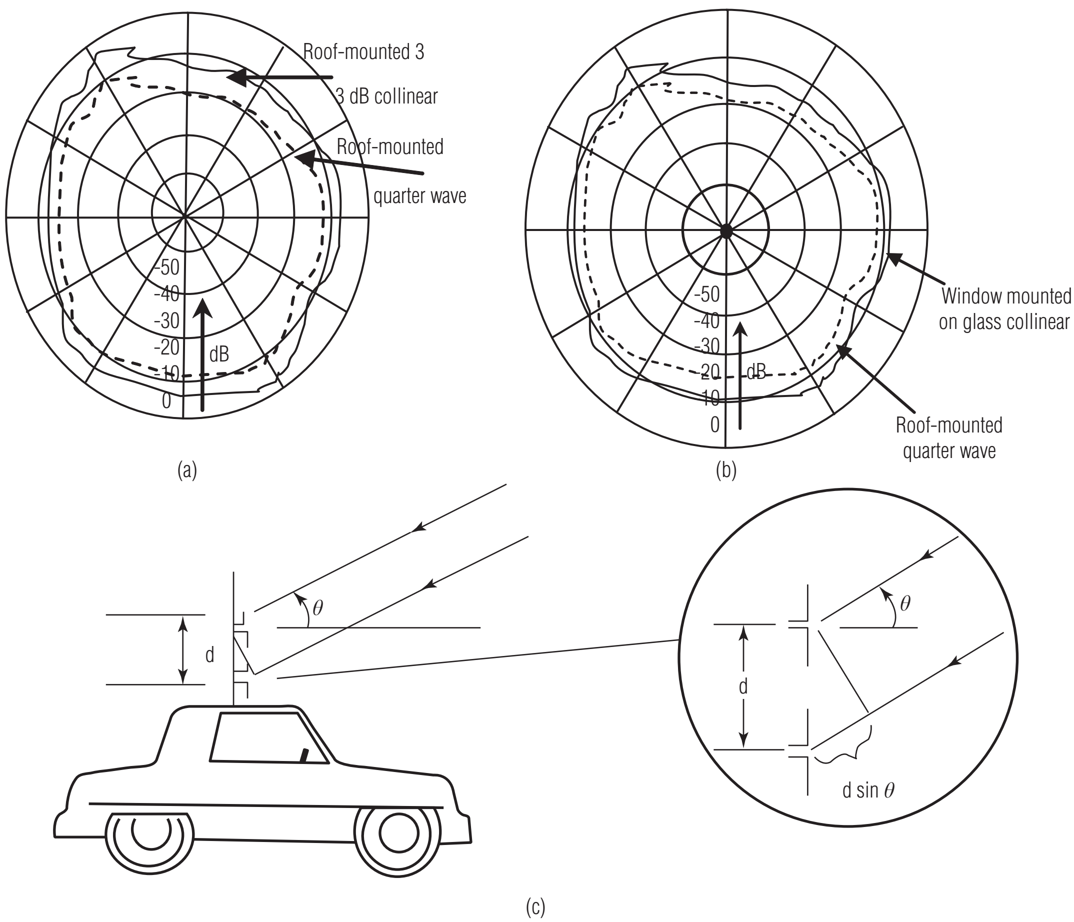


Figure 9.13 Mobile antenna patterns. (a) Roof-mounted 3-dB gain collinear antenna versus roof-mounted quarter-wave antenna; (b) window mounted "on-glass" gain antenna versus roof-mounted quarter-wave antenna; (c) vertical separation between two mobile antennas

The basic function of a macrocell base station antenna is to provide not only uniform coverage in the azimuth plane, but also directivity in the vertical plane, making the best possible use of the input power by directing it at the ground rather than the at the sky. When full omnidirectional coverage is required, vertical directivity is usually provided by creating a vertical array of dipoles, phased to give an appropriate pattern (Figure 9.14). This is usually called a collinear antenna and has the appearance of a simple monopole. Such antennas are commonly used for private mobile radio systems. More commonly in cellular mobile systems , however , some limited azimuth directivity is required in order to divide the coverage into sectors as shown in Figure 9.15. The choice of the azimuth beamwidth is a trade-off between allowing sufficient overlap between sectors, permitting smooth handovers, and controlling the interference reduction between co-channel sites, which is the main point of sectorization.

This can be achieved using either mechanical downtilt (where the antenna is simply pointed slightly downwards) or electrical downtilt (where the phase weighting of the individual elements within the panel is chosen to produce a downward pointing pattern) with the antenna axis maintained vertical.

Electrical downtilt is usually preferred as it can produce relatively even coverage in the azimuth plane. The combination of a downtilted radiation pattern with the macrocell path-loss models has the effect of increasing the effective path-loss exponent. This causes the power received

Figure 9.14 Elevation pattern for omni-directional collinear antenna

Figure 9.15 Typical macrocell sector antenna

from the base station to fall off more abruptly at the edge of the cell, reducing the impact of interference and permitting the available spectrum to be reused more efficiently.

9.4.2 Macrocell antenna design

Modern macrocell antennas usually achieve the desired radiation pattern by creating an array of individual elements in the horizontal and vertical directions, built together into a single panel antenna. The array is typically divided into several subarrays. The array elements are fed via a feed network, which divides power from the feed to excite the elements with differing amplitudes and phases to produce the desired pattern. The overall elevation pattern, $G(\theta)$, of such an array is given by

$$G(\theta) = g_0(\theta)\sum_{n=1}^{N/M}\sum_{m=1}^{M} I_{nm} \times \exp\left(j\phi_{nm}\right) \times \exp\left(jkd_{nm}\sin\theta\right) \times \exp\left(-jkd\phi_r\right) \tag{9.10}$$

where
$g_0(\theta)$ is the radiation pattern of an individual array element (or subarray)

I_{nm} and ϕ_{nm} are the amplitude and the phase of the excitation of the array element numbered $m + (n - 1)M$, respectively

$d_{nm} = d(m + (n - 1)M)$ [m] is the distance along the array of the n; m element,

where
M is the number of element rows in a subarray
m is the row number
N is the total number of rows
n is the sub-array number.

$$\varphi_r = m\sin\theta_{\text{sub}} + (n-1)M\sin\theta_{\text{tilt}} \tag{9.11}$$

where
θ_{sub} is the wave front downtilt associated with a subarray and
θ_{tilt} is the desired electrical downtilt of the whole antenna.

The excitation phases and amplitudes are chosen to maximize gain in the desired direction and to minimize side-lobes away from the main-lobe, particularly, in directions which will produce interference to neighbouring cells. In the vertical plane, it is important to minimize side-lobes above the main-lobe and to ensure that the first null below the main-lobe is filled to avoid the presence of a coverage hole close to the base station.

Downtilt is a critical parameter for optimizing the coverage area of a network. It is often desirable to vary the downtilt as the network evolves from providing wide-area coverage to providing high capacity over a limited area. This can be achieved by varying the element feed phases in an appropriate network. For example, the antenna illustrated in Figure 9.16 allows multiple operators to share the same antenna, while being able to provide independent variable electrical downtilt per operator according to the needs of each operator's network.

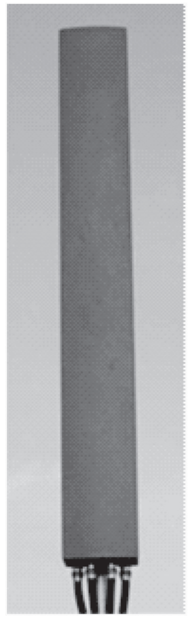


Figure 9.16 A multi-operator antenna with independently selectable electrical down tilt

Array elements can be composed of dipoles operating over corner reflectors over a corner reflector. It is more common in modern antennas, however, to use patch antennas to reduce the antenna panel thickness. These are often created with metal plates suspended over a ground plane rather than printed on a dielectric to maximize efficiency and to avoid arcing arising from the high RF voltages, which can be developed in high-power macrocells. Another important practical consideration in the antenna design is to minimize the creation of passive intermodulation products (PIMs). These arise from nonlinearities in the antenna structure, which create spurious transmission products at frequencies which may be far removed from the input frequency. They occur due to rectification of voltages at junctions between dissimilar metals or at locations where metal corrosion has occurred, so that the choice of metals and the bonding arrangements at junctions must be carefully considered to minimize such issues.

9.4.3 Macrocell antenna diversity

Figure 9.17 shows typical macrocell antenna. The first is a three-sector system, where each sector consists of two panels arranged to provide spatial diversity to overcome narrowband fading, with a spacing following principles which will be described in handset diversity antennas.

A more compact arrangement is produced by using polarization diversity, where each panel provides two orthogonally polarized outputs as will be described in polarization diversity. This is typically achieved by dual orthogonal feeds to patch antenna elements.

9.5 Microcell antennas

The large numbers of individual rays, which can contribute to microcell propagation, makes it clear that the cell shape is not determined directly by the radiation pattern of the base

Figure 9.17 Typical macrocell antenna

station antennas, since each of the rays will emerge with a different power. Nevertheless, it is still important to ensure that power is radiated generally in the right directions so as to excite desirable multipath modes and to avoid wasting power in the vertical direction.

In determining the practical antenna pattern, the interactions between the antenna and its immediate surroundings are also very important. These objects may include walls and signs, which may often be within the near field of the antenna, so that accurate prediction and analysis of these effects require detailed electromagnetic analysis using techniques such as the finite-difference time-domain method.

Some typical examples of practical microcell antennas are shown in Figure 9.18. All are designed to be mounted on building walls and to radiate in both directions along the streets they are serving. An alternative approach is to use a directional antenna, such as a Yagi-Uda antenna (Fig. 9.19), which may help with minimizing interference to other cells. Directional antennas are also useful for containing cell coverage along roads which are not lined with buildings in a sufficiently regular pattern.

Figure 9.18 Microcell antennas (used for 3G)

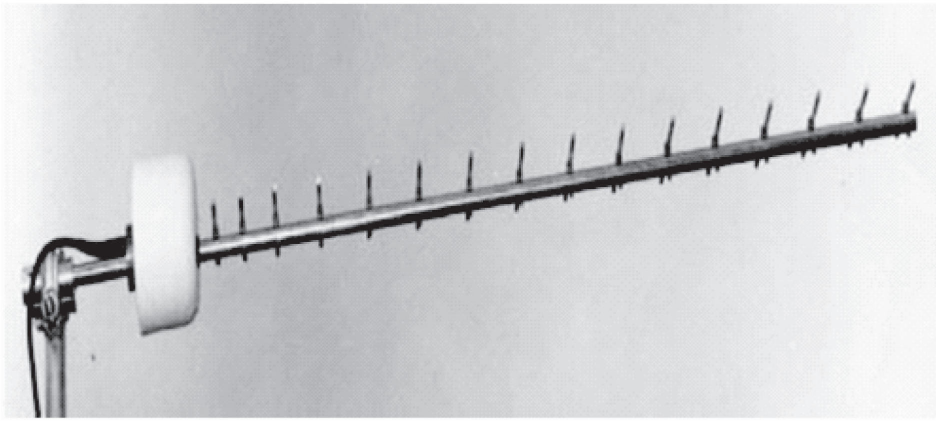


Figure 9.19 Yagi antenna for 900 MHz

It is common practice to use microcell antennas for either outdoor or in-building environments, and often antenna manufacturers do not distinguish between these applications. Although some microcell antennas can be used in in-building, there are other types that due to their size and construction would be aesthetically inappropriate for indoor use. Yagi-Uda, omnidirectional shrouded, ceiling-mount, and monopole antennas are often employed.

9.6 Picocell antennas

A practical ceiling-mounted antenna for indoor coverage at 900 MHz is shown in Figure 9.20 (picocell antenna). Particular requirements of indoor antennas include very wide beamwidth, consistent with a discrete appearance. Hence this particular antenna has been designed to look similar to a smoke detector. Linear polarization is currently used almost universally for indoor communications, but there are potential benefits in the use of circular polarization. This has

Features

- Light weight, slim profile
- Indoor or outdoor use
- Weatherproof and corrosion resistant
- Built-in swivel mount
- Fast and easy installation
- Choice of polarization and frequency

Figure 9.20 Picocell antenna and its features

been shown to substantially reduce the fade depth and RMS delay spread due to the rejection of odd order reflections as well as reducing polarization mismatch loss. Similarly, reduction in antenna beamwidth has been shown to substantially reduce the delay spread in line-of-sight situations, but this effect must be traded against the difficulty of providing a reasonably uniform coverage area.

Increasingly, indoor antennas and the associated feed powers also typically have to be compliant with specific requirements on radiated power density and specific absorption. It is also increasingly desirable to achieve a high level of integration between the systems and technologies deployed by different operators, so wideband and multiband indoor antennas are increasingly of interest, providing, for example, WLAN, 2G, and 3G technologies in single antenna housing. Printed antennas, including micro-strip patches, are attractive for indoor antenna designs, with wire antennas being more useful at lower frequencies. Biconical antennas have been proposed for millimetre-wave systems, giving a good uniformity coverage.

9.7 Femtocell

A femtocell is a small cellular base station designed for use in residential or small business environments. It connects to the service provider's network via broadband (such as DSL or cable) and typically supports 2–4 active mobile phones in a residential setting, and 8–16 active mobile phones. In the indoor areas where the service coverage is limited, a femtocell can be used to extend service coverage. It also decreases backhaul costs, since it routes your mobile phone traffic through the IP network.

Although WCDMA is discussed much, the same concept is applicable to GSM, CDMA2000, TD-SCDMA, WiMAX, and LTE solutions.

A femtocell is sometimes referred to as a "home base station," "access point base station," "3G access points," "small cellular base station," and "personal 2G-3G base station."

The femtocell attracts the mobile operators by improving both coverage and capacity indoors. These can reduce both capital expenditure and operating expense. There may also be opportunity for new services. Consumers benefit from improved coverage and potentially better voice quality and battery life. Depending on the carrier, they may also be offered more attractive tariffs, for example, discounted calls from home. The advantage of fixed-mobile convergence can also be obtained from femtocells. A distinction of a femtocell from other architectures is that it will work with existing handsets but a new access point with licensed spectrum is required for its installation. On the other hand, most other FMC architectures need a dual mode handset that works with existing unlicensed spectrum wireless access points.

Many service providers such as AT&T, Vodafone, Verizon, and Mobile TeleSystems and Sprint Nextel have launched femtocell service. Note that the 3GPP refers to 3G femtocells as Home Node Bs (HNBs). The range of a microcell is below 2 km, that of a picocell is below 200 m, and a femtocell has a range of nearly 10 m.

9.7.1 Mode of operation

Femtocells are sold by a mobile network operator (MNO) to its residential or enterprise customers. A femtocell is typically the size of a residential gateway or smaller, and connects to the user's broadband line. Integrated femtocells, which include both a DSL router and a femtocell, also exist. Once plugged in, the femtocell connects to the MNO's mobile network, and provides extra coverage in a range of typically 30–50 m for residential femtocells, (depending on the existing coverage and output power, usually 20 mW which is five times less than a WiFi router). From a user's perspective, it is plug and play, there is no specific installation or technical knowledge required – anyone can install a femtocell at home.

In most cases, the user must then declare which mobile phone numbers are allowed to connect to his/her femtocell, usually via a web interface provided by the MNO. This needs to be done only once. When these mobile phones arrive under coverage of the femtocell, they switch over from the macrocell (outdoor) to the femtocell automatically. Most MNOs provide a way for the user to be aware of this change, by having a different network name appearing on the mobile phone. All communications will then automatically go through the femtocell. When the user leaves the femtocell coverage (whether in a call or not) area, his phone hands over seamlessly to the macro network. Femtocells require specific hardware, so existing WiFi or DSL routers cannot be upgraded to a femtocell.

Once installed in a specific location, most femtocells have protection mechanisms so that a location change will be reported to the MNO. Whether the MNO allows femtocells to operate in a different location depends on the MNO's policy. International location change of a femtocell is not permitted because the femtocell transmits licensed frequencies which belong to different network operators in different countries.

9.7.1.1 The main advantages of the femtocell for the mobile users

The main advantages for an end-user are as follows:

- "5 bar" coverage when there is no existing signal or poor coverage
- Higher mobile data capacity, which is important if the end-user makes use of mobile data on their mobile phone (may not be relevant to a large number of subscribers who instead use Wi-Fi where femtocell is located)
- Depending on the pricing policy of the MNO, special tariffs at home can be applied for calls placed under femtocell coverage
- For enterprise users, having femtos instead of DECT phones enable them to have a single phone, so a single contact list, etc.

9.8 Space diversity antennas

Antenna diversity is also known as space diversity. Space diversity antennas are used to improve the quality and reliability of a wireless. We generally use space diversity antennas for mobile communication.

9.8.1 Handset diversity antennas

Practical handset diversity antennas are of increasing interest to increase reliability when users are slow moving, to provide an increase in downlink signal-to-noise ratios to increase download data rates, and to increase overall channel capacity via MIMO techniques. For practical implementation, the fundamental challenge is the antenna size. One approach is to combine an external antenna such as a loaded whip or sleeve dipole with a compact internal antenna, such as a Planar Inverted-F Antenna (PIFA) or patch. Although the internal antenna is likely to provide lower mean effective gain (MEG), this still allows it to overcome the majority of fading nulls encountered at the "main" element. With a 3-dB reduction in MEG for one antenna, a two-branch diversity system loses only around 1 dB of diversity gain at 1% fade probability in a Rayleigh environment.

Additionally, the dissimilar patterns and polarization states mean that the fading correlation coefficient is usually much smaller than that might be expected from simple theory based on the element spacing alone. The use of two similar antennas spaced apart by a small fraction of wavelength is far more effective than might be expected. The mutual coupling between the elements interacts with the spatial field patterns to produce low cross-correlation even with a spacing of only 0.05–0.10 wavelengths

9.8.2 Base station space diversity

The angular distribution of scattering at the base station antenna may be very different from the mobile case, particularly for macrocells. Figure 9.21 illustrates a typical geometry for a macrocell in a built-up area.

If it is assumed that the main scatterers contributing to the signal are those within the first Fresnel zone around the direct ray from the mobile to the base, then the separation of the scatterers will clearly reduce as the base height is increased. The main scatterers are likely to be located close to the mobile, so the angular distribution at the base station may be rather narrow.

Despite the large required spacings, horizontal space diversity is very commonly applied in cellular base stations to allow compensation for the low transmit power possible from hand-portables compared to the base stations. Vertical spacing is rarely used; this is because of the large spacing's required to obtain low cross-correlation and because the different heights of the antennas can lead to significant differences in the path loss for each antenna, which degrades the diversity effect.

In microcell environments, both the base and mobile antennas are submerged among scatterers, so that the angular spread of scatterers is very high. However, there is a high probability of encountering a strong line-of-sight component. The usefulness of space diversity will depend on the particular geometry of the scatterers in the cell. In picocells, the angles of arrival will be distributed even more widely in three dimensions, particularly when propagation takes place between floors.

9.8.3 Polarization diversity

Both reflection and diffraction are polarization sensitive and can produce a rotation of the polarization of the scattered wave compared to the incident wave. The compound effect of multiple instances of these processes in the propagation path depolarizes a vertically polarized transmission, producing a significant horizontally polarized component at the receiver. This allows polarization diversity when two collocated but differently polarized antennas are used as the branches of a diversity receiver.

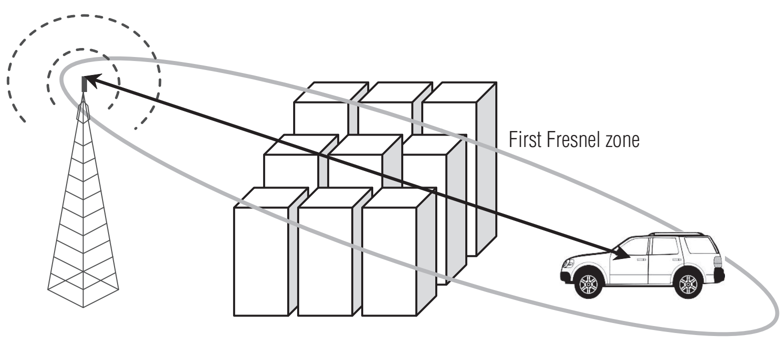


Figure 9.21 Effect of base antenna height on scatter distribution in macrocells

Collocation is attractive to reduce the aesthetic impact of base station antennas and to allow a very compact solution in the hand-held case. Base station polarization diversity also helps to reduce the polarization mismatch which may be produced by hand-held users, who tend to hold their hand-held with an average angle of round 45° to the vertical, although this is mainly significant in line-of-sight cases.

9.8.4 Time diversity

The Bessel function decorrelation of the fading channel implies that diversity can be obtained from a single antenna by transmitting the same signal multiple times, spaced apart in time sufficiently so that the channel fading is de-correlated, that is, at least around 0.5 l between antenna locations when the repeated signal is received. This is rarely used in practice; however, as the retransmission of information reduces the system capacity and introduces a transmission delay. Nevertheless, the principle is applied to improving efficiency in coded modulation schemes, which apply interleaving to spread errors across fades, allowing better potential for error correction.

9.8.5 Frequency diversity

In wideband channels, two frequency components spaced wider than the coherence bandwidth experience uncorrelated fading, providing another means of obtaining diversity. As in time diversity, the simple retransmission of information on two frequencies would be inefficient.

9.9 Base station systems

Smart antenna systems are employed to overcome multipath fading, extend range, and increase capacity by using diversity or beam-forming techniques in wireless communication systems. Understanding of the smart base antenna performance mechanisms for various environments is important to design cost-effective systems and networks.

Capacity and reliability of wireless communication systems are limited by three major impairments: multipath fading, delay spread, and co-channel interference. Smart antenna techniques have been investigated to increase capacity and extend range by overcoming these impairments with more than one antenna. The performance of smart antenna techniques is determined by multipath fading and co-channel interference, and is limited by the antenna configuration and the condition of propagation environment. Digital-cellular voice communication assumes equal-data traffic on the reverse-link (or uplink) and the forward-link (or downlink), but the reverse-link is more challenging due to the limited mobile-terminal transmit power. Diversity techniques are employed singly or in combination to improve reverse-link performance by overcoming multipath fading.

9.9.1 Smart base station antenna

The smart base station consists of a mobile transmitter unit, antenna assembly, receiver assembly, and data acquisition assembly as shown in Figure 9.22. An assistant operator with the mobile transmitter unit is equipped with a cellular phone for the two-way voice communication with the operator at the base station.

Figure 9.23 illustrates the overall concept of the simultaneous measurement with the space, polarization, and angle diversity systems in multipath environments.

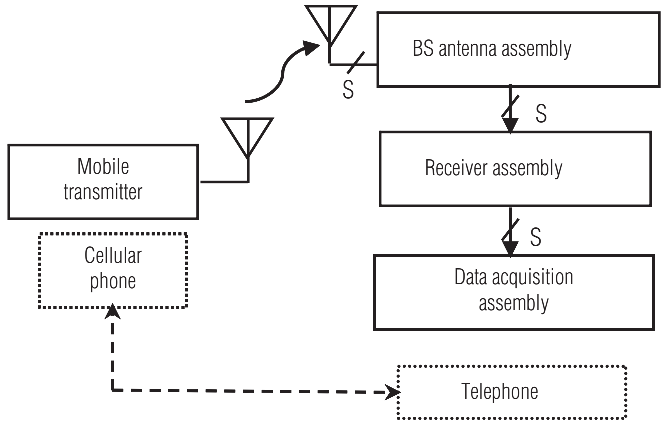


Figure 9.22 Configuration of the smart base station hardware

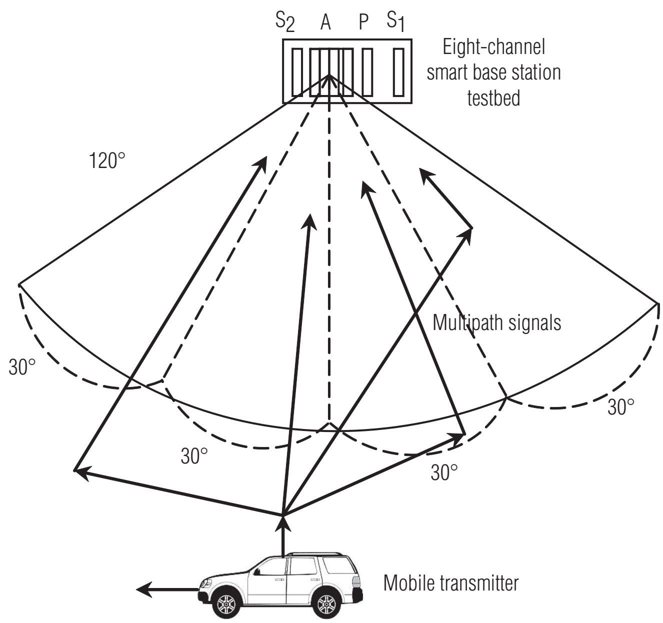


Figure 9.23 Eight-channel smart base station bed

9.9.2 Base station antenna assembly

The practical smart base station antenna configuration is shown in Figure 9.24. The base station antenna assembly consists of one panel antenna with 4 × 30° beams, two sector antennas, and one ±45° slanted dual polarized antenna. The 4 × 30° panel antenna covers 120°, sector antennas cover 95°, and the dual polarized antenna covers 90°. All have a vertical beamwidth of about 15°. The three antenna subsystems have a total of eight outputs connected to an eight-channel receiving system with the channel designations shown in Table 9.2. Channel S_1 is used as a reference channel through the measurement and calibration process.

Table 9.2 Assigned channel antenna

Channel number	1	2	3	4	5	6	7	8
Symbol	S_1	P_1	P_2	A_1	A_2	A_3	A_4	S_2

where, S is the space diversity antenna,
P is the polarization diversity antenna,
A is the angle diversity antenna

Figure 9.23 shows antenna assembly with the 95° sector (S_1), dual polarized (P), 4 × 30° narrowbeam (A), and 95° sector antennas (S_2) on top of the six-story high building. The height of the antenna assembly is 30 m from the ground. The order of the narrow beams is counted from outer left (A_1) to outer right (A_4). The polarization of P_1 is 45° slanted to the right and that of P_2 is slanted to the left. The separation between the 95° sector antennas is about 3 m (8.5 wavelengths at 842 MHz).

Figure 9.24 Configuration of smart base station antenna

9.10 Effects of cell-site antenna parameters

There are several ways of exploiting the effects of various cell-site antenna parameters. For example, antenna radiation pattern, antenna beamwidth, antenna gain, antenna height, separation between transmitting and receiving antennas, and antenna tilting affect the cellular system design. There are some other aspects of antenna parameters, such as reorienting the direction antenna patterns, changing the antenna beamwidth, and decreasing the antenna height at the cell site, which may also affect the radio coverage but reduce the system interference. The antenna pattern can be omnidirectional, directional, or any other shape in both the horizontal and vertical planes. The antenna patterns seen in cellular systems are different from the patterns seen in the free space. For example, if a mobile unit travels around a cell site in areas with many buildings, the omnidirectional antenna will not duplicate the omnidirectional radiation pattern. The radiation pattern is distorted in an urban or suburban environment. So the design of the antenna pattern should be based on the terrain contour, the building and forest density, and other conditions within a given area. The signal reflection mechanism in mobile radio environment is illustrated in Figure 9.25.

For a 120° directional antenna, the front-to-back ratio is about 10 dB less than the front-lobe. This condition exists because the strong signal radiates in front, bouncing back from the surroundings, so that some part of the signal energy can be received from the back of the antenna. In such a case, the beamwidth of the directional antenna used at the cell site has no correlation between its measured values in the free space and the mobile radio environment. Control of secondary lobe formation in an antenna radiation pattern is very critical in the implementation of a directional antenna. A skewed configuration of the number of directional antennas at the same cell site results in smoother pattern as compared to a pattern accompanied with ripples and deep null formation in case of all antennas facing outward.

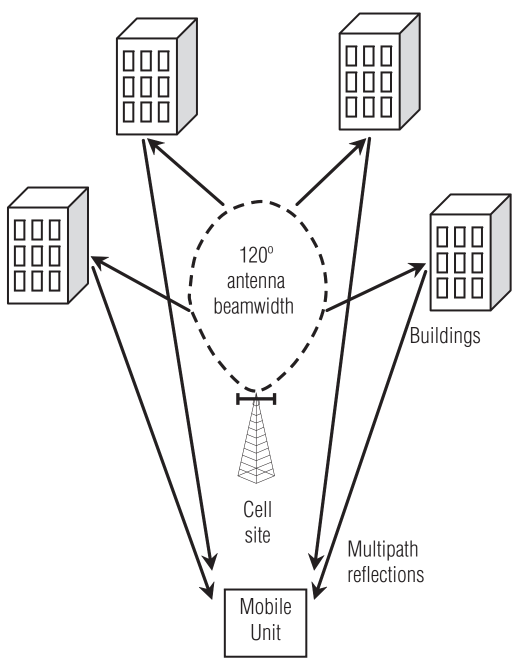


Figure 9.25 Signal reflection mechanisms in a mobile radio environment

The antenna gain enhances the radiated power from the antenna. The height of the cell-site antenna can affect the area and shape of the coverage in the system. Effective antenna height should be taken into consideration for received signal analysis within a cell. The effective antenna height and the actual antenna height are same only when the mobile unit is traveling on flat ground. It is easy to decrease antenna height to control coverage in a flat-terrain area. For decreasing antenna height in a hilly area, the signal strength contour is different from the situation of transmitting power decrease. Therefore, a decrease in antenna height becomes very difficult to control the overall coverage plan. Some areas within the cell may have a high attenuation, while others may not. Separation between two transmitting antennas at the same cell site should be minimized to avoid the inter-modulation.

The minimum separation between a transmitting antenna and a receiving antenna must be ensured in order to avoid receiver desensitization. For example, if the separation between a transmitting antenna and a receiving antenna is 15 m horizontally, the signal isolation obtained from the free-space formula is 56 dB. For better performance, the transmitting and receiving antennas are not mounted in the same horizontal plane, but rather these are mounted on the same vertical pole, if they are omnidirectional antennas. In case of directional antennas, this restriction can be moderated because of the directive patterns.

9.11 Summary

- A base station (BS) is the interface between wireless phones and traditional wired phones. These are mounted on tall towers because from this high point it is easier to stay in communication with cell phone users and also to overcome obstacles.
- There are two types of base station antennas used, omnidirectional and directional antennas.
- Omnidirectional antennas are generally used in low traffic volume areas. The classical omnidirectional $\lambda/2$ antennas are of a ground plane or $\lambda/4$-skirt nature. The names indicate how the antenna is decoupled from the mast.
- Directional antennas are useful in remote locations where high gain is required and the direction to a desired transmitter/receiver is known.
- Base station antenna (BSA) series omni antennas are engineered to provide lasting performance in the most demanding field conditions.
  - YA series Yagis are built to precise specifications to perform in extreme weather conditions.
  - FB series omni antennas are designed for optimum performance in extreme weather conditions.
- A macrocell provides the largest area of coverage within a mobile network. It provides radio coverage over varying distances depending on the frequency used, the number of calls made, and the physical terrain.
- Microcell antennas are the large numbers of individual rays which can contribute to microcell propagation; the cell shape is not determined directly by the radiation pattern of the base station antennas since each of the rays will emerge with a different power.
- A practical ceiling-mounted antenna for indoor coverage at 900 MHz is known as picocell antenna.
- A femtocell is a small cellular base station designed for use in residential or small business environments. It connects to the service provider's network via broadband (such as DSL or cable) and typically supports 2–4 active mobile phones in a residential setting, and 8–16 active mobile phones.

- Smart antenna systems are employed to overcome multipath fading, extend range, and increase capacity by using diversity or beam-forming techniques in wireless communication systems.
- Capacity and reliability of wireless communication systems are limited by three major impairments: multipath fading, delay spread, and co-channel interference.
- The smart base station consists of a mobile transmitter unit, antenna assembly, receiver assembly, and data acquisition assembly.
- An assistant operator with the mobile transmitter unit is equipped with a cellular phone for the two-way voice communication with the operator at the base station.
- The base station antenna assembly consists of one panel antenna with 4 × 30° beams, two sector antennas, and one ±45° slanted dual polarized antenna.
- The 4 × 30° panel antenna covers 120°, sector antennas cover 95°, and the dual polarized antenna covers 90°.

Example problem 9.1

A cell-site transmitter generates a +15 dBm RF signal and is connected to an antenna using an RF coaxial cable that induces a 3-dB loss. The cable has two connectors at its either end that induce a loss of 2 dB each. What is the signal level at the input of the antenna?

Solution

Data given:
The RF output signal level of cell-site transmitter = +15 dBm
Signal loss due to one connector of RF coaxial cable = 2 dB

- **Step 1.** To determine total loss in cable connectors
  Number of connectors on a coaxial cable = 2
  Therefore, signal loss due to both connectors = 2 × 2 = 4 dB
- **Step 2.** To determine signal loss due to cable and connectors
  Signal loss due to RF coaxial cable = 3 dB
  Total signal loss due to cable and connectors = 3 dB + 4 dB = 7 dB
- **Step 3.** To determine signal level at the input of the antenna
  Signal level at the input of the antenna = +15 dBm – 7 dB = +8 dBm

---

Example problem 9.2

A cell-site transmitter generates a +12 dBm RF signal and is connected to an antenna using an RF coaxial cable that induces a 2-dB loss. The cable has two connectors at its either end that induce a loss of 3 dB each. What is the signal level at the input of the antenna?

Solution

Data given:
The RF output signal level of cell-site transmitter = +15 dBm
Signal loss due to one connector of RF coaxial cable = 2 dB

- **Step 1.** To determine total loss in cable connectors
  Number of connectors on a coaxial cable = 2
  Therefore, signal loss due to both connectors = 2 × 3 = 6 dB
- **Step 2.** To determine signal loss due to cable and connectors
  Signal loss due to RF coaxial cable = 3 dB
  Total signal loss due to cable and connectors = 3 dB + 6 dB = 9 dB
- **Step 3.** To determine signal level at the input of the antenna
  Signal level at the input of the antenna = +12 dBm – 9 dB = +3 dBm

Review questions

1. Describe the structure of macrocell antenna with design principles.
2. Describe the structure of microcell antenna with design principles.
3. What is a femtocell?
4. Define space diversity technique.
5. Explain mobile high-gain antennas in detail.
6. What are directional antennas? Explain directional antennas for interference in detail.
7. Explain space diversity antennas in detail.
8. What is the use of broadband umbrella-pattern antenna?
9. Draw a simple diagram of diversity antenna spacing concept in cell site.
10. Concerning the cell-site antenna, explain the start-up configuration and abnormal antenna configuration of the start-up system. (Refer Section 9.2.5.2)
11. Explain in detail the importance of consideration of cell-site antennas. (Refer Section 9.2)
12. How can a high gain broadband umbrella pattern antenna be constructed for cell site? (Refer Section 9.2.5.4)
13. Explain horizontally and vertically oriented space diversity antennas. (Refer Section 9.8)
14. What are the antennas used at cell site? Explain them. (Refer Section 9.2)
15. Draw the directional antenna configuration for 120°sector (90 channels) and explain how the interference is reduced. (Refer Section 3.4)

Exercise problems

1. A cell-site transmitter generates a +10 dBm RF signal and is connected to an antenna using an RF coaxial cable that induces a 1-dB loss. The cable has two connectors at its either end that induce a loss of 3 dB each. What is the signal level at the input of the antenna? (Ans: 3 dBm)
2. A cell-site transmitter generates a +11 dBm RF signal and is connected to an antenna using an RF coaxial cable that induces a 2-dB loss. The cable has two connectors at its either end that induce a loss of 2 dB each. What is the signal level at the input of the antenna? (Ans: 5 dBm)
3. A cell-site transmitter generates a –9 dBm RF signal and is connected to an antenna using an RF coaxial cable that induces a 3-dB loss. The cable has two connectors at its either end that induce a loss of 1 dB each. What is the signal level at the input of the antenna? (Ans: –14 dBm)

Objective type questions and answers

1. In space diversity antennas, if h is the height of antenna and d is separation distance, then efficiency η is
   (a) $\eta = h^2/d$ (b) $\eta = h/d$ (c) $\eta = h/d^2$ (d) $\eta = (h/d)^2$
2. Several vertically stacked umbrella pattern can form a(n)
   (a) abnormal antenna (b) high-gain antenna
   (c) interference reduction antenna (d) mobile antenna
3. Minimum separation of cell-site receiving antennas will have one of the following advantage
   (a) can avoid intermodulation (b) reduces ISI
   (c) Reduces noise (d) controls distortion
4. The two-branch space diversity antennas can ____ in mobile communication
   (a) reduce fading (b) increase fading (c) reduce gain (d) increase ISI
5. The gain of glass-mounted antennas will be
   (a) 1–10 dB (b) 3 dB (c) 4 dB (d) 7 dB
6. Separation between two transmitting antennas at the same cell site should be minimized to avoid
   (a) co-channel interference (b) adjacent channel interference
   (c) intermodulation (d) receiver desensitization
7. In abnormal antenna configuration, an omni cell-site is equipped with 45 channels, for which ………….. transmitting antennas are used.
   (a) 13 (b) 5 (c) 3 (d) 9
8. One transmitting antenna for two receiving antennas is applied in directional antenna arrangement. Then the type of cell sectoring is a …….
   (a) 60° sector (b) 120° sector (c) 30° sector (d) 90° sector
9. Cellular base-station receiving antennas are usually mounted in such a way so as to obtain __________ diversity.
   (a) frequency (b) polarization (c) space (d) horizontal
10. The radiation resistance of a half-wave dipole antenna situated in free space and fed at the centre is approximately
    (a) 50 Ω (b) 70 Ω (c) 300 Ω (d) 1 MΩ

Answers: 1. (a), 2. (b), 3. (a), 4. (a), 5. (b), 6. (c), 7. (c), 8. (b), 9. (c), 10. (b).

Open book questions

1. What is known as abnormal antenna configuration?
2. Write short notes on location antennas.
3. Why are high-gain omnidirectional or directional antennas used at the cell site?
4. What is the significance of quarter-radius rule in the selection of the cell site?
5. Why is the in-line arrangement of the two horizontally oriented space-diversity mobile antennas used?
6. Which antenna parameters can affect the cellular system design so as to improve the signal quality by reducing co-channel interference?
7. Define the following to the antennas table:
   (i) ERP
   (ii) Equivalent aperture
   (iii) 120° sector cell

8. Describe the effects of cell-site antennal heights and signal coverage cells.
9. What are the types of antennas used for coverage and interference reduction? Explain them.
10. Draw the cell-site antenna for omnicells for 45 and 90 channels and explain them.
11. State the differences between roof-mounted and glass-mounted antennas.
12. What are the advantages of using umbrella-pattern antennas at cell site?
13. Derive the relation between the received power and electrical field of the antenna.
14. Assume a receiver is located 10 km from a 50 W transmitter. The carrier frequency is 6 GHz and free space propagation is assumed $G_t = 1$ and $G_r = 1$. Find the following:
    (i) The power of the receiver.
    (ii) Magnitude of the electric field at the receiving antenna.
    (iii) The rms voltage applied to the receiver input assuming that the receiving antenna has purely real importance of 50 Ω and is matched to the receiver.

Key equations

1. The signal-to-co-channel interference ratio (C/I) at the desired mobile receiver is given by

$$\frac{C}{I} \approx \frac{C}{\sum_{i=1}^{N_I} I_i}$$

2. The overall elevation pattern, $G(\theta)$ of such an array is given by

$$G(\theta) = g_0(\theta) \sum_{n=1}^{N/M} \sum_{m=1}^{M} I_{nm} \times \exp(j\varphi_{nm}) \times \exp(jkd_{nm} \sin\theta) \times \exp(-jkd\varphi_r)$$

Further reading

Dey, S., Venugopalan, P., Jose, K. A., Aanandan, C. K., Mohanan, P., and Nair, K. G. "Bandwidth Enhancement by Flared Microstrip Dipole Antenna," *IEEE Antennas and Propagation Society International Symposium, 1991*, AP-S, Digest, 24–28 June 1991, Vol. 1, pp. 342–345, London, Ont.

Evtioushkine, G. A., Kim, J. W., and Han, K. S. "Very Wideband Printed Dipole Antenna Array," *Electronics Letters*, Vol. 34, 1998, 2292–2293.

Kim, J. I., Lee, B. M., and Yoon, Y. J. "Wideband Printed Dipole Antenna for Multiple Wireless Services," *IEEE Radio and Wireless Conference*, 19–22 August 2001, pp. 153–156, Waltham, MA, USA.

Kiminami, K., Hirata, A., and Shiozawa, T. "Double-Sided Printed Bow-Tie Antenna for Uwb Communications," *IEEE Antenna Wireless Propagation Letters*, Vol. 3, 2004, 152–153, Salt Lake City, UT, USA.

Levine, E., Shtrikman, S., and Treves, D. "Double-Sided Printed Arrays With Large Bandwidth," *IEE Proceedings*, Part H, Vol. 135, No. 1, 1988, 54–59.

Lin, Y. D., and Tsai, S. N. "Coplanar Waveguide-Fed Uniplanar Bow-Tie Antenna," *Electronics Letters*, Vol. 45, 1997, 305–306, Provo, USA.

Mobile Radio Propagation and Modeling 10

10.1 Introduction

The design of spectrally efficient wireless communication system requires a detailed understanding of the radio propagation environment. The characteristics of the radio channel vary greatly with the operating frequency, and the mode of propagation, for example, line-of-sight (LoS) radio links, diffraction/scatter, and satellite links. The emphasis is on land mobile radio channels that are typical of terrestrial cellular mobile radio systems. A typical cellular radio system consists of a collection of fixed base stations (BSs) that define the radio coverage areas or cells. The height and placement of the BS antennas affects the proximity of local scatterers at the BS. A non-line-of-sight (NLoS) radio propagation path exists between a BS and mobile station (MS) due to natural and man-made objects between the BS and MS. As a consequence the radio waves must propagate via reflections, diffraction, and scattering. Plane waves arrive in different directions and delays at MS as shown in Figure 10.1. This property is called multipath propagation. The mobile radio channel is characterized by two types of fading, large-scale fading and small-scale fading. Large-scale fading is the slow variation of the mean signal power over time. This depends on the presence of obstacles in the signal path and in the position of mobile unit. The small-scale fading is a flat fading, that is, there is no inter-symbol interference.

10.2 Basics of mobile radio propagation

The carrier wavelength used in Ultra High Frequency (UHF) mobile radio applications typically range from 15 to 60 cm. Therefore, small changes in the differential propagation delays due to MS mobility will cause large changes in the phases of the individually arriving plane waves. Hence, the plane waves arriving at the MS and BS antennas will experience constructive and destructive addition depending on the location of the MS. If there is mobility in the MS or if there are changes in the scattering environment, the spatial variations in the amplitude and phase of the composite received signal will manifest themselves as time variations, a phenomenon called envelope fading. This depends on the velocity of MS.

Radio channels are reciprocal in the sense that, if a propagation path exists, it carries energy equally well in both directions. However, the spatial distribution of arriving plane waves may be significantly different in each direction. An MS in a typical macrocellular environment is usually surrounded by local scatterers and the plane waves will arrive from many directions without a direct LoS component. Two-dimensional isotropic scattering where the arriving plane waves

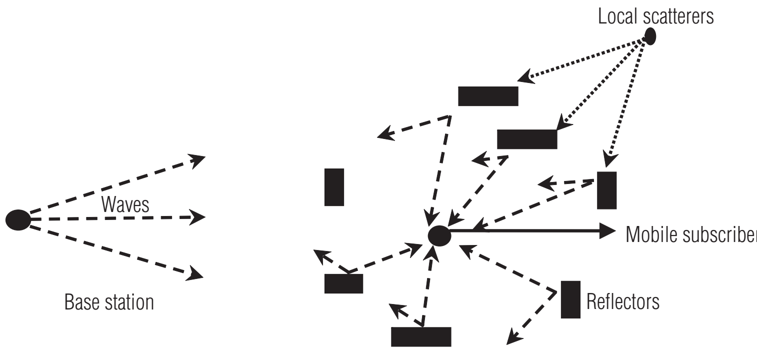


Figure 10.1 Typical macrocellular radio propagation environment

arrive in from all directions with equal probability is a very commonly used scattering model for the forward channel in a macrocellular system.

The BSs in macrocells are relatively free from local scatterers so that the plane waves tend to arrive from one direction with a fairly small angle of arrival. In a microcellular environment, the BS antennas are often placed below the skyline of buildings and are surrounded by local scatterers, such that the plane waves will arrive at the BS with a larger angle of arrival. Not all times do LoS paths exist between MS and BS. In such cases, often there exists a dominant reflected or diffracted path between the MS and the BS.

At frequencies below 1 GHz, antennas normally consist of a wire or wires of a suitable length coupled to the transmitter via a transmission line. At these frequencies it is relatively easy to design an assembly of wire radiators which form an array so as to beam the radiation in a particular direction. For distances large in comparison with the wavelength and the dimensions of the array, the field strength in free space decreases with an increase in distance. A plot of the field strength as a function of spatial angle is known as the radiation pattern of the antenna.

Antennas can be designed to have radiation patterns which are not omni directional, and it is convenient to have a figure of merit to quantify the ability of the antenna to concentrate the radiated energy in a particular direction. The directivity D of an antenna is defined as

$$D = \frac{\text{Power density at a distance in the direction of maximum radiation}}{\text{Mean power density at a distance } d} \tag{10.1}$$

This is a measure of the extent to which the power density in the direction of maximum radiation exceeds the average power density at the same distance. The directivity involves knowing the power actually transmitted by the antenna and this differs from the power supplied at the terminals by the losses in the antenna itself. From the system designer's viewpoint, it is more convenient to work in terms of terminal power. Power gain G is defined as

$$G = \frac{\text{Power density at a distance } d \text{ in the direction of maximum radiation}}{P_{\mathrm{T}} / 4\pi d^2} \tag{10.2}$$

where P_T is the power supplied to the antenna.

So, given P_T and G, it is possible to calculate the power density at any point in the far field that lies in the direction of maximum radiation. The power gain is unity for an isotropic antenna, that is, one which radiates uniformly in all directions, and an alternative definition of power gain is therefore the ratio of power density, from the specified antenna, at a given distance in the direction of maximum radiation, to the power density at the same point, from an isotropic antenna which radiates the same power.

10.3 Free-space propagation model

The power received by an antenna that is separated from a transmitter antenna by a distance d in free space is given by the Friis free-space equation

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} = P_0 \left(\frac{\lambda}{4\pi d} \right)^2 \tag{10.3}$$

where
- $P_r(d)$ = received power, which is a function of d
- P_t = transmitted power
- $P_0 = P_t G_t G_r$
- G_t = transmitter antenna gain
- G_r = receiver antenna gain
- d = transmitter to receiver separation in metres
- λ = wavelength in metres
- $EIRP = P_t G_t$

Free-space gain is defined as

$$G_{FS} = \left(\frac{\lambda}{4\pi d} \right)^2 = \frac{1}{PL_{FS}} \tag{10.4}$$

where PL_{FS} is the free-space path loss and usually expressed in decibels:

$$\begin{aligned} PL_{FS}(\mathrm{dB}) &= 20\log\left(\frac{4\pi d}{\lambda}\right)(\mathrm{dB}) \\ &= PL(d_0) + 20\log\left(\frac{d}{d_0}\right) \end{aligned} \tag{10.5}$$

where, d_0 is a reference distance for $PL(d_0)$.

10.3.1 Path-loss model in obstructed environments

In obstructed environments, the path loss is often modelled as

$$PL(d) = PL(d_0) + 10n\log\left(\frac{d}{d_0}\right) \tag{10.6}$$

Table 10.1 Path-loss exponents for various environments

Environment	Path-loss exponent, *n*
Free space	2
Urban area cellular radio	2.7–3.5
Shadowed urban cellular radio	3–5
In building LoS	1.6–1.8
Obstructed in building	4–6
Obstructed in factories	2–3

Where the first term is the free-space path loss at a reference distance d_0 and the path loss exponent n is empirically obtained by using a fitting curve from the measured data. Table 10.1 shows the typical path-loss exponents obtained in various mobile radio environments.

10.3.2 Log-normal shadowing

When measured path losses are plotted on a graph, the data points at a given distance show deviation due to the vastly different surrounding environment for each measurement. The measured path loss $PL(d)$ at a particular location is log-normally distributed (normal in decibels) about the mean distance path loss and is expressed as

$$PL(d) = \overline{PL}(d) + X_\sigma \tag{10.7}$$

where X_σ is a zero mean Gaussian distributed random variable (in decibels) with a standard deviation σ^2(in decibels).

Example problem 10.1

If 100 W is applied to a unit gain antenna with a 600 MHz carrier frequency, find the received power in dBm at a free-space distance of 200 m from the antenna. What is P_r(10 km)? Assume unity gain for the receiver antenna.

Solution

Given:

Transmitter power $P_t = 100$ W
Carrier frequency $f_c = 600$ MHz
Transmitter antenna gain $G_t = 1$
Receiver antenna gain $G_r = 1$
Transmitter to receiver separation in metres (d) = 200 m
The received power can be determined

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2}$$

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8 \text{m/s}}{600 \text{ MHz}} = 0.5 \text{ m}$$

$$P_r(d) = \frac{100(1)(1)(0.5)^2}{(4\pi)^2(200)^2} = 3.95 \times 10^{-3} \text{mW}$$

$$P_r(\text{dBm}) = 10\log P_r(\text{mW}) = 10\log(3.95 \times 10^{-3} \text{mW}) = -24.03 \text{ dBm}$$

The received power at 10 km can be expressed in terms of dBm using equation

$$P_r(d) = P_r(d_0) + 10n\log(\frac{d}{d_0})$$

where $d_0 = 200$ m and $d = 10$ km

$$P_r(10 \text{ km}) = P_r(200) + 20\log\left(\frac{200}{10{,}000}\right) = -24.03 \text{ dBm} - 33.979 \text{ dBm} = -58 \text{ dBm}$$

Example problem 10.2

Assume a receiver is located 20 km from a 100 W transmitter. The carrier frequency is 1,000 MHz, free-space propagation is assumed, $G_t = 1$, and $G_r = 3$, Find the power at the receiver.

Solution

Given:

Transmitter power $P_t = 100$W
Carrier frequency $f_c = 600$MHz
Transmitter antenna gain $G_t = 1$
Receiver antenna gain $G_r = 3$
Transmitter to receiver separation in metres (d) = 20 km
The received power can be determined

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8 \text{m/s}}{1{,}000 \text{ MHz}} = 0.3 \text{ m}$$

$$P_r(\text{dBm}) = 10\log\frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2} = 10\log(\frac{100 \times 1 \times 3 \times (0.3)^2}{(4\pi)^2 \times (20{,}000)^2}) = -93.69 \text{ dBm}$$

10.3.3 Direct propagation

When there exists a visible straight path between the BS antenna and the mobile terminal antenna and no other multipath component exists as shown in Figure 10.2, the received signal power attenuation conforms to free-space propagation. This direct signal component is called

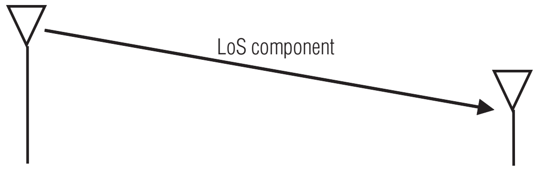


Figure 10.2 Direct propagation

as a LoS component. If the received signal has only the LoS component, it does not experience multipath fading. However, there always exist multipath components other than the LoS component between the transmitter and the receiver in real terrestrial communication systems. Under real propagation conditions, various terrestrial objects such as mountains, buildings, and hills frequently block the LoS component.

10.3.4 Refraction

A radio wave transmitted into ionized layers is always bent. This bending of radio waves is called *refraction*. Consider the radio wave, shown in Figure 10.3, travelling through the earth's atmosphere at a constant speed. As the wave enters the denser layer of charged ions, its upper portion moves faster than its lower portion. The abrupt speed increase in the upper part of the wave causes it to bend back towards the earth. This bending is always towards the propagation medium where the radio wave's velocity is the least.

The amount of refraction a radio wave undergoes depends on three main factors:

1. The ionization density of the layer
2. The frequency of the radio wave
3. The angle at which the radio wave enters the layer

Layer density

Figure 10.4 shows the relationship between radio waves and ionization density. Each ionized layer has a middle region of relatively dense ionization with less intensity above and below. As a radio wave enters a region of increasing ionization, a velocity increase causes it to bend back towards the earth. In the highly dense middle region, refraction occurs more slowly because the

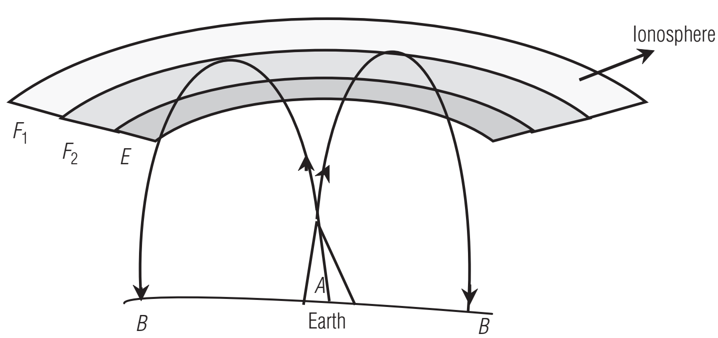


Figure 10.3 Radio wave reflection

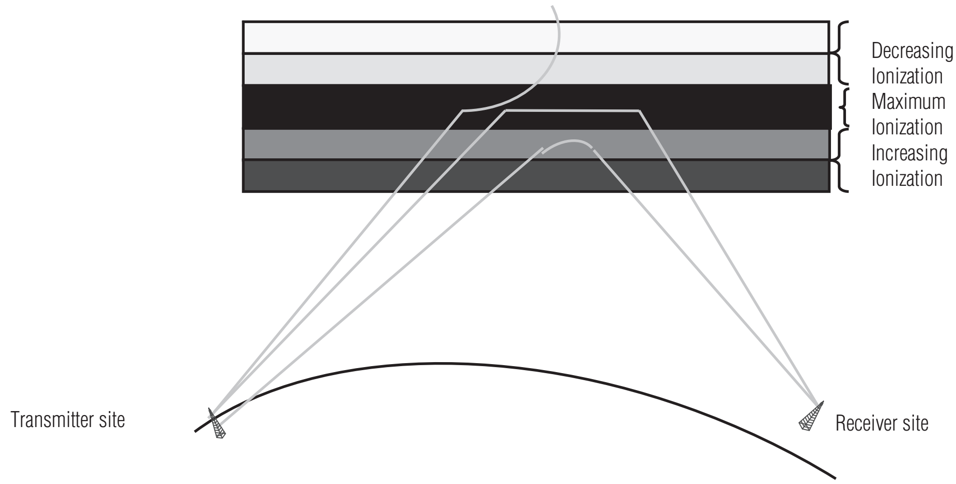


Figure 10.4 Effects of ionospheric density on radio waves

ionization density is uniform. As the wave enters the upper less dense region, the velocity of the upper part of the wave decreases and the wave is bent away from the earth.

Frequency

The lower the frequency of a radio wave, the more rapidly the wave is refracted by a given degree of ionization. Figure 10.5 shows three separate waves of differing frequencies entering the ionosphere at the same angle. You can see that the 5-MHz wave is refracted quite sharply while the 20-MHz wave is refracted less sharply and returns to earth at a greater distance than the 5-MHz wave.

Notice that the 100-MHz wave is lost into space. For any given ionized layer, there is a frequency, called the *escape point,* at which energy transmitted directly upward will escape into space. The maximum frequency just below the escape point is called the critical frequency. In this example, the 100-MHz wave's frequency is greater than the critical frequency for that ionized layer.

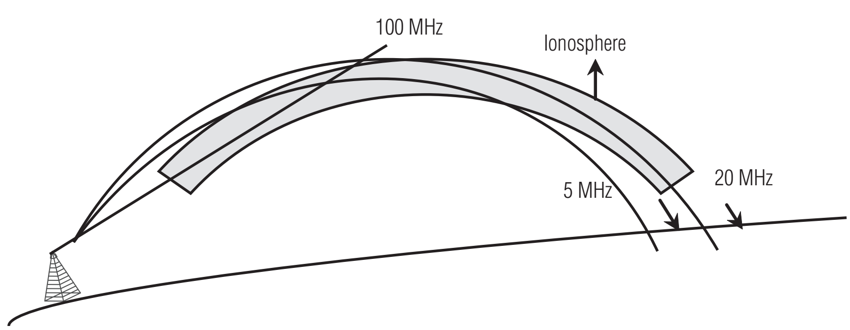


Figure 10.5 Frequency versus refraction and distance

The critical frequency of a layer depends upon the layer's density. If a wave passes through a particular layer, it may still be refracted by a higher layer if its frequency is lower than the higher layer's critical frequency.

Angle of incidence and critical angle

When a radio wave encounters a layer of the ionosphere, that wave is returned to earth at the same angle (nearly) as its *angle of incidence*. The angle of incidence beyond which the wave escapes into space is called critical angle. The critical angle for radio waves depends on the layer density and the wavelength of the signal. Figure 10.6 shows three radio waves of the same frequency entering a layer at different incidence angles. In Figure 10.6, the angle of incidence of wave B is shown as critical angle. We can observe that wave A, whose angle of incidence is more than critical angle is lost into space and wave C, whose angle of incidence is less than critical angle is refracted to earth.

As the frequency of a radio wave is increased, the critical angle must be reduced for refraction to occur. Notice in Figure 10.7 that the 2-MHz wave strikes the ionosphere at the critical angle for that frequency and is refracted. Although the 5-MHz line (broken line) strikes the ionosphere at a less critical angle, it still penetrates the layer and is lost as the angle is lowered, a critical angle is finally reached for the 5-MHz wave and it is refracted back to earth.

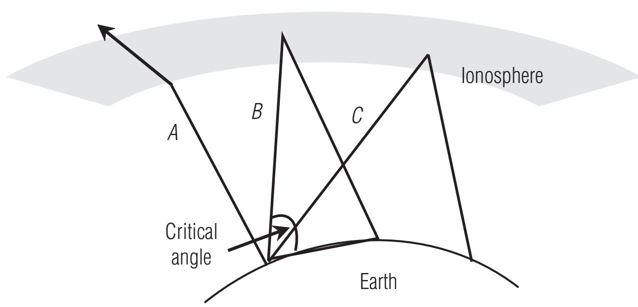


Figure 10.6 Incidence angles of radio waves

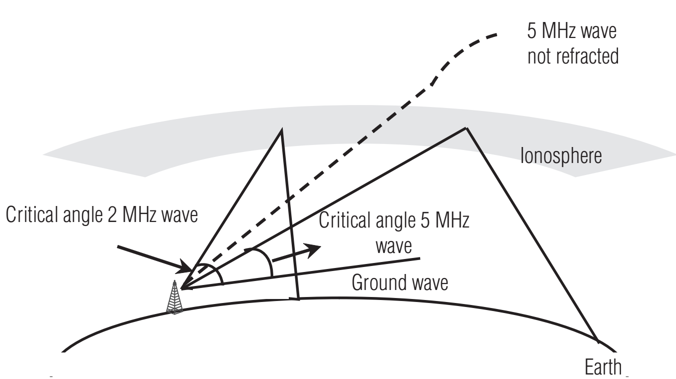


Figure 10.7 Effect of frequency on the critical angle

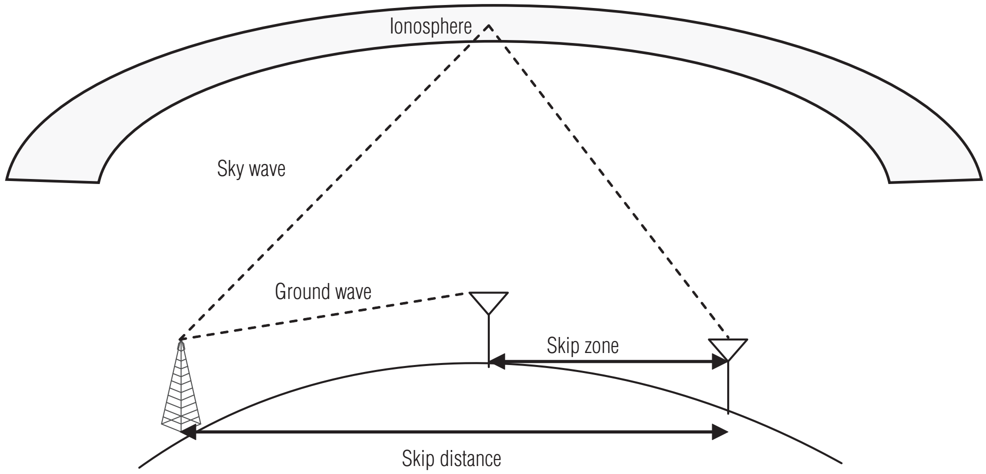


Figure 10.8 Relationship among skip zone, skip distance, and ground wave

10.3.5 Skip distance and zone

Recall from the previous study that a transmitted radio wave separates into two parts, the sky wave and the ground wave. With those two components in mind, we will now briefly discuss *skip distance* and *skip zone*.

Skip distance

Look at the relationship among the sky wave skip distance, skip zone, and ground wave coverage as shown in Figure 10.8. The *skip distance* is the distance from the transmitter to the point where the sky wave first returns to the earth. The skip distance depends on the wave's frequency and angle of incidence, and the degree of ionization.

Skip zone

The *skip zone* is a zone of silence between the point where the ground wave is too weak for reception and the point where the sky wave is first returned to earth. The outer limit of the skip zone varies considerably, depending on the operating frequency, the time of day, the season of the year, sunspot activity, and the direction of transmission. At very-low, low, and medium frequencies, a skip zone is never present.

However, in the high frequency spectrum, a skip zone is often present. As the operating frequency is increased, the skip zone widens to a point where the outer limit of the skip zone might be several thousand miles away. At frequencies above a certain maximum, the outer limit of the skip zone disappears completely, and no F-layer propagation is possible. Occasionally, the first sky wave will return to earth within the range of the ground wave. In this case, severe fading can result from the phase difference between the two waves (the sky wave has a longer path to follow).

10.3.6 Reflection

Reflection occurs when radio waves are "bounced" from a flat surface. There are basically two types of reflection that occur in the atmosphere: earth reflection and ionospheric reflection. Figure 10.9 shows two waves reflected from the earth's surface.

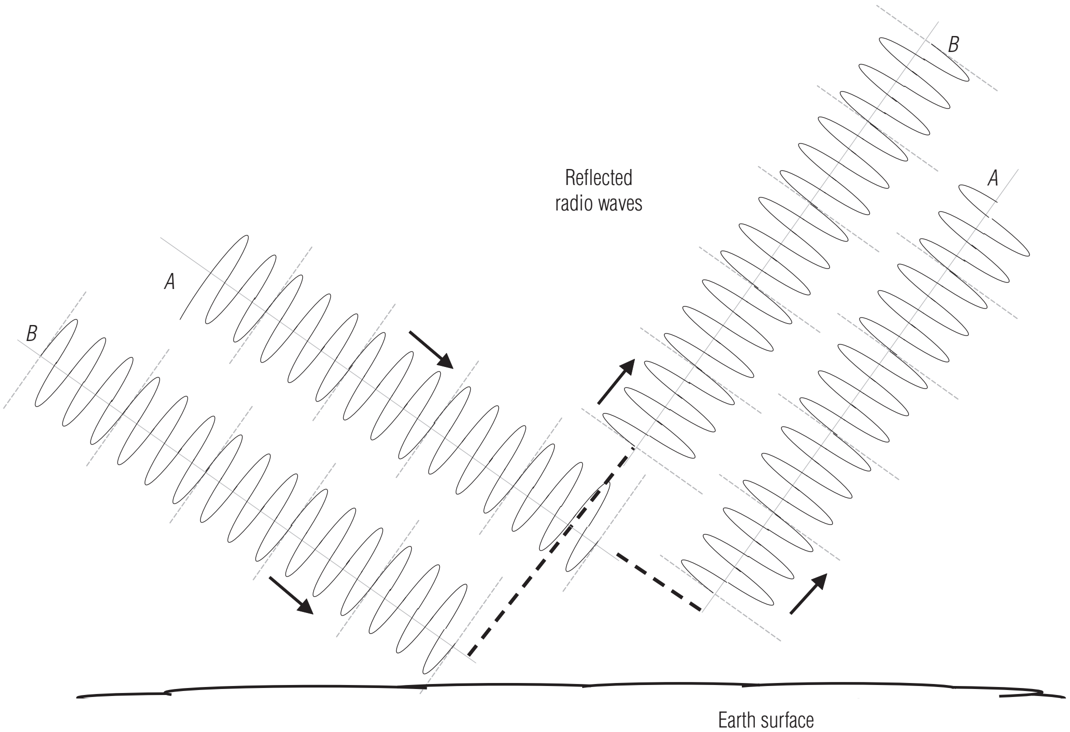


Figure 10.9 Phase shift of reflected radio waves

Waves A and B bounce off the earth's surface like light off of a mirror. Notice that the positive and negative alternations of radio waves A and B are in phase before they strike the earth's surface. However, after reflection the radio waves are approximately 180° out of phase. The amount of phase shift that occurs is not constant. It varies, depending on the wave polarization and the angle at which the wave strikes the surface.

Normally, radio waves reflected in phase produce stronger signals, while those reflected out of phase produce a weak or fading signal. Ionospheric reflection occurs when certain radio waves strike a thin, highly ionized layer in the ionosphere. Although the radio waves are actually refracted, some may be bent back so rapidly that they appear to be reflected. For ionospheric reflection to occur, the highly ionized layer can be approximately no thicker than one wavelength of the wave. Since the ionized layers are often several miles thick, ionospheric reflection mostly occurs at long wavelengths (low frequencies).

The reflected path between BS antenna and the mobile terminal antenna is shown in Figure 10.10.

10.3.7 Diffraction

Diffraction is the ability of radio waves to turn sharp corners and bend around obstacles. Figure 10.11 shows the diffraction results in a change of direction of part of the radio-wave energy around the edges of an obstacle. Radio waves with long wavelengths compared to the diameter of an obstruction are easily propagated around the obstruction.

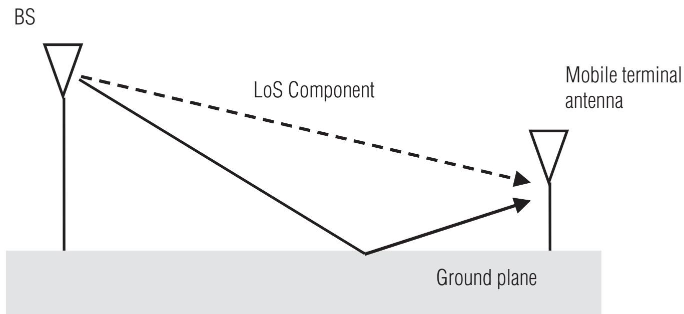


Figure 10.10 Reflection

However, as the wavelength decreases, the obstruction causes more and more attenuation, until at very-high frequencies a definite *shadow zone* develops. The shadow zone is basically a blank area on the opposite side of an obstruction in LoS from the transmitter to the receiver as shown in Figure 10.11(b). Diffraction can extend the radio range beyond the horizon. By using high power and low-frequencies, radio waves can be made to encircle the earth by diffraction. Direct propagation is shown in Figure 10.12.

10.3.8 Scattering

Scattering occurs when the wave travels through or reflected from an object with dimensions smaller than the wavelength. If the surface of the scattering object is random, the signal energy

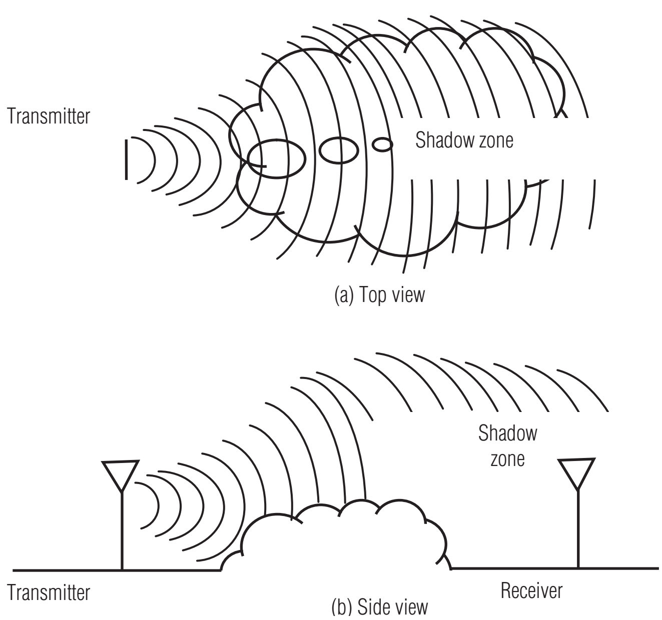


Figure 10.11 Diffraction around an object

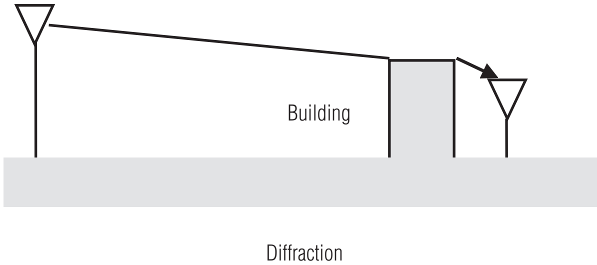


Figure 10.12 Direct propagation

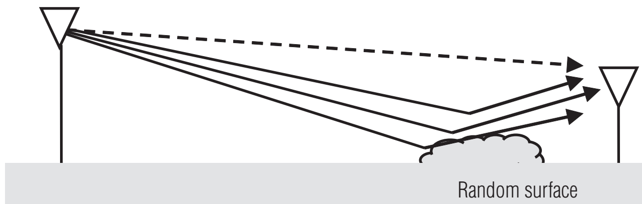


Figure 10.13 Scattering

is scattered in many directions as shown in Figure 10.13. Rough surfaces, small objects, or other irregularities in the channel cause scattering.

10.4 Link budget design

A calculation of signal powers, noise powers, and/or signal-to-noise ratios for a complete communication link is a link budget, and it is a useful approach to the basic design of a complete communication system. Such calculations are usually fairly simple, but they can give very revealing information as to the system performance, provided appropriately accurate assumptions are made in calculating the individual elements of the link budget.

The maximum acceptable path loss is usually split into two components:

1. Distance-dependent path-loss model
2. Fade margin, which is included to allow the system some resilience against the practical effects of signal fading beyond the value predicted by the model.

Thus,

Maximum acceptable propagation loss (dB) = Predicted loss + Fade margin

The greater the fade margin, the greater the reliability and quality of the system, but this will constrain the maximum system range. The path loss relates the transmit and received powers assuming no transmitter/receiver loss or gain (between isotropic antennas), that is,

$$L(\text{dB}) = 10\log(P_t/p_r) \tag{10.8}$$

The path loss can be split into the sum of the free-space loss and the so-called excess loss,

$$L(\text{dB}) = L_{\text{fs}} + L_{\text{excess}} \tag{10.9}$$

where the excess loss is given by,

$$L_{\text{excess}}(\text{dB}) = 20\log(e_0/e) \tag{10.10}$$

where e_0 is the field strength at the received antenna under free-space conditions and e is the actual field strength both in linear units (V/m).

So far we have assumed that the link does not contain gains or losses other than the path loss. If the gains and losses at both ends of the link are taken into account, the received power is given by

$$p_{\text{r}} = \frac{p_{\text{t}}g_{\text{t}}g_{\text{r}}}{l_{\text{t}}ll_{\text{r}}} \tag{10.11}$$

given in linear units or, in logarithmic units,

$$P_r(\text{dBW or dBm}) = P_t + G_t + G_r - L_t - L - L_r \tag{10.12}$$

where l_t, or alternatively, L_t when using dB, is the loss at the transmit side, for example, cables, etc., and g_t (G_t) is the transmit antenna gain.

A frequently used parameter for describing the radiated power is the EIRP (equivalent isotropic radiated power) defined as

$$\text{EIRP} = p_{\text{t}}g_{\text{t}}/l_{\text{t}} \tag{10.13}$$

in linear units or, in dB,

$$\text{EIRP} = P_{\text{t}} + G_{\text{t}} - L_{\text{t}} \tag{10.14}$$

Link budgets are usually computed in terms of the received power.

10.5 Propagation models

A propagation model, is an empirical mathematical formulation for the characterization of radio wave propagation as a function of frequency, distance, and other conditions. A single model is usually developed to predict the behavior of propagation for all similar links under similar constraints. The following are two types of propagation models:

1. Outdoor propagation models
2. Indoor propagation models

10.5.1 Outdoor propagation models

There are many empirical outdoor propagation models such as Longley–Rice model, Durkin's model, Okumura model, Hata model, and so on. Longley–Rice model is the most commonly used model within a frequency band of 40 MHz to 100 GHz over different terrains. Certain modifications over the rudimentary model like an extra urban factor (UF) due to urban clutter near the receiver are also included in this model. Below, we discuss some of the outdoor models, followed by a few indoor models too.

10.5.1.1 Okumura model

The Okumura model is a radio propagation model used in urban areas for signal prediction. The frequency coverage of this model is in the range 200–1,900 MHz and at distances of 1–100 km. It can be applicable for BS effective antenna heights (h_t) ranging from 30 to 1,000 m. Okumura used extensive measurements of base station-to-mobile signal attenuation throughout Tokyo to develop a set of curves giving median attenuation relative to free space (A_{mu}) of signal propagation in irregular terrain. The empirical pathloss formula of Okumura at distance *d* parameterized by the carrier frequency f_c is given by

$$P_L(d)\,\mathrm{dB} = L(f_c, d) + A_{mu}(f_c, d) - G(h_t) - G(h_r) - G_{AREA} \tag{10.15}$$

where $L(f_c, d)$ is free-space path loss at distance *d* and carrier frequency f_c, $A_{mu}(f_c, d)$ is the median attenuation in addition to free-space path loss across all environments, $G(h_t)$ is the BS antenna height gain factor, $G(h_r)$ is the mobile antenna height gain factor, G_{AREA} is the gain due to type of environment. The values of $A_{mu}(f_c, d)$ and G_{AREA} are obtained from Okumura's empirical plots. Okumura-derived empirical formulas for $G(h_t)$ and $G(h_r)$ are as follows:

$$G(h_t) = 20\log_{10}(h_t/200), \quad 30 < h_t < 1{,}000 \tag{10.16}$$

$$G(h_r) = 10\log_{10}(h_r/3), \quad h_r \le 3\text{ m} \tag{10.17}$$

$$G(h_r) = 20\log_{10}(h_r/3), \quad 3 < h_r < 10\text{ m} \tag{10.18}$$

Correlation factors related to terrain are also developed in order to improve the models accuracy. Okumura's model has a 10–14 dB empirical standard deviation between the path loss predicted by the model and the path loss associated with one of the measurements used to develop the model.

10.5.1.2 Hata model

The Hata model is an empirical formulation of the graphical path-loss data provided by Okumura, and is valid from 150 to 1,500 MHz. The standard Hata formula for median path loss in urban area is given by

$$PL_{50}(\text{urban})(\text{dB}) = 69.55 + 29.16\log f_c - 13.82\log h_{te} - a(h_{re}) + (44.9 - 6.55\log h_{te})\log d \tag{10.19}$$

where

f_c = Frequency in (MHz) from 150 to 1,500 MHz
h_{te} = Effective transmitter (base station) antenna height (in metres) ranging from 30 to 200 m
h_{re} = Effective receiver (mobile) antenna height (in metres) ranging from 1 to 10 m
d = Tx–Rx separation distance (in kilometres)
$a(h_{re})$ = Correction factor for effective mobile antenna height

For a small-to-medium sized city, the mobile antenna correction factor is given by

$$a(h_{re}) = (1.1\log f_c - 0.7)h_{re} - (1.56\log f_c - 0.8) \text{ dB} \tag{10.20}$$

and for a large city, it is given by

$$a(h_{re}) = 8.29(\log 1.54h_{re})^2 - 1.1 \text{ dB} \quad \text{if } 150 \le f_c \le 200 \text{ MHz} \tag{10.21}$$

$$a(h_{re}) = 3.2(\log 11.75h_{re})^2 - 4.97 \text{ dB} \quad \text{if } 200 \le f_c \le 1500 \text{ MHz} \tag{10.22}$$

For the path loss in a suburban area, the standard Hata formula is modified as

$$PL_{50}(\text{dB}) = PL_{50}(\text{Urban}) - 2\left[\log(f_c/28)\right]^2 - 5.4 \tag{10.23}$$

and for path loss in open rural areas, the formula is modified as

$$PL_{50}(\text{dB}) = PL_{50}(\text{Urban}) - 4.78(\log f_c)^2 - 18.33\log f_c - 40.98 \tag{10.24}$$

10.5.1.3 PCS extension to Hata model

The European Co-operative for Scientific and Technical research (EURO-COST) organized the COST-231 working committee to develop an extension version of Hata model to cover the PCS bands, forming a Hata model version that is valid to 2,000 MHz [Eur91]:

$$PL_{50}(\text{Urban})(\text{dB}) = 46.3 + 33.9\log f_c - 13.82\log h_{re} - a(h_{re}) \tag{10.25}$$

where, f_c = 1,500–2,000 MHz
h_{te} = 30–200 m
h_{re} = 1–10 m
d = 1–20 km
C_M = 0 dB for medium-sized city and suburban areas
3 dB for metropolitan centers

Example problem 10.3

If the heights of transmitting and receiving antennas are 40m and 3m respectively over a distance of 15 km in a dense urban mobile environment determine the propagation path loss for a radio signal at 900 MHz using Hata propagation path-loss model. If the free-space propagation path loss is 110.5 dB for the given system parameters, how is Hata propagation path loss comparable with that of free-space propagation path loss?

Solution

Frequency of transmission, f_c = 900 MHz (given)
Transmitting antenna height, h_{te} = 40 m (given)
Receiving antenna height, h_{re} = 3 m (given)
Distance between T_x and R_x, d = 15 km (given)
To find propagation path loss, PL (dB) using Hata model
The propagation path loss using Hata propagation path-loss model is given by
PL (dB) = 69.55 + 29.16 log f_c – 13.82 log h_{te} + (44.9 – 6.55 log h_{te}) log d
PL (dB) = 69.55 + 29.16 log 900 – 13.82 log 40 + (44.9 – 6.55 log 40) log 15 = 174.020 dB
To comparison PL (dB) with free-space propagation path loss
Free-space propagation path loss, L_f (dB) = 110.5 dB (given)
Difference between two path loss values = 174.05 – 110.5 = 63.520 dB

10.5.2 Indoor propagation models

The indoor radio channel differs from the traditional mobile radio channel in different ways – the distances covered are much smaller, and the variability of the environment is much greater for smaller range of Tx–Rx separation distances. Features such as lay-out of the building, the construction materials, and the building type strongly influence the propagation within the building. Indoor radio propagation is dominated by the same mechanisms as outdoor: reflection, diffraction, and scattering with variable conditions. In general, indoor channels may be classified as either LoS or obstructed.

Partition losses inside a floor (Intra-floor)

The internal and external structure of a building formed by partitions and obstacles vary widely. Partitions that are formed as a part of building structure are called hard partitions , and partitions that can be moved and which do not span to the ceiling are called soft partitions. Partitions vary widely in their physical and electrical characteristics, making it difficult to apply general models to specific indoor installations.

Partition losses between floors (Inter-floor)

The losses between floors of a building are determined by the external dimensions and materials of the building, as well as the type of construction used to create the floors and the external surroundings. Even the number of windows in a building and the presence of tinting can impact the loss between floors.

Log-distance path-loss model

It has been observed that indoor path loss obeys the distance power law given by

$$PL(\text{dB}) = PL(d_0) + 10n\log_{10}(d/d_0) + X_\sigma \tag{10.26}$$

where n is path loss exponent depends on the building and surrounding type, and X_σ represents a normal random variable in dB having standard deviation of σ dB.

10.5.3 Multipath propagation

In wireless telecommunications, multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. Causes of multipath include atmospheric ducting, ionospheric reflection and refraction, and reflection from water bodies and terrestrial objects such as mountains and buildings. The effects of multipath include constructive and destructive interference, and phase shifting of the signal. In digital radio communications (such as GSM) multipath can cause errors and affect the quality of communications.

10.5.4 Small-scale fading and multipath

Multipath signals are received in a terrestrial environment, that is, where different forms of propagation are present and the signals arrive at the receiver from transmitter via a variety of paths. Therefore, there would be multipath interference, causing multipath fading. Adding the effect of movement of either Tx or Rx or the surrounding clutter to it, the received overall signal amplitude or phase changes over a small amount of time. Mainly this causes the fading.

Fading

The term fading, or, small-scale fading, means rapid fluctuations of the amplitudes, phases, or multipath delays of a radio signal over a short period or short travel distance. This might be so severe that large scale radio propagation loss effects might be ignored.

Multipath fading effects

In principle, the following are the main multipath effects:

1. Rapid changes in signal strength over a small travel distance or time interval.
2. Random frequency modulation due to varying Doppler shifts on different multipath signals.
3. Time dispersion or echoes caused by multipath propagation delays.

Factors influencing fading

The following physical factors influence small-scale fading in the radio propagation channel:

- ***Multipath propagation***: Multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. The effects of multipath include constructive and destructive interference, and phase shifting of the signal.
- ***Speed of the mobile***: The relative motion between the BS and the mobile results in random frequency modulation due to different Doppler shifts on each of the multipath components.
- ***Speed of surrounding objects***: If objects in the radio channel are in motion, they induce a time varying Doppler shift on multipath components. If the surrounding objects move at a greater rate than the mobile, then this effect dominates fading.
- ***Transmission bandwidth of the signal***: If the transmitted radio signal bandwidth is greater than the bandwidth of the multipath channel (quantified by coherence bandwidth), the received signal will be distorted.

10.5.5 Small-scale multipath measurements

A wideband pulsed bistatic radar usually transmits a repetitive pulse of width T_{bb} s, and uses a receiver with a wide bandpass filter (BW = $2/T_{bb}$ Hz). The signal is then amplified, envelope detected, and displayed and stored on a high-speed oscilloscope. Immediate measurements of the square of the channel impulse response convolved with the probing pulse can be taken. If the oscilloscope is set on averaging mode, then this system provides a local average power delay profile. The system set-up is shown in Figure 10.14.

This system is subject to interference noise. If the first arriving signal is blocked or fades, severe fading occurs, and it is possible that the system may not trigger properly.

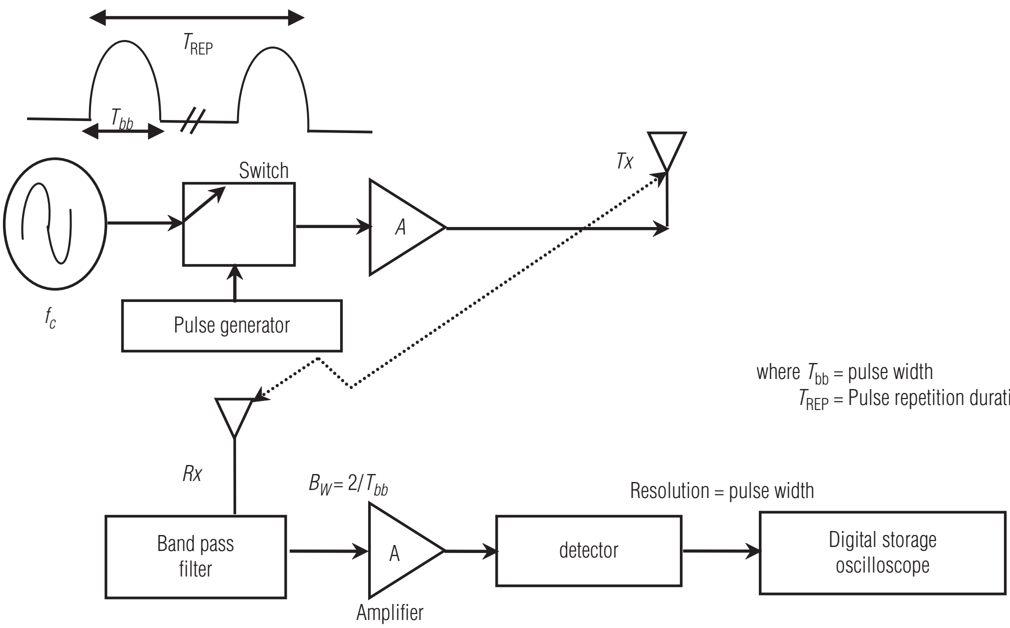


Figure 10.14 Direct RF pulsed channel IR measurement

Frequency domain channel sounding

In this case we measure the channel in the frequency domain and then convert it into time domain impulse response by taking its inverse discrete Fourier transform (IDFT). A vector network analyser controls a swept frequency synthesizer. An S-parameter test set is used to monitor the frequency response of the channel. The sweeper scans a particular frequency band, centred on the carrier, by stepping through discrete frequencies. The number and spacing of the frequency step impacts the time resolution of the impulse response measurement. For each frequency step, the S-parameter test set transmits a known signal level at port 1 and monitors the received signal at port 2. These signals allow the analyser to measure the complex response, $S_{21}(\omega)$, of the channel over the measured frequency range. The $S_{21}(\omega)$ measure is the measure of the signal flow from transmitter antenna to receiver antenna (i.e., the channel). This system is suitable only for indoor-channel measurements. This system is also non-real-time. Hence, it is not suitable for time-varying channels unless the sweep times are fast enough. Frequency domain channel IR measurement is shown in Figure 10.15.

10.5.6 Multipath channel parameters

To compare the different multipath channels and to quantify them, some parameters are defined. They all can be determined from the power delay profile. These parameters can be broadly divided into two types.

Time-dispersion parameters

These parameters include the mean excess delay, rms delay spread, and excess delay spread. The mean excess delay is the first moment of the power delay profile and is defined as

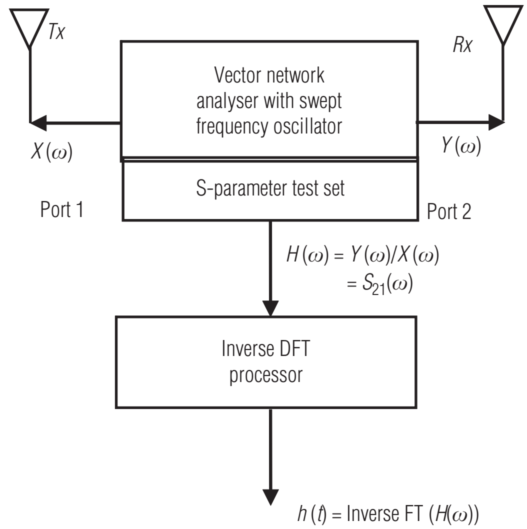


Figure 10.15 Frequency domain channel IR measurement

$$\tau = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} = \frac{\sum P(\tau_k)\tau_k}{\sum P(\tau_k)} \tag{10.27}$$

where a_k is the amplitude, τ_k is the excess delay, and $P(\tau_k)$ is the power of the individual multipath signals.

The mean square excess delay spread is defined as

$$\tau^2 = \frac{\sum_k P(\tau_k)\tau_k^2}{\sum_k P(\tau_k)} \tag{10.28}$$

Since the rms delay spread is the square root of the second central moment of the power delay profile, it can be written as

$$\sigma_\tau = \sqrt{\bar{\tau}^2 - (\bar{\tau})^2} \tag{10.29}$$

As a rule of thumb, for a channel to be at fading, the following condition must be satisfied:

$$\frac{\sigma_\tau}{T_S} \leq 0.1 \tag{10. 30}$$

where T_S is the symbol duration. For this case, no equalizer is required at the receiver.

Frequency dispersion parameters

To characterize the channel in the frequency domain, the following two parameters are necessary

1. ***Coherence bandwidth***: This is a statistical measure of the range of frequencies over which the channel can be considered to pass all the frequency components with almost equal gain and linear phase. Such a channel is said to be flat.

Practically, coherence bandwidth is the minimum separation over which the two frequency components are affected differently. If the coherence bandwidth is considered to be the bandwidth over which the frequency correlation function is above 0.9, then it is approximated as.

$$B_C \approx \frac{1}{50\,\sigma_\tau} \tag{10.31}$$

where σ_τ is the rms delay spread

However, if the coherence bandwidth is considered to be the bandwidth over which the frequency correlation function is above 0.5, then it is defined as

$$B_C \approx \frac{1}{5\sigma_\tau} \tag{10.32}$$

The coherence bandwidth describes the time dispersive nature of the channel in the local area. A more convenient parameter to study the time variation of the channel is the coherence time. This variation may be due to the relative motion between the mobile and the BS or the motion of the objects in the channel.

2. ***Coherence time***: This is a statistical measure of the time duration over which the channel impulse response is almost invariant. When channel behaves like this, it is said to be slow faded. Essentially it is the minimum time duration over which two received signals are affected differently. For an example, if the coherence time is considered to be the bandwidth over which the time correlation is above 0.5, then it can be approximated as

$$T_C \approx \frac{9}{16\pi f_m} \tag{10.33}$$

where f_m is the maximum Doppler spread given by $f_m = \frac{v}{\lambda}$.

Example problem 10.4

Assuming the speed of a vehicle to be equal to 60 mph (88 ft/s), carrier frequency, f_c = 860 MHz, and rms delay spread τ_d = 2 μs, calculate coherence time and coherence bandwidth. At a coded symbol rate of 19.2 Kbps what kind of symbol distortion will be experienced? What type of fading will be experienced by the channel?

Solution

v = 60 mph (=88 ft/s)

$$\lambda = \frac{c}{f} = \frac{9.84 \times 10^8}{860 \times 10^6} = 1.1442 \text{ ft}$$

$$\text{Maximum Doppler shift} = \frac{v}{\lambda} = \frac{88}{1.1442} = 77 \text{ Hz}$$

$$T_c = \frac{1}{2\pi f_m} = \frac{1}{2\pi \times 77} = 0.0021 \text{ s}$$

$$T_s = \frac{10^6}{19{,}200} = 52 \ \mu\text{s}$$

The symbol interval is much smaller compared to the channel coherence time. Symbol distortion is, therefore, minimal. In this case fading is slow.

$$\text{Coherence bandwidth} = B_c \approx \frac{1}{2\pi\tau_d} = \frac{1}{2\pi \times 2 \times 10^{-6}} = 79.56 \text{ kHz}$$

This shows that the channel is a wide band system in this multipath situation and experiences selective fading only over 6.5 per cent (79.56/1228.8 = 0.0648 ~ 6.5 per cent) of its bandwidth (B_w = 1228.8 kHz).

10.6 Types of small-scale fading

The type of fading experienced by the signal through a mobile channel depends on the relation between the signal parameters (bandwidth, symbol period) and the channel parameters (rms delay spread and Doppler spread). Hence, we have four different types of fading. There are two types of fading due to the time dispersive nature of the channel.

10.6.1 Fading effects due to multipath time delay spread

Flat fading

Such type of fading occurs when the bandwidth of the transmitted signal is less than the coherence bandwidth of the channel. Equivalently, if the symbol period of the signal is more than the rms delay spread of the channel, then the fading is at fading.

So we can say that at fading occurs when

$$B_S << B_C$$

where B_S is the signal bandwidth and B_C is the coherence bandwidth.
Also,

$$T_S >> \sigma_\tau$$

where T_S is the symbol period and σ_τ is the rms delay spread. In such a case, mobile channel has a constant gain and linear phase response over its bandwidth.

Frequency selective fading

Frequency selective fading occurs when the signal bandwidth is more than the coherence bandwidth of the mobile radio channel or equivalently, the symbols duration of the signal is less than the rms delay spread.

$$B_S >> B_C$$

and

$$T_S >> \sigma_\tau$$

At the receiver, we obtain multiple copies of the transmitted signal, all attenuated and delayed in time. The channel introduces inter-symbol interference. A rule of thumb for a channel to have at fading is if

$$\frac{\sigma_\tau}{T_S} \leq 0.1$$

Frequency-selective fading affects unequally the different spectral components of a radio signal. Selective fading is usually significant only relative to the bandwidth of the overall wireless communication channel. If the signal attenuates over a portion of the bandwidth of the signal, the fading is considered to be selective in frequency domain. Frequency selective fading on the received signal occurs when a radio channel has a constant gain and linear phase response, but the channel bandwidth is less than that of the transmitted signal. Under such conditions, the channel impulse response has a multipath delay spread which is greater than the reciprocal bandwidth of the transmitted signal.

The received signal includes multiple versions of the transmitted signal which are faded and delayed in time, and hence the received signal is distorted. Frequency selective fading is due to time dispersion of the transmitted symbols within the channel, and the channel induces inter-symbol interference. Because this effect varies by frequency, fading is different at different frequencies and it is extremely difficult to counter its impact or compensate for the signal loss. Frequency-selective fading channels are also known as wideband channels since the bandwidth of the transmitted signal is wider than the bandwidth of the channel impulse response.

Let us consider a mobile receiver moves directly away from the transmitting antenna but towards a reflecting surface. If the two signals are in phase at a given point, they will add. As the mobile moves forward a distance of $\lambda_c/4$, the direct path is increased and the reflected path is reduced by the same amount, resulting in a total phase shift of 180°, resulting into partial cancellation of the signal (which means the signal may fade up to 50 dB in worst cases). When the mobile moves another distance of $\lambda_c/4$, the signals are once again in phase. Thus, the fades occur each time the mobile moves a distance of $\lambda_c/2$. Given the frequency of the signal and the speed of the mobile, it is easy to estimate the time between fades. The time between fades is given by

$$T_f = (\lambda_c/2)/V_m = \lambda_c/(2V_m) = c/(2\, f_c\, V_m)$$

where V_m is the speed of the mobile, T_f is time between fades, and f_c is cell phone operating frequency.

Example problem 10.5

A mobile subscriber travels at a uniform speed of 100 km/h. Compute the time between fades if the mobile uses

(a) a cellphone operating at 900 MHz
(b) a PCS phone operating at 1,800 MHz
(c) and comment on the results obtained.

Solution

Speed of the mobile, V_m = 100 km/h (given) = 100×10^3 m/3,600 s = 27.7 m/s
Time between fades is given by the expression, $T_f = c/(2\, f_c\, V_m)$
where c is the speed of radio waves = 3×10^8 m/s

(a) To compute time between fades for a mobile operating at 900 MHz
Frequency of operation, f_c = 900 MHz or 900×10^6 Hz (given)
Time between fades, $T_f = (3 \times 10^8$ m/s / $(2 \times 900 \times 10^6$ Hz $\times$ 27.7m/s)
Hence, time between fades at 900 MHz = **6 ms**

(b) To compute time between fades for a mobile operating at 1,800 MHz
Frequency of operation, f_c = 1,800 MHz or $1{,}800 \times 10^6$ Hz (given)
Time between fades, $T_f = (3 \times 10^8$ m/s / $(2 \times 1{,}800 \times 10^6$ Hz $\times$ 27.7 m/s)
Hence, time between fades at 1,800 MHz = **3 ms**
It is observed that the rapidity of the fading increases with the frequency of the transmission at the same speed of the mobile vehicle.

10.6.2 Fading effects due to Doppler spread

The fading effects due to Doppler spread can also be classified as fast fading and slow fading. Depending on how rapidly the transmitted baseband signal changes as compared to the rate of change of the channel, a wireless channel is classified as a fast-fading or slow-fading channel. In fast-fading conditions, as the mobile unit moves down a street in an urban environment, rapid fluctuations in received signal strength occur over distances of about one-half a wavelength. The channel impulse response changes rapidly within the symbol duration. That is, the coherence time of the channel is smaller than the symbol period of the transmitted signal. This causes frequency dispersion, also called time-selective fading, due to Doppler spreading. This results into signal distortion which increases with increasing Doppler spread relative to the bandwidth of the transmitted signal. Therefore, a signal undergoes fast fading if the symbol period $T_s > T_c$, the time between fades. Fast fading occurs when the rate of change of the channel characteristics is faster than the rate of change of the information data signal, and results in distortion.

Fast fading

In a fast-fading channel, the channel impulse response changes rapidly within the symbol duration of the signal. Due to Doppler spreading, signal undergoes frequency dispersion leading to distortion.

Therefore, a signal undergoes fast fading if

$$T_S >> T_C$$

where T_C is the coherence time and

$$B_S >> B_D$$

where B_D is the Doppler spread. Transmission involving very low data rates suffers from fast fading.

Slow fading

In such a channel, the rate of the change of the channel impulse response is much less than the transmitted signal. We can consider a slow-faded channel as in which is almost constant over atleast one symbol duration. Hence,

$$T_S << T_C$$

$$B_S >> B_D$$

We observe that the velocity of the user plays an important role in deciding whether the signal experiences fast or slow fading.

10.6.3 Impulse response model of a multipath channel

Mobile radio channel may be modelled as a linear filter with time-varying impulse response in continuous time. To show this, consider time variation due to receiver motion and time-varying impulse response h (d, t) and x (t), the transmitted signal. The received signal y (d, t) at any position d would be

$$y(d,t) = x(t) * h(d,t) = \int_{-\infty}^{\infty} x(\tau) h(d,t-\tau) d\tau \tag{10.34}$$

For a casual system: $h(d,t) = 0,$ for $t < 0$ and for a stable system $\int_{-\infty}^{\infty} |h(d,t)| d\tau \ < \infty$

Applying causality condition in the above equation, $h(d,\ t - \tau\ ) = 0$ for $t - \tau < 0 \Rightarrow \tau > t$, that is, $h(d,t-\tau)$ is defined for $\tau < t$. Hence, the integral limits are changed to

$$y(d,t) = \int_{-\infty}^{t} x(\tau) h(d,t-\tau) d\tau \tag{10.35}$$

Since the receiver moves along the ground at a constant velocity v, the position of the receiver is $d = vt$, that is,

$$y(vt,t) = \int_{-\infty}^{t} x(\tau) h(vt,t-\tau) d\tau \tag{10.36}$$

Since v is a constant, y (vt, t) is just a function of t. Therefore, the above equation can be expressed as

$$y(t) = \int_{-\infty}^{t} x(\tau)h(vt, t-\tau)d\tau = x(t)*h(vt,t) = x(t)*h(d,t) \tag{10.37}$$

It is useful to discretize the multipath delay axis of the impulse response into equal time delay segments called excess delay bins, each bin having a time delay width equal to ($\tau_{i+1}-\tau_i$) $=\Delta\tau$ and $\tau_i = i\Delta\tau$ for $i\in\{0, 1, ... N-1\}$, where N represents the total number of possible equally spaced multipath components, including the _rst arriving component. The useful frequency span of the model is $2/\Delta\tau$. The model may be used to analyse transmitted RF signals having bandwidth less than $2/\Delta\tau$.

If there are N multipaths, maximum excess delay is given by $N\Delta\tau$.

$$\{y(t) = x(t)*h(t,\tau_i) \quad i = 0,1,....N-1\} \tag{10.38}$$

Band pass channel impulse response model is

$$x(t) \rightarrow h(t,\tau) = \mathrm{Re}\{h_{\mathrm{b}}(t,\tau)e^{j\omega_c t}\} \rightarrow y(t) = \mathrm{Re}\{r(t)e^{j\omega_c t}\} \tag{10.39}$$

Baseband equivalent channel impulse response model is given by

$$c(t) \rightarrow \frac{1}{2}h_{\mathrm{b}}(t,\tau) \rightarrow r(t) = c(t)*\frac{1}{2}h_{\mathrm{b}}(t,\tau) \tag{10.40}$$

Average power is

$$\overline{x^2(t)} = \frac{1}{2}|c(t)|^2 \tag{10.41}$$

The baseband impulse response of a multipath channel can be expressed as

$$h_{\mathrm{b}}(t,\tau) = \sum_{i=0}^{N-1} a_{\mathrm{i}}(t,\tau)\exp\left[j\left(2\pi f_{\mathrm{c}}\tau_i(t) + \phi_i(t,\tau)\right)\right]\delta\left(\tau - \tau_i(t)\right) \tag{10.42}$$

where $a_{\mathrm{i}}(t, \tau)$ and $\tau_i(t)$ are the real amplitudes and excess delays, respectively, of the i^{th} multipath component at time t. The phase term $2\pi f_{\mathrm{c}}\tau_i(t) + \phi_i(t,\tau)$ in the above equation represents the phase shift due to free-space propagation of the i^{th} multipath component, plus any additional phase shifts which are encountered in the channel.

If the channel impulse response is wide sense stationary over a small-scale time or distance interval, then

$$h_{\mathrm{b}}(\tau) = \sum_{i=0}^{N-1} a_i \exp[j\theta_i]\delta(\tau - \tau_i) \tag{10.43}$$

For measuring $h_{\mathrm{b}}(\tau)$, we use a probing pulse to approximate $\delta(t)$ that is,

$$p(t) \approx \delta(\tau - \tau) \tag{10.44}$$

Power delay profile is taken by spatial average of $|h_b(t,\tau)|^2$ over a local area. The received power delay profile in a local area is given by

$$p(\tau) \approx k\overline{|h_b(t,\tau)|^2} \tag{10.45}$$

10.7 Statistical models for multipath propagation

Many multipath models have been proposed to explain the observed statistical nature of a practical mobile channel. Both the first-order and second-order statistics have been examined in order to find out the effective way to model and combat the channel effects. The most popular of these models are Rayleigh model, which describes the NLoS propagation.

The Rayleigh model is used to model the statistical time varying nature of the received envelope of a flat-fading envelope. The main first order and second order statistical models are discussed below. Two-ray NLoS multipath resulting in Rayleigh fading is shown in Figure 10.16

10.7.1 NLoS propagation: Rayleigh fading model

Let there be two multipath signals $S1$ and $S2$ received at two different time instants due to the presence of obstacles as shown in Figure 10.16. Now there can either be constructive or destructive interference between the two signals. Let E_n be the electric field and θ_n be the relative phase of the various multipath signals. So we have

$$\tilde{E} = \sum_{n=1}^{N} E_n e^{j\theta_n} \tag{10.46}$$

Now if $N \to \infty$ (i.e., are sufficiently large number of multipaths) and all the E_n are IID distributed, then by central limit theorem we have,

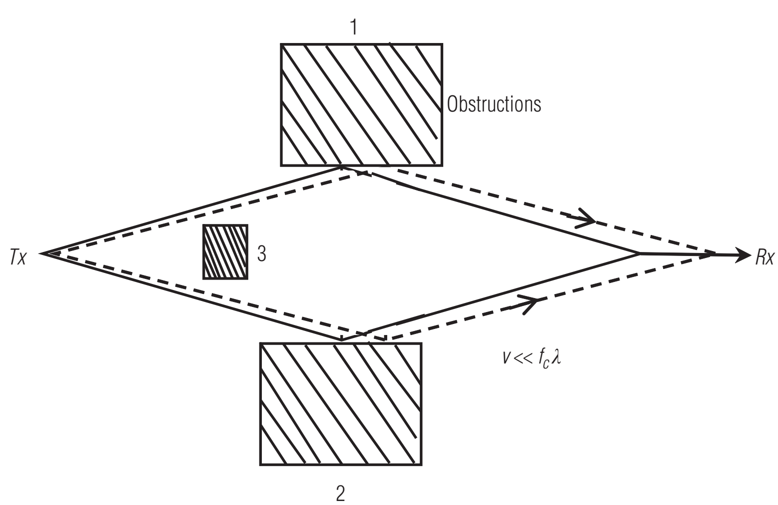


Figure 10.16 Two-ray NLoS multipath resulting in rayleigh fading

$$\lim_{N\to\infty} \tilde{E} = \lim_{N\to\infty} \sum_{n=1}^{N} E_n e^{j\theta_n}$$
$$= \mathrm{Z}_{\mathrm{r}} + jZ_{\mathrm{i}} = \mathrm{Re}^{j\varphi} \tag{10.47}$$

where Z_r and Z_i are Gaussian random variables. For the above case

$$R = \sqrt{Z_r^2 + Z_i^2} \tag{10.48}$$

$$\varphi = \tan^{-1} \frac{Z_i}{Z_r} \tag{10.49}$$

For all practical purposes we assume that the relative phase θ_n is uniformly distributed.

$$E\left[e^{j\theta_n}\right] = \frac{1}{2\pi} \int_0^{2\pi} e^{j\phi} d\theta = 0 \tag{10.50}$$

It can be seen that E_n and θ_n are independent. So,

$$E\left[\tilde{E}\right] = E\left[\sum E_n e^{j\theta_n}\right] = 0 \tag{10.51}$$

$$E\left[\left|\tilde{E}\right|^2\right] = E\left[\sum E_n e^{j\theta_n} \sum E_n^* e^{-j\theta_n}\right] = E\left[\sum_m \sum_n E_n E_m e^{j(\theta_n - \theta_m)}\right] = \sum_{n=1}^{N} E_n^2 = P_0 \tag{10.52}$$

where P_0 is the total power obtained. To find the cumulative distribution function (CDF) of R,

$$F_R(r) = P_r(R \le r) = \iint_A f_{Z_i,Z_r}(Z_i, Z_r) dZ_i dZ_r \tag{10.53}$$

where A is determined by the values taken by the dummy variable r. Let Z_i and Z_r be zero mean Gaussian random variables. Hence, the CDF can be written as

$$F_R(r) = \iint_A \frac{1}{\sqrt{2\pi\sigma}^2} e^{\frac{-\left(Z_r^2 + Z_i^2\right)}{2\sigma^2}} dZ_i dZ_r \tag{10.54}$$

Let $Z_r = p \cos(\theta)$ and $Z_i = p \sin(\theta)$ So we have

$$F_R(r) = \int_0^{2\pi} \int_0^{2\pi} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-p^2}{2\sigma^2}} p dp d\theta$$
$$= 1 - e^{\frac{-r^2}{2\sigma^2}} \tag{10.55}$$

Equation (10.55) is valid for all $r \ge 0$. The probability distribution function (PDF) can be written as

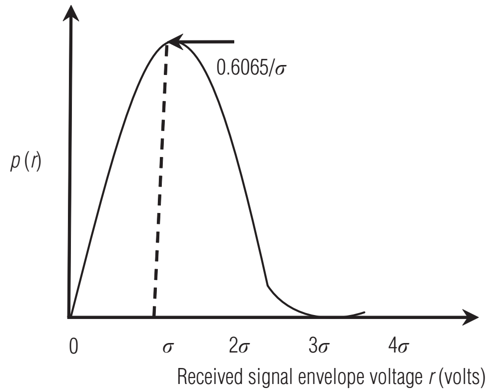


Figure 10.17 Rayleigh probability density function

$$f_R(r) = \frac{r}{\sigma^2} e^{\frac{-r^2}{2\sigma^2}} \tag{10.56}$$

and is shown in Figure 10.17 with different σ values. This equation also is valid for all $r \geq 0$. Above distribution is known as Rayleigh distribution and it has been derived for slow fading. However, if Doppler frequency, f_D <<1 Hz, we call it as quasi-stationary Rayleigh fading. We observe the following

$$E[R] = \sqrt{\frac{\pi}{2}}\sigma$$

$$E\left[R^2\right] = 2\sigma^2$$

$$\operatorname{var}[R] = \left(2 - \frac{\pi}{2}\right)\sigma^2$$

$$\operatorname{median}[R] = 1.77\sigma$$

10.7.2 LoS propagation: Rician fading model

Rician fading is the addition of a direct LOS path to all the normal multipaths. The Rician probability density function is shown in Figure 10.18.

$$f_R(r) = \frac{r}{\sigma^2} e^{\frac{-(r^2 + A^2)}{2\sigma^2}} I_0\left(\frac{A_r}{\sigma^2}\right) \tag{10.57}$$

for all $A \geq 0$ and $r \geq 0$. Here, A is the peak amplitude of the dominant signal and I_0 is the modified Bessel function of the first kind and zeroth order. A factor K is defined as

$$K_{\text{dB}} = 10 \log \frac{A_r}{\sigma^2} \tag{10.58}$$

As $A \rightarrow 0$ then $K_{\text{dB}}\ \Omega \rightarrow \infty$.

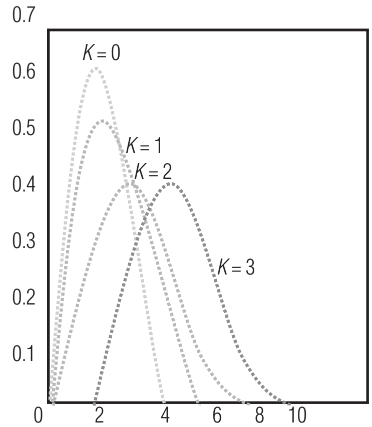


Figure 10.18 Rician probability density function

10.7.3 Generalized model: Nakagami distribution

A generalization of the Rayleigh and Rician fading is the Nakagami distribution. Nakagami-probability density function is shown in Figure 10.19. The schematic representation of level crossing with a Rayleigh fading envelope at 10 Hz Doppler spread as shown in Figure 10.20.
Its PDF is given as,

$$f_{\mathrm{R}}(r) = \frac{2r^{m-1}}{\Gamma(m)}\left(\frac{m^m}{\Omega^m}\right)e^{\frac{-mr^2}{\Omega}} \tag{10.59}$$

where
$\Gamma(m)$ is the gamma function
Ω is the average signal power and
m is the fading factor, which is always greater than or equal to 0.5

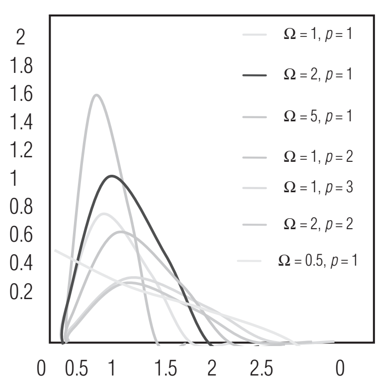


Figure 10.19 Nakagami-probability density function

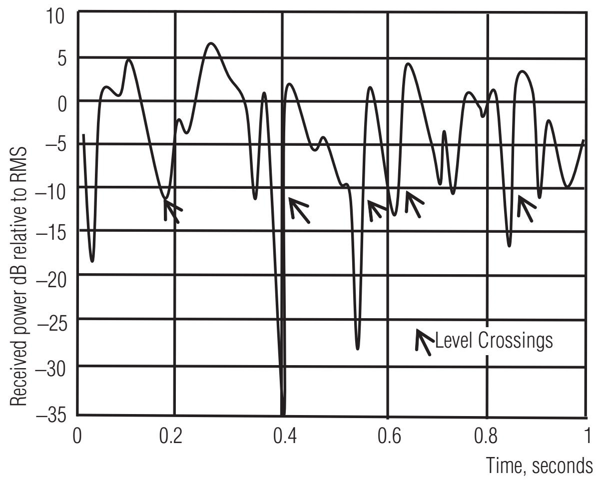


Figure 10.20 Schematic representation of level crossing with a Rayleigh fading envelope at 10Hz Doppler spread

When m=1, Nakagami model is the Rayleigh model.
When

$$m = \frac{(M+1)^2}{2M+1} \tag{10.60}$$

where,

$M = A/2\sigma$

Nakagami fading is the Rician fading.

As $m \to \infty$ Nakagami fading is the impulse channel and no fading occurs.

10.7.4 Second-order statistics

To design better error control codes, we have two important second-order parameters of fading model, that is, the level-crossing rate (LCR) and average fade duration (AFD). These parameters can be utilized to assess the speed of the user by measuring them through the reverse channel. The LCR is the expected rate at which the Rayleigh fading envelope normalized to the local rms amplitude crosses a specific level "R" in a positive going direction.

$$N_{\mathrm{R}} = \int_0^{\infty} \dot{r} p(R, \dot{r}) \, d\dot{r} = \sqrt{2\pi} f_{\mathrm{D}} \rho e^{-\rho^2} \tag{10.61}$$

where N_R is the time derivative of $r(t)$, f_{D} is the maximum Doppler shift, and ρ is the value of the specified level R, normalized to the local rms amplitude of the fading envelope. The other important parameter, AFD, is the average period time for which the receiver power is below a specified level R.

$$\tau = \frac{1}{N_{\mathrm{r}}} P_{\mathrm{r}}(r \le R) \tag{10.62}$$

As

$$P_{\mathrm{r}}(r \le R) = \int_0^R p(r)\,dr = 1 - e^{-\rho^2} \tag{10.63}$$

Therefore,

$$\bar{\tau} = \frac{1 - e^{-\rho^2}}{\sqrt{2\pi}\, f_{\mathrm{D}} \rho\, e^{-\rho^2}} = \frac{e^{-\rho^2} - 1}{\sqrt{2\pi}\, f_{\mathrm{D}} \rho} \tag{10.64}$$

Apart from LCR, another parameter is fading rate, which is defined as the number of times the signal envelope crosses the middle value (r_{m}) in a positive going direction per unit time. The average rate is expressed as

$$N(r_{\mathrm{m}}) = \frac{2\upsilon}{\lambda} \tag{10.65}$$

Another statistical parameter, sometimes used in the mobile communication, is called as depth of fading. It is defined as the ratio between the minimum value and the mean square value of the faded signal. Usually, an average value of 10 per cent as depth of fading gives a marginal fading scenario.

10.8 Summary

- The characteristics of the radio channel depend on operating frequency and the mode of propagation. For example, LoS radio links, diffraction/scattering, and satellite links.
- When there exists a visible straight path between the BS antenna and the mobile terminal antenna, no other multipath component exists. This direct signal component is called as a LoS component.
- The maximum frequency just below the escape point is called the critical frequency.
- At very-low, low, and medium frequencies, a skip zone is never present. At high frequency spectrum, a skip zone is present.
- Longley–Rice model is the most commonly used outdoor propagation models within a frequency band of 40 MHz to 100 GHz.
- In wireless telecommunications, multipath is the propagation of radio signals reaching the receiving antenna by two or more paths.
- Fading means rapid fluctuations of the amplitudes, phases, or multipath delays of a radio signal over a short period.
- The depth of fading is defined as the ratio between the minimum value and the mean square value of the faded signal.
- Propagation models have traditionally focused on predicting the average received signal strength at a given distance from the transmitter, as well as the variability of the signal strength in close spatial proximity to a particular location.
- Large-scale propagation models are based on signal strength over large T–R separation distances (several hundreds or thousands of metres).
- Reflection, diffraction, and scattering are three basic propagation mechanisms which impact propagation in a mobile communication system.

- Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities (edges).
- A mobile radio channel may be modelled as a liner filter with a time varying impulse response, where the time variation is due to receiver motion in space.
- The filtering nature of the channel is caused by the summation of amplitudes and delays of the multiple arriving waves at any instant of time.
- In a fast-fading channel, the impulse response changes rapidly within the symbol duration.
- In a slow-fading channel, the channel impulse response changes at a rate much slower than the transmitted baseband signal $S(t)$.

Review questions

1. Write short notes on skip distance and skip zone.
2. Explain in detail about link budget.
3. What is outdoor propagation model? Explain Hata model.
4. What is small-scale fading? Write the factors influencing fading.
5. Explain small-scale multipath measurements.
6. Explain impulse response model of a multipath channel.
7. Explain non-line-of-sight propagation for Rayleigh fading model.
8. What are the types of small-scale fading?
9. In what ways does the received signal get affected by multipath propagation in mobile communications?
10. List the factors that influence small-scale fading.
11. What are the causes of fast and slow fading? Distinguish between them.
12. Describe at least two problems that can occur as a result of multipath interference.
13. Explain about the transmission medium. (Refer Section 10.2)

Exercise problems

1. If 120 W is applied to a unit gain antenna with a 500 MHz carrier frequency, find the received power in dBm at a free-space distance of 150 m from the antenna. What is P_r (15 km)? Assume unity gain for the receiver antenna.
2. Assume a receiver is located 15 km from a 150 W transmitter. The carrier frequency is 1,500 MHz, free-space propagation is assumed, $G_t = 1$ and $G_r = 4$, Find the power at the receiver.
3. In a WLAN the minimum SNR required is 12 dB for an office environment. The background noise at the operational frequency is 115 dBm. If the mobile terminal transmits power is 100 mW, what is the coverage radius of an access point if there are three floors between the mobile transmitter and the access point? (Ans: 54 m)
4. Find the received power for the link from a synchronous satellite to a terrestrial antenna. Use the following data: height = 60,000 km; satellite transmit power = 4 W; transmit antenna gain=18 dBi; receive antenna gain = 50 dBi; and transmit frequency = 12 GHz.
5. For a flat Rayleigh fading channel determine the number of fades per second for $\rho = 1$ and average fade duration, when the maximum Doppler frequency is 20 Hz. What is the maximum velocity of the mobile if the carrier frequency is 900 MHz? (Ans: 24 km/h)
6. A mobile station travelling at a speed of 60 km/h transmits at 900 MHz. If it receives or transmits data at a rate of 64 Kbps, is the channel fading slow or fast?

Objective type questions and answers

1. The distance from the transmitter to the point where the sky wave first returns to the earth is called
   (a) near distance (b) far distance (c) skip distance (d) none
2. The skip distance depends on
   (a) waves frequency (b) angle of incidence
   (c) degree of ionization (d) all of the above
3. The zone of silence between the point where the ground wave is too weak for reception and where sky wave is first returned is called
   (a) dead zone (b) skip zone (c) near zone (d) far zone
4. The outer limit of the skip zone varies depending on
   (a) operating frequency (b) sunspot activity (c) time of day (d) all of the above
5. A blank area on the opposite side of an obstruction in line of sight from transmitter to receiver is called as
   (a) shadow zone (b) skip zone (c) near zone (d) far zone
6. For a communication, link budget determines
   (a) signal powers (b) noise powers
   (c) signal-to-noise ratio (d) all of the above
7. Okumura model is applicable for base station effective antenna heights ranging from
   (a) 30–1,000 m (b) 1,000–1,500 m
   (c) 1,500–2,000 m (d) 2,000–2,500 m
8. Longley–Rice is most commonly used in the frequency band of
   (a) 1 MHz to 40 MHz (b) 40 MHz to 100 GHz
   (c) 100 GHz to 200 GHz (d) 200 GHz to 300 GHz
9. Rapid fluctuations of the amplitudes, phases, or multipath delays of a radio signal over a short period is called
   (a) interference (b) fading (c) multipath (d) none
10. Frequency selective fading occurs when the signal bandwidth is
    (a) more than coherence bandwidth
    (b) less than coherence bandwidth
    (c) more than non-coherence bandwidth
    (d) less than non-coherence bandwidth
11. ________________ occurs when the radio path between a transmitter and receiver is obstructed by a surface with sharp irregular edges.
12. ___________ results from the presence of objects between the transmitter and the receiver.
13. __________ refers to the phenomenon by which multiple copies of a transmitted signal are received at the receiver, due to the presence of multiple radio paths.
14. Two main reasons that contribute to the rapid fluctuations of the signal amplitude in mobile communications are

Answers: 1. (c), 2. (d), 3. (b), 4. (d), 5. (a), 6. (d), 7. (a), 8. (b)9. (b), 10. (a), 11. diffraction, 12. shadow fading, 13. multipath, 14. multipath fading and Doppler effect.

Open book questions

1. Write short notes on diffraction.
2. Explain Okumura model.
3. How is propagation path loss related to the received signal power in mobile communications?
4. State the reasons for attenuations of the signal strength of electromagnetic waves upon reflection.
5. Explain about mobile fading characteristics.
6. What is the difference between frequency selective fading and flat fading?

Key equations

1. The directivity D of an antenna is defined as

$$D = \frac{\text{Power density at a distance in the direction of maximum radiation}}{\text{Mean power density at a distance } d}$$

2. Power gain G is defined as

$$G = \frac{\text{Power density at a distance } d \text{ in the direction of maximum radiation}}{P_{\mathrm{T}} / 4\pi d^2}$$

3. The power received by an antenna that is separated from a transmitter antenna by a distance d in free space is given by the Friis free-space equation

$$P_{\mathrm{r}}(d) = \frac{P_{\mathrm{t}} G_{\mathrm{t}} G_{\mathrm{r}} \lambda^2}{(4\pi)^2 d^2} = P_0 \left(\frac{\lambda}{4\pi d} \right)^2$$

4. Free-space gain is defined as

$$G_{\mathrm{FS}} = \left(\frac{\lambda}{4\pi d} \right)^2 = \frac{1}{PL_{\mathrm{FS}}}$$

5. The empirical path-loss formula of Okumura at distance d parameterized by the carrier frequency f_{c} is given by

$$P_{\mathrm{L}}(d)\,\mathrm{dB} = L(f_c, d) + A_{\mathrm{mu}}(f_c, d) - G(h_t) - G(h_r) - G_{\mathrm{AREA}}$$

6. The standard Hata formula for median path loss in urban area is given by

$$PL_{50}(\text{urban})(\text{dB}) = 69.55 + 29.16 \log f_c - 13.82 \log h_{te} - a(h_{re}) + (44.9 - 6.55 \log h_{te}) \log d$$

7. The mean excess delay is the first moment of the power delay profile and is defined as

$$\tau = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} = \frac{\sum P(\tau_k)\tau_k}{\sum P(\tau_k)}$$

8. The mean square excess delay spread is defined as

$$\tau^2 = \frac{\sum_k P(\tau_k)\tau^2_k}{\sum_k P(\tau_k)}$$

9. A generalization of the Rayleigh and Rician fading is the Nakagami distribution PDF which is given as

$$f_R(r) = \frac{2r^{m-1}}{\Gamma(m)}\left(\frac{m^m}{\Omega^m}\right)e^{\frac{-mr^2}{\Omega}}$$

10. The LCR is the expected rate at which the Rayleigh fading envelope normalized to the local rms amplitude crosses a specific level "*R*" in a positive going direction.

$$N_R = \int_0^\infty \dot{r}p(R,\dot{r})\,d\dot{r} = \sqrt{2\pi}\,f_D\rho\,e^{-\rho^2}$$

Further reading

Fernando Pérez Fontán, Perfecto Mariño Espiñeira "Modeling the Wireless Propagation Channel A simulation Approach with MATLAB", 2008, New Delhi.

Kostov, N. "Mobile Radio Channels Modeling in MATLAB," *Radio Engineering*, Vol. 12, No. 4, December 2003, pp. 12–16.

Rappaport, T. S. *Wireless Communications*, PHI Pvt. Ltd, pp. 177-248, New Delhi, 2007.

Parsons, J. D *The Mobile Radio Propagation Channel*. John Wiley, 2000, 2nd edition, New York.

Kim, Byung-ki. "Smart Base Station Antenna Performance for Several Scenarios – An Experimental and Modeling Investigation," PhD Thesis, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, 7 May 2002.

Simmons L. D, and Ace III, F. L., Nonresident Training Course on, "Antennas and Wave Propagation", Vol. 7, Naval Education and Training Professional Development and Technology Center, United States, October 1995.

Cell Coverage for Signal and Traffic 11

11.1 Introduction

Coverage is the extent of a geographical area in which a wireless network provider offers cellular service for mobile phone users. It is usually a measure of cell phone connectivity and depends on several factors such as orography (i.e., study of mountains), buildings, technology, and radio frequency. Some frequencies provide better coverage while other frequencies penetrate better through obstacles, such as buildings in cities. The ability of a mobile phone to connect to a base station depends on the strength of the signal which may be boosted by higher power transmissions, taller antenna masts, and better antennae. This is also a problem when designing networks for large metropolitan areas with modern skyscrapers. Also, signals do not travel deep underground. Hence, specialized transmission solutions are used to deliver mobile phone coverage into areas such as underground parking garages and subway trains.

The extent of coverage for a given building is represented in the form of percentage of the area covered at a certain height above the ground level. The task is to cover the whole area with a minimum number of cell sites. Since 100 per cent cell coverage of an area is not possible, the cell sites must be engineered so that the holes are located in the no-traffic locations. The coverage area will be examined in environments such as in an open area, in a suburban area, in an urban area, over flat terrain, over hilly terrain, over water, and through foliage areas.

This chapter discusses the prediction model, which is a point-to-point model, followed by a study of propagation over water or flat open area and the loss due to foliage. The cell-site antenna heights and signal coverage cells are then briefly discussed. Finally, the chapter concludes with an overview of near and long-distance propagation. The results generated from the prediction model will differ depending on the coverage area being used. There are many field-strength prediction models that provide more or less an area-to-area prediction.

11.2 Point-to-point model

The free-space propagation model does not apply in a mobile radio environment. Propagation path loss depends on distance of the mobile from its serving cell-site, carrier frequency of transmission, f_c (or wavelength λ_c), the antenna heights of cell-site and mobile unit and the local terrain characteristics such as buildings and hills. Mobile radio propagation path loss is uniquely different from that experienced in other communication media. The signal received by a mobile unit remains constant only over a small operating area and varies as the mobile unit moves. This is mainly due to the variations in the terrain conditions and presence of man-made structures surrounding the mobile unit.

Area-to-area signal prediction models are not useful for cellular communication systems because of the large uncertainty of the prediction. It usually provides an accuracy of prediction with a measured standard deviation of about 8 dB, which means that 68 per cent of the actual path-loss data are within ±8 dB of the predicted path-loss value.

This implies that the uncertainty range for predicting the path loss is too large. When plotting received signal strengths at a given radio-path distance, the deviation from predicted values is approximately 8 dB. This is applicable in various operational geographical areas, including near the cell-site mainly due to the close-in buildings around it, as well as at a distant location due to the large variation along different radio paths.

Moreover, at a distance from the cell site, radio paths may be line-of-sight, partial line-of-sight, or out-of-sight. Thus, the received signals could be strong, normal, or weak accordingly. Therefore, the standard deviation of 8 dB is always found along the predicted path-loss curve at any distance.

The point-to-point propagation prediction model reduces the uncertainty range by including the terrain contour profiles in the path-loss predictions. In a non-obstructive condition, the direct radio path from the cell-site to the mobile unit is not obstructed by the terrain contour but the radio path may be obstructed by man-made structures. In the mobile radio environment, line-of-sight condition is generally not available. Under these conditions, the antenna height gain is calculated for every location of the mobile unit it covers. Taking into account the antenna-height gain at various mobile locations, the path-loss slope will have a standard deviation of only less than 2–3 dB instead of 8 dB as observed in area-to-area prediction model.

In obstructive condition (i.e., the direct path from the cell-site to the mobile unit is obstructed by the terrain contour), first the area-to-area prediction model is applied. This is then followed by diffraction loss. When the heavy foliage is close in at the mobile unit, the loss due to the foliage can be obtained from the diffraction loss. In either ease, according to the theoretical model, the 40 dB per decade path loss slope applies.

The mobile point-to-point prediction model provides a standard deviation of less than 3 dB from the predicted value. It is very useful for designing a mobile cellular system with a radius of 15 km or less for each cell. Since the path loss or received signal strength data follows the log-normal distribution, 68 per cent of predicted values obtained from a point-to-point propagation model are within 2–3 dB. In irregular terrain conditions, the signal received by a mobile varies due to the change in terrain conditions and the presence of natural or man-made obstructions.

The features of point-to-point prediction model are as follows:

- In point-to-point model, the large range of uncertainty is reduced.
- For prediction of path loss, it applies the detailed terrain contour information.
- In the analysis of point-to-point method, the predicted values were placed at *x*-axis and actual measured values at *y*-axis and it was a simple method.
- A line of 45° as shown in Figure 11.1 is drawn between these two values and this 45° line is the errorless prediction line.
- This prediction model is very useful for cellular mobile communication.
- But the largest difference between the predicted and measured values is around 3 dB and the accuracy is less than that of area-to-area prediction model.
- The occurrence of handoffs in cellular mobile communication can be predicted with more accuracy in this model.

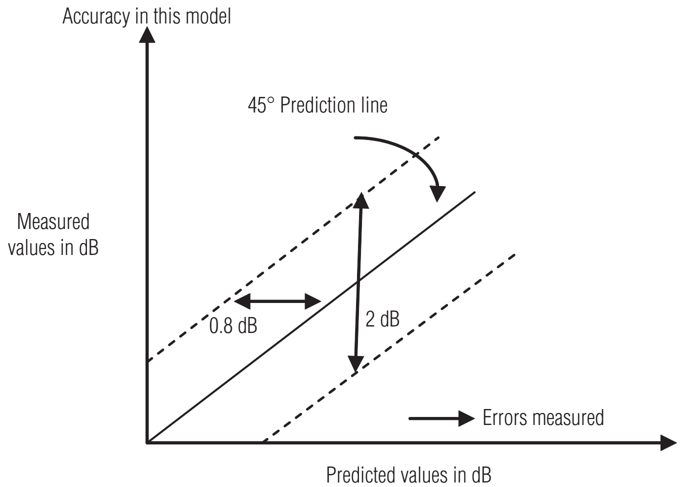


Figure 11.1 Point-to-point prediction curves in no obstruction

11.3 Propagation over water or flat open area

The mobile signal propagation over water or flat open area has to be given proper design setups because probability of occurrence of interference is high if necessary arrangements are not done. For fresh as well as the sea water, the value of relative permittivity (ε_r) is same whereas their conductivities and dielectric constant (ε_c) are different from one another.

Figure 11.2 shows two antennas; one at mobile unit and the other at sea level with two reflection points. One point (P_1) is close to mobile unit and the other point (P_2) is away from the mobile unit.

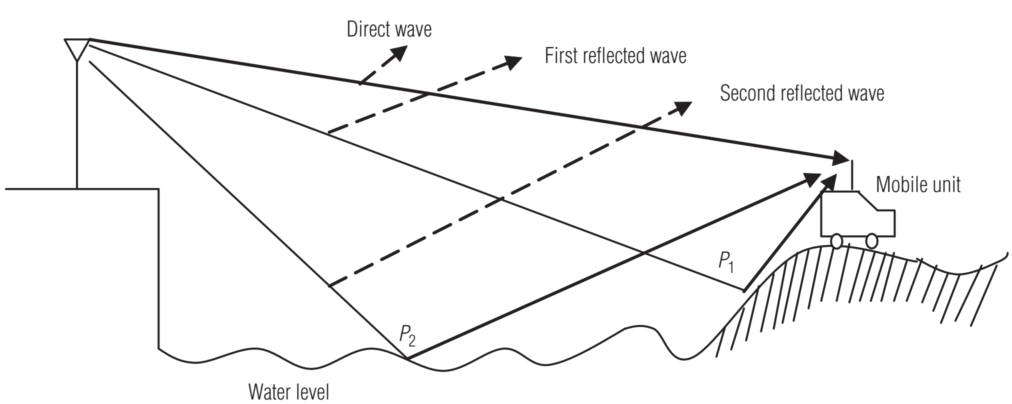


Figure 11.2 Propagation over water

11.3.1 Propagation between two fixed stations

Consider two fixed stations that use point-to-point communication between them. Here, the received power P_r will be,

$$P_r = P_t\left(\frac{1}{\frac{4\pi d}{\lambda}}\right)^2 \left|1 + be^{-j\phi_v}.\exp(j\Delta\phi)\right|^2$$

where

P_t is transmitted power

λ is the wavelength and d distance between two stations

b and ϕ_v are the amplitude and phase of the complex reflection coefficient, respectively.

Note: The phase difference $\Delta\phi$, phase differenceis due to the path difference Δd (between direct and reflected waves)

The path loss for the land-to-mobile propagation over the land is 40 dB per decade and it is different for the land-to-mobile propagation over water. In land-to-mobile propagation over water, path loss is 20 dB per decade.

11.4 Foliage loss

Foliage loss is the loss that occurs due to trees. It includes many parameters and variations pertaining to the size; the density; the distribution of leaves, branches, trunks; the height of the trees relative to the antenna height; and so on. Figure 11.3 illustrates the problem.

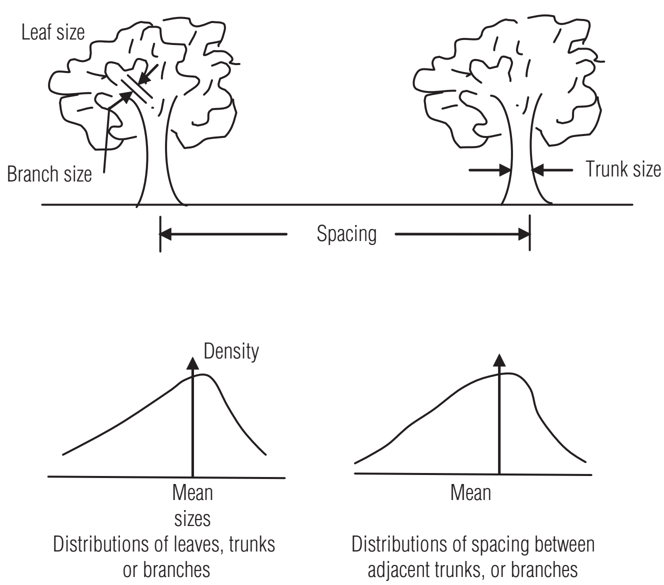


Figure 11.3 Foliage environment

Consider three levels: trunks, branches, and leaves. Each level includes a distribution of sizes of trunks, branches, and leaves and also of the density and spacing between the adjacent trunks, branches, and leaves. The texture and thickness of the leaves are also taken into account. However, the estimate of the signal reception due to foliage loss does not need any degree of accuracy for system design.

However, a rough estimate should be sufficient. In tropical zones, the signal can hardly penetrate due to the large and thick sizes of tree leaves. It will then propagate from the top of the tree and deflect to the mobile receiver. This calculation will also be included.

When the foliage is uniformly heavy and the path lengths are short, the foliage loss can be treated as a wire-line loss, in decibels per foot or decibels per metre. Decibels per octaves or decibels per decade are used when the path length is long and the foliage is non-uniform. In general, foliage loss is proportional to the frequency to the fourth power ($\sim f^4$).

11.5 Cell-site antenna heights and signal coverage cells

The term effective antenna height refers to the height of the centre of radiation of an antenna above the effective ground level.

Antenna height unchanged

If the antenna height is unchanged then the whole signal-strength map can be linearly updated according to the change in power of the cell-site transmitter. If the transmitted power increases by 3 dB, just add 3 dB to each grid in the signal-strength map. The relative differences in power among the grids remain the same.

Antenna height changed

If the antenna height changes ($\pm\Delta h$), then the whole signal-strength map obtained from the old antenna height cannot be updated with a simple antenna gain formula as

$$\Delta g = 20\log\frac{h_1'}{h_1}$$

where h_1 is the old actual antenna height and h_1' is the new actual antenna height. However, we can still use the same terrain contour data along the radio paths (from the cell-site antenna to each grid) to figure out the difference in gain resulting from the different effective antenna heights in each grid.

$$\Delta g' = 20\log\frac{h_e'}{h_e} = 20\log\frac{h_e \pm \Delta h}{h_e}$$

where, h_e is the old effective antenna height and h_e' is the new effective antenna height. The additional gain (increase or decrease) will be added to the signal-strength grid based on the old antenna height.

Antenna location changed

If the location of the antenna is changed, then the point-to-point program has to start all over again. The old point-to-point terrain contour data are no longer useful. The old effective antenna height seen from a distance will be different when the location of the antenna changes and there is no relation between the old effective antenna height and the new effective antenna height. Therefore, every time the location of the antenna changes, the new point-to-point prediction calculation starts.

Visualization of the effective antenna height

The effective antenna height changes when the location of the mobile unit changes. Therefore, we can visualize the effective antenna height as always changing up or down while the mobile unit is moving. This kind of picture should be kept in mind as shown in Figure 11.4(a). In addition, the following facts would be helpful.

Case 1: The mobile unit is driven up a positive slope (up to a high spot). The effective antenna height increases if the mobile unit is driving away from the cell-site antenna, and it decreases if the mobile unit is approaching the cell-site antenna.

Case 2: The mobile unit is driven down a hill. The effective antenna height decreases if the mobile unit is driving away from the cell-site antenna, and it increases if the mobile unit is approaching the cell-site antenna.

Visualization of signal coverage cells

A physical cell is usually visualized as a signal-reception region around the cell site. Within the region, there are weak spots called holes. This is always true when a cell covers a relative flat terrain.

However, if a cell covers a hilly area, then the coverage patterns of the cell will look like those shown in Figure 11.4(b). Here, the two cell sites are separated by a river. Because of the shadow loss due to the river bank, cell site A cannot cover area A', but cell site B can. The same situation

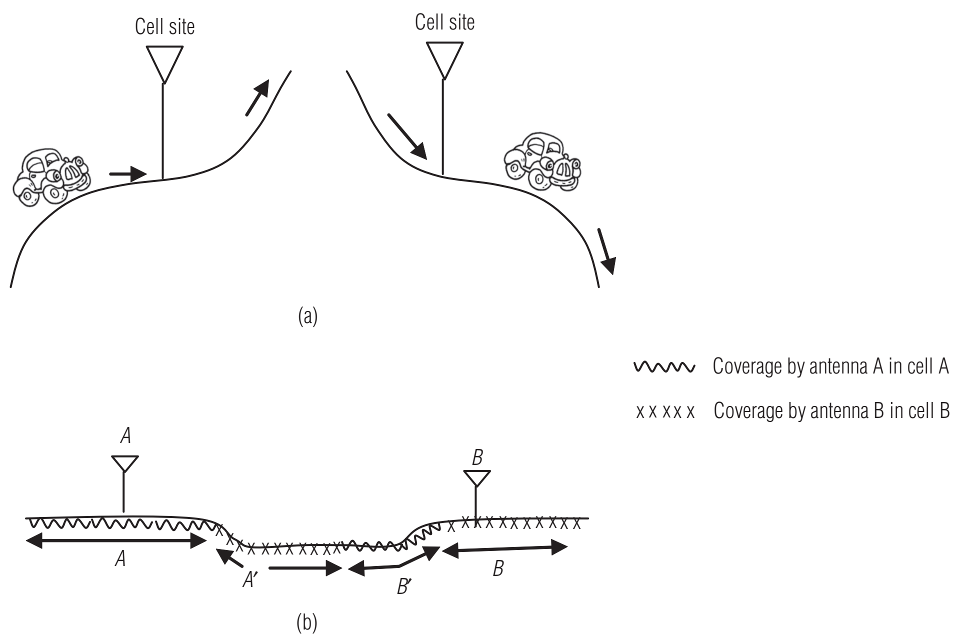


Figure 11.4 (a) Signal coverage due to effective antenna heights (b) Signal coverage served by two cell sites

applies to cell site B in area B'. Now every time the vehicle enters area A', a handoff is requested as if it were in cell B.

Therefore, in most cases, the holes in one cell are covered by the other sites. As long as the processing capacity at the mobile telephone switching office (MTSO) can handle excessive handoff, this overlapped arrangement for filling the holes is a good approach in a non-interference condition.

Example problem 11.1

When the cell-site antenna height is 30 m (100 ft), the mobile unit 8 km (5 m) away sees this cell-site effective antenna height as 60 m. Now if the antenna height is changed to 45 m, then find the changed effective antenna height seen by the same mobile unit.

Solution

Given

Old cell site antenna height, $h_1 = 30$ m

New cell site antenna height, $h_1' = 45$ m

Effective antenna height seen by the mobile unit, $h_e = 60$ m

The new cell-site effective antenna height, h_e' seen from the mobile unit can be derived

$$h_e' = h_e + \left(h_1' - h_1\right) = 60 + (45 - 30) = 75\,\text{m}$$

11.6 Near and long-distance propagation

In this section, the cases of near-in and long-distance propagation are considered.

11.6.1 Near-in distance propagation

A high-gain omnidirectional antenna has narrow beamwidth in the vertical plane within the radius of 1 mi (mile). The amount of signal received at the mobile unit will be reduced when the mobile unit is 1 mi away (it is said to be in the shadow region outside the main beam of the antenna).

Because of the antenna's vertical pattern, the elevation angle will be larger resulting in a weak reception level. There will be significant difference in the signal reception level say in the range of 10–20 dB due to road structures that would be in-line and perpendicular orientations.

Also the nearby surrounding objects of the cell site could influence the signal level reception, when the mobile unit is available within the 1 mi radius. As the distance of objects is more than 1 mi from the cell site, they may not affect the signal as with the previous case. For the land-to-mobile propagation, the height of antenna at cell site will definitely affect the signal reception. Here, the antenna height at base station is an important criterion to consider.

Near-in propagation has also to be considered as that of the far-in distance case. A suburban area can be assumed for analysis. The signal received level is –61.7 dBm for a suburban area at 1 mi intercept, and it is dependent on the reference set of the parameters, including antenna height.

11.6.2 Long-distance propagation

In cellular mobile communication, for short- and long-distance propagation, there are some advantages and disadvantages. In the case of a high cell site, it can cover large areas even in a noisy environment. Apart from the interference generated from co-channels and adjacent channels in the cell site, the long-distance propagation also affects the interference. As the traffic increases, the noise limited system becomes interference limited.

Considering an area within cell radii of 50 and 200 mi

Within a 50 mi radius, for the high cell site, the low-atmospheric region would influence the ground-wave path to travel in a non-straight line. This effect is more dominant over sea water than over a flat land profile. The signal may be very strong at one point of measurement and weak at another because the wave path can bend either upward or downward.

Within a radius of 200 mi, the case is different. The tropospheric signal propagation will prevail at a frequency of 800 MHz for long-distance propagation and the signal may sometimes reach 200 mi (320 km) away. This is due to the sudden changes in dielectric constant values of the troposphere which is 10 km above the earth's surface. The change in temperature also will result with a change in dielectric constant value. In tropospheric signal propagation, the wave may be divided by the mechanisms of reflection and refraction. In addition, moisture also affects signal propagation. The pressure of water vapour will decrease as the antenna height increases. If the value of refractive index starts to decrease with height over a particular range, then the rays get curved downwards. This condition is called as duct propagation or trapping.

Thus, the two cases discussed above on signal reflections due to flat terrain and hilly structure show how reflection takes place. The height of antenna above the hilly structure is accounted along with the height of the hill.

In both the cases, there are reflections but if the contour structure is with more irregularities, the amount of reflection is also severe and will result in more propagation path loss.

Consider Figure 11.5 for measurement of antenna heights. The effective antenna height h_e can be calculated from the point where the reflected ground plane and antenna meet. The effective antenna height is 40 m in case (a) and 200 m in case (b). The actual antenna height h_a is 100 m.

Example problem 11.2

From Figure 11.5, find the antenna height gain for both cases (a) and (b).

solution

In case (a) $h_a = 100$ m

$h'_a = 140$ m

$\Delta h = 40$ m

In case (b) $h_a = 100$ m

$h'_a = 300$ m

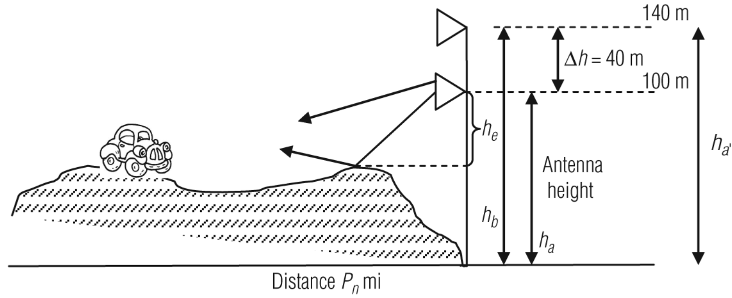
$\Delta h = 200$ m

The antenna height gain (ΔG) measurement for case (a) and case (b) is given by

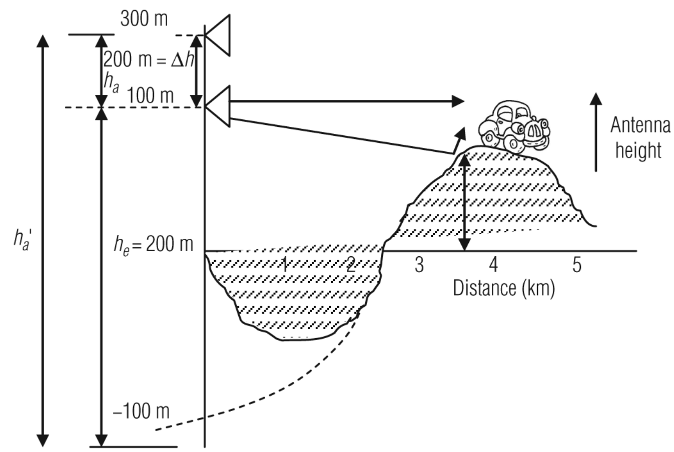
$$\text{Gain} = \Delta G = 20\log\,(h_e/h_a)$$

For case (a) $\Delta G = 20\log\,(40/100) = -8$ dB (negative gain value)

For case (b) $\Delta G = 20\log\,(200/100) = 6$ dB (positive gain value)



Case (a)



Case (b)

Figure 11.5 Antenna height gain

From the above example, it is evident that as the mobile unit travels along the road according to the land profile whether hilly or quasi-flat terrain the gain value changes. In addition, the effective antenna height changes as the mobile unit moves.

If the changes in antenna height gain are not considered as the mobile travels then the path loss slope will 8 dB standard deviation. Otherwise it would be in range 2–3 dB.

In Figure 11.6(a), if the height $h_1 << H$ then the length l of the floor will be approximately equal to the length of the moving vehicle. In the case of free propagation shown, a strong signal reception is possible.

In Figure 11.6(b), the length l when compared to previous length is longer and the antenna set-up remains the same. Since the l is longer, the ground reflection takes place by which there are two waves – direct and reflected – are generated.

The stronger the reflected wave, the larger will be the path loss. This wave occurs at a small angle θ. Thus, owing to a small incident angle, a large reflection coefficient is possible due to reflection mechanism. In addition, the presence of a stronger reflected wave tends to weaken the direct wave in propagation.

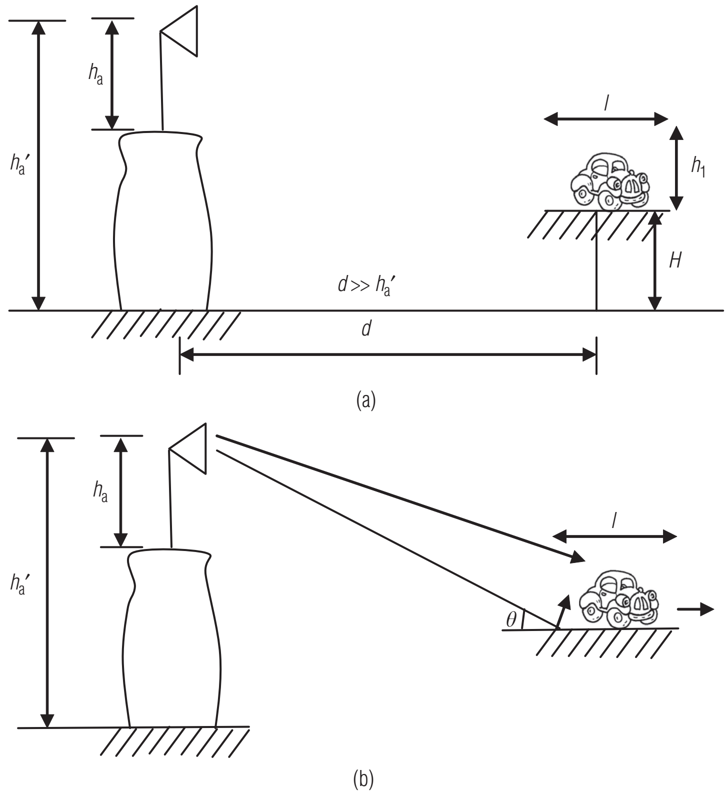


Figure 11.6 Effective antenna heights

11.7 Summary

- The extent of coverage for a given building is represented in the form of percentage of the area covered at a certain height above the ground level.
- Signal coverage can be predicted by coverage prediction models and is usually applied to a start-up system.
- The mobile signal propagation over water or flat open area has to be given proper design set-ups because probability of occurrence of interference is high.
- For freshwater and the sea water, the value of relative permittivity (ε_r) is same whereas their conductivities are different from one another.
- Foliage loss is due to the sizes of leaves, branches, and trunks; the density and distribution of leaves, branches, and trunks; and the height of the trees relative to the antenna heights.
- The amount of signal received at the mobile unit will be reduced when the mobile unit is 1 mi away thus it is said to be in shadow region.
- If the power of the cell-site transmitter changes, the whole signal-strength map can be linearly updated according to the change in power.
- The effective antenna height changes when the location of the mobile unit changes. Therefore, we can visualize the effective antenna height as always changing up or down while the mobile unit is moving.

Review questions

1. Explain point-to-point model and mention the types of point-to-point model.
2. Write short notes on foliage loss.
3. Explain briefly long-distance propagation.
4. Explain propagation over water or flat open area.
5. Write short notes on cell-site antenna height.
6. Discuss in detail point-to-point path-loss prediction model. Discuss the factors that effect the accuracy of prediction. (Refer Section 11.2)
7. If h_1=100 m, h_2 = 6 m, d = 5 km, and H = 100 m, use approximate method to find incident angle, elevation angle, ground reflection, and reflection point. (Refer Section 11.5)
8. Explain mobile propagation through obstructive path. (Refer Section 11.3)
9. Explain the effect of propagation of mobile signals over water. (Refer Section 11.3)
10. Calculate the power received in land-to-mobile propagation over water. (Refer Section 11.3)
11. Explain in detail about near-distance propagation. (Refer Section 11.6.1)
12. Discuss the merits of point-to-point model. (Refer Section 11.2)
13. Explain the various steps involved in finding antenna-height gain. (Refer Section 11.5.2)
14. Explain in detail about long-distance propagation? (Refer Section 11.5.2)
15. Explain about foliage loss in detail. (Refer Section 11.4)

Objective type questions and answers

1. Foliage areas are
   (a) natural terrains (b) man-made structures
   (c) open areas (d) urban areas
2. In a standard local mean spread, in the curve of signal strength versus distance from transmitting antenna, the measured standard deviation would be
   (a) 18 dB (b) 5 dB (c) 8 dB (d) 0.8 dB
3. Foliage loss is due to
   (a) tall buildings (b) indoor structures
   (c) tall trees, leaves, etc. (d) outdoor structures
4. A cell site located at a height covers a signal in
   (a) smaller area (b) larger area
   (c) short-distance propagation (d) long-distance propagation
5. The gradual bending of rays due to changing effective dielectric constant of atmosphere is the following effect
   (a) troposphere reflection (b) troposphere refraction
   (c) moistness (d) diffraction
6. Signal coverage can be found by _________ and generally applied to________
   (a) coverage prediction models, end system
   (b) coverage prediction models, start up system
   (c) point-to-point, end system
   (d) point-to-point model, terrain contour
7. The path that is not obstructed by terrain profile and man-made structure is
   (a) non-obstructive path (b) indirect path
   (c) line-of-sight path (d) none

8. Consider the following statements and mention whether it is true or false:
   - In man-made structures, there are sparse structures
   - In man-made structures, there are dense man-made structures

   (a) first statement is true (b) second statement is true
   (c) both are true (d) both are false
9. In mobile-to mobile-propagation, the propagation channel acts as
   (a) repeater (b) attenuator (c) filter (d) amplifier

Answers: 1. (a), 2. (c), 3. (c), 4. (b), 5. (b), 6. (b), 7. (c), 8. (c), 9. (c).

Open book questions

1. Write short notes on propagation in near-in distance.
2. Mention the merits of point-to-point model.
3. State the major factors causing propagation path loss.
4. What is meant by foliage? Define foliage loss.
5. Distinguish between large-scale propagation path-loss models and small-scale propagation path-loss models.
6. How is location of cell-site and mobile unit influenced by foliage loss?
7. Explain ground incident angle, elevation angle, ground reflection, and reflection point with respect to signal coverage.
8. If $h_1 = 50$ m, $h_2 = 3$ m, $d = 5$ km, and $H = 100$ m, use approximate method to find the incident angle, elevation angle, ground reflection, and reflection point.
9. Determine the phase difference between direct path and reflected path.
10. Why there is constant standard deviation along the path-loss curve?
11. The distance between two fixed stations is 40 km. The effective height at one end h_1 is 100 m above sea level. Find the height h_2 under the two conditions?
12. Determine the maximum and minimum spectral frequency received from a stationary transmitter which has a central frequency of exactly 1,950 MHz. Assume that the receiver is travelling at speed of 5 km/h.
13. Describe all physical circumstances that relate to a stationary transmitter and a moving receiver such that the Doppler shift at the receiver is equal to
    (i) 0 Hz
    (ii) f_{dmax}
    (iii) $-f_{\text{dmax}}$
    (iv) $f_{\text{dmax/2}}$

Key equations

1. If two fixed stations use point-to-point communication, the received power will be

$$P_r = P_t \left(\frac{1}{\frac{4\pi d}{\lambda}} \right)^2 \left|1 + be^{-j\phi_v}.\exp(j\Delta\phi)\right|^2$$

2. If the antenna height changes ($\pm\Delta h$), then the whole signal-strength map obtained from the old antenna height cannot be updated with a simple antenna gain formula as

$$\Delta g = 20\log\frac{h_1'}{h_1}$$

Further reading

Claude E Shannon, Warren Weaver, "The Mathematical Theory of Communication", University of Illinois Press, Urbana, 1949.

Rappaport T. S, "Wireless Communications", PHI Pvt. Ltd; New Delhi, 2007.

Garg V. K., Wilkes J. E., "Wireless and Personal Communications Systems: Fundamentals and Applications", PHI Pvt. Ltd, New Delhi, 1996.

Parsons, J. D. *The Mobile Radio Propagation Channel*. John Wiley & Sons Ltd.

Frequency Management and Channel Assignment 12

12.1 Introduction

Achieving optimum system capacity with a limited frequency spectrum is one of the main research issues in cellular communications. In a cellular system, frequency management and channel assignment are essential in order to achieve the basic objectives of spectrum utilization as well as adaptability to traffic density. In this chapter, the various aspects leading to an efficient and effective frequency planning of cellular systems are discussed in detail.

Depending upon the system parameters, the allocated frequency spectrum is divided into a number of frequency channels. These available frequency channels are then divided into the subsets that can be assigned to each cell. Different strategies are followed for the assignment of these channel sets to cells. Fixed channel assignment (FCA) technique and dynamic channel allocation techniques are covered in detail. Frequency management includes operations such as designation of set-up and voice channels, numbering the channels, and grouping voice channels into subsets.

The main objective of channel-assignment is to stabilize the fluctuations in the probability of call blockage over the entire coverage area of a cellular network over a period of time. The channel assignment does the allocation of specific channels to cell sites and mobile units. It can be done in two ways:

- Short-term assignment, where one channel assignment per call is handled by mobile telephone switching office (MTSO).
- Long-term assignment, where a fixed channel set consisting of one or more subsets are assigned to cell site on a long-term basis.

This chapter introduces numbering of the radio channels, traffic and channel assignment, non-FCA, the simulation process followed, and the results obtained.

12.2 Numbering the radio channels

Many cellular mobile systems operate on 666 channels. Each channel consists of two frequency channel bandwidths (mobile transmit/uplink or reverse channel and cell-site transmit/downlink or forward channel) to allow duplex operation. These two channel bandwidths must be separated in frequency in order to avoid interference. The frequency separation between the uplink and downlink channels is termed as channel spacing (or) duplex spacing. In the present 800 MHz band cellular system, the separation between the mobile transmit and the cell-site transmit is specified as 45 MHz. The sub-sections that follow describe the numbering of radio channels by discussing the frequency management chart and grouping of channels into subsets.

12.2.1 Frequency management chart

The total channels available are 832 in number. However, most mobile units and systems are still operating on 666 channels.

Figure 12.1 shows the arrangement of 666 frequency channels in block A and block B systems, each containing 333 channels. Out of these 333 available channels in each system, 312 channels are used for voice communication and 21 channels are used for controlling the system. These 21

	1A	2A	3A	4A	5A	6A	7A	1B	---	7B	1C	2C	3C	4C	5C	6C	7C	
	1	2	3	---	---	---	---	---	---	---	---	---	---	---	---	---	21	
	22	23	24	---	---	---	---	---	---	---	---	---	---	---	---	---	42	
	43	44	45	---	---	---	---	---	---	---	---	---	---	---	---	---	63	
	64	65	66	---	---	---	---	---	---	---	---	---	---	---	---	---	84	
	85	86	87	---	---	---	---	---	---	---	---	---	---	---	---	---	105	
	106	107	108	---	---	---	---	---	---	---	---	---	---	---	---	---	126	
	127	128	129	---	---	---	---	---	---	---	---	---	---	---	---	---	147	
	148	149	150	---	---	---	---	---	---	---	---	---	---	---	---	---	168	
	169	170	171	---	---	---	---	---	---	---	---	---	---	---	---	---	189	
	190	191	192	---	---	---	---	---	---	---	---	---	---	---	---	---	210	
	211	212	213	---	---	---	---	---	---	---	---	---	---	---	---	---	231	
	232	233	234	---	---	---	---	---	---	---	---	---	---	---	---	---	252	
Block A system	253	254	255	---	---	---	---	---	---	---	---	---	---	---	---	---	273	
	274	275	276	---	---	---	---	---	---	---	---	---	---	---	---	---	294	
	295	296	297	---	---	---	---	---	---	---	---	---	311	312	---	---	---	
	313	314	315	---	---	---	---	---	---	---	---	---	---	---	---	---	333	Set-up channel sets
	334	335	336	---	---	---	---	---	---	---	---	---	---	---	---	---	354	
	355	359	360	---	---	---	---	---	---	---	---	---	---	---	---	---	375	
Block B system	376	377	378	---	---	---	---	---	---	---	---	---	---	---	---	---	396	
	397	398	399	---	---	---	---	---	---	---	---	---	---	---	---	---	417	
	418	419	420	---	---	---	---	---	---	---	---	---	---	---	---	---	438	
	439	440	441	---	---	---	---	---	---	---	---	---	---	---	---	---	459	
	460	461	462	---	---	---	---	---	---	---	---	---	---	---	---	---	480	
	481	482	483	---	---	---	---	---	---	---	---	---	---	---	---	---	501	
	502	503	504	---	---	---	---	---	---	---	---	---	---	---	---	---	522	
	523	524	525	---	---	---	---	----	---	---	---	---	---	---	---	---	543	
	544	545	546	---	---	---	---	---	---	---	---	---	---	---	---	---	564	
	565	566	567	---	---	---	---	---	---	---	---	---	---	---	---	---	585	
	586	587	588	---	---	----	---	---	---	---	---	---	---	---	---	---	606	
	607	608	609	---	---	---	---	---	---	---	---	---	---	---	---	---	627	
	628	629	630	----	---	---	---	----	---	---	---	---	---	---	---	---	648	
	649	650	651	----	---	---	---	----	---	---	---	---	665	666	---	---	---	

Figure 12.1 Frequency management chart

channels are called as control channels or set-up channels. Therefore, a total of 42 channels are used for controlling the system.

In channel 1, the two frequencies available for mobile and cell-site transmit are

1. 825.030 MHz (mobile transmit)
2. 870.030 MHz (cell-site transmit)

In channel 666, the two frequencies available for mobile and cell-site transmit are

1. 844.98 MHz (mobile transmit)
2. 889.98 MHz (cell-site transmit)

Each market (i.e. each city) has two systems for a duopoly market policy with each block having 333 channels.

The 42 set-up channels also called as control channel sets are assigned as follows:

1. Channels 313–333 in block A
2. Channels 334–354 in block B

The voice channels are assigned as follows:

1. Channels 1–312 (312 voice channels) in block A
2. Channels 355–666 (312 voice channels) in block B

The 42 set-up channels (control channel sets) are assigned in the middle of all the assigned channels to facilitate scanning of those channels by frequency synthesizers. In the new additional spectrum allocation of 10 MHz, an additional 166 channels are assigned.

Since a first channel is assigned below 825 MHz (or 870 MHz), in the future, additional channels will be numbered up to 849 MHz (or 894 MHz) and will then circle back. The last channel number is 1023. There are no channels between 799 and 991 channels. New additional spectrum allocations are shown in Figure 12.2.

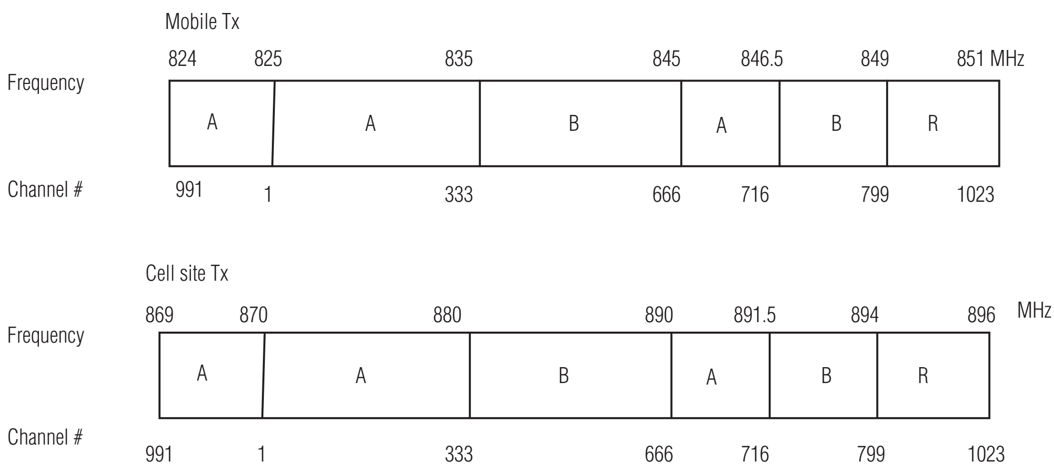


Figure 12.2 New additional spectrum allocations

12.2.2 Grouping into subsets

Since there are 21 set-up channels for each block, it is logical to group 312 voice channels into 21 subsets of 15 voice channels each (the last subgroup has 12 voice channels only). Each subset then consists of 16 channels – –15 voice channels and one control channel. In each set, the closest adjacent channel is 21 channels away as shown in Figure 12.1. The channel separation is provided in such a way that it is sufficient to meet the adjacent channel isolation requirement.

The 16 channels in each subset can be mounted on a frame and connected to a channel combiner. Wide separation between adjacent channels is required for meeting the requirement of minimum isolation.

12.3 Set-up channels

The set-up channels are also called as control channels. They are designated to set-up calls in the system. But even without set-up channels, a system could work where all the channels are in either block A or block B which will be used as voice channels. If a frequency reuse technique is applied to a cellular system, the set-up channels act as control channels.

The set-up channels are classified with respect to their application. They are

- access channels
- paging channels

12.3.1 Access channels

Access channels are used for calls originating from mobile. When a mobile set scans all the 21 set-up channels (in block A), two conditions are considered:

- If no set-up channels are operational in block A, then the mobile unit switches automatically to block B.
- If there is a strong set-up channel with no message detected then within the second setup, it will be selected by the scanner.

12.3.2 Paging channels

Paging channels are used for calls originating from land. Every cell site is assigned its own control or set-up channels. For example, *FOCC* is the forward set-up channel in which every cell site are mainly used to page the mobile unit with control message of same mobile station.

- The same message is transmitted by different set-up channels and there is no simulcast interference.
- A better algorithm is used to page from all the cell sites.

12.4 Traffic and channel assignment

The vehicular traffic density of a coverage area is a critical element and must be determined before a system is designed. This traffic pattern in busy hours can be confined to different zones within the service area. If the traffic pattern predominates over the simple signal coverage pattern, then the cell-site selection will be based on the traffic pattern. Choice of the initial cell sites should be based on the signal covered in zones of heavy vehicular traffic. This means that the cell site would most likely be located at the centre of those zones.

If call traffic data are collected while the system is operating, then we can update the call traffic data at each cell site to correlate with the vehicular traffic data. This information will be useful for determining whether new cell splitting is needed. If it is, then we must determine how many radios should be installed at the new site and where it is to be located. These decisions are all related to frequency channel assignment.

12.4.1 Fixed channel assignment

In FCA, each cell assigns its own frequency channel to the mobile subscribers within its cell. Channel assignment is primarily based on causing least co-channel and adjacent channel interference in the cellular system. The channel assignment for each voice call is determined by MTSO on a short-term basis. In a FCA, the set-up and voice channels are usually assigned to the cell site for relatively long periods. Channels in a channel set are usually 21 channels apart and must meet minimum frequency spacing requirements of a multi-channel transmitter combiner. Channels are usually numbered in order of increasing frequency. Regardless of the number of channels in a channel set, the highest channel set is frequency adjacent to the lowest channel set.

The following are the advantages of FCA:

- Fixed parameters (power, frequency) for transceivers.
- Good performance under uniform- and/or high-traffic loads as cells independently decide their channel allocation decisions.
- If each cell is allocated to a pre-determined set of voice channels then the call is *blocked* and all the channels are occupied.
- Borrowing strategy: A cell is allowed to borrow channels from a neighbouring cell if all of its own channels are occupied.
- Mobile switching centre (MSC) supervises the borrowing procedure to ensure no disrupting calls or interference with any of the calls in progress in the donor cell.

12.4.2 Dynamic channel assignment

In dynamic channel assignment (DCA), the central common pool maintains all the available channels. Channels are assigned dynamically as new requests for radio resource (for a fresh originating call or handoff of existing call) arrive in the system. This also implies that when the use of assigned channel is completed, the channel currently in use is returned to the central pool.

In order to achieve optimum system capacity with limited frequency spectrum, many DCA schemes have been proposed to allocate the channels more efficiently. In a cellular system, a mobile subscriber moves from one cell to another and continuation of communication link is ensured with suitable handoff mechanism. This demands for additional and flexible radio resources utilization. One way to ascertain minimum blocking probability is to increase the number of channels per cell. Then every cell would expect to have a large number of channels. However, because a limited frequency band is allocated for cellular communication, there is an upper limit to the maximum number of channels, thereby restricting the number of available channels that can be assigned to each cell. Another way is non-uniform FCA based on the amount of traffic expected to be served in different cells as per the statistical traffic data. Another alternative arrangement could be dynamic assignment of channels to different cells, as per the current demand. This may be done from a central pool of channels, or a combination of both FCA and DCA.

The following are the advantages of DCA:

- No fixed channels are assigned to each cell.
- Out of the available channels, any channel can be assigned to any cell on need basis.
- The serving base station (BS) requests a channel from the MSC whenever a call request is made.
- Consideration of the likelihood of future blocking in the cell, the frequency use of the candidate cell, the reuse distance of the channel, and other cost functions.
- MSC needs to collect real-time data on channel occupancy, traffic distribution, and received signal strength indicator (RSSI) of all channels on a continuous basis, which increases storage and computational load on the system.

12.4.3 Channel sharing scheme

When a particular cell needs more channels in order to meet the increased traffic demand, the channels of another sector at the same cell site can be shared to meet the short-term overload traffic. Channel sharing can be done from one of the two adjacent sectors of the neighbouring cells in a sectored cellular system configuration. Shared channels can be returned back when the channels become available in the shared sector. This scheme is called the ordered channel assignment scheme with rearrangement.

An alternate scheme is channel assignment with sharing and reassignment. This scheme makes sure that channel-sharing arrangement causes minimum impact on call-blocking probability in neighbouring cells. Reassignment of shared channels is done in order to provide maximum assistance to the neighbouring cells to meet the temporary increased traffic demand. The channel can also be ordered based on which channels provide better performance. Channel sharing always increases the trunking efficiency of the channels.

12.4.4 Channel-borrowing scheme

The channel-borrowing scheme is used for slow growing systems on a long-term basis as an alternate to the costly cell-splitting technique to handle increased traffic. One approach to address increased traffic of either mobile originating calls or handoff calls in a cell is to borrow free available channels from neighbouring cells. A simple channel-borrowing scheme implies that if all channels assigned to a cell have already been used, then additional channels as per the current need can be borrowed from any other cell (preferably adjacent cells) that has some free unused channels. In addition, the central cell site can also borrow channels from neighbouring cells. The extent of borrowing channels depends on the traffic density in the area.

12.5 Channel assignment algorithms

There are several algorithms available for non-FCA. They are listed below:

- *Fixed channel algorithm* (FCA): This algorithm is the most commonly adopted in many cellular systems. Here, each cell assigns its own radio channels to the vehicles within its cell.
- *Dynamic channel algorithm* (DCA): Here no fixed channels are assigned to each cell. Therefore, any channel in a composite of 312 radio channels can be assigned to the mobile unit. This

means that a channel is assigned directly to a mobile unit. On the basis of overall system performance, the DCA can also be used during a call.

- *Hybrid channel algorithm* (HCA): This is a combination of FCA and DCA. A portion of the total frequency channels will use FCA and the rest will use DCA.
- *Borrowing channel algorithm* (BCA): It uses FCA as a normal assignment condition. When all fixed channels are occupied, then the cell borrows channels from the neighbouring cells.
- *Forcible-borrowing channel algorithm* (FBCA): In this case, if a channel is in operation and the situation warrants it, then channels must be borrowed from the neighbouring cells and at the same time another voice channel will be assigned to continue the call in the neighbouring cell.

Channel cannot be borrowed frequently from adjacent cells.

12.6 Simulation process and results

On the basis of FBCA, FCA, and BCA algorithms, a seven-cell reuse pattern with an average blocking of 3 per cent is assumed.

The simulation model is described as follows:

- Randomly select the cell (among 41 cells).
- Determine the state of the vehicle in the cell (idle, off-hook, on-hook, and handoff).

Method of implementation

There are many different ways of implementing FBCA. In a general sense, FBCA can also be applied while accounting for the forcible borrowing of the channels within a fixed-channel set to reduce the chance of co-channel assignment in a reuse cell pattern.

Reuse distance

The FBCA algorithm is based on assigning a channel dynamically but obeying the rule of reuse distance. The distance between the two cells is reuse distance, which is the minimum distance at which no co-channel interference would occur. If all the channels in the neighbouring cells cannot be borrowed because of interference problems, the FBCA stops.

12.6.1 Blocking

Two types of blocking are possible in FBCA algorithm:

1. *Average blocking*: It happens mostly in non-uniform traffic.
2. *Handoff blocking*: It happens mostly in uniform traffic.

Average blocking

Two average blocking cases illustrating this simulation are shown in Figure 12.3. In a uniform traffic condition (Fig. 12.3(a)), the 3 per cent blocking of both BCA and FBCA will result in a load increase of 28 per cent, compared to 3 per cent blocking of FCA. There is no difference between BCA and FBCA when a uniform traffic condition exists. In a non-uniform traffic distribution (Fig. 12.3(b)), the load increase in BCA drops to 23 per cent and that of FBCA increases to 33 per cent, as at an average blocking of 3 per cent. The load increase can be utilized in another way by reducing the number of channels. The percent increase in load is same as the percent reduction in the number of channels.

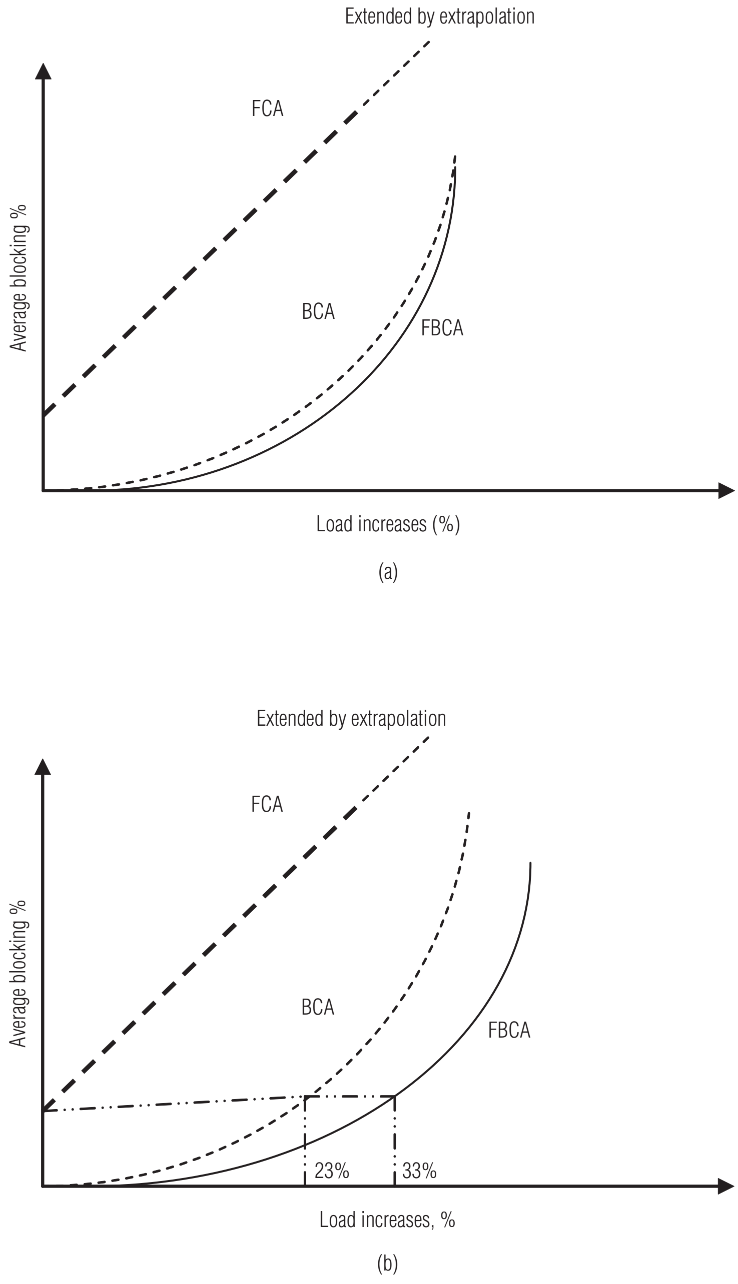


Figure 12.3 Averaging blocking in spatially (a) uniform and (b) non-uniform traffic distribution

Handoff blocking

Handoff blocking is not considered as the regular cell blocking which can only occur at the call set-up stage. In both BCA and FBCA, load is increased almost equally to 30 per cent, as compared to FCA at 3 per cent handoff blocking in uniform traffic (shown in Fig. 12.4(a)).

For a non-uniform traffic distribution, the load increase of both BCA and FBCA at 4 per cent blocking is about 50 per cent (Fig. 12.4(b)), which is a big improvement, considering the reduction in interference and blocking. Otherwise, there would be multiple effects from interference in several adjacent cells.

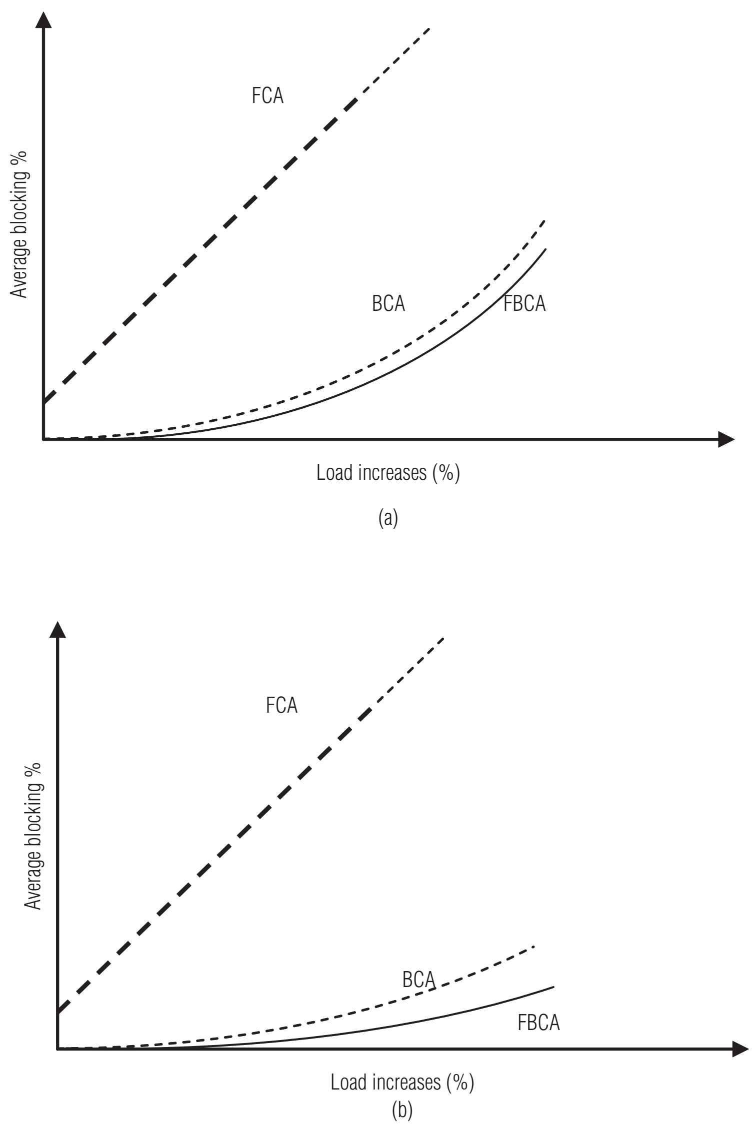


Figure 12.4 Handoff blocking in spatially (a) uniform and (b) non-uniform traffic distribution

12.7 Summary

In a cellular system, frequency management and channel assignment are essential in order to achieve the basic objectives of spectrum utilization as well as adaptability to traffic density.

- Frequency management includes operations such as
  - designating the set-up and voice channels
  - numbering the channels
  - grouping voice channels into subsets, and so on
- Channel assignment does the allocation of specific channel to the cell sites and mobile units:
  - The channel assignment can be done in two ways. They are as follows:
    - *Short-term assignment*: During a call, a particular channel is assigned to a mobile unit on a short-term basis, that is the channel is assigned only during the call duration. After the call, the channel will be altered.
    - *Long-term assignment*: A fixed channel set consisting of one or more subsets is assigned to a cell site on a long-term basis.
  - The set-up channels are classified as access channels and paging channels.
- *Fixed channel assignment/(algorithm)* (FCA): This algorithm is the most commonly adopted in many cellular systems. Here, each cell assigns its own radio channels to the vehicles within its cell.
- *Dynamic channel assignment/(algorithm)* (DCA): Here, no fixed channels are assigned to each cell. Therefore, any channel in a composite of 312 radio channels can be assigned to the mobile unit. This means that a channel is assigned directly to a mobile unit. On the basis of overall system performance, DCA can also be used during a call.
- *Hybrid channel assignment/(algorithm)* (HCA): This is a combination of FCA and DCA. A portion of the total frequency channels will use FCA and the rest will use DCA.
- *Borrowing channel assignment/ (algorithm)* (BCA): It uses FCA as a normal assignment condition. When all fixed channels are occupied, then the cell borrows channels from the neighbouring cells.
- *Forcible-borrowing channel assignment/(algorithm)* (FBCA): Here, if a channel is in operation and the situation warrants it, channels must be borrowed from the neighbouring cells and at the same time, another voice channel will be assigned to continue the call in the neighbouring cell.
- Channel cannot be borrowed frequently from adjacent cells.
- There are two types of blocking possible in FBCA algorithm.
  - *Average blocking*: It happens mostly in non-uniform traffic.
  - *Handoff blocking*: It happens mostly in uniform traffic.
- Queuing of handoff calls can increase traffic capacity.

Example problem 12.1

A full-duplex wireless cellular system is allocated a total spectrum of 20 MHz and each simplex channel has 25 kHz RF bandwidth. Determine the following:

(a) Total number of full-duplex channels available.
(b) Number of channels per cell site if $K = 4$ cell reuse pattern is employed.

Solution

Given data:
Total allocated RF spectrum bandwidth = 20 MHz
Channel bandwidth per simplex channel = 25 KHz

(a) To determine number of full-duplex channel
Channel bandwidth per simplex channel = 25 kHz
Number of channels in a duplex link = 2
Therefore, duplex channel bandwidth = 25 × 2 = 50 kHz
Number of full-duplex channels = total bandwidth/duplex channel bandwidth
Number of full-duplex channels = 20 MHz/50 kHz
Hence, *total number of duplex channels = 400 channels.*

(b) To determine number of channels per cell site
Number of cells in one cluster, $K = 4$ (given)
Number of channels per cell site = total number of channels/K = 400/4 =100.
Hence, *number of channels per cell-site = 100 channels.*

Example problem 12.2

A full-duplex wireless cellular system is allocated a total spectrum of 25 MHz and each simplex channel has 15 kHz RF bandwidth. Determine the following:

(a) Total number of full-duplex channels available.
(b) Number of channels per cell site if $K = 7$ cell reuse pattern is employed.

Solution

Given data:
Total allocated RF spectrum bandwidth = 25 MHz
Channel bandwidth per simplex channel = 15 KHz

(a) To determine number of full-duplex channel
Channel bandwidth per simplex channel = 15 kHz
Number of channels in a duplex link = 2
Therefore, duplex channel bandwidth = 15 × 2 = 30 kHz
Number of full-duplex channels = total bandwidth/duplex channel bandwidth
Number of full-duplex channels = 25 MHz/30 kHz
Hence, *total number of duplex channels = 833 channels.*

(b) To determine number of channels per cell site
Number of cells in one cluster, $K = 7$ (given)
Number of channels per cell site = total number of channels/K = 833/7 =119.
Hence, *number of channels per cell site = 119 channels.*

Example problem 12.3

Calculate the number of set-up and voice channels per cell for a cellular system having a total spectrum allocation of 60 MHz which uses two 25 kHz simplex channels to provide full duplex set-up and voice channels. Assume that the system is designed with nine-cell frequency-reuse pattern and 1 MHz of the total spectrum is exclusively allocated for set-up channels.

Solution

Given data:
Total allocated RF spectrum bandwidth = 60 MHz
Channel bandwidth per simplex channel = 25 kHz
Number of cells in one cluster = 9
Allocated RF bandwidth for set-up channels = 1 MHz

- To determine duplex channel bandwidth
  Channel bandwidth per simplex channel = 25 kHz
  Number of channels in a duplex link = 2
  Therefore, duplex channel bandwidth = $25 \times 2 = 50$ kHz
- To determine number of full-duplex channels
  Number of full-duplex channels = Total bandwidth/duplex channel bandwidth
  Number of full-duplex channels = 60 MHz/50 kHz
  Hence, total number of duplex channels = 1,200 channels
- To determine total number of set-up channels
  Duplex channel bandwidth = 50 kHz (As calculated in Step 1)
  Total number of available set-up channels = 1 MHz/50 kHz = 20
- To distribute number of set-up channels per cell
  Total available 20 number of set-up channels can be distributed among nine cells in a cluster as *7 cells can have 2 set-up channels each*, and remaining *2 cells can then have 3 set-up channels each* which means a total $(7 \times 2 + 2 \times 3)$ of 20 set-up channels in a system.
- To determine total number of voice channels
  Available RF bandwidth for voice channels = 60 MHz – 1 MHz= 59 MHz
  Total number of available voice channels = 59 MHz/50 kHz = 1180
- To distribute number of voice channels per cell
  Total 1180 number of available voice channels can be distributed among nine cells in a cluster as *8 cells can have 131 voice channels each*, and remaining *1 cell can then have 132 voice channels* which means a total $(8 \times 131 + 1 \times 132)$ of 1180 voice channels in a system.

Review questions

1. What is meant by frequency management and channel assignment?
2. What is known as FOCC?
3. Define blocking? What are the types of blocking? What is their significance?

4. Explain briefly about FCA.
5. What is the importance of frequency management chart?
6. What are the methods for reducing interference?
7. Write the procedure to allot the channels for the travelling mobile units. (Refer Section 12.4)
8. Explain the channel assignment to the cell sites based on the adjacent channels. (Refer Section 12.4)
9. Give the structure of the channels in 800 MHz system with frequency ranges. (Refer Section 12.2.1)
10. Explain how set-up channels act as control channels in a cellular system? (Refer Section 12.3)
11. What are the advantages of reuse-partition scheme? (Refer Section 12.2)
12. Explain the following,
    (i) Channel sharing (Refer Section 12.4.3)
    (ii) Channel borrowing (Refer Section 12.4.4)
    (iii) Underlay and overlay (Refer Section 12.3)
    (iv) Set-up channel (Refer Section 12.3.2)
    (v) Paging channel (Refer Section 12.2.1)
    (vi) Voice channel (Refer Section 12.2.1)
13. How a reuse-partition scheme reduces the number of cell sites? Explain it with suitable examples. (Refer Section 12.2.1)
14. Differentiate between FCA and non-FCA in detail. (Refer Section 12.4.1)
15. Discuss the concept of frequency management concern to the numbering the channels and grouping into the subset. (Refer Sections 12.2 and 12.3)
16. Explain in brief the grouping of voice channels into subsets. (Refer Section 12.2.2)
17. Explain how paging channels are used for the land originating calls? (Refer Section 12.3.2)
18. What do you understand by non-FCA? Describe the corresponding algorithms. (Refer Section 12.5)
19. Describe the grouping of the voice, set-up, and paging channels. (Refer Sections 12.2.2, 12.3, and 12.3.2)
20. Explain how the 666 channels are divided into groups? (Refer Section 12.2.1)
21. What are the different techniques to utilize the frequency spectrum? Explain with brief explanation. (Refer Section 12.2.4)
22. Explain the forcible-borrowing channel assignment algorithm and its implementation? (Refer Section 12.5.5)

Objective type questions and answers

1. The main function of the frequency management is
   (a) increasing gain
   (b) increasing power
   (c) dividing total number of channels into subsets
   (d) adding the given number of channels
2. Numbering the channel is done by the following channel
   (a) RVC (b) RCC (c) FVC (d) FCC

3. What basis in a fixed channel set that consists of one or more subsets is assigned to a cell site?
   (a) short-term basis (b) in dynamic basis
   (c) long-term basis (d) temporary basis
4. What is the method of channel assignment to a mobile unit during call in progress?
   (a) short-term basis (b) in dynamic basis
   (c) long-term basis (d) temporary basis
5. Allocation of specific channels to a cell site is known as
   (a) frequency management (b) frequency allotment
   (a) channel assignment (d) channel modelling
6. What is the other name of set-up channels?
   (a) reverse channels (b) forward channels
   (c) control channels (d) traffic channel
7. In set-up channels every two-way channel contains a ___________bandwidth.
8. In normal case, ____ set-up channels can be used for paging and accessing.
9. All the set-up channels carry only ________.
10. The access channel can be designated by ___ as a channel apart from set-up channels in the cell site.

Answers: 1. (c), 2. (d), 3. (c), 4. (a), 5. (c), 6. (c), 7. 30 MHz, 8. One, 9. Data information, 10. MTSO.

Open book questions

1. Explain the different channel assignment algorithms in detail.
2. Explain the significance of FBCA algorithm.
3. What is the need of set-up channels? Classify them.
4. Why is it necessary to form frequency channel groups?
5. How is voice channels assigned for establishment of voice calls?
6. Which channel-assignment approach can be effectively deployed to handle increased traffic situation?
7. On what basis channels are assigned in an overlapped cell-based system?
8. Explain in detail access channels and operational techniques.
9. The U.S. AMPS system is allocated 50 MHz of spectrum in the 800 MHz range and provides 832 channels. Forty-two of those channels are control channels. The forward channel frequency is exactly 45 MHz greater than the reverse channel frequency.
   (i) Assume a base station transmits control information on channel 352, operating at 880.560 MHz. What is the transmission frequency of the subscriber unit on transmitting on channel 352?
   (ii) The A-side and B-side cellular carriers evenly split AMPS channels. Find the number of voice channels and the number of control channels for each carrier.
   (iii) Suppose that you are chief engineer of a cellular system using seven-cell reuse. Purpose a channel assignment strategy for a uniform distribution of user throughout your cellular system specifically, assume that each cell has three control channels (1200 sector is employed) and specify the number of voice channels you would assign to each control in your system.

(iv) For an ideal hexagonal cellular layout which has identical cell coverage, what is the distance between the centers of two nearest co-channel cells for seven-cell reuse and for four-cell reuse?

10. What are the common principles of channel allocation schemes?

Further reading

Beach M., *Propagation and System Aspects*, University of Bristol, Future Communication Systems course, April 1994.

Classen, F., Meyr, H., and Sehier, P. "Maximum likelihood open loop carrier synchronizer for digital radio," Proceedings ICC'93, pp. 493–497, 1993.

Lee, W. C. Y. "Elements of Cellular Radio System," IEEE Transactions on Vehicular Technology, 35 (May 1986): pp. 48–56.

Lee, W. C. Y. *Mobile Cellular Telecommunications System*. New York: McGraw-Hill, 1989.

Van Nee, R., and Prasad R. *OFDM for Wireless Multimedia Communications*, Artech House, Boston, pp. 80–81, 2000.

Modulation Techniques 13

13.1 Introduction

Most communication systems fall into one of the two categories: bandwidth efficient or power efficient. Bandwidth efficiency describes the ability of a modulation scheme to accommodate data within a limited bandwidth. Power efficiency describes the ability of the system to send reliable information at the lowest practical power level.

The aim of modulation is to transfer a source data over a channel in a way most suitable for this channel. That is, the original data should be translated into a form that is compatible with the channel. Since the scope of this book is wireless communication, only radio channel is considered. The modulation process can be divided into two stages: *baseband modulation* and *bandpass modulation*. In this way, the *baseband modulation* consists of translating the original data (analogue or digital) into waveforms of low frequency and *bandpass modulation* consists of modifying the characteristics of high-frequency carrier wave, or simply *carrier*, in accordance with waveforms obtained at the output of the baseband modulation process.

The types of modulation techniques are classified into analogue *modulation* and *digital modulation*. The aim of *analogue modulation* is to transfer analogue signal, such as speech or TV signal, over bandpass channel and in this case there is an infinite number of possible states of analogue signal to modulate some parameter of a carrier. The changing of the carrier parameter in this case is continuous in time in accordance with the changing of original analogue signal. The examples of analogue modulation are *amplitude modulation* (AM) and *frequency modulation* (FM). In the case of **digital modulation**, a digital bit stream should be transferred over the bandpass channel. The examples of different digital modulation types are *amplitude-shift keying* (ASK), *frequency-shift keying* (FSK), and *phase shift keying* (PSK). Usually, the term *shift keying* stands for modulation in names of different modulation types when we are referring to digital modulation. Shifting here means the changing (modulation) of some parameter.

Most first generation cellular systems such as the advanced mobile telephone system (AMPS) use analogue FM. Digital modulation schemes, however, are the obvious choice for current wireless systems, especially if data services such as wireless multimedia are to be supported. Digital modulation can also improve spectral efficiency, because digital signals are more robust against channel impairments. Spectral efficiency is a key attribute of wireless systems that must operate in a crowded radio frequency spectrum. To achieve high spectral efficiency, the modulation schemes must have high bandwidth efficiency as measured in units of bits per second per Hertz of bandwidth. Various wireless communication systems such as cellular telephones operate on the principle of frequency reuse, where the carrier frequencies are reused at geographically separated locations. The link quality in these systems is limited by co-channel interference.

Hence, modulation schemes that are identified must be both bandwidth efficient and capable of tolerating high levels of co-channel interference.

13.2 Frequency modulation

In FM, the instantaneous frequency of the radio-frequency wave (i.e. carrier) is varied in accordance with the modulating signal. The amount by which the frequency departs from the average is controlled by the amplitude of the modulating signal. This variation is referred to as the frequency deviation of the frequency-modulated wave. Figure 13.1 illustrates the FM. The frequency variation of the carrier is represented by "1"s or "0"s. A "0" is represented by the original carrier frequency and "1" by a much higher frequency (the cycles are spaced closer together).

According to the definition of the FM, the frequency of the carrier signal is varied linearly with the instantaneous value of the modulating signal $m(t)$. With the sinusoidal modulating signal $m(t) = A_m \cos(2\pi f_m t)$, the FM can be represented mathematically by the expression

$$X(t) = A\cos[2\pi f_c t + 2\pi k_f \int_0^t m(t)\,dt] \tag{13.1}$$

where k_f is the frequency sensitivity of the FM signal measured in Hz/V.

The relationship between the amplitude of the modulating signal and the bandwidth of the FM transmitted signal is given by

$$\beta = (k_f A_m)/f_m = \Delta f/f_m \tag{13.2}$$

where β = modulation index of FM signal
A_m = peak value of the modulating signal
f_m = maximum frequency of the modulating signal
Δf = peak frequency sensitivity of the FM signal

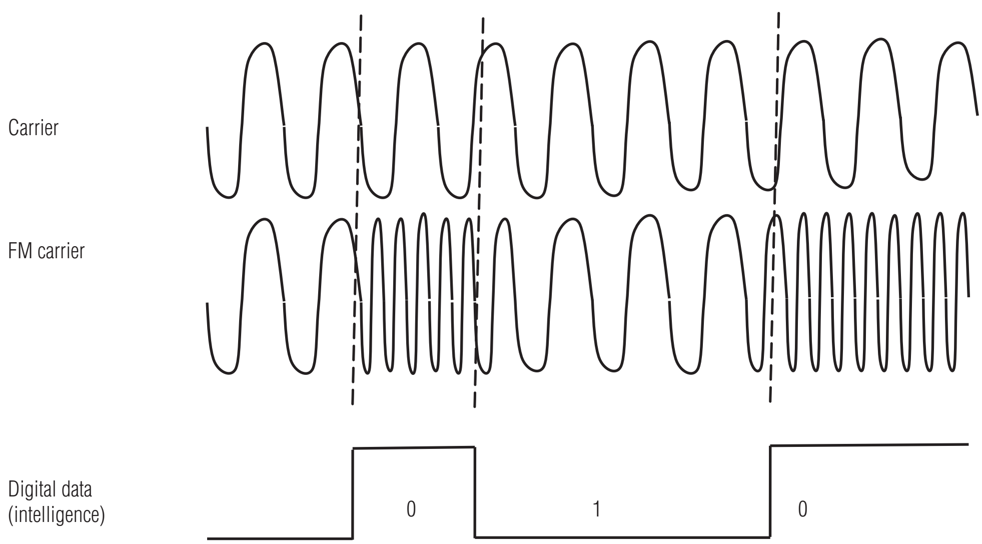


Figure 13.1 Frequency modulation on an RF carrier

The RF bandwidth of FM depends on the value of β. For $\beta < 1$, the narrowband FM is generated and the spectral width is $2f_m$. If $\beta >> 1$, then wideband FM is generated where the spectral occupancy is slightly greater than $2(\Delta f)$. In general, the approximate bandwidth of an FM signal is

$$W \approx 2(\Delta f) + 2f_m = 2(\Delta f)\left(1 + \frac{1}{\beta}\right) \tag{13.3}$$

This relation is known as Carson's rule. Unfortunately, typical analogue cellular radio systems use a modulation index in the range of $1 \le \beta \le 3$ where Carson's rule is not accurate. Furthermore, the message waveform $m(t)$ is not a pure sinusoid. Hence, Carson's rule does not directly apply in this situation. In analogue cellular systems, the waveform $m(t)$ is obtained by first companding the speech waveform and then hard limiting the resulting signal. The purpose of the limiter is to control the peak frequency deviation, Δf. The limiter introduces high frequency components that must be removed with a low pass filter (LPF) prior to modulation. To estimate the bandwidth occupancy, we first determine the ratio of the frequency deviation, Δf, corresponding to the maximum amplitude of $m(t)$ and the highest frequency component B that is present in $m(t)$. These two conditions are the most extreme cases and the resulting ratio, $D = \Delta f/B$, is called the deviation ratio. Then replace β by D and f_m by B in Carson's rule, giving

$$W \approx 2(\Delta f) + 2B = 2(\Delta f)\left(1 + \frac{1}{D}\right) \tag{13.4}$$

This approximation will overestimate the bandwidth requirements. A more accurate estimate of the bandwidth requirements must be obtained from simulation or measurements.

13.3 FM detection techniques

FM is widely used for radio transmissions for a wide variety of applications from broadcasting to general point-to-point communications. Frequency modulation offers many advantages, particularly in mobile radio applications where its resistance to fading and interference is of great advantage. It is also widely used for broadcasting in very high frequencies where it is able to provide a medium for high quality audio transmissions.

We know that the instantaneous frequency of an FM signal is proportional to the amplitude of the modulating signal. Therefore, an FM demodulator should produce an output that is proportional to this instantaneous frequency. The process of removing the information signal from the carrier is termed demodulation or detection.

There is a wide variety of techniques and circuits that can be used including the Foster–Seeley, ratio detectors using discrete components, integrated circuits (ICs) using the phase locked loop (PLL), and quadrature detectors. In recent years, the Foster–Seeley discriminator and the ratio detector have been less widely used. The main reason for this is that they require the use of wound inductors and these are expensive to manufacture. Other types of FM demodulator have overtaken them, mainly as a result of the fact that the other FM demodulator configurations lend themselves more easily to being incorporated into integrated circuits.

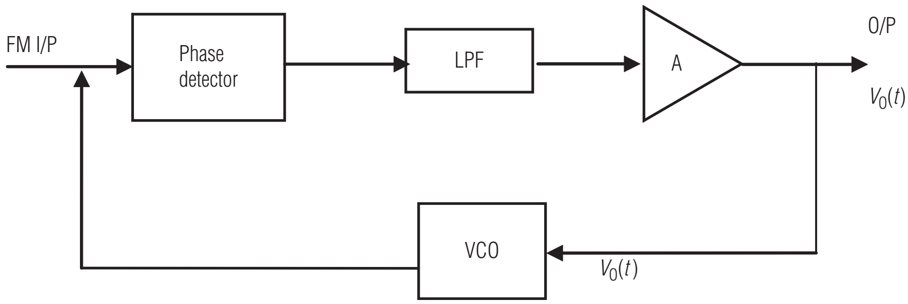


Figure 13.2 PLL block diagram

13.3.1 Phase locked loop

The PLL is a non-linear feedback loop used in signal demodulation applications. PLL FM demodulator or detector is a form of FM demodulator that can be easily made from the variety of PLL ICs and are found in many types of radio equipment ranging from broadcast receivers to high performance communication equipment.

The block diagram of a PLL FM demodulator is shown in Figure 13.2. The phase detector, voltage controlled oscillator (VCO), and LPF are the basic elements in a PLL.

The phase detector, which is basically a balance modulator, is used to compare the FM input to the output of a VCO. If there is a difference in phase or frequency, the phase detector output varies in proportion to the difference. The frequency component is selected by the LPF, which also removes much of the noise. The filtered signal amplified through amplifier A acts as a control voltage to the VCO, where it results in FM of the VCO frequency. This control voltage is called the error signal. The error voltage forces the VCO frequency to reduce the amount of phase or frequency difference between the VCO and the input.

At some point, the error voltage causes the VCO frequency equal to the input frequency then the PLL is said to be in locked condition. When the loop is in lock, the VCO frequency follows or tracks the incoming frequency. For example, when the instantaneous frequency increases, the control voltage will cause the VCO frequency to increase. The control voltage to the VCO will endeavour to keep the VCO frequency locked to the incoming carrier, and thus will be an exact copy of the original message.

13.3.2 Quadrature detector

The quadrature detector is the most widely used FM detector. In the quadrature demodulator, the modulated carrier is passed through an LC tank circuit that shifts the unmodulated carrier signal by 90°. This phase shift is either greater or less than 90° depending on the direction of deviation. The quadrature FM detection is mainly used in FM radio systems and TV audio demodulation.

Figure 13.3 shows a quadrature FM detector. The frequency modulated signal is applied through a very small capacitor (C_s) to the parallel tank circuit, which is adjusted to resonate at the centre carrier frequency. At resonance, the tank circuit acts as a resistance with a high value. The C_s has a very high reactance compared to the tank circuit impedance. Therefore, the output across the tank circuit is very close to the 90° at the carrier frequency and leads the output.

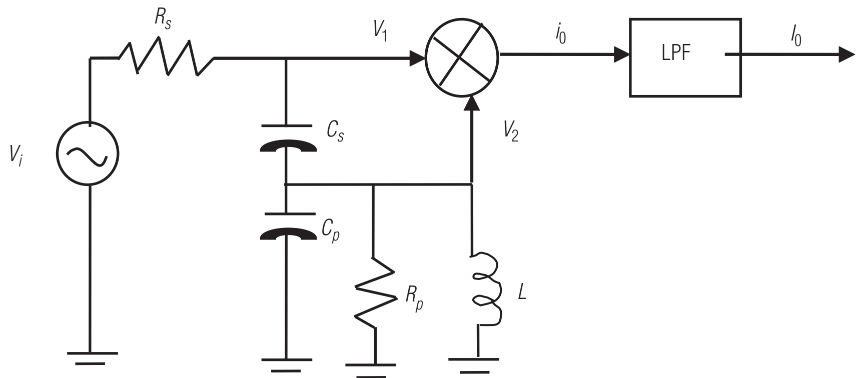


Figure 13.3 Quadrature demodulator block diagram

The two-quadrature signals are then fed to a phase detector. The phase detector compares the phase of the IF signal (V_1) to V_2. The output (I_0) of the phase detector is a series of pulses whose width varies with the phase shift between the two signals. Then, the output signal is averaged in a RC low-pass filter to extract the original modulating signal.

13.4 Digital modulation

Digital modulation is the transmission of digitally modulated analogue carrier signals in wireless communication systems. Digital modulation systems offer several outstanding advantages over the analogue modulation systems such as easier and faster signal processing, greater noise immunity, and robustness to channel impairments. Digital modulation is a simple case of transmitting digital data using analogue signals. The process of modulation involves operation on one of the three characteristics of the analogue carrier signal: amplitude, frequency, and phase.

There are two major categories of digital modulation. One category uses a constant amplitude carrier and carries the information in phase or frequency variations, known as PSK or FSK. The vast majority of frequency hopping wireless LAN and spread spectrum-based systems today employs simple FSK modulation schemes. The other category conveys the information in carrier amplitude variations and is known as ASK.

13.4.1 Amplitude shift keying

In ASK, the two binary values, logic 1 and logic 0, of the information data are represented by two different amplitudes of the carrier signal, $c(t) = A_c \cos(2\pi f_c t)$.

The ASK signal can be mathematically expressed as

$$X_{\text{ASK}}(t) = \begin{Bmatrix} A_c \cos(2\pi f_c t) \text{for binary } 1 \\ 0 \text{ for binary } 0 \end{Bmatrix} \tag{13.5}$$

The example of ASK modulation is depicted in Figure 13.4(a). This is the simplest form of a digital-modulation scheme. However, it is an inefficient digital-modulation technique. ASK is susceptible to sudden amplitude variations due to noise, multipath propagation, and interference.

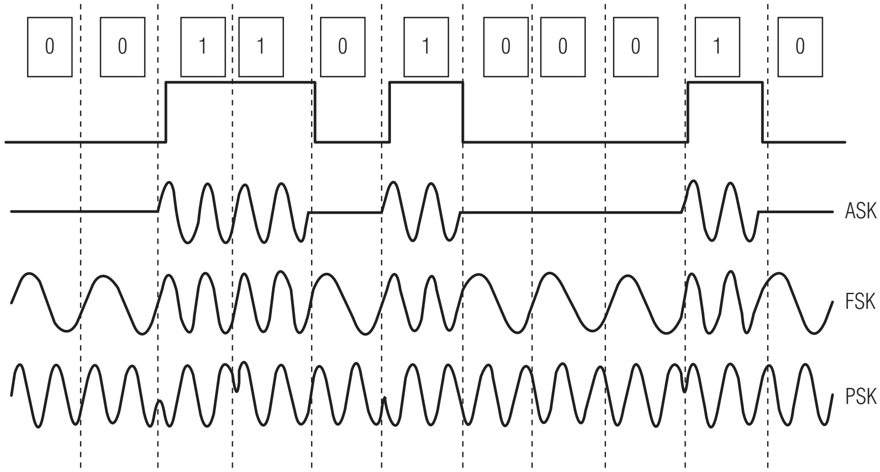


Figure 13.4 Digital modulation types (a) ASK (b) FSK (c) PSK

So ASK is not used in wireless transmission. It requires low bandwidth and is typically used only up to 1200 bps data rate on voice-grade lines as in telemetry. It is used to transmit digital data over optical fibre for LED transmitters, and wireless infrared transmissions using a directed beam or diffused light in wireless LANs applications.

The probability of error (P_e) is the main parameter which improves the performance of a digital modulation scheme. P_e of a ASK modulation scheme is given as

$$P_e = \frac{1}{2} erf_c \sqrt{\frac{E_b}{4N_0}} \tag{13.6}$$

where N_0 = noise density
E_b = bit energy

Therefore, P_e in ASK decreases with increase in the ratio $E_b/4N_0$. Hence, the P_e depends only upon the signal energy and not on its shape or any other parameter.

13.4.2 Frequency shift keying

The most common form of FSK is binary FSK (BFSK), in which the two binary values, logic 1 and logic 0, of the information data are represented by two different frequencies of the carrier signal, $c(t) = A_c \cos(2\pi f_c t)$ near the carrier frequency f_c.

The FSK signal for one-bit duration can be mathematically expressed as

$$X_{\text{BFSK}}(t) = \begin{Bmatrix} A_c \cos(2\pi f_1 t) & \text{for binary 1} \\ A_c \cos(2\pi f_2 t) & \text{for binary 0} \end{Bmatrix} \tag{13.7}$$

where f_1 and f_2 are typically offset from the carrier frequency f_c, by an equal but opposite values. The example of FSK modulation is depicted in Figure 13.4(b). FSK is less susceptible to error than ASK but it requires larger bandwidth compared to ASK. The peak frequency offset is constant and always at its maximum value, and the highest fundamental frequency is equal to half the information bit rate. The P_e of a FSK modulation scheme is given as

$$P_e = \frac{1}{2} erfc\sqrt{\frac{E_b}{2N_0}} \tag{13.8}$$

Therefore, the P_e in FSK decreases with increase in the ratio $E_b/2N_0$.

M-ary frequency shift keying

In FSK or BFSK, the carrier is shifted between two frequencies. M-ary or multiple frequency shift keying (MFSK) is a higher level version of the FSK modulation technique, in which more than two frequencies are used. In MFSK, each signalling element represents more than one bit. The MFSK signal for one-signal-element duration can be expressed as

$$X_{\text{MFSK}}(t) = A_c \cos(2\pi f_i t) \quad \text{for } 1 \le i \le M \tag{13.9}$$

where $f_i = f_c + (2i - 1 - M) f_d$; f_c being the carrier signal frequency, f_d being the difference frequency, and M being the number of different signal elements. It occupies a bandwidth of $2\,M f_d$. The bandwidth efficiency of an MFSK decreases with increasing M. MFSK signals are bandwidth inefficient. However, since all the M signals are orthogonal, there is no crowding in the signal space and hence the power efficiency increases with M. MFSK is less susceptible to errors. It finds applications in wireless LANs.

13.4.3 Phase shift keying

PSK is a constant amplitude digital modulation scheme. When higher data rates are required in a band-limited channel that cannot be achieved with FSK, PSK is often used. *PSK* comprises the manipulation of a carrier's phase in accordance with the transmitted bit stream. The general expression for PSK is

$$X_{\text{PSK}}(t) = A_c \cos(2\pi f_c t + \phi) \tag{13.10}$$

The P_e of a PSK modulation scheme is given as

$$P_e = \frac{1}{2} erfc\sqrt{\frac{E_b}{N_0}} \tag{13.11}$$

Therefore, the P_e in PSK decreases with increase in the ratio E_b/N_0.

Binary phase shift keying

Binary phase shift keying (BPSK) is the simplest form of a PSK. BPSK is an M-ary modulation scheme, with $M = 2$. In BPSK, phase of the sinusoidal carrier is changed according to the data bit to be transmitted. In this, the two binary values, logic 1 and logic 0, of the information data are represented by two different phases separated by 180° of the carrier signal, $c(t) = A_c \cos(2\pi f_c t)$.

The PSK signal for one-bit duration can be mathematically expressed as

$$X_{\text{BPSK}}(t) = \left\{ \begin{array}{l} A_c \cos(2\pi f_c t) \ \text{ for binary 1} \\ A_c \cos(2\pi f_c t + \pi) \ \text{ for binary 0} \end{array} \right\} \tag{13.12}$$

In a BPSK, the binary symbol "1" and "0" modulate the phase of the carrier. When the symbol is changed, then the phase of the carrier will also be changed by an amount of 180° (i.e. π radians). The representation of a BPSK signal is shown in Figure 13.5. BPSK has good

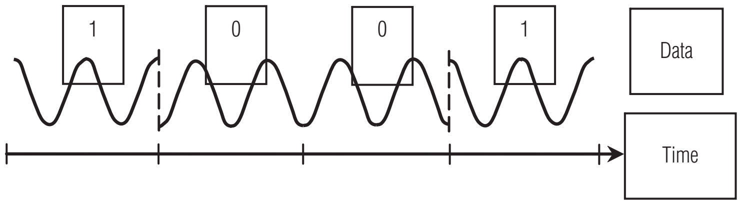


Figure 13.5 Representation of a BPSK signal

signal-to-noise ratio (SNR) but must be demodulated synchronously, which means a reference carrier signal is required to be received to compare with the phase of the received signal, which in turn makes the demodulator complex.

Figure 13.6 illustrates the BPSK modulator. BPSK can be considered as a form of ASK where each non-return to zero (NRZ) data bit of value "0" is mapped into "–1" and each NRZ "1" is mapped into "+1". The resulting signal is passed through a filter to limit its bandwidth and then multiplied by the carrier signal cos ($2\pi f_c t$).

The signal space representation for BPSK is shown in Figure 13.7. It uses two phases which are separated by 180° and can also be termed as 2-PSK. It does not matter exactly where the constellation points are positioned, and in Figure 13.7 they are shown on the real axis, at 0° and 180°.

The performance of a digital modulation scheme depends upon the P_e. The P_e for BPSK is same as that of PSK and can be given as

$$P_e = \frac{1}{2} erfc \sqrt{\frac{E_b}{N_0}} \tag{13.13}$$

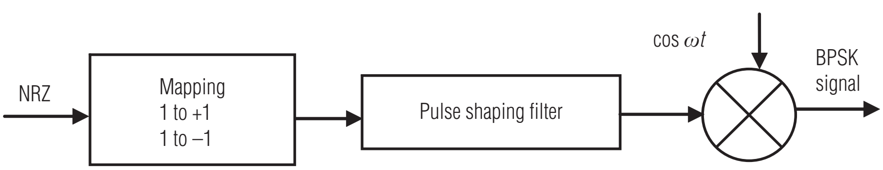


Figure 13.6 BPSK modulator

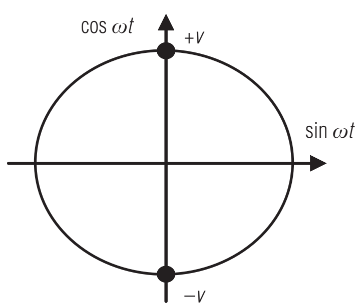


Figure 13.7 Signal constellations for BPSK

In addition, the SNR is given by

$$\mathrm{SNR} = \left(\frac{\mathrm{E}_b}{\mathrm{N}_0}\right) \tag{13.14}$$

If the length of each symbol is T, then the bit energy is given by

$$E_b = A^2 T \tag{13.15}$$

where A = amplitude of the carrier.

The bandwidth efficiency of the BPSK is given as

$$\frac{R}{B_w} = \frac{1}{2} \tag{13.16}$$

where R = data rate and B_w = bandwidth.

Quadrature phase shift keying

In quadrature phase shift keying (QPSK), the information bits are encoded into the phase of the carrier and the amplitude of the signal remains constant. If we define four signals, each with a phase shift differing by $\pi/2$ or 90°, then we have QPSK. In QPSK, there are four possible phases and therefore two bits of information are conveyed within each time slot. The representation of QPSK is shown in Figure 13.8.

The rate of change (baud) in this signal determines the signal bandwidth, but the throughput or bit rate for QPSK is twice the baud rate. The QPSK signal for one symbol duration consisting of two bits, each of which can be mathematically expressed as

$$X_{\mathrm{QPSK}}(t) = \begin{cases} A_c \cos(2\pi f_c t + \pi/4) & \text{for} \quad 11 \\ A_c \cos(2\pi f_c t + 3\pi/4) & \text{for} \quad 01 \\ A_c \cos(2\pi f_c t - 3\pi/4) & \text{for} \quad 00 \\ A_c \cos(2\pi f_c t - \pi/4) & \text{for} \quad 10 \end{cases} \tag{13.17}$$

Figure 13.9 shows the QPSK modulation scheme. The input binary bit stream [b_k], $b_k = \pm 1$; $k = 0, 1, 2, \ldots$, arrives at the modulator at a rate $1/T_b$ bps and is separated into two data streams, $d_I(t)$ and $d_Q(t)$, containing odd and even bits respectively.

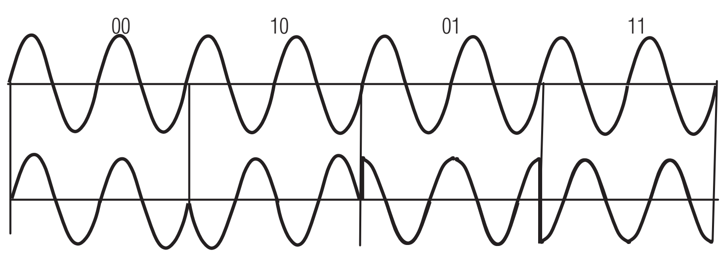


Figure 13.8 Representation of a QPSK signal

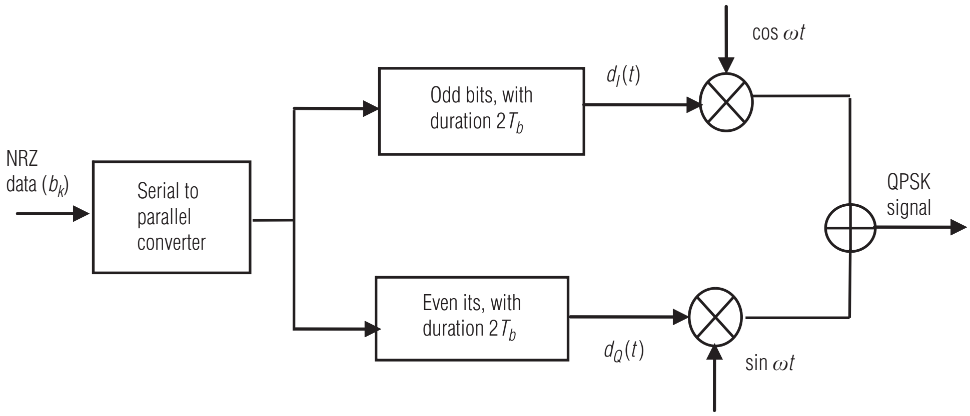


Figure 13.9 QPSK modulator

The modulated QPSK signal $s(t)$ is given by

$$X_{\text{QPSK}}(t) = \frac{1}{\sqrt{2}}\left[d_I(t)\cos(2\pi f_c t) - d_Q(t)\sin(2\pi f_c t)\right] \tag{13.18}$$

It can be seen from Figure 13.10 that the constellation points of QPSK can be represented as two orthogonal sets of BPSK constellation points. It uses four points on the constellation diagram, equispaced around a circle. With four phases, QPSK can encode two bits per symbol as shown in the figure. The mathematical analysis shows that QPSK can be used either to double the data rate compared with a BPSK system while maintaining the same bandwidth of the signal, or to maintain the data-rate of BPSK but halving the bandwidth needed.

The P_e for QPSK, OQPSK, and $\pi/4$-QPSK is same as that of BPSK.

Offset-quadrature phase shift keying

Offset quadrature phase shift keying (OQPSK) is a modified form of QPSK. OQPSK is obtained by introducing a shift or offset equal to one bit delay (T_b) in the quadrature signal $Q(t)$ of QPSK. In OQPSK, the timing of the pulse stream $d_I(t)$ and $d_Q(t)$ is shifted such that the alignment of the two pulse streams is offset by T_b. In this, the pulse stream $d_I(t)$ and $d_Q(t)$ are staggered and

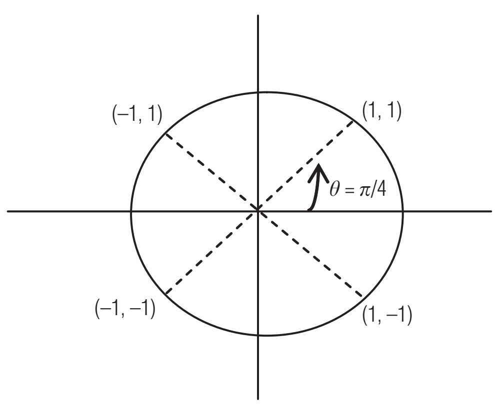


Figure 13.10 QPSK constellation diagram

thus do not change states simultaneously. The possibility of the carrier-changing phase by π degree is eliminated, since only one component can make a transition at one time. Changes are limited to 0° and $\pm\pi/2$ every T_b seconds. Since the phase transitions of 180° are avoided in OQPSK, the signal envelope does not pass through zero as it does in QPSK. The high-frequency components associated with the collapse of the signal envelope are not reinforced. Thus, out-of-band interference is avoided. The OQPSK signal can be expressed as

$$X_{\text{OQPSK}}(t) = \frac{1}{\sqrt{2}}\left[d_I(t)\cos\left(2\pi f_c t + \frac{\pi}{4}\right) + d_Q(t - T_b)\sin\left(2\pi f_c t + \frac{\pi}{4}\right)\right] \tag{13.19}$$

This ensures that I and Q signals have signal transitions at the time instants separated by $T_b/2$, where T_b is the symbol period. The examples of QPSK and OQPSK signals are shown in Figure 13.11. The phase shift of the QPSK signal is π. The maximum possible phase change in OQPSK signal is $\pi/2$. Unlike the QPSK where transitions can occur in every symbol, in OQPSK they can occur more frequently: every half-symbol. Figure 13.12 shows the OQPSK modulation scheme.

The disadvantage of OQPSK is that, the modulated signal transitions are ±90° maximum. It has no 180° phase shift and this result in reduction of out-of-band radiations. However, the abrupt phase transitions still remain. Another disadvantage of OQPSK is that changes in the output phase occur at twice the data rate in either I or Q channels. Consequently, the minimum bandwidth is twice that of QPSK for a given transmission data rate.

π/4-QPSK modulation

$\pi/4$-QPSK comprises two QPSK rotated by $\pi/4$. In this the phase is restricted to fluctuate between $\pm\pi/4$ and $\pm(3\pi)/4$ rather than the $\pm\pi/2$ phase changes for OQPSK. In BPSK and QPSK, the input data stream is encoded in the absolute position in the constellation. In $\pi/4$-QPSK, the input data stream is encoded by the changes in the amplitude and direction of the phase shift. The example of $\pi/4$-QPSK signal is represented in Figure 13.13.

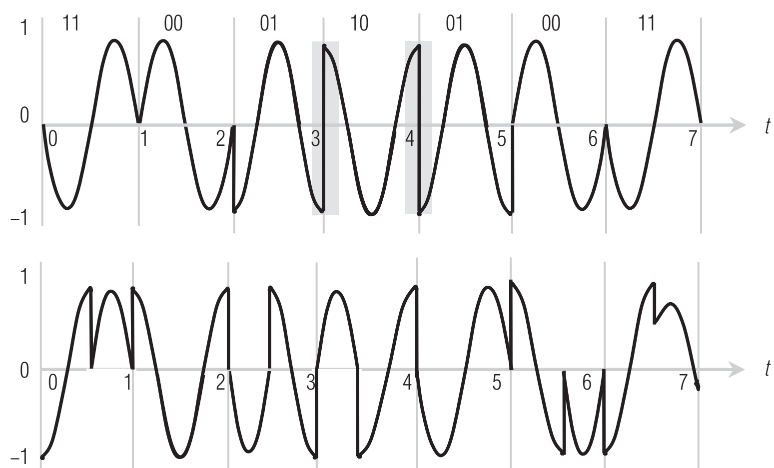


Figure 13.11 (a) QPSK signal (b) OQPSK signal

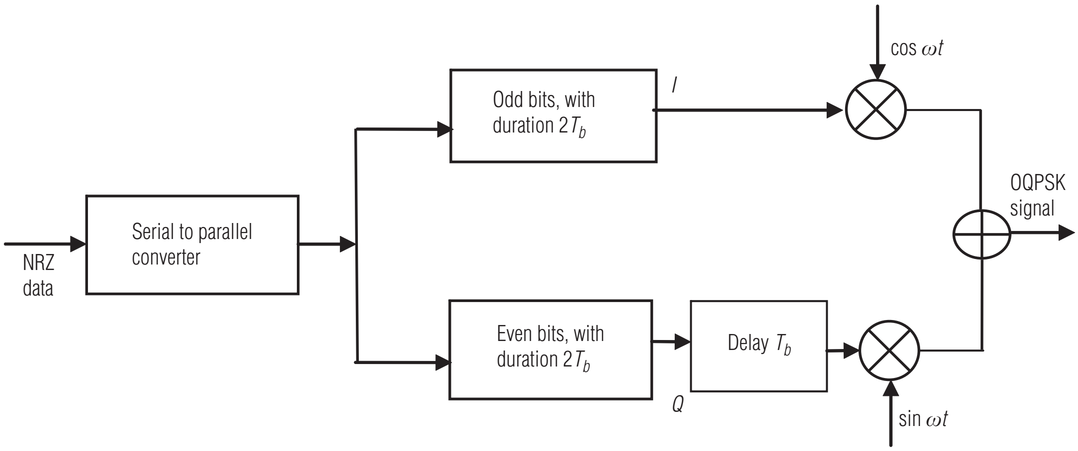


Figure 13.12 OQPSK modulator

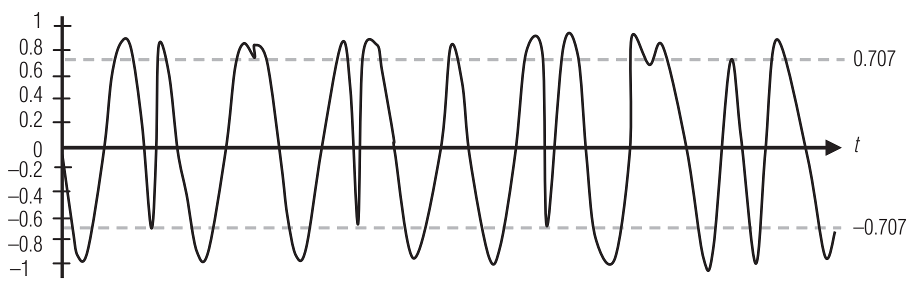


Figure 13.13 $\pi/4$-QPSK signal

The main advantage of $\pi/4$-QPSK is that of the reduced envelope fluctuation as compared to that of QPSK. The $\pi/4$-QPSK digital modulation technique is widely used in most second-generation cellular systems such as United States digital cellular (IS 54) TDMA systems, personal communications service (PCS) systems, and Japanese digital cellular (JDC) standards.

The constellation diagram of $\pi/4$-QPSK is depicted in Figure 13.14. It uses two QPSK constellations offset by $\pm\pi/4$. Signalling elements are selected in turn from the two QPSK constellations. Transition must occur from one constellation to the other one. This ensures that there will always be a phase change for each input symbol. It consists of symbols corresponding to eight phases. These eight-phase points are formed by superimposing two QPSK signal constellations offset by 45° relative to each other.

DPSK modulation

Differential encoding of baseband signals is that where a binary sequence is differentially encoded as the other sequence. Differential phase shift keying (DPSK) is differentially coherent modulation method. DPSK does not need a synchronous (coherent) carrier at the demodulator. The input sequence of binary bits is modified such that the next bit depends upon the previous bit. Therefore, in the receiver the previous received bits are used to detect the present bit.

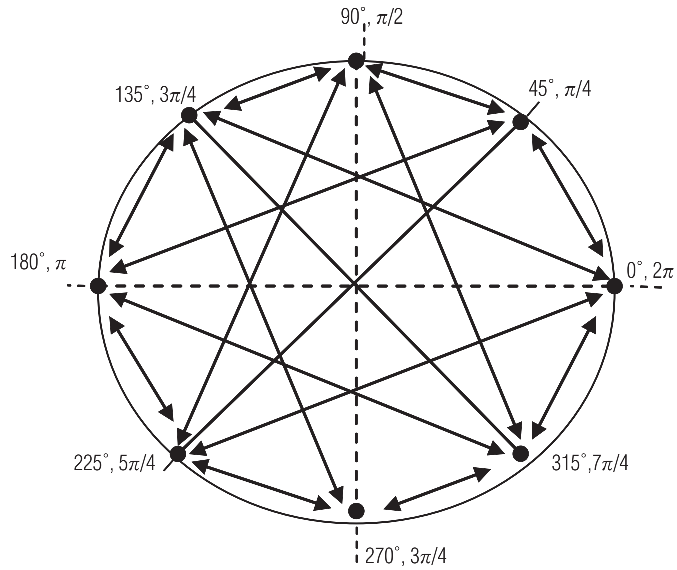


Figure 13.14 $\pi/4$-QPSK constellations

The DPSK waveforms are shown in Figure 13.15. Depending on an arbitrary sequence *d(t), b(t)* and $b(t - T_b)$ are found.

From the above waveform, it is clear that $b(t - T_b)$ is the delayed version of *b(t)* by one period T_b. While drawing the waveforms the value of $b(t - T_b)$ is not known initially in the first interval. Therefore, it is assumed to be zero and then waveforms are drawn. The P_e for DPSK can be given as

$$P_e = \frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right) \tag{13.20}$$

The main drawback of DPSK is that the bit error rate (BER) is higher than that of BPSK and noise interference is more.

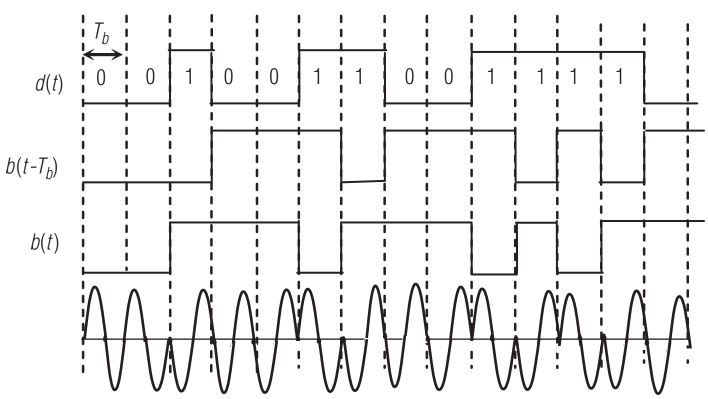


Figure 13.15 DPSK waveforms

M-PSK modulation

When we define PSK with M phases, it is called M-ary PSK. In M-ary PSK, every "b" (where $M = 2^b$) bits of the binary bit stream are coded as a signal that is transmitted in the form $A \sin(\omega t + \theta_j)$, $j = 1, \ldots, M$.

The BPSK transmits one bit at a time with two symbols. Hence the phase shift is

$$\text{Phase shift in BPSK} = \frac{\text{phase of PSK signal}}{\text{number of symbols}} = \frac{2\pi}{2} = \pi \text{ or } 180°$$

The QPSK transmits two bits at a time with four symbols. Hence the phase shift is

$$\text{Phase shift in QPSK} = \frac{\text{phase of PSK signal}}{\text{number of symbols}} = \frac{2\pi}{4} = \frac{\pi}{2} \text{ or } 90°$$

Therefore, for the "b" successive bits there will be $M = 2^b$ possible symbols (called as M-PSK). Then the phase shift can be

$$\text{Phase shift in QPSK} = \frac{\text{phase of PSK signal}}{\text{number of symbols}} = \frac{2\pi}{M}$$

In M-PSK modulation, the amplitude of the transmitted signal remains constant, thereby yielding a circular constellation. The symbol error probability for M-PSK under additive white Gaussian noise (AWGN) conditions is given as

$$\therefore P_b = 2Q\left[\sin\left(\frac{\pi}{M}\right).\sqrt{\frac{2bE_b}{N_0}}\right] \tag{13.21}$$

Using Gray code, a symbol error is likely to result in only one out of "b" bit errors.

$$\therefore P_b = \frac{2}{b}.Q\left[\sin\left(\frac{\pi}{M}\right).\sqrt{\frac{2bE_b}{N_0}}\right] \tag{13.22}$$

For 8-PSK, $M = 8$, $b = 3$, and

$$P_b = \frac{2}{3}Q\left[\sin\frac{\pi}{8}.\sqrt{\frac{6E_b}{N_0}}\right] = \frac{2}{3}Q\left[0.937\sqrt{\frac{E_b}{N_0}}\right] \tag{13.23}$$

For 16-PSK, $M = 16$, $b = 4$, and

$$P_b = \frac{Q}{2}\left[\sin\frac{\pi}{16}.\sqrt{\frac{8E_b}{N_0}}\right] = \frac{Q}{2}\left[0.552\sqrt{\frac{E_b}{N_0}}\right] \tag{13.24}$$

13.4.4 Minimum shift keying

QPSK is affected with the discontinuous phase transition. To avoid the discontinuity in the phase transition, there was a motivation for designing of a continuous phase modulation schemes. Minimum shift keying (MSK) is one such scheme of a continuous envelope digital modulation. This method involves a trade-off between the width of the main lobe of the power spectrum and the side lobe of the power content.

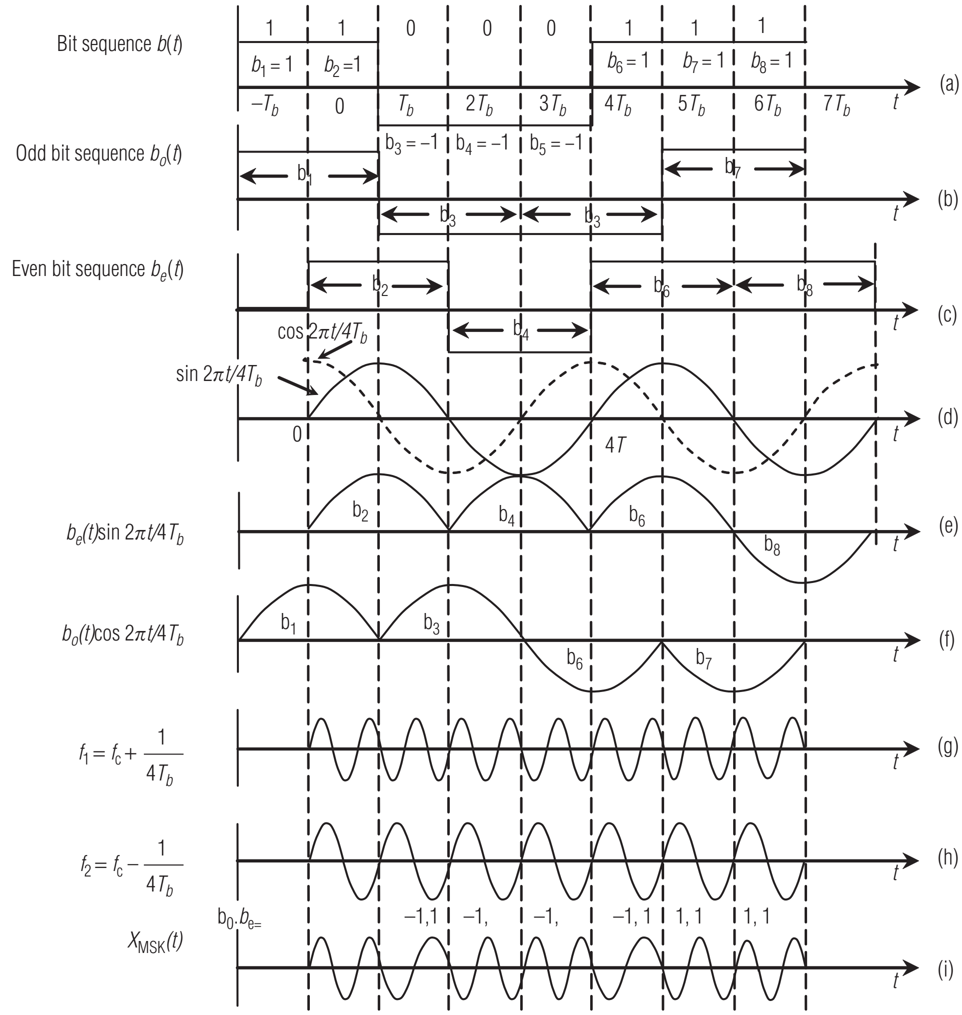


Figure 13.16 MSK waveforms

In MSK the output waveform is continuous in phase, hence there are no abrupt changes in amplitude. The side lobes of MSK are very small, hence band pass filtering is not required to avoid interchannel interference. Figure 13.16 shows the waveform of MSK. Figure 13.16(a) shows the corresponding NRZ waveform, $b(t)$. Figures 13.16(b) and (c) show waveforms of odd and even bits, $b_o(t)$ and $b_e(t)$. Figure 13.16(d) shows the two offset waveforms. Figures 13.16(e) and (f) show the product waveforms.

The two frequencies f_1 and f_2 in MSK must satisfy the following equations:

$$f_1 = f_c + \frac{1}{4T_b} \quad \text{and} \quad f_2 = f_c - \frac{1}{4T_b} \tag{13.25}$$

Figures 13.16(g) and (h) show the waveforms of f_1 and f_2. Figure 13.16(i) shows the final MSK waveform.

MSK is a binary modulation with symbol interval T_b and frequency deviation is equal to ±1/4 T_b. In addition, there is phase continuity of the modulated RF carrier at the bit transitions. RF phase varies linearly exactly ±90° with respect to the carrier over one bit period T_b. The spacing between f_1 and f_2 is the minimum that can be used to allow successful detection of the received signal at the receiver. That is why it is called MSK digital-modulation technique.

MSK can be considered as OQPSK with one significant change. Each channel I and Q is pulse shaped with a half sinusoid, and the carrier signal is multiplied by a sinusoidal function. In this case, the MSK signal can be expressed as

$$X_{\text{MSK}}(t) = I(t)\cos\left(\frac{\pi t}{2T_b}\right)\left[\cos(2\pi f_c t) + Q(t - T_b)\cos\left(\frac{\pi t}{2T_b}\right)\sin(2\pi f_c t)\right] \tag{13.26}$$

The P_e for MSK is same as that of PSK.

MSK is used in the global system for mobile communications (GSM) cellular standard. The main advantage of a MSK is that it has a better BER performance than conventional binary FSK for a given SNR value. Its disadvantage is that it requires synchronous circuits for modulators and demodulators and it is more expensive. The MSK has the same bit error probability as ordinary BPSK. The error probability for an ideal MSK system is approximately the same as BPSK in power efficiency.

Gaussian minimum shift keying

Gaussian minimum shift keying (GMSK) is a binary digital modulation technique, which can be considered as a special case of MSK. The difference from the MSK is that in GMSK the Gaussian pulse is used for pulse shaping instead of half sinusoid. The spectrum of a Gaussian signal has low side lobes and a narrower main lobe than the spectrum of a rectangular pulse. Figure 13.17 illustrates the impulse response of a Gaussian filter.

Instead of bandwidth of low-pass Gaussian filter, it is more convenient to use the *normalized bandwidth (BT)*, which is the product of the filter bandwidth B and the symbol period T. From the figure the Gaussian filter with $BT = \infty$ indicates the rectangular filter. The small value of BT provides improvement in bandwidth efficiency, but degrades the power efficiency. Generally, the values of BT from 0.3 to 0.5 are used for GMSK. The Gaussian response is interesting in that the

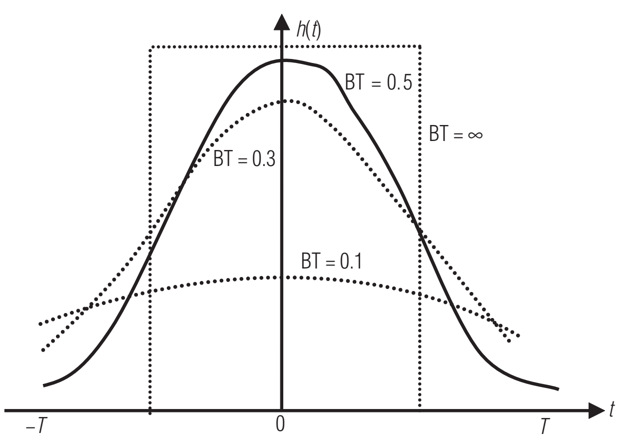


Figure 13.17 Impulse response of the Gaussian filter

frequency response has the same shape as the impulse response. The relation between standard deviation "σ" of the impulse response and the filter bandwidth B is

$$B = \frac{\log_e 2}{\pi\sigma} = \frac{0.2206}{\sigma} \tag{13.27}$$

In the case of MSK, the BER performance for coherent detection is the same as that for differentially encoded BPSK with coherent detection. BER for GMSK is degraded due to inter-symbol interference (ISI) by the pre-modulation Gaussian filter. The BER performance of GMSK with coherent detection under AWGN conditions is given by

$$P_e = erfc\left[\sqrt{\frac{2\beta E_b}{N_0}}\right] \tag{13.28}$$

where β is the degradation factor due to pre-modulation.

$\beta = 1$ corresponds to the performance for MSK. The BER under flat Rayleigh fading conditions is given as

$$P_e = 1 - \frac{1}{\sqrt{1 + \frac{1}{\beta\left(\frac{E_b}{N_0}\right)_{\text{avg}}}}} \tag{13.29}$$

where $(E_b/N_0)_{\text{avg}}$ represents an average value.

The Figure 13.18 shows the GMSK modulator. The digital data stream is first shaped with a Gaussian filter before being applied to the modulator. This has the advantage of reducing sideband power, which in turn reduces out-of-band interference between signal carriers in adjacent frequency channels. However, this causes the ISI making it more difficult to discriminate between different transmitted data values and requiring more complex channel equalization

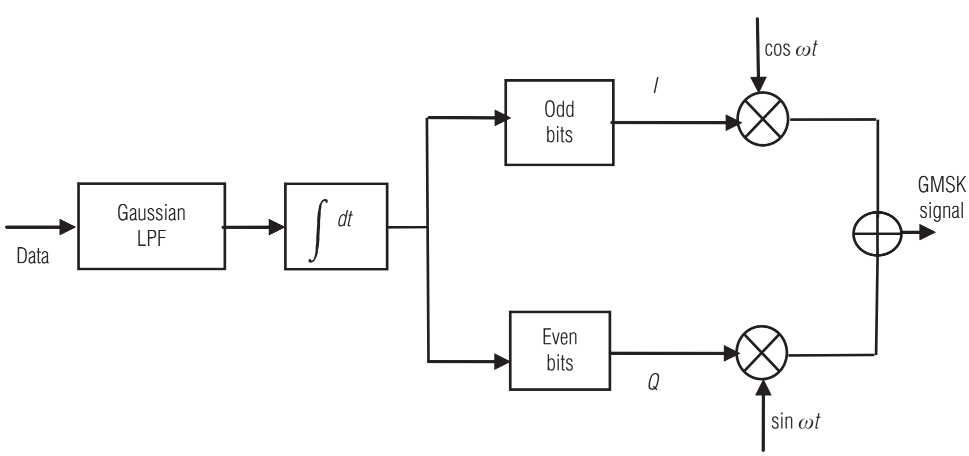


Figure 13.18 GMSK modulator

algorithms such as an adaptive equalizer at the receiver. An important application of GMSK is GSM, which is a time-division multiple access (TDMA) system.

13.4.5 Quadrature amplitude modulation

Quadrature amplitude modulation (QAM) is another member of the digital modulation family, which can be considered as an extension of QPSK or a combination of ASK and PSK, except the digital information is contained in both the amplitude and the phase of the modulated signal. QAM is used in applications including microwave digital radio, digital video broadcasting-cable (DVB-C), and modems.

In QAM, two different signals are sent simultaneously on the same carrier frequency, one shifted by the other by 90° with respect to the other. Figure 13.19 illustrates the QAM modulation scheme. Each QAM symbol corresponds to the block of "k" bits. The data flow is split in blocks by "$k/2$" bits; this process generates two independent signals. One data block is ASK modulated on a carrier frequency f_c, and the other data block is FSK modulated by the same carrier signal shifted by 90°. The two modulated signals are added and transmitted over the channel. This is achieved by multilevel AM of each data block.

The M-QAM signal can be expressed as

$$X_{\text{QAM}}(t) = d_1(t)\cos(2\pi f_c t) + d_2(t)\sin(2\pi f_c t) \tag{13.30}$$

where $M = 2^k$ is the modulation cardinality and "k" is the number of bits per symbol. Each M-QAM symbol corresponds to the block of k bits.

There are a finite numbers of allowable amplitude-phase combinations for a given system. If a two-level ASK is used, then each of the two data blocks can be in one of the two states and the combined data block can be in one of four states, that is $M = 4$. This is essentially a QPSK. If a four-level ASK is used then the combined data stream data block can be in one of the 16 states, that is $M = 16$ and it is called as 16-QAM. Similarly, 64-QAM and 256-QAM can be implemented.

Figure 13.20 shows a constellation diagram of 16-QAM. This constellation consists of a square lattice of signal points. Due to more effective use of signal space in QAM, it is possible to increase the minimum distance between the constellation points in comparison with PSK, which leads to better performance.

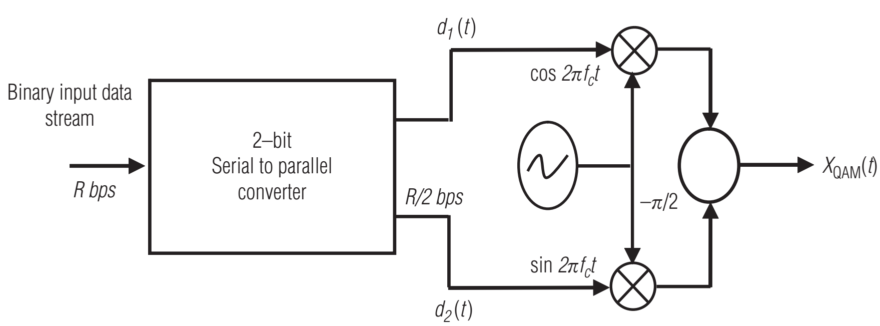


Figure 13.19 QAM modulator

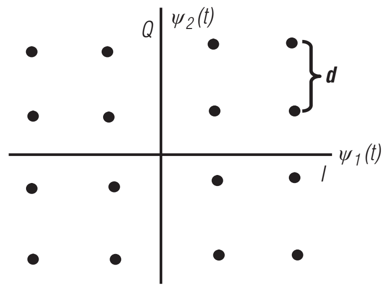


Figure 13.20 16-QAM constellation

The disadvantage of 16-QAM compared to 16-PSK is that the signal amplitude has an inherent variation, regardless of filtering. This may cause problems in non-linear amplifiers and on fading channels.

13.4.6 Orthogonal frequency division multiplexing

Orthogonal frequency division multiplexing (OFDM) is a multicarrier digital modulation technique, where a single data stream is transmitted over a number of lower rate subcarriers. OFDM is the extension of the *frequency division multiplexing* (*FDM*) technique. The basic idea of the FDM is to divide the available bandwidth into many narrow sub-bands and to use a large number of parallel narrow-band subcarriers rather than a single wide-band carrier to transfer the information.

The main advantage of FDM is the robustness against frequency-selective fading and narrowband interference. The narrowband interference will affect only one or two subcarriers of the whole bunch of subcarriers. The other subcarriers will not be corrupted by the interference. Another advantage of the FDM is the much greater symbol length in comparison with wideband carrier. Due to this fact, the ISI in the case of FDM is comparatively short, which helps to cope with this problem.

The main disadvantage of FDM is that the spectral efficiency is quite low due to guard intervals. To eliminate this problem, it is possible to use the *orthogonal subcarriers*. The use of orthogonal subcarriers allows the overlapping of subcarriers' spectra. Due to the orthogonality, it is possible to recover the individual subcarrier's signals despite the overlapping and thus there is no need of guard intervals as in FDM. The usage of the orthogonal subcarriers also helps to decrease the implementation complexity of both transmitter and receiver.

In FDM, each subcarrier needs a separate pair of matched filters at the transmitter and the receiver to make it possible to eliminate the inter-carrier interference. In OFDM, the inter-carrier interference is eliminated due to the orthogonality of the subcarriers and there is no need to use separate filters for each subcarrier. The idea of using orthogonal subcarriers was suggested more than 40 years ago, but in practice it was not used for a long time mostly due to its complexity. The detailed description of OFDM is given in Chapter 16.

13.4.7 Performance of digital modulation schemes

The performance of different digital modulations schemes discussed in the above sections is summarized in Table 13.1 with respect to the BER.

Table 13.1 BERs of different digital modulation schemes

Modulation scheme	Bit error rate
ASK	$\frac{1}{2} erfc \sqrt{\frac{E_b}{4N_0}}$
FSK	$\frac{1}{2} erfc \sqrt{\frac{E_b}{2N_0}}$
PSK BPSK QPSK MSK	$\frac{1}{2} erfc \sqrt{\frac{E_b}{N_0}}$
DPSK	$\frac{1}{2} \exp\left(-\frac{E_b}{N_0}\right)$
GMSK	$erfc\left[\sqrt{\frac{2\beta E_b}{N_0}}\right]$

13.5 Synchronization

The demodulation of a signal requires that the reference signal at the receiver be synchronized with the transmitter signal in phase and frequency as it is received at the input of the receiver. Both coherent and non-coherent demodulations require symbol timing at the receiver to be synchronized in phase and frequency with the received signal.

Carrier synchronization can be achieved by sending a pilot tone before sending the message signals. Since the pilot tone has a strong spectral line at the carrier frequency, the receiver can easily lock on it and generates a local coherent carrier. However, this requires extra transmission bandwidth. Carrier synchronization also can be achieved with a *carrier recovery circuit*, which extracts the phase and frequency information from the noisy received signal and use it to generate a clean sinusoidal reference signal.

Symbol synchronization usually is achieved by a *clock (symbol timing) recovery circuit*, which uses the received signal to control the local oscillator.

13.5.1 Carrier recovery

Carrier recovery is the process of extracting a reference carrier from a received signal. The PSK signals have no spectral line at carrier frequency. Therefore, a non-linear device is needed in the carrier recovery circuit to generate such a line spectrum. Figure 13.21 illustrates the carrier recovery for PSK.

The carrier signal changes phase for every bit interval in PSK. If the received signal is multiplied by an integer, *N*, we can convert all phase changes to multiples of 360°. Then, the new signal has no phase changes and we can recover it using a narrowband phase-locked loop (PLL). The PLL consists of the phase detector, the LPF and the VCO, tracks and locks onto the frequency and phase. After the PLL recovers the carrier signal, it is divided by *N* to recover the carrier at the proper frequency.

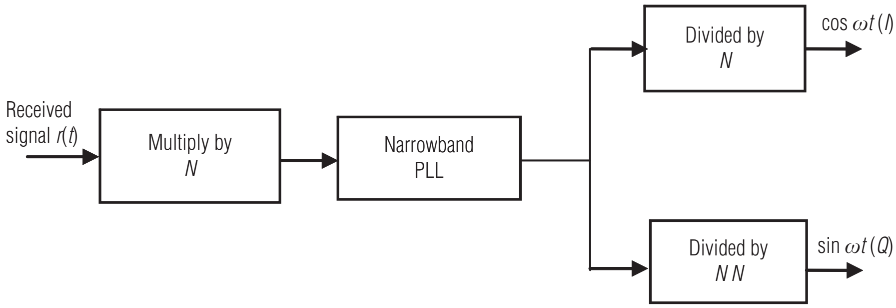


Figure 13.21 Carrier recovery for PSK

By a suitable choice of digital dividing circuits, it is possible to get a precise 90° difference in the output of two divider and thus generate both cos ωt and sin ωt signals needed by the receiver.

For BPSK, we would need an N of 2, but an N of 4 would be used to enable the sine and cosine terms to be generated. For QPSK and its derivatives, an N of 4 is necessary. For $\pi/4$-QPSK, an N of 8 would be needed. The advantage of $\pi/4$-QPSK is that it is not important to know the absolute carrier phase. Only the change in carrier phase from one symbol to the next is important.

MSK is a form of frequency modulation. Therefore, a different method of carrier recovery is required. Figure 13.22 shows the carrier recovery for MSK. The MSK signal has frequency f and deviation $\Delta f = 1/2T_b$. First the signal is multiplied by 2, thus doubling the deviation and generating strong frequency components at $2f + 2\Delta f$ and $2f - 2\Delta f$. Here two PLLs are used to recover these two signals.

$$X_1(t) = \cos(2\pi ft + \pi\Delta ft) \tag{13.31a}$$

$$X_2(t) = \cos(2\pi ft - \pi\Delta ft) \tag{13.31b}$$

We then take the sum and difference of $s_1(t)$ and $s_2(t)$ to generate the desired $I(t)$ and $Q(t)$ signals.

$$I(t) = X_1(t) + X_2(t) = 2\cos 2\pi ft\ \cos \pi\Delta ft \tag{13.32a}$$

$$Q(t) = X_1(t) - X_2(t) = 2\sin 2\pi ft\ \sin \pi ft \tag{13.32b}$$

13.5.2 Clock recovery

The clock or symbol timing recovery is usually a closed-loop synchronizer, which attempts to lock a local clock signal onto the received data stream by the use of comparative measurements

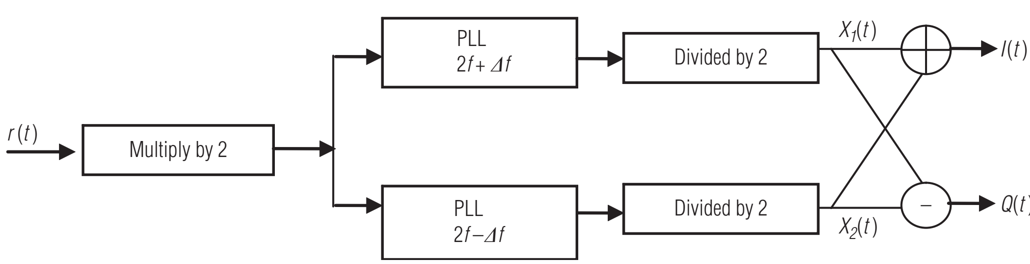


Figure 13.22 Carrier recovery for MSK

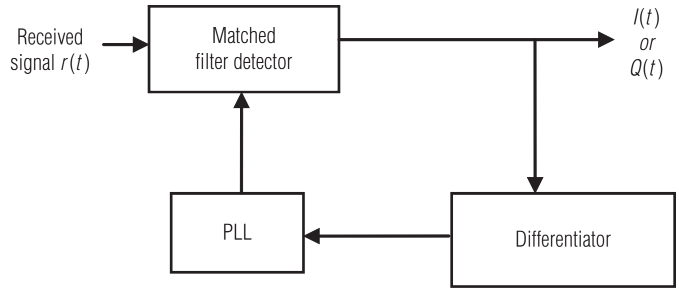


Figure 13.23 Generalized data timing recovery circuit

on the local and received signals. Figure 13.23 illustrates a generalized symbol timing recovery circuit. Most communication systems transmit a sequence of 1s and 0s in an alternating pattern to enable the receiver to maintain bit synchronization. A PLL operating at bit timing is used to maintain timing. Once the PLL is synchronized onto the received 101010 … pattern, it will remain synchronized onto any other patterns except for long sequences of all 0s or all 1s.

MSK uses an additional circuit to achieve symbol timing. The $s_1(t)$ and $s_2(t)$ signals are multiplied together and low-pass filtered (Fig. 13.24).

The output of a low-pass filter is a clock signal at one-half of the transmitted bit rate. The one-half bit rate clock is the correct rate for demodulation of the signal since I and Q signals are at one-half of the bit rate.

$$\begin{aligned} X_1(t)X_2(t) &= \cos(2\pi ft + \pi\Delta ft) \times \cos(2\pi ft - \pi\Delta ft) \\ &= 0.5\cos 4\pi ft + 0.5\cos 2\pi\Delta ft \end{aligned}$$

The low-pass filtered signal is

$$\left[X_1(t)X_2(t)\right] = 0.5\ \cos 2\pi\Delta ft = 0.5\ \cos\frac{\pi t}{T_b} \tag{13.33}$$

13.6 Performance of digital modulation in slow-flat fading channels

The performance of a digital modulation scheme is degraded by many transmission impairments including fading, delay spread, Doppler spread, co-channel and adjacent channel interference, and noise. Fading causes a very low instantaneous received SNR when the channel exhibits a deep fade. Delay spread causes ISI between the transmitted symbols, and a large Doppler spread is

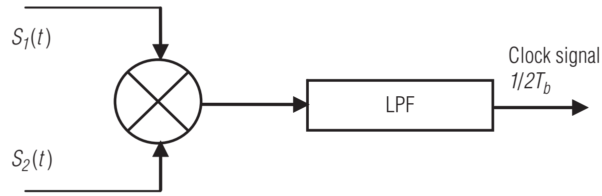


Figure13.24 MSK data timing recovery circuit

indicative of rapid channel variation and necessitates a receiver with a fast convergent algorithm. Co-channel interference, adjacent channel interference, and noise are all additive distortions that degrade the BER performance by reducing the SNR.

Here, we derive the BER performance of digital signalling on frequency non-selective (flat) fading channel with AWGN. Flat fading channel models are appropriate for narrow-band land mobile radio systems or mobile satellite systems. Flat fading channels affect all frequency components of a narrow-band signal in exactly the same way and, therefore, do not introduce amplitude or phase distortion into the received signal. Frequency selective channels distort the transmitted signal. Flat fading channel will significantly degrade the BER performance unless appropriate countermeasures are taken. Diversity and coding techniques are well-known methods for combating fading. The basic idea of diversity systems is to provide the receiver with multiple replicas of the same information-bearing signal, where the replicas are affected by uncorrelated fading. Coding techniques introduce a form of time diversity into the transmitted signal, which can be exploited to mitigate the effects of fading. These techniques will be discussed in Chapter 14.

In a flat fading channel, the signal undergoes a multiplicative variation. In general, this multiplicative factor is complex, that is the signal amplitude as well as phase is affected. If we further assume that the fading is slow, then the amplitude attenuation and phase shift of the received signal can be considered constant over at least symbol duration. Therefore, the received equivalent low-pass complex signal $\tilde{r}(t)$ can be written as

$$\tilde{r}(t) = ze^{-j\phi}\tilde{s}(t) + \tilde{n}(t) \tag{13.34}$$

where z = the gain of the signal

ϕ = the phase shift of the signal caused by the channel

$\tilde{n}(t)$ = the equivalent low-pass complex additive Gaussian noise

The received signal may be coherently or non-coherently detected, depending on whether it is possible to accurately estimate the phase shift. In either case, the average error probability can be evaluated by averaging the error probability for a fixed amplitude z over the entire range of z. That is,

$$P_e = \int_0^{\infty} P_e(\gamma_b)\, p(\gamma_b)\, d\gamma_b \tag{13.35}$$

where $P_e(\gamma_b)$ is the probability, if error at a specific value of SNR γ_b, γ_b is the z^2E_b/N_0, and $p(\gamma_b)$ is the probability density function of γ_b. The E_b and N_0 are constant and represent the average energy per bit and noise power density in a non-fading AWGN channel, and z^2 is used to represent the instantaneous power values of the channel.

13.7 Performance of digital modulation in frequency selective mobile channels

In the wireless environment, the frequency selective fading which is produced by multipath time delay spread ultimately causes ISI. If there exists ISI it will lead to irreducible BER in the mobile transmitters and receivers. It is also notable that even if the mobile radio channel is not frequency selective, the doppler effect creates irreducible BER because of random spectral spreading. These factors are thus limiting factors on the data rate and the BER value on transmitter side in frequency selective channels. A better simulation software can be used to analyze this type of frequency selective fading channel effects.

In frequency selective channels, the irreducible error is caused mainly by errors which is produced by ISI that gets interfered with actual signal component at the receiver's sampling time instant. This happens in the following circumstances:

- When a non-zero value of delay spread "*d*" causes ISI effects.
- When undelayed signal (main signal) component is eliminated via multipath cancellations.
- If the sampling time (*t*) of the mobile receiver is shifted due to the delay spread (*d*).

It is finally observed that for a small delay spread "*d*" the resultant flat fading will be strong cause for any error bursts. For large delay spread case the ISI and timing errors will be the strong error mechanisms. Thus proper receiver designing and prediction of channel will help to reduce the interference to an extent in frequency selective mobile channels.

13.8 Summary

- The modulation and demodulation process applicable to cellular systems is studied in this chapter. Since baseband signals can be transmitted only over short distances with wires and require very long antennas to transmit them without wires, the baseband signals are modulated onto radio frequency carriers.
- Modulation is the process of varying one or more properties, such as amplitude, frequency, and phase of a high-frequency periodic waveform, called the carrier signal, with a modulating signal which typically contains information to be transmitted.
- FM is widely used for radio transmissions for a wide variety of applications from broadcasting to general point-to-point communications.
- Digital modulation is used to provide more information capacity, compatibility with digital data services, higher data security, better quality communications, and quicker system availability.
- The basic forms of the three modulation methods used for transmitting digital signals are
  - ASK
  - FSK
  - PSK
- Digital communication systems have certain distinction in comparison to their analogue counter part due to their signal-space representation.
- The important modulation techniques for wireless communication such as BPSK, QPSK, OQPSK, $\pi/4$-DQPSK, MSK, and GMSK were taken up at length. A relatively new modulation technology, OFDM, has also been implemented.
- OFDM has been proposed for advanced cellular systems. The application of OFDM to wireless systems is due to the increasing need for higher bit rate, higher bandwidth data transmission over radio-based communication systems.
- QAM is a hybrid of amplitude and phase modulation. QAM is used in applications including microwave digital radio, DVB-C, and modems.
- The demodulation of a signal requires that the receiver be synchronized with the transmitter signal as it is received at the input of the receiver. The synchronization must be carried out in carrier, bit, and word.

Example problem 13.1

A digital communication system uses BPSK modulation scheme for information transmission. Calculate the P_e in AWGN with a 12.48 dB of $E_b/N_{0.}$

Solution

Given data: $\frac{E_b}{N_0} = 12.48\,\text{dB},$

The P_e of a BPSK system in AWGN is given as

$$P_e = \frac{1}{2} erfc\sqrt{\frac{E_b}{N_0}}$$

$$\frac{E_b}{N_0} = 12.48\,\text{dB} = 10^{\frac{12.48}{10}} = 17.7011$$

$$\therefore P_e = \frac{1}{2} erfc\sqrt{17.7011} = 1.341\times 10^{-9}$$

Example problem 13.2

A BPSK modulation is used in a Gaussian channel with $N_0 = 10^{-8}$ W/Hz and $E_b/N_0 = 4.68$ dB. Calculate the amplitude of the carrier signal for a data rate of 200 kbps.

Solution

Given data: $N_0 = 10^{-8}$ W/Hz

$$\frac{E_b}{N_0} = 4.68\text{ dB} = 10^{\frac{4.68}{10}} = 2.937$$

We have

$$\frac{E_b}{N_0} = \frac{A^2T}{2N_0} = \frac{A^2}{2N_0R}$$

$$\therefore A = \sqrt{\left[2N_0\left(\frac{E_b}{N_0}\right)R\right]}$$

$$\therefore A = \sqrt{2\times 10^{-8}\times 2.937\times 2\times 10^5} = 0.1084\,\text{V}$$

Example problem 13.3

A GSM system uses the GMSK modulation with a data rate of 198.5 kbps and channel bandwidth of 220 kHz. Calculate (a) the frequencies transmitted at the carrier frequency of 750 MHz, (b) the frequency shift between binary 1 and binary 0, and(c) the bandwidth efficiency in bps/Hz.

Solution

Given data: Channel bandwidth, $B_w = 220$ kHz

Data rate, $R = 198.5$ kbps

(a) The frequency shift is

$\Delta f = f_H - f_L = 0.5\ R = 0.5 \times 198.5$ kHz $= 99.25$ kHz

(b) The maximum and minimum frequencies are

$f_H = f_c + 0.25$ R = 750 MHz + 0.25 × 198.5 kHz = 750.0496 MHz,
$f_L = f_c - 0.25$ R = 750 MHz − 0.25 × 198.5 kHz = 749.950 MHz

(c) The bandwidth efficiency

$R/Bw = 198.5 / 220 = 0.90\text{bps} / \text{Hz}$

Example problem 13.4

Determine the 3 dB bandwidth for a Gaussian low-pass filter that is used to generate 0.3 GMSK with a channel data rate of 270 kbps. What is the bit error probability for GMSK if $E_b/N_0 = 6$ dB?

Solution

$$T_b = \frac{1}{R_b} = \frac{1}{270 \times 10^3} = 3.7\mu\text{s}$$

$$BT_b = 0.3 \Rightarrow B = 0.3/T_b = 0.3/3.7 \times 10^{-6} = 81.08 \text{ kHz}$$

The 3 dB bandwidth is 81.08 kHz. To determine the 99.99 per cent power bandwidth, we use Table 13.1 to find that 1.41 R_b is the required value. The occupied RF spectrum for a 99.99% power bandwidth will be RF bandwidth = 1.41, $R_b = 1.41 \times 270 \times 10^3 = 380.7$ kHz, $\beta = 0.89$. For $E_b/N_0 = 6$ dB (3.9811), we get

$$P_e = erfc\left[\sqrt{\frac{2\beta E_b}{N_0}}\right] = erfc(2.662) \approx 11.5 \times 10^{-5}$$

Example problem 13.5

Determine the bit rate if a modulator transmits symbols at a rate of 20,400 per second. Each symbol has 64 different possible states.

Solution

Number of bits per symbol: $b = \log_2 M = \log_2 64 = \log_2 2^6 = 6$
Bit rate of the modulator = 6 × 20,400 = 122.4 kbps

Example problem 13.6

A data stream with data rate $R_b = 128$ Mbps is transmitted on an RF channel with a bandwidth of 32 MHz. Assuming Nyquist filtering and Gaussian channel, determine the modulation scheme that should be used to also find the required E_b/N_0 if the probability of the bit error is 3×10^{-5}.

Solution

Required bandwidth efficiency

$$\frac{R}{B_w} = \frac{128}{32} = 4\,\text{bps/Hz}$$

Since the channel is bandwidth limited $\frac{R}{B_w} = \frac{152}{38} = \log_2 M$
where $M = 2^4 = 16$.
16-QAM should be used as it is more efficient than 16-PSK. For a rectangular constellation, with a Gaussian channel and matched filter reception, the bit error probability is given as

$$P_b = \frac{3}{4} Q\left[\sqrt{\frac{4E_b}{5N_0}}\right] = 3 \times 10^{-5}$$

$$\therefore \frac{E_b}{N_0} = 19.5 = 12.9\,\text{dB}$$

Review questions

1. State the advantages of digital modulation schemes.
2. State the types of modulation schemes used in mobile communication.
3. What is coherent detector?
4. State the advantage of using GMSK rather than MSK.
5. State the difference between MSK and GMSK.
6. Explain about modulation techniques in detail.
7. Bring out the salient features of the MSK modulation scheme.
8. Explain Gaussian MSK and M-ary QAM.
9. Explain the performance of different modulation scheme with respect to BER.
10. A digital signalling system is required to operate at 9.6 kbps. If a signal element encodes 4-bit word, what is the minimum bandwidth of the channel?
11. Given a channel with intended capacity of 20 Mbps, the bandwidth of the channel is 3 MHz. What is the SNR required to achieve this capacity?
12. The receiver in a communications system has a received signal power of –134 dBm, received noise power spectral density of –174 dBm/Hz, and a bandwidth of 2000 Hz. What is the maximum rate of error-free information for the system?

Objective type questions and answers

1. Which of the following gives maximum probability error?
   (a) ASK (b) FSK (c) PSK (d) DPSK
2. The BER of BPSK is
   (a) $\frac{1}{2} erfc\sqrt{\frac{E_b}{2N_0}}$ (b) $\frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right)$ (c) $\frac{1}{2} erfc\sqrt{\frac{E_b}{4N_0}}$ (d) $\frac{1}{2} erfc\sqrt{\frac{E_b}{N_0}}$
3. The symbol error probability for M-PSK under AWGN conditions is given as
   (a) $P_b = \frac{2}{b}.Q\left[\sin\left(\frac{\pi}{M}\right).\sqrt{\frac{2bE_b}{N_0}}\right]$ (b) $P_b = \frac{2}{b}.Q\left[\sin\left(\frac{\pi}{M}\right).\sqrt{\frac{2E_b}{N_0}}\right]$
   (c) $P_b = \frac{2}{b}.Q\left[\sin\left(\frac{\pi}{M}\right).\sqrt{\frac{2bE_b}{3N_0}}\right]$ (d) $P_b = \frac{2}{b}.Q\left[\sin\left(\frac{\pi}{M}\right).\sqrt{\frac{2bE_b}{5N_0}}\right]$
4. The phase shift of QPSK when symbol change occurs
   (a) 2π (b) π (c) $\pi/2$ (d) $3\pi/2$
5. Which of the following gives minimum probability error?
   (a) ASK (b) FSK (c) PSK (d) DPSK
6. The output S/N ratio of a matched filter, fed at its input by a rectangular pulse of amplitude A and duration T is given as (assume τ as noise power spectral density)
   (a) $\frac{2A^2T}{\tau}$ (b) $A.T$ (c) A/T (d) A/τ
7. The impulse response of a matched filter to a data pulse tg(t) with $0 \le t \le T$ is
   (a) $\delta^{\otimes}(t)$ (b) $g^{\otimes}(t)$ (c) $g^{\otimes}(t-T)$ (d) $g^{\otimes}(T-t)$

Answers: 1. (a), 2. (d), 3. (a), 4. (c), 5. (c), 6. (a), 7. (d).

Open book questions

1. Write the advantages of MSK over QPSK.
2. What is linear modulation?
3. Define non-linear modulation.
4. Mention some merits of MSK.

Key equations

1. The ASK signal can be mathematically expressed as

$$X_{\text{ASK}}(t) = \begin{Bmatrix} A_c \cos(2\pi f_c t) \text{ for binary } 1 \\ 0 \text{ for binary } 0 \end{Bmatrix}$$

2. The P_e of a ASK modulation scheme is given as

$$P_e = \frac{1}{2} erf_c \sqrt{\frac{E_b}{4N_0}}$$

3. The FSK signal for one-bit duration can be mathematically expressed as

$$X_{\text{BFSK}}(t) = \begin{Bmatrix} A_c \cos(2\pi f_1 t) \text{ for binary 1} \\ A_c \cos(2\pi f_2 t) \text{ for binary 0} \end{Bmatrix}$$

4. The P_e of a FSK modulation scheme is given as

$$P_e = \frac{1}{2} erfc \sqrt{\frac{E_b}{2N_0}}$$

5. The P_e of a PSK modulation scheme is given as

$$P_e = \frac{1}{2} erfc \sqrt{\frac{E_b}{N_0}}$$

6. The P_e for BPSK is same as that of PSK and can be given as

$$P_e = \frac{1}{2} erfc \sqrt{\frac{E_b}{N_0}}$$

7. The probability of error (P_e) for DPSK can be given as

$$P_e = \frac{1}{2} \exp\left(-\frac{E_b}{N_0}\right)$$

8. The BER performance of GMSK with coherent detection under AWGN conditions is given by

$$P_e = erfc\left[\sqrt{\frac{2\beta E_b}{N_0}}\right]$$

9. The M-QAM signal can be expressed as

$$X_{\text{QAM}}(t) = d_1(t)\cos(2\pi f_c t) + d_2(t)\sin(2\pi f_c t)$$

Further reading

Blake, R. *Electronic Communication Systems*. Delmar, Singapore: Thomson Asia Pvt Ltd, 2002.

Forney, G. D. Jr., “Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference,” IEEE Trans. Inform. Theory, IT-18, 363–378, May 1972.

Lathi, B. P., and Ding, Z. *Modern Digital and Analog Communication Systems*, 4th ed. NY: Oxford University Press, 2009.

Lucky, R. W. “Automatic equalization for digital communications,” Bell Syst. Tech. J.,44, 547–588, April 1965.

Lucky, R. W. "Techniques for adaptive equalization of digital communication," Bell Syst. Tech. J., 45, 255–286, February 1966.

Proakis, J. G. *Digital Communications*, 3rd ed., New York: McGraw-Hill,1995.

Sklar, B. *Digital Communications: Fundamentals and Applications*, 2nd ed. Singapore: Pearson Education, Inc., 2005.

Tranter, W. H. et.al., *Principles of Communication Systems Simulation*. Singapore: Pearson Education, Inc., 2004.

Signal Processing in Wireless Systems 14

14.1 Introduction

Signal processing techniques such as diversity, equalization, speech, and channel coding have been successfully used in communication systems to improve the quality of communications. With the recent exploding research interest in wireless communications, the application of signal processing to this area is becoming increasingly important. Indeed, it is the advances in the signal processing technology that make most of today's wireless communications possible and hold the key to future services. The application of signal processing techniques to wireless communications is an emerging area that has recently achieved dramatic and important results. It holds the potential for even greater results in the future as an increasing number of researchers from the signal processing and communications areas participate in this expanding field.

From an industrial viewpoint, advanced signal processing technology can not only dramatically increase wireless system capacity, but also can improve communication quality including the reduction of the effects of all types of interference. In this chapter, we first define diversity as a commonly used technique in mobile radio systems to combat signal fading to improve the signal to noise ratio (SNR) of the system. We then discuss the basic principle of diversity, different diversity schemes; space, polarization, frequency and time diversity, etc.; and the roles of different diversity signal combining techniques.

We then examine the concept of equalization. Equalization means equalizing all parts of the received signal (with respect to time delay and frequency) to the same attenuation level. This helps mitigate the inter-symbol interference (ISI) and fading effect. Then we focus on the different equalization techniques (linear and non-linear). Equalization techniques which can combat and/or exploit the frequency selectivity of the wireless channel are of enormous importance in the design of high data rate wireless systems. Finally, the detailed descriptions of different speech and channel coding schemes with reference to the Global System for Mobile Communications (GSM) are also included in this chapter.

Equalization, diversity, and channel coding are three techniques that can be used independently to improve the received signal quality.

14.2 Diversity

Diversity refers to a technique for improving the transmission of a signal, by receiving and processing multiple versions of the same transmitted signal. The multiple received versions can be the result of signals following different propagation paths (spatial diversity), being transmitted at different times (time diversity) or frequencies (frequency diversity).

Diversity is a commonly used technique in mobile radio systems to combat signal fading. If several replicas of the same information carrying signal are received over multiple channels with comparable strengths which exhibit independent fading, then there is a good likelihood that at least one or more of these received signals will not be faded at any given instant. This makes it possible to deliver adequate signal level to the receiver. Without diversity techniques, in noise-limited conditions, the transmitter would have to deliver a much higher power level to protect the link during the short intervals when the channel is severely faded. In mobile radio, the power available on the reverse link is severely limited by the battery capacity of hand-held subscriber units. Diversity methods play a crucial role in reducing transmitted power needs. Also, cellular communication networks are mostly interference limited and mitigation of channel fading through use of diversity can translate into reduced variability of carrier to interference ratio (*C*/*I*), which in turn means lower *C*/*I* margin and hence better reuse factors and higher system capacity.

14.2.1 Principle of diversity

The principle of diversity systems is to send copies of the same information signal through several different channels. Performance enhancement is achieved as these channels fade independently; thus, fading will affect only a part of the transmission. The diversity techniques operate over time, frequency, or space, but the basic idea remains the same. By sending signals that carry the same information through different paths, multiple independently faded replicas of data symbols are obtained at the receiver end and more reliable detection can be achieved.

Diversity exploits the random nature of radio propagation by finding independent (or at least highly uncorrelated) signal paths for communication.

There are many techniques for obtaining independently fading branches and these can be subdivided into two main classes. The first are explicit techniques where explicit redundant signal transmission is used to exploit diversity channels. An example of explicit diversity is the use of dual polarized signal transmission and reception in many point-to point radios. Clearly, such redundant signal transmission involves a penalty in frequency spectrum or additional power.

In the second class, there are implicit diversity techniques: the signal is transmitted only once, but the decorrelating effects in the propagation medium such as multipaths are exploited to receive signals over multiple diversity channels. A good example of implicit diversity is the RAKE receiver (discussed in Chapter 20) in code division multiple access (CDMA) systems, which uses independent fading of resolvable multipath to achieve diversity gain. Figure 14.1 illustrates the principle of diversity where two independently fading signals are shown along with the diversity output signal (thick line) which selects the stronger signal. The fades in the resulting signal have been substantially smoothed out while yielding higher average power.

The envelope cross-correlation ρ between these signals is a measure of their independence.

$$\rho = \frac{E\left[[r_1 - \overline{r}_1][r_2 - \overline{r}_2]\right]}{\sqrt{E[r_1 - \overline{r}_1]^2 \, E[r_2 - \overline{r}_2]^2}} \tag{14.1}$$

where E is the expectance and r_1 and r_2 represent the instantaneous envelope levels of the normalized signals at the two receivers, and $\overline{r}_1$ and $\overline{r}_2$ are their respective means.

It has been shown that a cross-correlation of 0.7 between signal envelopes is sufficient to provide a reasonable degree of diversity gain. Depending on the type of diversity employed, these diversity channels must be sufficiently *separated* along the appropriate diversity dimension. For

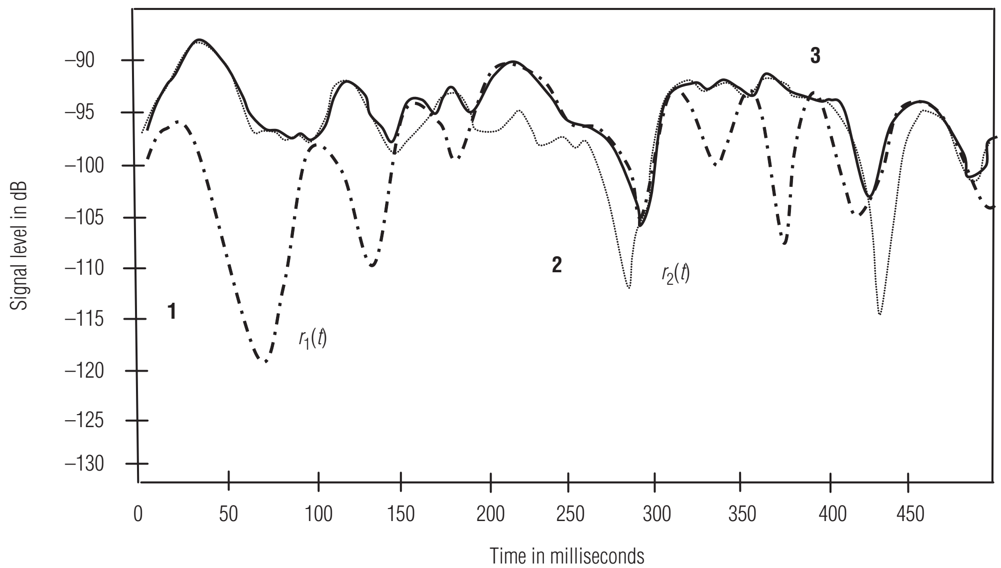


Figure 14.1 Example of diversity combining. The two independently fading signals are 1 and 2. Signal 3 is the result of selecting the strongest signal

spatial diversity, the antennas should be separated by more than the *coherence distance* to ensure a cross-correlation of less than 0.7.

14.2.2 Diversity schemes

Diversity techniques are widely used to improve wireless link performance and received signal quality. In wireless systems, the most effective method of counteracting the effects of channel fading is to use diversity schemes in the transmission and reception of the signal. Diversity is based on the fact that individual channels experience independent fading patterns. The short-term multipath fading can severely reduce transmission accuracy.

There are many ways to obtain diversity. Diversity over time can be obtained via *coding* and *interleaving*: information is coded and the coded symbols are dispersed over time in different coherence periods so that different parts of the code words experience independent fades. Analogously, one can also exploit diversity over frequency if the channel is frequency selective. In a channel with multiple transmitting or receiving antennas spaced sufficiently, diversity can be obtained over space as well. In a cellular network, macro diversity can be exploited by the fact that the signal from a mobile can be received at two base stations. Since diversity is such an important resource, a wireless system typically uses several types of diversity. There are several techniques for obtaining diversity branches. The most important of these are discussed in the following section.

Space diversity

Space diversity, also known as *antenna diversity*, is one of the most popular forms of receiver diversity schemes widely used in wireless systems. It is easy to implement and does not require additional frequency spectrum resources. Space diversity is exploited on the reverse link at the base station receiver by spacing antennas apart so as to obtain sufficient decorrelation.

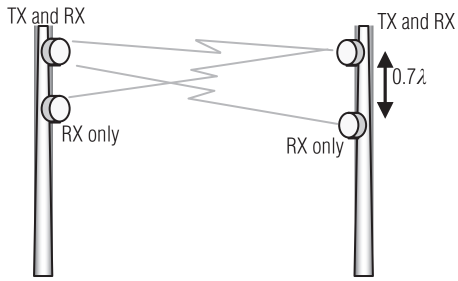


Figure 14.2 Space diversity topology

The key for obtaining minimum uncorrelated fading of antenna outputs is adequate spacing of the antennas. The required spacing depends on the degree of multipath angle spread. For example, if the multipath signals arrive from all directions in the azimuth, as is usually the case at the mobile, antenna spacing (coherence distance) of the order of 0.5λ to 0.8λ is quite adequate.

On the other hand, if the multipath angle spread is small, as in the case of base stations, the coherence distance is much larger. Also, empirical measurements show a strong coupling between antenna height and spatial correlation. Larger antenna heights imply larger coherence distances. Typically, 10λ–20λ separation is adequate to achieve $\rho = 0.7$ at base stations in suburban settings when the signals arrive from the broadside direction. The coherence distance can be 3–4 times larger for end fire arrivals. The end-fire problem is averted in base stations with tri-sectored antennas as each sector needs to handle only signals arriving ±60° off the broadside. The coherence distance depends strongly on the terrain. The space diversity topology is shown in Figure 14.2. There is a statistical relationship between the spatial separation of two receiving antennas and the likelihood of both receiving faded signals simultaneously.

Polarization diversity

In *polarization diversity*, horizontally and vertically polarized signals or left and right circular polarized waves are transmitted simultaneously. Since these are uncorrelated in the mobile radio path, one of these signals will provide strong received level after fading. In mobile radio environments, signals transmitted on orthogonal polarizations exhibit low fade correlation and, therefore, offer potential for diversity combining. Polarization diversity can be obtained in two ways:

- ***Explicit polarization diversity*** – In explicit polarization diversity, the signal is transmitted and received in two orthogonal polarizations.
- ***Implicit polarization diversity*** **–** In the implicit polarization technique, the signal is launched in a single polarization, but is received with cross-polarized antennas.

Frequency diversity

Another technique to obtain decorrelated diversity branches is to transmit the same signal over different frequencies. The frequency separation between carriers should be larger than the coherence bandwidth. The coherence bandwidth, of course, depends on the multipath delay spread of the channel. The larger the delay spread, the smaller the coherence bandwidth and more closely we can space the frequency diversity channels. Clearly, frequency diversity is an explicit diversity technique and needs additional frequency spectrum. A common form of frequency diversity is multicarrier (also known as multitone) modulation.

This technique involves sending redundant data over a number of closely spaced carriers to benefit from frequency diversity, which is then exploited by applying interleaving and channel coding/forward error correction (FEC) across the carriers. Another technique is to use frequency hopping wherein the interleaved and channel coded data stream is transmitted with widely separated frequencies from burst to burst. The wide frequency separation is chosen to guarantee independent fading from burst to burst. Some examples of systems employing frequency diversity include spread spectrum systems, such as frequency-hopping spread spectrum (FHSS) and multi-carrier spread spectrum (MCSS) systems.

Time diversity

In *time diversity*, the desired message is transmitted repeatedly over several time periods having separation between adjacent time slots larger than the channel coherence time. In mobile communications channels, the motion of the mobile together with scattering in the vicinity of the mobile causes time selective fading of the signal with Rayleigh fading statistics for the signal envelope. Signal fade levels separated by the coherence time show low correlation and can be used as diversity branches, if the same signal can be transmitted at multiple instants separated by the coherence time.

The coherence time depends on the Doppler spread of the signal, which in turn is a function of the mobile speed and the carrier frequency. Time diversity is usually exploited via interleaving, FEC coding, and automatic request for repeat (ARQ). These are sophisticated techniques to exploit channel coding and time diversity. One fundamental drawback with time diversity approaches is the delay needed to collect the repeated or interleaved transmissions. If the coherence time is large, as for example, when the vehicle is slow moving, the required delay becomes too large to be acceptable for interactive voice conversation.

14.2.3 Micro diversity

In Micro diversity or antenna diversity the signal from antennas mounted at separate locations are combined. Typically, these antennas are located on the vehicle or at the same base station tower and their spacing are few wavelengths. The received signal amplitude is correlated for antennas separated by a distance *d*. The received multipath signal becomes practically uncorrelated if antennas at the mobile are spaced by more than, say, half a wavelength.

In the analysis of this correlation, it is assumed that the mobile antenna is mounted at low height and close to all kinds of reflecting and scattering objects. The base station antenna, however, is mostly located well above such obstacles. Hence, at the base station all multipath waves arrive from approximately the same direction. If the antenna is moved over a certain small distance *d*, the phase shift is almost identical for all arriving waves. This is in sharp contrast to the situation at the mobile where motion over half a wavelength leads to almost uncorrelated signal phases. To ensure effective antenna diversity at the base station, antennas must be separated much farther than the fraction of the wavelength required for diversity at the mobile.

14.2.4 Macro diversity

In the field of wireless communication, macro diversity or site diversity is a kind of space diversity scheme using several receiver antennas and/or transmitter antennas for transferring the same signal. The distance between the transmitters is much longer than the wavelength as opposed to micro diversity where the distance is in the order of or shorter than the wavelength. In a cellular network or a wireless LAN, macro diversity implies that the antennas are typically situated in different base station sites or access points. Receiver macro diversity is a form of antenna combining and requires an infrastructure that mediates the signals from the local antennas or receivers to a central receiver or decoder.

Transmitter macro diversity may be a form of simultaneous broadcasting, where the same signal is sent from several nodes. If the signals are sent over the same physical channel (e.g., the channel frequency and spreading sequence), the transmitters are said to form a single frequency network, a term used especially in the broadcasting world. The aim is to combat fading and to increase the received signal strength and signal quality in exposed positions in between the base stations or access points. Macro diversity may also facilitate efficient broadcasting and multicasting services, where the same frequency channel can be used for all transmitters sending the same information. The diversity scheme may be based on transmitter (downlink) macro diversity and/or receiver (uplink) macro diversity.

14.2.5 Diversity signal combining techniques

The use of diversity at receiver needs combining the outputs of statistically independent fading channels, in accordance with a criterion that leads to improved receiver performance. It is assumed that the wireless communication channel is described by a frequency flat, slow fading Rayleigh channel. It implies that all the frequency components constituting the transmitted signal are characterized by the same random attenuation and phase shift. The fading remains essentially unchanged during the transmission of each symbol and the fading phenomenon is described by Rayleigh distribution. Several techniques have been studied for diversity combining. Some of these techniques are selection combining, maximal ratio, equal gain, and square law combining. They can be used with each of the aforementioned diversity schemes.

Selection combining

Selection combining is the simplest and perhaps the most frequently used form of diversity combining. With selective diversity, one best signal is chosen based on the received signal strengths from the set of diversity branches. The receiver monitors the SNR value of each diversity branch and selects the one with the maximum SNR value for signal detection. It is much easier to implement without much performance degradation. It is generally employed for the reverse link transmission where the diversity branches can be physically located in different base stations. Figure 14.3 depicts a simple block diagram of the selective diversity combining technique.

Consider *M* number of independent fading signals received by multiple receiver antennas. There are *M*-branch receivers comprising of coherent demodulators. The output of demodulators is presented to a logic circuit, which selects the particular branch receiver output having the largest SNR value of the received signal. Conceptually, selection diversity combining is the simplest

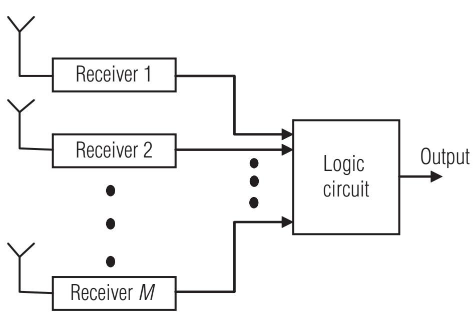


Figure 14.3 Selective diversity combining technique

form of "space diversity on receive" technique. The selection combining procedure requires that the receiver outputs are monitored continuously. At each instant of time, the receiver with the largest instantaneous SNR value is selected. From a practical implementation point of view, such a selective combining procedure is difficult.

There is an alternate method by adopting a scanning version of the selective combining procedure. Firstly, the receiver with the strongest output signal is selected and maintained till its instantaneous SNR value does not drop below a pre-defined threshold SNR value. Then, a new receiver that offers the strongest output signal is selected and the selection procedure is repeated. This method has an advantage that it requires only one receiver and is very simple to implement. The performance of the selective combining technique is not optimum because it ignores the information available from all the diversity branches. But there is exception for the particular selected receiver branch that produces the instantaneous SNR value greater than the pre-defined threshold SNR value. The probability, P_M (S_i), that all independent diversity branches receive signals, which are simultaneously less than specified threshold SNR, S_i, is given by

$$P_M\left(S_i\right)=\left(1-e^{S_i/S_r}\right)^M \tag{14.2}$$

where

S_r is the average received SNR value

M is the number of independent diversity branches

It can be seen that selective diversity can greatly improve the performance of the bit error rate (BER). The performance improvement is more significant when M is increased from 1 to 2 than it is increased from 2 to 4 and 4 to 8.

Maximal ratio combining

Maximal ratio combining is considered to be the optimum technique of combining in which the diversity branches are weighted before summing them, each weight being proportional to the signal strength of the received branch. This technique assumes that the receiver is able to accurately estimate the amplitude fading and carrier phase distortion for each diversity channel. With knowledge of the complex channel gains, the receiver coherently demodulates the received signal from each branch. After removing the phase distortion, the coherently detected signal is then weighted by the corresponding amplitude gain. The weighted received signals from all the M branches are then summed and applied to the decision device. Figure 14.4 describes the block diagram of a maximal ratio combiner.

Consider the transmission of a digital modulated signal over flat slow Rayleigh fading channels using coherent demodulation with M^{th} order diversity. It is assumed that the channel fading processes and the additive white Gaussian noise (AWGN) processes are mutually statistically independent. For a slow fading channel, the complex gain can be assumed to be a complex constant over each symbol interval. The demodulator in each channel is optimum for an AWGN channel using filters matched to the orthonormal functions.

Theoretically, the maximal ratio combiner is the optimum among linear diversity combining techniques in the context that it produces the largest possible value of instantaneous SNR value. In fact, the instantaneous output SNR value of the maximal ratio combiner can be large even when the SNR values of the individual branches are small. Practically, it is difficult to achieve the exact setting, and significant instrumentation is needed to adjust the complex weighting parameters to their exact values.

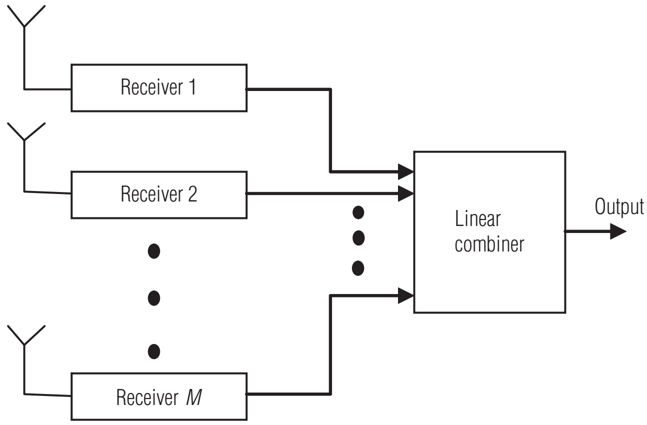


Figure 14.4 Maximal ratio diversity combining technique

Equal gain combining
In this, all the signals are weighted equally after coherent demodulation which removes the phase distortion. All the weighting parameters have their phase angles set opposite to those of their respective multipath branches, and their magnitudes are set equal to some constant value. The coherently detected signals from all the *M* branches are simply added and applied to the decision device. As the receiver does not need to estimate the amplitude fading, the receiver design is not complex. Equal gain diversity combining also relies on the ability to estimate the phase of the different diversity branches and to combine the signals coherently. Owing to hardware limitations or physical separation of the diversity receivers, it is difficult to implement it practically. The performance of an equal gain combiner is only marginally inferior to a maximal ratio combiner and superior to a selection diversity combiner. Among the four linear combining techniques, maximal ratio combining offers the best performance, followed by equal gain combining.

Square law non-linear combining
Square law combining offers the opportunity to obtain an advantage in diversity without requiring phase estimation. Unlike maximum ratio combining, square law combining is applicable to certain modulation schemes such as orthogonal modulations. This includes direct sequence CDMA signals or frequency shift keying in which different but approximately orthogonal sequences or frequencies are used to represent different data symbols.

14.2.6 Transmit diversity
Transmit diversity (TD) is one of the key contributing technologies to define the International Telecommunications Union (ITU) endorsed 3G systems, WCDMA and CDMA2000. Spatial diversity is introduced into the signal by transmitting through multiple antennas. The antennas are spaced far enough so that the signals emanating from them can be assumed to undergo independent fading. In addition to diversity gain, antenna gain can also be incorporated through channel state feedback. This leads to the categorization of TD methods into open loop and closed loop methods.

Most mobile communication channels must combat the effects of fading caused by multipath propagation. An important way of quantifying fading is in terms of a measure called the *coherence* *bandwidth* B_c, which indicates the amount of bandwidth that will fade in a correlated fashion at any instant of time. To define this correlation, consider the *delay spread* of the channel, T_d, which is the maximum duration of the mobile communication channel. The time index is a measure of the time of arrival relative to the first multipath component at time 0. Often, the "direct path" arrives first and subsequent paths represent paths reflected at increasing distances from the receiver. The coherence bandwidth is given by

$$B_c \approx \frac{1}{T_d} \tag{14.3}$$

The coherence bandwidth is approximated at typically path powers less than 5–10 per cent of the total power. Considering a communication system with bandwidth B_w. If $B_c > B_w$, the channel between the transmitter and receiver is called a *flat fading channel*, and if $B_c < B_w$ the channel is called a *frequency selective channel*. Flat fading channels are problematic for systems without time diversity, because a deep fade can result in a received signal that is below the background noise level, making communication unreliable. The worst types of channel conditions for many communication systems are slowly changing flat fading channels. This is due to the length of time the receiver cannot reliably demodulate the bits sent by the transmitter.

TD can improve the receiver performance in the presence of flat fading. It reduces the impact of fading by offering multiple independent copies of the digitally modulated waveform at the receiver, where the chance that all copies are simultaneously in a fade is very small. TD in radio communication using signals that originate from two or more independent sources has been modulated with identical information bearing signals. It may vary in their transmission characteristics at any given instant. It can help overcome the effects of fading, outages, and circuit failures. When using diversity transmission and reception, the amount of received signal improvement depends on the independence of the fading characteristics of the signal as well as circuit outages and failures.

14.3 Equalization

Equalization means equalizing all parts of the received signal with respect to time delay and frequency to the same attenuation level. This is because of the fact that the various components of the signals suffer different levels of attenuation or fading when travelling through the medium. This helps mitigate the ISI, fading effect, and so on. An equalizer is usually implemented at the baseband or at the IF section in a receiver. In a typical wireless system, the RF communication occurs in a pass band of bandwidth B with a centre frequency f_c. Most of the processing such as coding/decoding, modulation/demodulation, and synchronization are performed in the baseband.

Equalization techniques which can combat and/or exploit the frequency selectivity of the wireless channel are of enormous importance in the design of high data rate wireless systems. The purpose of an equalizer is to reduce the ISI as much as possible to maximize the probability of correct decisions.

14.3.1 Fundamentals of equalization techniques

Equalization techniques have found wide use in wireless systems for combating the effect of ISI. The general idea of equalization is to predict the ISI that will be encountered by a transmission and accordingly modify the signal to be transmitted so that the signal reaching the receiver will

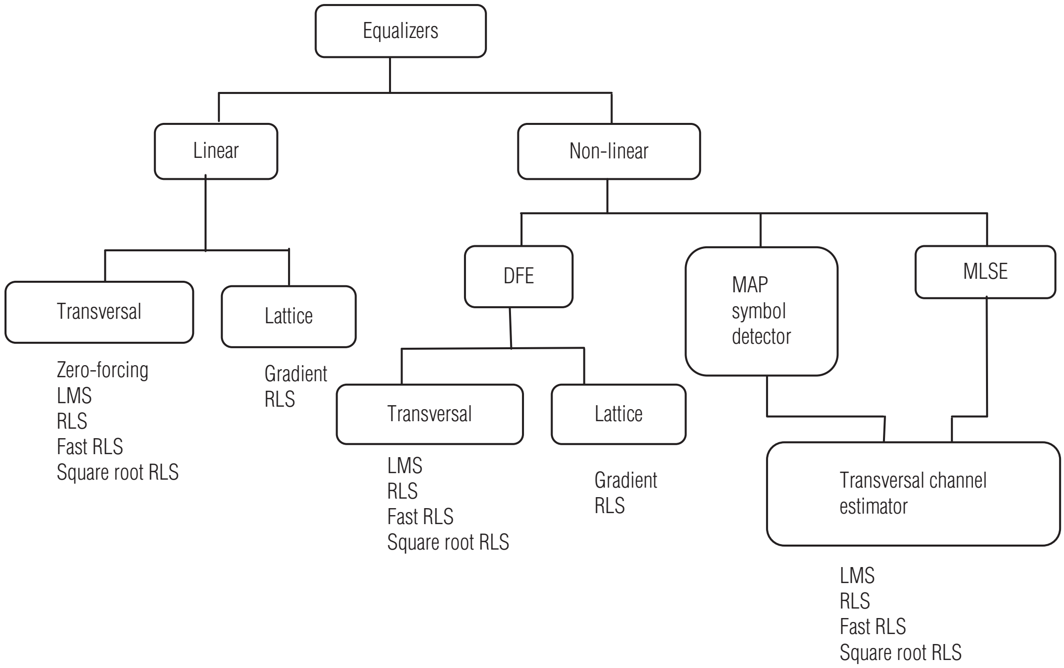


Figure 14.5 Types of equalizers

Where LMS - Least Mean Square
RLS - Recursive Least Square
DFE - Decision Feedback Equalizer
MLSE - Maximum Likelihood Sequence Estimation
MAP - Mobile Application Part

represent the information the transmitter wants to send. Equalization techniques can be broadly classified as linear and non-linear (see Fig. 14.5). These categories are determined from output of an adaptive equalizer used for subsequent control of the equalizer. In general, the analogue signal $d(t)$ is processed by the decision-making device in the receiver. The decision maker determines the value of the digital data bit being received and applies a slicing or thresholding operation (a non-linear operation) in order to determine the value of $d(t)$. If $d(t)$ is not used in the feedback path to adapt the equalizer, the equalization is linear. On the other hand, if $d(t)$ is fed back to change the subsequent outputs of the equalizer, the equalization is non-linear. Many filter structures are used to implement linear and non-linear equalizers.

Equalization is used to overcome ISI due to channel time dispersion, and diversity is used to overcome flat fading.

14.3.2 Linear equalization

A *linear equalizer* is of the simplest type and can be implemented as a finite impulse response (FIR) filters. In this, the present and the past values of the received signal are linearly weighted by the filter coefficient and summed up to produce the output. The linear equalizer can be implemented either as the simple transversal filter or as a complicated lattice filter. Linear equalizers increase the noise present in near vicinity of the spectrum, so they are not very effective on channels having severe distortion. The ISI can be completely removed, without taking into consideration the resulting noise enhacement. Symbol spaced equalizers and fractionally spaced equalizers are examples of linear equalizers.

A symbol spaced linear equalizer (SSLE) consists of a tapped delay line that stores samples from the input signal. Once per symbol period, the equalizer outputs a weighted sum of the values in the delay line and updates the weights to prepare for the next symbol period. This class of equalizer is called "symbol spaced" because the sample rates of the input and output are equal. The algorithms for the weight setting and error calculation blocks are determined by the adaptive algorithm chosen. The new set of weights depends on the current set of weights, the input signal, the output signal, and for adaptive algorithms other than constant modulus algorithm, a reference signal whose characteristics depend on the operation mode of the equalizer. In typical applications, the equalizer begins in training mode to gather information about the channel, and later switches to decision-directed mode.

A fractionally spaced equalizer (FSE) is a linear equalizer that is similar to a symbol spaced linear equalizer. By contrast, however, a fractionally spaced equalizer receives Z input samples before it produces one output sample and updates the weights, where Z is an integer. In many applications, Z is 2. The output sample rate is $1/T_b$, while the input sample rate is Z/T_b, where T_b is the bit duration. The weight updating occurs at the output rate, which is slower. Figure 14.6 illustrates a general approach to implement an equalizer using a linear equalizer circuit. The linear equalizer consists of a tapped delay line, which stores the input samples and then outputs a weighted sum of the stored values once per symbol interval.

Some algorithms like Wiener algorithm are used to minimize the error, and the tap coefficients are updated in preparation for the next symbol interval. If the time delay τ is equal to symbol interval T_s (Symbol rate = R_s), then the linear equalizer is called the symbol spaced equalizer. Such equalizers are optimum if preceded by receiving filters matched to the operating characteristics of the channels. However, in such practical cases, the channel characteristics are unknown and consequently the equalizers are sub-optimum.

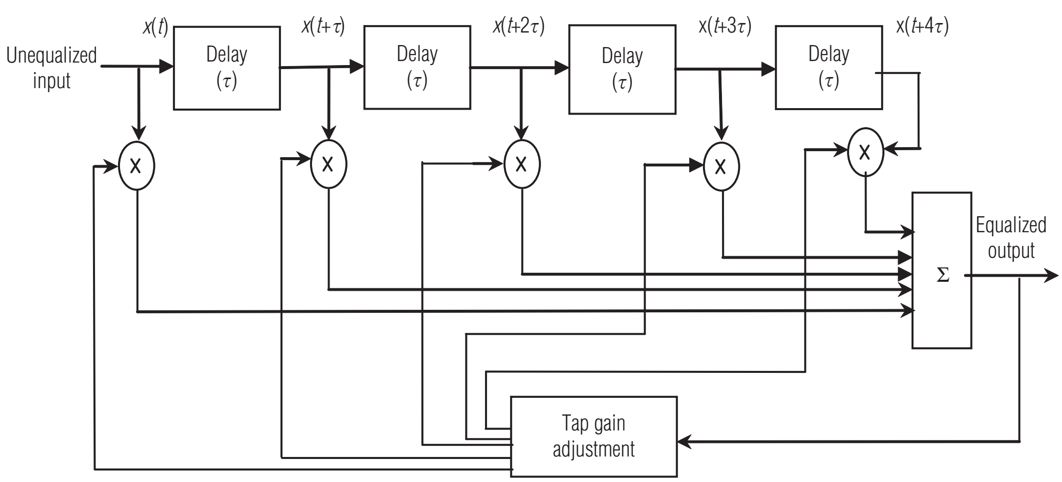


Figure 14.6 Block diagram on linear equalizer

14.3.3 Non-linear equalization

Linear equalizers have the drawback of enhancing channel noise while trying to eliminate ISI, a characteristic known as noise enhancement. As a result, satisfactory performance is unattainable with linear equalizers for channels having severe amplitude distortion. Linear equalization techniques are not preferred for wireless communication systems; whereas non-linear techniques, such as DFE, data directed estimation (DDE), and MLSE, are commonly used for wireless systems. Of the non-linear techniques, the choice for use in wireless systems is usually DFE since MLSE requires an increased computational complexity and knowledge of the channel characteristics.

When channel distortion is too severe, then non-linear equalizers are used. The basic limitation of a linear equalizer such as transversal filter is the poor performance on the channel having spectral nulls. These equalizers do not perform well on channels that have deep spectral nulls in the pass band. The most commonly used non-linear equalizers are DFE andMLSE equalizer.

Decision feedback equalization

A decision feedback equalizer (DFE) is a non-linear equalizer that contains a forward filter and a feedback filter. The forward filter is similar to the symbol spaced equalizer while the feedback filter contains a tapped delay line whose inputs are the decisions made on the equalized signal. The purpose of a DFE is to cancel ISI while minimizing noise enhancement. By contrast, noise enhancement is a typical problem with the aforementioned linear equalizers.

Figure 14.7 depicts a typical structure of DFE comprising of two filters, referred to as the forward and the feedback equalizers. The received signal is the input to the forward equalizer. The input to the feedback equalizer is the stream of the detected symbols. The tap gains of this section are the estimates of the channel sampled impulse response, including the forward equalizer. Due to past samples, this section cancels the ISI. *Decision directed mode* means that the equalizer uses a detected version of its output signal when adapting the weights. Adaptive equalizers typically start with a training sequence and switch to decision-directed mode after exhausting all symbols in the training sequence.

A delay can be taken into account by padding or truncating data appropriately. The DFE is particularly useful for channels with severe amplitude distortions and has been widely used in wireless communications. There is an improved performance since the addition of the feedback filter allows more freedom in the selection of feed forward coefficients. The exact inverse of the

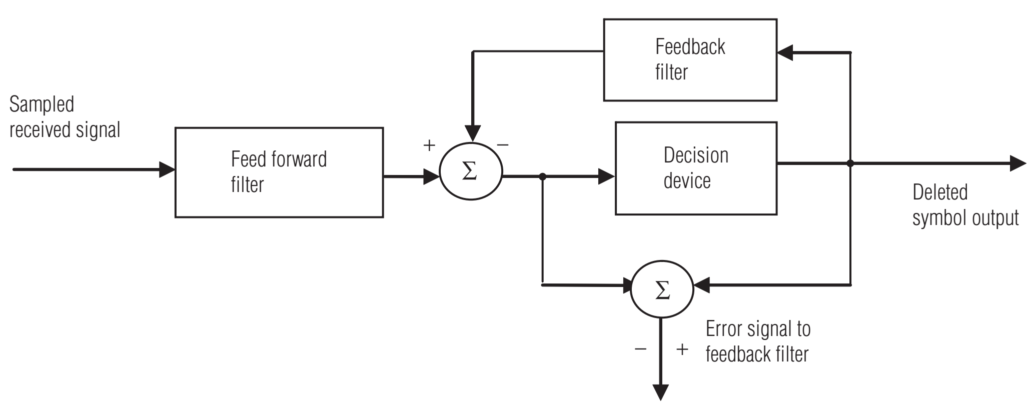


Figure 14.7 Block diagram of DFE

channel response need not be synthesized in the feed forward filter, therefore excessive noise enhancement is avoided and sensitivity to sampler phase is decreased.

The advantage of a DFE implementation is the feedback filter, which is additionally working to remove ISI, which operates on noiseless quantized levels and thus its output is free from channel noise. A drawback of the DFE structure surfaces when an incorrect decision is applied to the feedback filter. The DFE output reflects this error during the next few symbols as the incorrect decision propagates through the feedback filter. Under this condition, there is a greater likelihood of more incorrect decisions following the first one, producing a condition known as error propagation. On most channels of interest, the error rate is so low enough that the overall performance degradation is less significant.

Maximum likelihood sequence estimation equalizer

Theoretically, an MLSE equalizer yields the best possible performance but is computationally intensive. In this equalizer, the training sequence is used to estimate the channel multipath characteristics and then this estimate of the channel is used to estimate the effects of ISI. Given the estimates of the discrete channel impulse response, an MLSE receiver uses a trellis diagram with the Viterbi algorithm to obtain maximum likelihood estimates of the transmitted symbols. It receives a baseband linearly modulated input signal and outputs the maximum likelihood sequence estimate of the signal, using an estimate of the channel modelled as a FIR filter.

The Viterbi algorithm, introduced by Viterbi in 1967, reduces the complexity of maximum likelihood decoding by systematically removing paths from consideration that cannot achieve the highest path metric.

The equalizer decodes the received signal first by applying the FIR filter, corresponding to the channel estimate, to the symbols in the input signal. Then it uses the Viterbi algorithm to compute the trace back paths and the state metric, which are the numbers assigned to the symbols at each step of the Viterbi algorithm. The metrics are based on Euclidean distance. The output is the maximum likelihood sequence estimate of the signal, which is a sequence of complex numbers corresponding to the constellation points of the modulated signal. A functional block schematic of the adaptive MLSE receiver is shown in Figure 14.8. It consists of two main parts, the adaptive channel estimator and the MLSE algorithm.

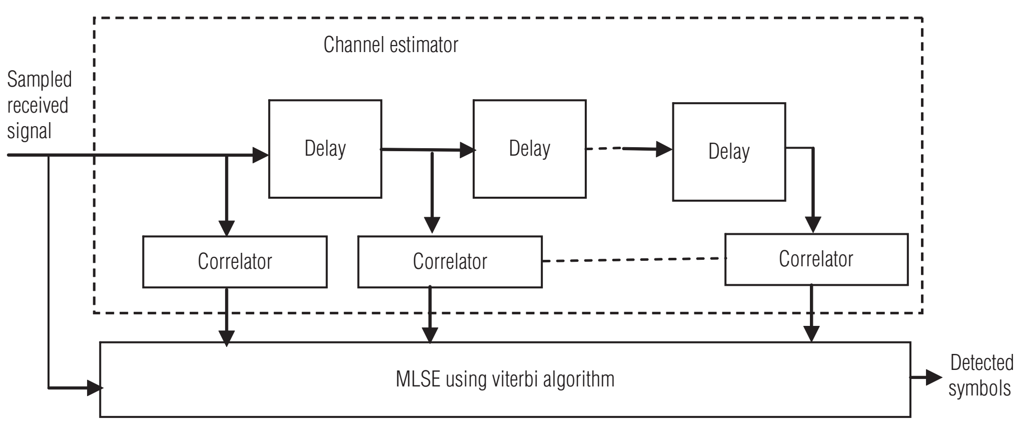


Figure 14.8 Functional schematic of adaptive MLSE receiver

The sampled impulse response of the channel, taken at symbol intervals, is measured with the adaptive channel estimator. It is then compared with the sequence of the sampled received signal with all possible received sequences and determines the most likely transmitted sequence of symbols. The MLSE is the optimal method of cancelling the ISI. However, the complexity of MLSE receiver grows exponentially with the length of the channel impulse response.

14.3.4 Adaptive equalization algorithms

The mobile fading channel is random and varies with time, so the equalizers must track the time-varying characteristics of the mobile channel and thus are called adaptive equalizers. Since an adaptive equalizer compensates for an unknown and time-varying channel, it requires a specific algorithm to update the equalizer coefficients and track the channel variations.

Three classic equalizer algorithms are primitive for most of today's wireless standards. These include the zero-forcing (ZF) algorithm, the least mean square (LMS) algorithm, and the normalized LMS (NLMS) algorithm.

Zero-forcing algorithm

The zero-forcing equalizer applies the inverse of the channel to the received signal, to restore the signal before the channel. It has many useful applications. The name zero forcing corresponds to bringing down the ISI to zero in a noise-free case. This will be useful when ISI is significant compared to noise.

For a channel with frequency response $F(f)$, the zero-forcing equalizer $C(f)$ is constructed by $C(f) = 1/F(f)$. Thus, the combination of channel and equalizer gives a flat frequency response and linear phase $F(f)\ C(f) = 1$.

In reality, zero-forcing equalization does not work in most applications, for the following reasons:

- Even though the channel impulse response has finite length, the impulse response of the equalizer needs to be infinitely long.
- The channel may have zeroes in its frequency response that cannot be inverted.
- At some frequencies, the noise may be very small. To compensate, the equalizer amplifies the noise making it a large value. As a consequence, any noise added after the channel gets boosted by a large factor and destroys the overall SNR.

The third reason is often the most important.

If the channel response (or channel transfer function) for a particular channel is $H(s)$, then the input signal is multiplied by the reciprocal of this. This is intended to remove the effect of channel from the received signal, in particular the ISI. The zero-forcing equalizer removes all ISI, and is ideal when the channel is noiseless. However, when the channel is noisy, the zero-forcing equalizer will amplify the noise greatly at frequencies f where the channel response $H(j2\pi f)$ has a small magnitude (i.e., near zeroes of the channel) in an attempt to invert the channel completely. A more balanced linear equalizer in this case is the *minimum mean square error (MSE) equalizer*, which does not usually eliminate ISI completely but instead minimizes the total power of the noise and ISI components in the output.

Least mean square algorithm

The LMS is an important member of the family of stochastic gradient algorithms. A significant feature of the LMS algorithm is its simplicity. The LMS algorithm adjusts the weights and biases of the linear network so as to minimize the MSE. LMS algorithm is the simplest algorithm based on minimization of the MSE between the desired equalizer output and the actual equalizer output. The LMS algorithm is a linear adaptive filtering algorithm that consists of two basic processes.

- ***Filter processes,*** which involve (1) computing the output of a transversal filter produced by a set of tap inputs; and (2) generating an estimation error by comparing this output to a desired response.
- ***An adaptive process,*** which involves the automatic adjustment of the tap weights of the filter in accordance with the estimated error.

Thus, the combination of these two processes working together constitutes a feedback loop around the LMS algorithm, as illustrated in the block diagram of Figure 14.9(a), and LMS algorithm filter is shown in Figure 14.9(b).

First, we have a transversal filter, around the LMS algorithm. This component is responsible for performing the filtering process. Second, we have a mechanism for performing the adaptive control process on the tap weight of the transversal filter, hence the designation adaptive weight control mechanism.

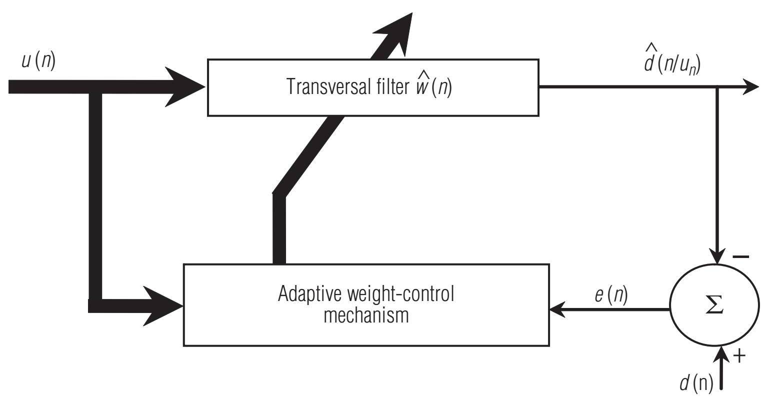


Figure 14.9(a) Block diagram of adaptive transversal filter

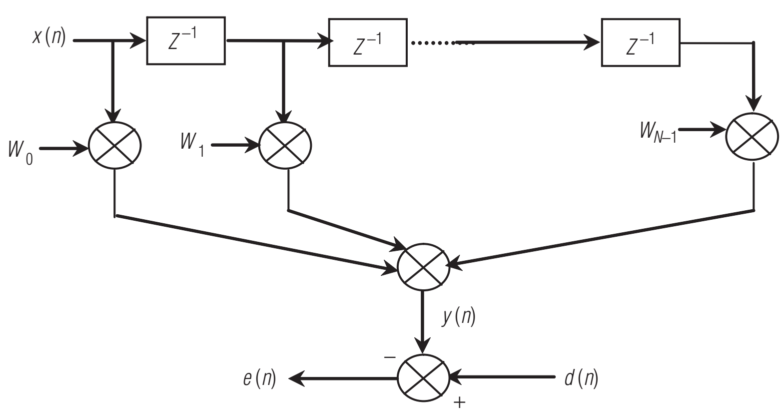


Figure 14.9(b) LMS algorithm filter

In practice, the minimization of the MSE is carried out recursively and may be performed by use of the stochastic gradient algorithm. It is the simplest equalization algorithm and requires only ($2N + 1$) operations per iteration. The filter weights are updated by the update equation. Let variable n denote the sequence of iteration. LMS is computed iteratively by considering $w_k(n)$ to denote the tap-weight vector of LMS filter, computed at iteration (time step) n. The adaptive linear operation of the filter is completely described by the recursive equation (assuming complex data).

Filter output:

$$y(n) = \hat{w}^{H}(n)\, x(n-i)$$

Estimation error:

$$e(n) = d(n) - y(n)$$

Tap-weight adaptation:

$$w_k(n+1) \equiv w_k(n) + \mu e_k(n)\, x(n-k) \tag{14.4}$$

where k is the k^{th} delay stage in the equalizer
- μ is the step-size parameter
- $w_k(n)$ is the tap-weight vector of LMS filter
- n is the iteration time step
- $e_k(n)$ is the the tap-input vector
- $x(n - k)$ is the input vector

The LMS equalizer maximizes the signal-to-distortion ratio at its output within the constraints of the equalizer filter length. If an input signal has a time dispersion characteristic that is greater than the propagation delay through the equalizer, then the equalizer will be unable to reduce distortion. The convergence rate of the LMS algorithm is slow due to the fact that there is only one parameter, the step size that controls the adaptation rate.

In reality, however, exact measurement of the gradient vector is not possible since this would require prior knowledge of both the correlation matrix R of the tap inputs and the cross-correlation vector ρ between the tap inputs and the desired response. Consequently, the gradient vector must be estimated from the available data. To prevent the adaptation from becoming unstable, the value of μ is chosen from

$$0 < \mu < 2 \Big/ \sum_{i \equiv 1}^{N} \lambda_i \tag{14.5}$$

where λ_i is the i^{th} Eigen value of the covariance matrix R.

Normalized LMS algorithm

In the LMS algorithm, the correction that is applied to $w_k(n)$ is proportional to the input sample $x(n - k)$. Therefore, when $x\,(n - k)$ is large, the LMS algorithm experiences gradient noise amplification. This problem is eliminated with the normalization of the LMS step size by $\|x(n)\|^2$ in the NLMS algorithm. Only when $x(n - k)$ becomes close to zero, the denominator term $\|x(n)\|^2$ in the NLMS equation becomes very small and the correction factor may diverge. So, a small positive number ε is added to the denominator term as the correction factor. Here, the step size is time varying and is expressed as

$$\mu(n) = \frac{\beta}{\left\| x(n) \right\|^2 + \varepsilon} \tag{14.6}$$

where

β is the positive step size

$\mu(n)$ is the step-size sequence

Therefore, the NLMS algorithm update equation takes the form of

$$w_k(n+1) = w_k(n) + \frac{\beta}{\left\| x(n) \right\|^2 + \varepsilon} e_k(n)\, x(n-k) \tag{14.7}$$

The NLMS algorithm is utilized, as its convergence speed is, generally, superior to that of LMS. The NLMS algorithm differs from the LMS algorithm in the incoming reference variable and is normalized by the magnitude of the incoming reference signal.

14.4 Review of speech coding

The efficient utilization of the allocated spectrum is the prime objective in the design of digital wireless communication systems. The systems rely on the use of speech coding to remove almost all the natural redundancy inherent in an analogue speech signal, while ensuring a high quality reproduction of the original speech signal at the receiver. The speech coding, also called *source coding*, makes the information signal compatible with digital processing. Encoding analogue voice into digital format enhances the quality of voice, improves the overall performance of the system in terms of spectral efficiency, and increases the system capacity.

Digital modulation techniques combined with robust speech coding schemes, which can inherently withstand higher BERs needing less channel coding, will be the ultimate choice for a spectrum-efficient mobile communication system. The carrier modulation and speech coding techniques are logically independent processes but are strongly interrelated. The improvement in either of them is towards achievement of a common goal of higher spectral efficiency.

Higher level modulation offers higher spectral efficiency but higher SNR is required to achieve a given BER that is difficult to achieve in a mobile environment. The required S/N can be reduced by using low bit rate speech coding and efficient error correction techniques prior to modulation. The important parameters of a speech coder are the transmitted bit rate, the speech quality, the robustness in the presence of fading and interference, and the complexity of implementation. Various speech coding techniques are mainly based on linear predictive coding (LPC) strategy.

14.4.1 Selection of speech coders for mobile communication

Speech coder can be defined as a hardware circuit that represents analogue waveforms with a sequence of binary digits. The fundamental idea behind coding schemes is to take advantage of the special features of human speech system, the statistical redundancy, and the shortcomings in human capability to receive sounds. The speech signal varies quite infrequently resulting in a high degree of correlation between consecutive samples. This short-term correlation is due to the nature of the vocal tract.

There also exists a long-term correlation due to the periodic nature of the speech. This statistical redundancy can be exploited by introducing prediction schemes, which quantize the prediction error instead of the speech signal itself. The shortcomings in human capability to receive sounds, on the other hand, lead to the fact that a lot of information in the speech signal is perceptually irrelevant. The perceptual irrelevancy means that the human ear cannot differentiate between changes of magnitude below a certain level and cannot distinguish frequencies below 16 Hz or above 20,000 Hz. This can be exploited by designing optimum quantization schemes, where only a finite number of levels are necessary.

Speech coding is the process for reducing the bit rate of digital speech representation for transmission or storage, while maintaining a speech quality that is acceptable for the application. There are three different speech coding methods, which use different features in different ways:

- Waveform coding
- Source coding
- Hybrid coding

The following sections will explain the speech coding methods in detail. Figure 14.10 shows the plot between the bit rate (Kbps) on a logarithmic axis and the speech quality classes of "poor to excellent" corresponding to the five point MOS scale values of 1 to 5, defined by the ITU. It may be noted that for low complexity and low delay, a bit rate of 32–64 Kbps is required. This suggests the use of waveform codecs. However, for low bit rate of 4–8 Kbps, hybrid codecs should be used. These types of codecs tend to be complex with high delay.

14.4.2 Speech codec attributes

The speech codec is able to achieve a much higher compression ratio, which results in a smaller amount of digital data transmission. Speech quality as produced by a codec is a function of transmission bit rate, complexity, delay, and quality. Therefore, it is mandatory to consider all

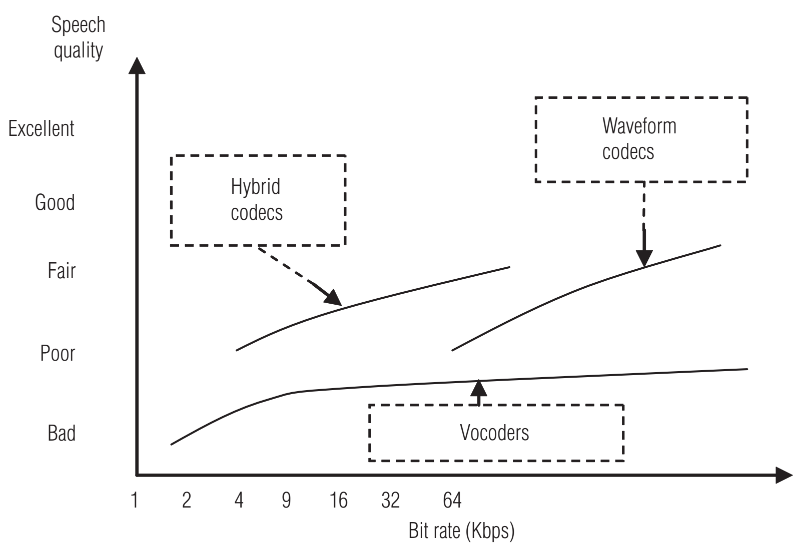


Figure 14.10 Quality of service versus bit rate

of these attributes when considering speech codecs. To have more delay, the codecs with lower bit rate can be considered as compared to the higher bit rate codecs. They are generally more complex to implement and often have lower speech quality than the higher bit rate codecs. The following are the speech codec attributes.

Transmission bit rate

Since the speech codec shares the communication channel with other data, the peak bit rate should be as low as possible so as not to use a disproportionate share of the channel. The codecs below 64 Kbps are primarily developed to increase the capacity of circuit multiplication equipment used for narrow bandwidth links. For the most part, they are fixed bit rate codecs, meaning they operate at the same rate regardless of the input. In the variable bit rate codecs, network loading and voice activity determine the instantaneous rate assigned to a particular voice channel. A two-state variable bit rate system is obtained by combining any of the fixed rate speech codecs with a voice activity detector (VAD). The lower rate could be either zero or some low rate needed to characterize slowly changing background noise characteristics. Either way, the bandwidth of the communications channel is only used for active speech.

Delay

The delay of a speech codec can have a great impact on its suitability for a particular application. The components of total system delay include the following:

- Algorithmic delay
- Processing delay
- Transmission delay

Algorithmic delay is mainly due to the delay introduced by frame size, look ahead, and multiplexing. Most low bit rate speech codecs process a frame of speech data at a time. The speech parameters are updated and transmitted for every frame. In addition, to analyse the data properly it is sometimes necessary to analyse data beyond the frame boundary. Hence, before the speech can be analysed, it is necessary to buffer a frame's worth of data. The resulting delay is referred to as *algorithmic delay*. This delay component cannot be reduced by changing the implementation, but all other delay components can.

The *processing delay* is the second major contribution for delay, which comes from the time taken by the encoder to analyse the speech and the decoder to reconstruct the speech. It depends on the speed of the hardware used to implement the coder. The sum of the algorithmic and processing delays is called the *one-way codec delay*.

The transmission delay, third component of delay, is the time taken for an entire frame of data to be transmitted from the encoder to the decoder. The total of the three delays (algorithmic, processing, and transmission delay) is the *one-way system delay*. In addition, frame interleaving delay adds an additional frame delay to the total transmission delay. Frame interleaving is necessary to combat channel fading and is part of the channel coding process.

Complexity

Special purpose hardware is used to implement the speech codecs, such as digital signal processing (DSP) chips. The following are the main attributes of DSP:

- The computing speed, in millions of instructions per second (MIPS)
- Random-access memory (RAM)
- Read-only memory (ROM)

To design a speech codec, the system designer should make estimation about how much of these resources are to be allocated. Here we have two types of speech codecs based on the complexity:

- Low complexity codecs, which use less than 15 MIPS
- High complexity codecs, which require 30 MIPS or more

More complexity results in higher costs and greater power usage. For portable applications, greater power usage means reduced time between battery recharges or use of larger batteries, which means more expense and weight.

Quality

Quality is the main attribute among all the speech codec attributes. In many applications, along with the general-purpose noise, that is, thermal noise, shot noise, etc., there is a large amount of background noise like car noise, street noise, and office noise. How well does the codec perform under these adverse conditions? What happens when there are channel errors during transmission? Are the errors detected or undetected? If undetected, the codec must perform even more robustly than when it is informed that entire frames are in error. How good does the codec sound when speech is encoded and decoded twice? All these questions must be carefully evaluated during the testing phase of a speech codec. The speech quality is often based on the 5-point MOS scale as defined by the International Telecommunication Union-Technical (ITU-T).

14.4.3 Linear-prediction based analysis-by-synthesis

In linear-prediction based analysis-by-synthesis (LPAS), coding predicting the present (or future) value of a discrete-time signal is done from a given set of past samples of the signal. The smaller the prediction error in a statistical sense, the more reliable the model will be. The *prediction error* is defined as the difference between the actual future value of the signal and the predicted value produced by the model. To be specific, consider a discrete-time signal represented by the set of samples $x(t), x(t-T_S), \ldots, x(t-NT_S)$, where T_s denotes the sampling period. The sampling rate f_s is related to the sampling period as $f_s = 1/T_s$. In the one-step form of linear prediction, the requirement is to estimate the present value of the signal $x(t)$ given the N past samples $x(t-T_S), x(t-2T_S), \ldots, x(t-NT_S)$.

Let $\hat{x}(t)$ be expressed as a linear combination of the N past samples through the formula

$$\hat{x} = \sum_{n=1}^{N} a_n x(t - nT_S) = a^T X \tag{14.8}$$

where superscript T denotes matrix transposition and the N-by-1 signal vector x and parameter vector a are, respectively,

$$X = \left[x(t-T_S), x(t-2T_S), \ldots, x(t-NT_S)\right]^T \tag{14.9}$$

and

$$a = \left[a_1, a_2, \ldots, a_N\right]^T \tag{14.10}$$

The parameters $a_{1,}\ a_2, \ldots, a_N$ in effect define the N degrees of freedom available to us in designing the predictive model. Equation (14.8) readily suggests the tapped delay line (TDL) shown in

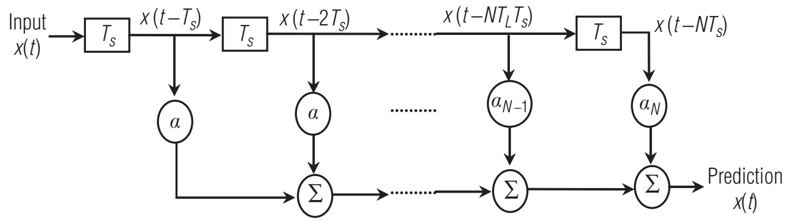


Figure 14.11 Structure of linear predictor

Figure 14.11 as the structure for the predictive model. The key question is how do we determine the filter coefficients? To answer this question, we need a statistical criterion for optimizing the design of the filter. A criterion widely used in practice is the MSE criterion, defined by

$$\text{MSE} = E[(x(t) - \hat{x}(t))^2] \tag{14.11}$$

where E denotes the statistical expectation operator and the difference $x(t) - \hat{x}(t)$ stands for the prediction error. It turns out that if $x(t)$ is the sample value of a stationary random process, then the optimum value of the parameter vector a is given by

$$a = R^{-1}r \tag{14.12}$$

where the $N \times N$ matrix R is the correlation matrix of the tap inputs of the predictive model and the $N \times 1$ vector r has its elements as the autocorrelation of the input signal $x(t)$ for lags $T_S, 2T_S, \ldots, NT_S$. The symbol R^{-1} in Equation (14.12) stands for the inverse of the correlation matrix. If, however, the physical process responsible for the generation of the signal $x(t)$ is *non-stationary* (i.e., its statistics vary with time), then we require the use of an adaptive procedure whereby the model parameters are allowed to vary with time.

14.4.4 Waveform coding

In general, waveform codecs are designed to be independent of signal. They map the input waveform of the encoder into a facsimile-like replica of it at the output of the decoder. Coding efficiency is quite modest. The coding efficiency can be improved by exploiting some statistical signal properties, if the codec parameters are optimized for most likely categories of input signals, while still maintaining good quality for other types of signals as well. The waveform codecs are further subdivided into *time-domain waveform* codec and *frequency-domain waveform* codec.

Time-domain waveform coding

Waveform codes attempt without using any knowledge of how the signal to be coded is generated to produce a reconstructed signal whose waveform is as close as possible to the original signal. That is, they should be signal independent. They are low complexity codes, which produce high quality speech at rates above 16 Kbps. Types of modulations used in waveform coding are pulse code modulation (PCM), adaptive PCM, differential PCM, adaptive differential PCM, delta modulation (DM), and adaptive DM.

Frequency-domain waveform coding

In the frequency-domain waveform codecs, the input signal undergoes some short-time spectral analysis. The signal is split into a number of frequency-domain sub-bands. The individual sub-band signals are then encoded by using different numbers of bits to fulfil the quality requirements of that band based on its prominence. The various schemes differ in their accuracies of spectral analysis and in the bit allocation principle (fixed, adaptive, and semi-adaptive). Two well-known representatives of this class are sub-band coding (SBC) and adaptive transform coding (ATC).

14.4.5 Vocoders

The source coding is also called a voice codec or vocoders. It is a hardware circuit that converts the human speech into a digital code and vice versa. Speech coding differs from other forms of audio coding, as speech is a much simpler signal than most other audio signals, a speech signal is limited to a bandwidth of 300–3,400 kHz (whereas audio signal is limited to a bandwidth of 0–20 kHz, e.g., audible range). There is a lot of statistical information available about the properties of speech. The pressure of sound waves radiated from the lips produces the human speech, although, with some sounds, significant energy emanates also from the nostrils, throat, etc.

In human speech, the air compressed by the lungs excites the vocal cord in two typical modes. When generating voice sounds, the vocal cord vibrates and generates quasi-periodic voice sounds. In the case of lower energy unvoiced sounds, the vocal cord does not participate in voice production and the source acts like a noise generator. The excitation signal is then filtered through the vocal apparatus, which behaves like a spectral shaping filter. This can be described adequately by an all pole transfer function that is constituted by the spectral shaping action of the vocal tract, lip radian characteristics, etc. In the case of vocoders, instead of producing a close replica of an input signal at the output, an appropriate set of source parameters is generated to characterize the input signal sufficiently close for a given period of time.

The following steps are used in this process:

1. First, the speech signal is partitioned into segments of 5–20 ms.
2. To minimize the prediction residual energy, the speech segments are subjected to spectral analysis to produce the coefficients of the all zero analysis filters. This process is based on the computation of the speech autocorrelation coefficients and then using either matrix inversion or iterative scheme.
3. Then the corresponding source parameters, that is, the excitation parameters and filter coefficients, are specified. These parameters are quantized and transmitted to the decoder to synthesize a replica of the original signal by exciting the all pole synthesis filter.

The quality of this type of scheme is predetermined by the accuracy of the source model rather than the accuracy of the quantization of the parameters. The speech quality is limited by the fidelity of the source model used. The main advantage of vocoders is their low bit rate, with the penalty of relatively low, synthetic speech quality. Vocoders can be classified into the frequency-domain and time-domain subclasses. When compared with the time-domain vocoders, the frequency-domain vocoders are more effective.

14.4.6 Hybrid coding

Hybrid coding method constitutes an attractive trade-off between waveform coding and source coding, both in terms of speech quality and transmission bit rate, although usually at the price of higher complexity. Combing waveform and source coding methods in order to improve the speech quality and reduce the bit rate falls into this broad category of speech coding method. The most important family of hybrid codecs, often reffered to as analysis-by-synthesis (Abs) codecs.

14.5 Review of channel coding

Channel coding is effective in combating independent random transmission errors over a noisy channel. Channel coding is a common strategy to make digital transmission more reliable, or, equivalently, to achieve the same required reliability for a given data rate at a lower power level at the receiver. This gain in power efficiency is called *coding gain*. For mobile communication systems, channel coding is often indispensable.

Channel coding adds redundancy information to the information data at the transmitter in a deterministic manner, following some logical relation with the original information. The receiver receives the encoded data with transmission errors. At the receiver, the original information can be decoded with the controlled redundancy based on the same logical relationship between the original information and redundant information. The redundancy causes channel coding to consume additional frequency bandwidth during transmission and may seem to be a waste of system resources. However, if an error correction code is designed properly using the frequency bandwidth and transmission power, the coded sequences can be transmitted at a faster rate.

When the transmission rate for the information bits remains the same as in the uncoded system, the transmission accuracy for the coded system is higher. This results in the coding gain, which translates to higher transmission accuracy, higher power, and higher spectral efficiency. In cellular systems, the traffic consists of compressed data and is very sensitive to transmission errors. Therefore, channel coding can be defined as the processing of coding discrete digital information in a form suitable for transmission, with an objective of enhanced reliability.

Channel coding is applied to ensure adequacy of transmission quality in terms of BER and frame error rate. Thus, channel coding provides excellent BER performance at low SNR values at the expense of reduction in the bandwidth efficiency of the wireless link in high SNR conditions. To improve the reliability of a digital transmission system further without increasing the required bandwidth or transmit power, channel coding can be combined with trellis-coded modulation. This is achieved by combining convolution code with a higher order modulation scheme at the transmitter and by combining the channel decoder and demodulator together at the receiver.

Figure 14.12 shows the classical channel coding setup for a digital transmission system. The channel encoder adds redundancy to digital data b_i from a data source. For simplicity, we will often speak of data bits b_i and channel encoder output bits c_i, keeping in mind that other data symbol alphabets than binary ones are possible and the same discussion applies to that case.

Speech coding scheme is used to save the bandwidth and improve the bandwidth efficiency, whereas channel coding is employed to improve the signal quality and to reduce the BER.

Owing to the inevitable presence of noise in the wireless channel, the transmitted data sequences are corrupted. This increases the bit error probability at the receiver. The design goal of channel coding is to detect and correct the bit errors in the received data sequence by adding some extra redundant bits into the transmitted data sequence. This is done by error control coding. Error control coding is the process of adding redundant information to a message to be transmitted that can then be used at the receiving end to detect and possibly correct errors in the transmission. There are two different approaches for error control coding

- Automatic repeat request (ARQ)
- Forward error correction (FEC)

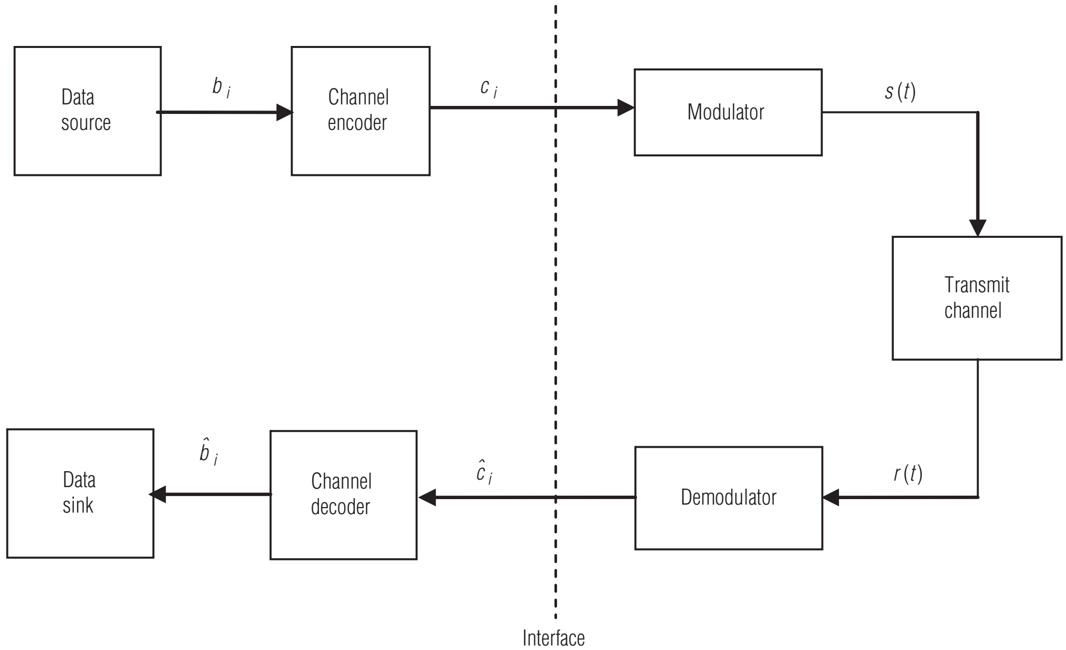


Figure 14.12 Block diagram for a digital transmission setup with channel coding

The error correction and detection codes can be broadly classified into three categories:

- Block codes; Hamming code, Bose–Chaudhuri–Hocquenghem (BCH) code, Reed–Soloman code, etc.
- Convolution codes; Viterbi code
- Turbo codes

14.5.1 Error correcting codes

Error controlling codes are used for correcting errors when messages are transmitted over a noisy channel or a stored data is retrieved. Error correcting codes are a kind of safety net– the mathematical insurance against the vagaries of an imperfect digital world. The main idea behind error correcting codes is to add some redundancy. This redundancy is added in a controlled manner. The encoded message when transmitted might be corrupted by noise of the channel. The original signal at the receiver's end can be recovered from the corrupted signal if the numbers of errors are within the limit for which the code has been designed.

The objectives of a good error control coding scheme include the following:

1. Error correcting capability in terms of the number of errors that it can correct
2. Fast and efficient encoding of the message
3. Fast and efficient decoding of the received message
4. Maximum transfer of information bits per unit time

14.5.2 Block codes

A block code operates on fixed length input blocks of information bits, which are known as message blocks. Here the general principle is to segment the information into blocks and add a parity check number, which is normally the product of information bits contained in the block. Block codes are FEC codes that help to detect and correct a limited number of errors. In a block encoder, k is the input information bits to the encoder unit and the encoder then uses different generator polynomials and adds r numbers of extra bits to k bits.

So the total number of output bits from the encoder will be $n = r + k$. The block code is referred to as an (n, k) code, and the rate of the code is defined as $R_c = k/n$. With a k input bit sequence, $2k$ distinct code words can be transmitted. For each code word, there is a specific mapping between the k message bits and the r check bits. The code is systematic because a part of the sequence in the code word coincides with the k message bits. As a result, it is possible to make a clear distinction in the code word between the message bits and the parity bits. The code is binary as these are constructed from bits and linear as each code word can be created by a linear modulo-2 addition of two or more code words.

To generate an (n, k) block code, the channel encoder accepts information data in successive k-bit blocks and adds $(n - k)$ redundant bits to each block that are algebraically related to the k message bits, where $k < n$, thereby producing an overall encoded block of n bits. The data rate at which the block encoder produces bits is given by

$$R_o = (n / k)R_s \tag{14.13}$$

Where R_o is the channel data rate at the output of block encoder
R_s is the source information data rate

Example problem 14.1

In GSM cellular communication system, a data block of 180 bits is encoded into 220 bits of code word on the control channel before sending it to a convolution encoder. Determine the number of parity check bits added and the code rate of the block encoder used.

Solution

Number of information bits, $k = 180$ bits (given)
Number of encoded bits, $n = 220$ bits (given)

Step 1: To determine the number of parity check bits added
Number of parity check bits $= n - k = 220 - 180 = 40$ bits

Step 2: To determine t he code rate of the block encoder used
The code rate of block encoder, $r = k/n = 180/220 = 0.82$

Block codes use algebraic technique properties to encode and decode blocks of data bits or symbols. The code words are generated in a systematic form such that the original k number of source information data bits are retained as it is and the $(n - k)$ parity check bits or redundant bits are either appended or prepended to information data bits. A simple operation of a block code is depicted in Figure 14.13.

Mostly, operation in a block encoder are based on linear feedback shift registers that are easy to implement and inexpensive. The parity check bits are generated using a generator polynomial or matrix. Codes generated by a polynomial are called *cyclic codes*, and the resultant block codes are called c*yclic redundancy* c*heck* (CRC) codes. The code word comprising of n bits of encoded

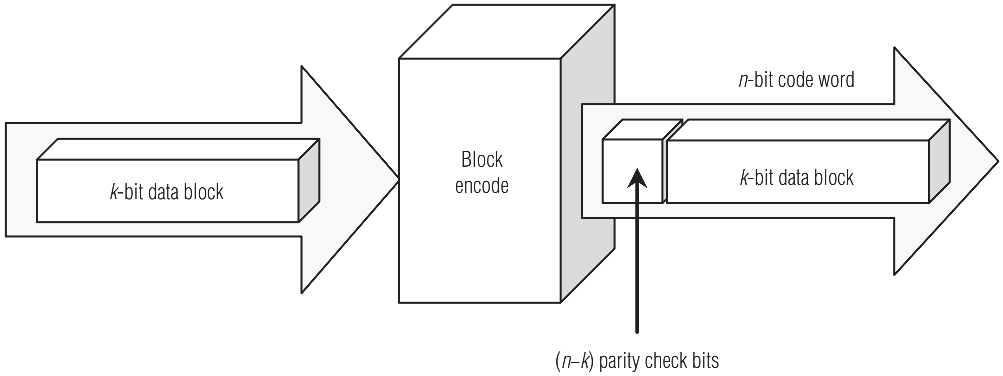


Figure 14.13 A simple block code operation

block of data is a transmitter over the channel. The received code word may be error free or modified due to channel error. Sometimes it may result in another valid code word, which cannot be detected as error, with a probability given by

$$P_{FD} \leq 2^{-(n-k)} \tag{14.14}$$

where P_{FD} is the probability of false detection.

It can be minimized by designing block codes in such a way that different code words have a large code distance. Block codes have certain limitations such as the following:

- Block codes have limited ability to handle large numbers of distributed errors in a wireless channel.
- Block codes use hard decisions that tend to destroy information. They do not achieve the performance as obtained through the use of soft decisions.
- The block encoder accepts a k-bit information data block and generates an n-bit code word. Thus, code words are generated on a block-by-block data basis. Clearly, a provision must be made in the block encoder to buffer an entire information data block before generating the associated code word.

14.5.3 Convolutional codes

Convolutional coding is a special case of error control coding. There are applications in which the information bits come in serially rather in large blocks, in which case the use of a buffer as in case of a block encoder may be undesirable. As for block codes, the convolution codes divide the bit stream from the source into k-bit blocks. Each k-bit block is then encoded into an n-bit block, but unlike block codes, the values of n bits depend not only on the values of the k bits in the corresponding source block but also on the values of the bits in the previous k-bit source blocks. Although convolution code is more complex, these are more powerful than block codes as they exploit past history. The idea is to make every code word symbol to be the weighted sum of the various input message symbols.

For a convolutional code at every moment k, the encoder delivers a block of N binary symbols 2 $c_k = (c_k,1,\ c_k,2,\ldots,\ c_k,N\ )$, a function of the block of K information symbols $d_k = (d_k,1,d_k,2,\ldots,\ d_k,K)$ present at its input along with m preceding blocks. Convolutional codes consequently introduce a memory effect of the order m. The quantity $\nu = m + 1$ is called the *constraint length of the code* and

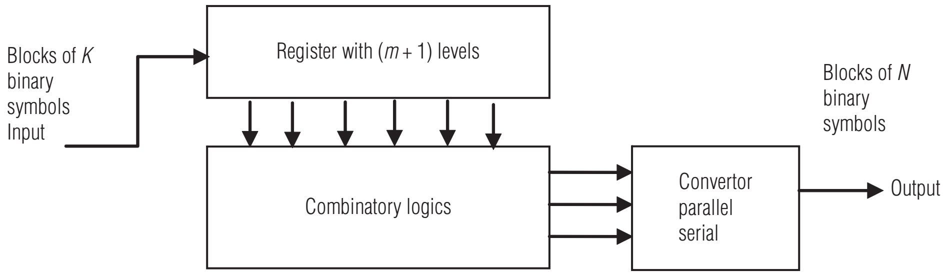


Figure 14.14 General diagram of a convolution encoder with an output of K/N and memory m

the ratio $R = K/N$ is called the *code rate*. If K information symbols at the encoder input are found explicitly in the coded block c_k, that is,

$$c_k = (d_k,1,\ldots,d_k,K,c_k,K+1,\ldots,c_k,N)$$

Then the code is known as *systematic*. In the contrary case, it is known as *non-systematic*. The general diagram of an encoder with output K/N and memory m is represented in Figure 14.14.

At every moment k, the encoder has m blocks of K information symbols in memory. These m K binary symbols define the S_k state of the encoder:

$$S_k = (d_k,d_{k-1},\ldots,d_{k-(m+1)})$$

If the input of the encoder is permanently fed by blocks of K information symbols, then the encoder output consists of N infinite sequences of coded symbols, which, for the output i, have the form:

$$(c_1,\ i,\ c_2,\ i,\ldots,c_k,\ i,\ldots)\ i = 1,\ldots,N$$

Let us note that convolutional codes are well adapted to code transmissions with continuous flow of data. Indeed, the sequences of data to be coded can have any length. To each coded sequence $i = 1, \ldots, N$

Generally, convolution coders are more powerful than block codes in terms of providing FEC, but are not useful for detection or ARQ schemes. At the receiver, FEC is performed using a maximum likelihood decoding algorithm that determines what data sequence would have been most likely transmitted, given the received sequence of bits. VLSI implementation of the Viterbi algorithm is the most common algorithm used for this purpose.

14.5.4 Turbo codes

Turbo codes may use serial (concatenated) and/or parallel recursive convolution codes. Traditionally, the design of good codes has been tackled by constructing codes with a great deal of algebraic structure, for which there are feasible decoding schemes. Such an approach is exemplified by the linear block codes and convolutional codes discussed in preceding sections. The difficulty with these traditional codes is that, in an effort to approach to the theoretical limit for Shannon's channel capacity, we need to increase the code word length of a linear block code or the constraint length of a convolutional code, which in turn causes the computational complexity of a maximum likelihood decoder to increase exponentially. Ultimately, we reach a point where complexity of the decoder is so high that it becomes physically unrealizable.

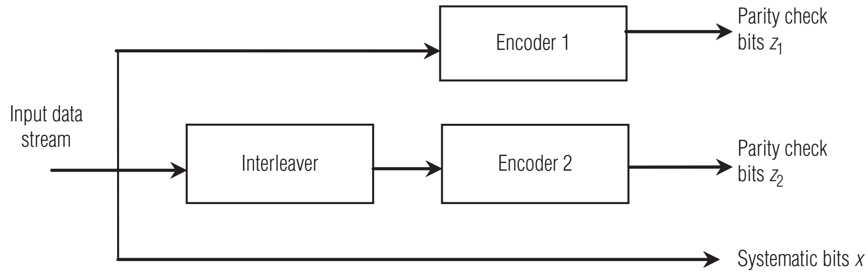


Figure 14.15 Block diagram of turbo encoder

Various approaches have been proposed for the construction of powerful codes with large equivalent block lengths structured in such a way that the decoding can be split into a number of manageable steps. Building on these previous approaches, the development of turbo codes and low-density parity-check codes has been by far most successful. Indeed this development has opened a brand new and exciting way of constructing good codes and decoding them with feasible complexity.

Turbo coding

In its most basic form, the encoder of a turbo code consists of two constituent systematic encoders joined together by means of an interleaver, as illustrated in Figure 14.15.

An interleaver is an input output mapping devise that permutes the ordering of a sequence of symbols from a fixed alphabet in a completely deterministic manner, that is, it takes the symbols at the input, produces identical symbols at the input, and produces identical symbols at the output but in a different temporal order. The interleaver can be of many types, of which two of them are periodic and pseudorandom. Turbo codes use a pseudorandom interleaver, which operates only on the systematic bits.

There are two reasons for the use of an interleaver in a turbo code:

- To tie together errors that is easily made in one-half of the turbo code to errors that are exceptionally unlikely to occur in the other half. This is indeed the main reason why the turbo code performs better than a traditional code.
- To provide robust performance with respect to mismatched decoding, this is a problem that arises when the channel statistics are not known or have been incorrectly specified.

Turbo decoding

Turbo codes derive their distinctive name from analogy of the decoding algorithm to the turbo engine principle. Figure 14.16 shows the basic structure of the turbo decoder. It operates on noisy versions of the systematic bits and two sets of parity check bits in two decoding stages to produce an estimate of the original message bits.

The first decoder, the interleaver, the second decoder, and the de-interleaver constitute a single hop feedback system. This arrangement makes it possible to iterate the decoding process in the receiver many times so as to achieve satisfactory performance. The inputs to the first decoder are the channel samples corresponding to the systematic information bits, the channel samples corresponding to the parity bits of the first encoder, and the extrinsic information about the systematic bits that were determined from the second decoder.

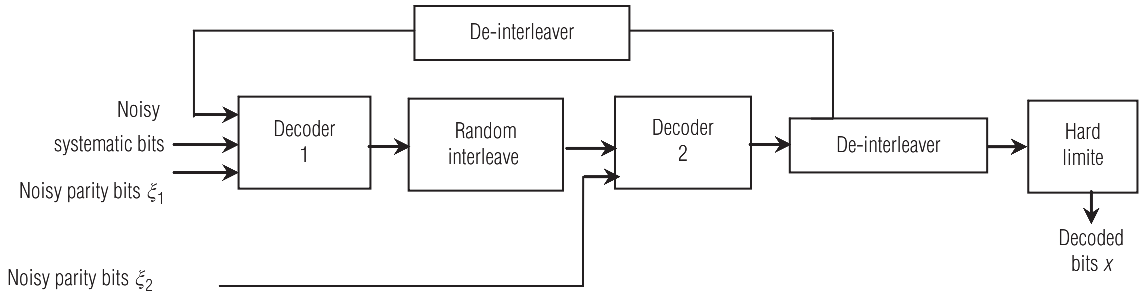


Figure 14.16 Block diagram of turbo decoder

Each decoding process yields a soft output decision for the next decoder. The key component is the soft-input, soft-output (SISO) decoder. To achieve the benefits of turbo code, several iterations are provided. Thus, the turbo codes are processing intensive and are applied to less delay sensitive applications such as data. They have been implemented in both software and hardware as a single integrated circuit, and theydo not suffer from the error at low BERs that have been attributed to other codes. Turbo codes are capable of providing high coding gain, even with high code rates.

14.5.5 Comparison between convolution and turbo codes

A turbo code is applicable where digital data is transmitted over a noisy channel because it spreads error uniformly over the interleaved duration. Owing to their strength in non-linear coding/decoding and feedback property, turbo codes are better than convolutional codes. Turbo codes can support the data rates achieved with other error correction codes while offering improved correction capability.

Figure 14.17 provides a comparison of a turbo code having constraint length $K = 4$ decoded with 4 iterations with a convolutional code of constraint length $K = 9$ decoded with a Viterbi decoder. Figure 14.17 shows the performance of a rate 1/3 turbo code compared to the corresponding rate 1/3 convolutional code. The results indicate that the turbo code outperforms the corresponding convolutional code with the same decoding complexity for data rates larger than 9.6 Kbps; the improvement in performance increases with the data rate for a fixed frame length of 20 ms.

The performance improvement is due to the fact that the number of bits in the 20 ms frames increases with the data rate and the performance of a turbo code improves with the number of bits in the frame. With a large number of bits in a frame, the interleaver separating the codes can randomize the errors more effectively.

- A turbo code is closer to the random code because it uses a pseudorandom interleaver to separate its own two convolutional encoders.
- In higher code rates or low SNR value conditions, turbo codes exhibit better performance than traditional convolutional codes.
- Convolutional codes and turbo codes perform better with soft decisions. However, convolutional codes can also work with hard decisions.
- Convolutional codes do not have an error floor, whereas turbo codes do have an error floor. It means that in turbo codes, the BER drops very quickly in the beginning, but eventually settles down and decreases at a much slower rate.

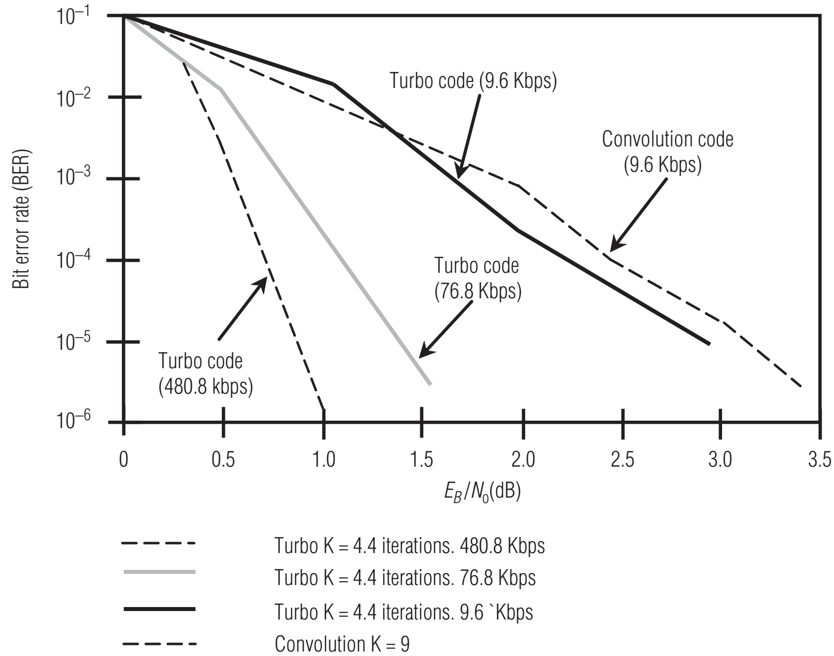


Figure 14.17 Performance of turbo code compared with convolutional code

- Both convolutional and turbo codes require the use of flush bits to initialize them to state 0 at the end of the incoming source information bits sequence. However, due to parallel encoding structure of turbo codes, it is not straightforward to flush the second encoder.
- Turbo codes are decodable, and hence they are of more practical importance.

Turbo codes are inherently block codes with the block size determined by the size of the turbo interleaver. These codes are used in 3G cellular technology for high-speed data rate applications.

14.6 Summary

- In this chapter, different signal processing techniques such as diversity, equalization, speech, and channel coding are introduced. They have been successfully used in communication systems to improve the quality of communications.
- Diversity is a commonly used technique in mobile radio systems to combat signal fading to improve the SNR of the system. The basic principle of diversity, different diversity schemes, and roles of different diversity signal combining techniques are discussed.
- Equalization is used to overcome ISI due to channel time dispersion.
- Equalization techniques are widely used to improve wireless link performance and received signal quality.
- Equalization techniques which can combat and/or exploit the frequency selectivity of the wireless channel are of enormous importance in the design of high data rate wireless systems.

- The efficient utilization of the allocated spectrum is the prime objective in the design of digital wireless communication systems. The systems rely on the use of speech coding to remove almost all the natural redundancy inherent in an analogue speech signal, while ensuring a high quality reproduction of the original speech signal at the receiver.
- Speech coding is the process for reducing the bit rate of digital speech representation for transmission or storage, while maintaining a speech quality that is acceptable for the application. Speech coding methods can be classified as waveform coding, source coding, and hybrid coding.
- Channel coding is a common strategy to make digital transmission more reliable, or, equivalently, to achieve the same required reliability for a given data rate at a lower power level at the receiver.
- There are two different approaches for error control coding: ARQ and FEC. ARQ is detection-only type coding in which transmission errors can only be detected by the receiver but not corrected. FEC allows not only detection of errors at the receiving end but correction of errors as well.

Review questions

1. What is the basic principle of diversity? Explain different diversity schemes.
2. What is diversity? How is it provided in a communication system?
3. Among the selection, equal gain, and maximal ratio combining, which scheme is the best and why?
4. What is meant by equalization? Explain adaptive linear equalization.
5. Explain the LMS algorithm.
6. List the main attributes of a speech coding.

Objective type questions and answers

1. The relation between the coherence bandwidth (B_c) and delay spread (T_d) is_______
   (a) $B_c \approx 1/T_d$ (b) $B_c \approx T_d$ (c) $B_c \approx 2T_d$ (d) $B_c \approx 4T_d$
2. The condition for a flat fading channel with respect to coherence bandwidth (B_c) and system bandwidth (B_w) is_______
   (a) $B_c > B_w$ (b) $B_c = B_w$ (c) $B_c < B_w$ (d) $B_c = 1/B_w$
3. The condition for a frequency selective channel with respect to coherence bandwidth (B_c) and system bandwidth (B_w) is_______
   (a) $B_c > B_w$ (b) $B_c = B_w$ (c) $B_c < B_w$ (d) $B_c = 1/B_w$
4. In a speech codec, the delay introduced by frame size, look ahead, and multiplexing is called the _______
   (a) processing delay (b) transmission delay
   (c) algorithmic delay (d) none
5. The advantage of adaptive multi rate codec (AMR) over enhanced full rate (EFR) is _______
   (a) greater spectral efficiency (b) better voice quality
   (c) operates under much worse conditions (d) all of the above
6. The waveform codecs use _______ approach in coding the speech signal.
   (a) time domain (b) frequency domain
   (c) either time or frequency domain (d) none

7. The channel coding approach which detects and corrects the transmission errors is _______
   (a) automatic repeat request (ARQ)
   (b) forward error correction (FEC)
   (c) Both (a) and (b)
   (d) none
8. Which of the following is the convolutional code _______
   (a) Hamming code
   (b) Viterbi code
   (c) Reed–Solomon code
   (d) BCH code

Answers: 1. (a), 2. (a), 3. (c), 4. (c), 5. (d), 6. (a), 7. (a), 8. (c), 9. (b).

Open book questions

1. What are the different diversity signal combining techniques?
2. What are the different factors that influence the equalizer algorithm?
3. What are the different speech coding methods?

Key equations

1. The coherence bandwidth is given by

$$B_c \approx \frac{1}{T_d}$$

2. The adaptive linear operation of the filter is completely described by the recursive equation

$$w_k(n+1) \equiv w_k(n) + \mu e_k(n) x(n-k)$$

3. The NLMS algorithm update equation takes the form of

$$w_k(n+1) = w_k(n) + \frac{\beta}{\|x(n))\|^2 + \varepsilon} e_k(n) x(n-k)$$

4. Mean square error (MSE) criterion is defined by

$$\text{MSE} = E[(x(t) - \hat{x}(t))^2]$$

5. The data rate at which the block encoder produces bits is given by

$$R_0 = (n/k)R_s$$

Further reading

Atal, B. S. "Predictive Coding of Speech Signals at Low Bit Rates," *IEEE Transactions on Communications*, Vol. 30, No. 4, 1982, pp. 600–614.

Atal, B. S., and Schroeder, M. R. "Stochastic Coding of Speech at Very Low Bit Rate," *Proceedings of International Conference* in *Communications*, Amsterdam, 1984, pp. 1610–1613.

Bahl, L. R., Cocke, J., Jelinek, F., and Raviv, J. "Optimal Decoding of Linear Codes for Minimizing Symbol Error Rate," *IEEE Transactions on Information Theory*, Vol. IT-20, No. 2.

Chen, J., Cox, R., Lin, Y., Jayant, N., and Melchner, M. "Coder for the CCITT 16 kbps Speech Coder Standard," *IEEE Journal of Selected Areas of Communications,* Vol. 6, 1988, pp. 353–363.

Furuskar, A., et al. "System Performance of EDGE: A Proposal for Enhanced Data Rates in Existing Digital Cellular System," *IEEE VTC 98*, pp. 1284–1289.

Hess, W. *Pitch Determination of Speech Signals*. Berlin: Springer Verlag, 1983.

Jarvinen, K., et al. "GSM Enhanced Full Rate Speech Codec," *IEEE GLOBECOM '97*.

Kleijn, W. B., Ramachandran, R. P., and Kroon, P. "Generalized Analysis by Synthesis Coding and Its Application to Pitch Prediction," *International Conference on Acoustics. Speech Signal Processing*, San Francisco, 1992, pp. 1337–1340.

Kroon, P., and Deprettere, E. F. "A Class of Analysis by Synthesis Prediction Coders for High Quality Speech Coding at Rates Between 4.8 and 16 kbps," *IEEE Journal of Selected Areas of Communications,* Vol. 6, 1988, pp. 353–363.

Spread Spectrum Techniques 15

15.1 Introduction

For wireless communication, radio spectrum is a limited resource. Radio spectrum usage must be carefully controlled. Cellular mobile communication uses various techniques to allow multiple users to access the same radio spectrum at the same time. Spread spectrum multiple access (SSMA) is another method that allows multiple users to access the mobile radio communications network. Spreading means increasing the bandwidth of the signal. Each user is allowed to use all the bandwidth, like time division multiple access (TDMA), and for the complete duration of the call, like frequency division multiple access (FDMA). Spread spectrum has been adopted as the air interface standard from 2G standard (IS95) onwards to 3G mobile system (IMT2000). The term spread spectrum today is one of the most popular in the radio engineering and communication community. Code division multiple access (CDMA) is a second-generation digital cellular scheme which is currently proving to be a successful multiple access method for cellular radio. We will describe its salient points, and leave the reader to pursue frequency-hopping (FH)/SSMA. CDMA though is conceptually more complex than FDMA and TDMA, it is not necessarily more difficult to implement because of the advances in microelectronics.

In this chapter, some basic concepts of spread spectrum techniques are discussed. These concepts are CDMA specific, and often not used in other technologies, so some explanation may be necessary. An understanding of these concepts will make reading this book much easier.

An important use of the spread spectrum concept in wireless communication systems is to allow multiple users occupy the same transmission band for simultaneous transmission of signals without considerable interference.

15.2 Spread spectrum technology

The spread spectrum technology was developed initially for military and intelligence applications. It became useful by spreading the information signal over a wide bandwidth to make jamming and interception more difficult. Its definition is as follows.

Spread spectrum technology is a wireless communication technique in which the user's original signal is transformed into another form that occupies a larger bandwidth than the original signal would normally need. The process of transformation is known as spreading.

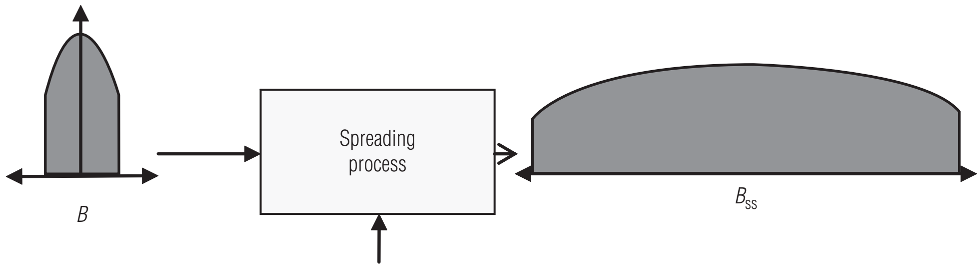


Figure 15.1 Spread spectrum process

The bandwidth spread is accomplished by means of a code which is independent of the data, and synchronized reception with the code at the receiver is used for *despreading* and subsequent data recovery. Spread-spectrum signals are highly resistant to narrowband interference. This technique decreases the potential interference to other receivers while achieving privacy. Spread spectrum concept is shown in Figure 15.1. A signal that occupies a bandwidth of B is spread out to occupy a bandwidth of B_{ss}. All signals are spread to occupy the same bandwidth B_{ss}. The input to the spreading process shown is bandwidth of an information bit (B). The input information bits are modulated by the spreading code and the resultant signal bandwidth is shown at the model output.

A signal with a narrow frequency band is now spread to a wider frequency band.

The spreading code is generated by a pseudo noise or pseudorandom noise (PRN) generator. Usually, the *pulse* in the *spreading code* is called *chip* and the *pulse in the data sequence* are called *bits (or symbols)*. The bits in the data sequence carry the information and the chips (such as a PRN code sequence) used in CDMA do not carry any information.

For example, if the original binary data sequence $D(t)$ is multiplied with a *spreading code* $C(t)$, *then the resultant spread code* $D(t)C(t)$ will have a much larger bandwidth than the original signal. This procedure is illustrated in the Figure 15.2.

In order to recover the transmitted data sequence at the destination mobile phone side, the same spreading code, as the one used for spreading at the transmission point, is used to perform correlation detection and this process is called *despreading*. Each user has its own spreading code. The identical code is used in both transformations on each end of the radio channel, spreading the original signal to produce a wideband signal, and despreading the wideband signal back to the original narrowband signal as shown in Figure 15.3.

Spreading Factor (SF): The rate at which the spread data (or PRN code) varies (chips per second) is called *chip rate*. The chip rate is always larger than the original data rate (or symbol rate), meaning that one symbol is represented by multiple chips. The ratio is known as the SF or processing gain (G_p).

SF = Chip rate/Symbol rate

or

= Transmission signal bandwidth/Original signal bandwidth.

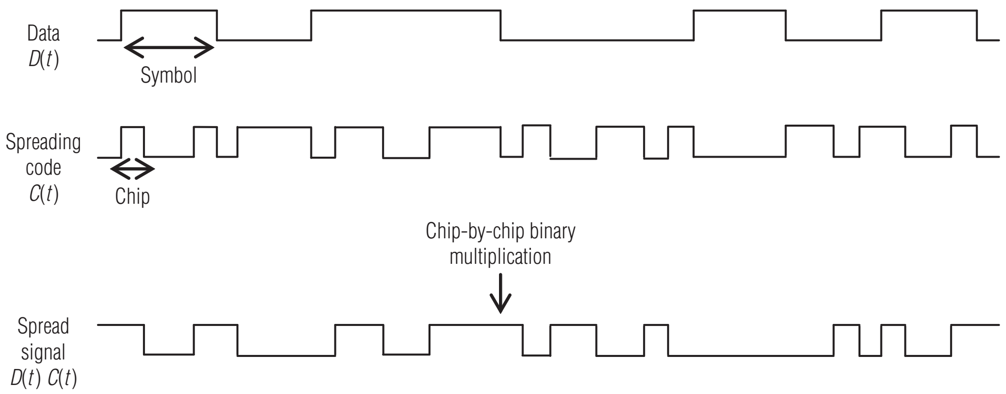


Figure 15.2 Spreading of data sequence

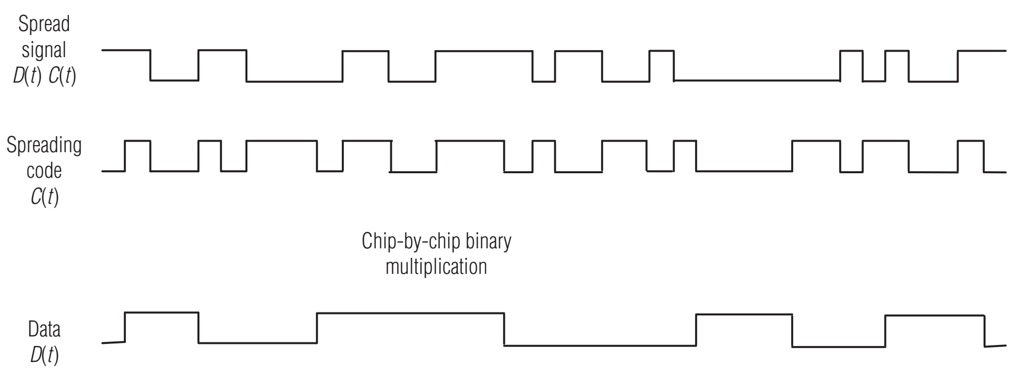


Figure 15.3 Despreading of data sequence

SF represents the number of chips used to spread one data symbol. The lower the SF, the more payload data a signal can convey on the radio interface.

As signals received by other users carry different spreading codes, the signal power is reduced eventually to 1/SF. Another 2G system based on CDMA (TIA/EIA/IS-95) is a direct sequence DSSS system in which the entire bandwidth of the carrier channel is made available to each user simultaneously in direct sequence code division multiple access (DS-CDMA). Here all users share the same frequency band and time-frame to communicate, and each user is identified by a spreading code uniquely assigned to the user.

For example, CDMA as defined in IS-95:

$$R_c = 1.2288 \text{ Mcps},$$
$$R_b = 9.6 \text{ kbps (max)},$$

resulting in process gain = 128 or 21.07 dB,where R_c is chip rate and R_b the bit rate.

15.2.1 Spreading codes properties

Spreading codes have low cross-correlation with other spreading codes. In the case of fully synchronized orthogonal codes, the cross-correlation is actually zero. This implies that several wideband signals can coexist on the same frequency without severe mutual interference. The energy of a wideband signal is spread over so large a bandwidth that it is just like background noise compared with the original signal. That is, its power spectral density is small.

When the combined wideband signal is correlated with the particular spreading code, only the original signal with the corresponding spreading code is despread, while all the other component original signals remain spread (Figure 15.4(a)). Thus, the original signal can be recovered in the receiver as long as the power of the despread signal is a few decibels higher than the interfering noise power. That is, the *carrier-to-interference ratio* (C/I) has to be large enough. Note that the power density of a spread signal can be much lower than the power density of the composite wideband signal, and the recovery of the original signal is still possible if the spreading factor is high enough. But if there are too many users in the cell generating too much interference, then the signal may get blocked and the communication becomes impossible, as depicted in Figure 15.4(b).

Note that a wideband carrier does not increase the capacity of the allocated bandwidth as such. In principle, a set of narrowband carriers occupying the same bandwidth would be able to convey as much data as the wideband signal. However, in a wideband system, the signals are more resistant to inter-cell interference, and thus it is possible to reuse the same frequency in adjacent cells. This means that the frequency reuse factor is one, while in typical GSM systems the value is at least four. That is, the same frequency can be reused at every fourth cell at most. This fact alone provides a substantial capacity gain over narrowband systems, although the capacity increase is not simply directly proportional to the reuse factor.

15.2.2 Applications of spread spectrum technology

The following are main applications of spread spectrum technology

- Anti-jamming
- Interference rejection

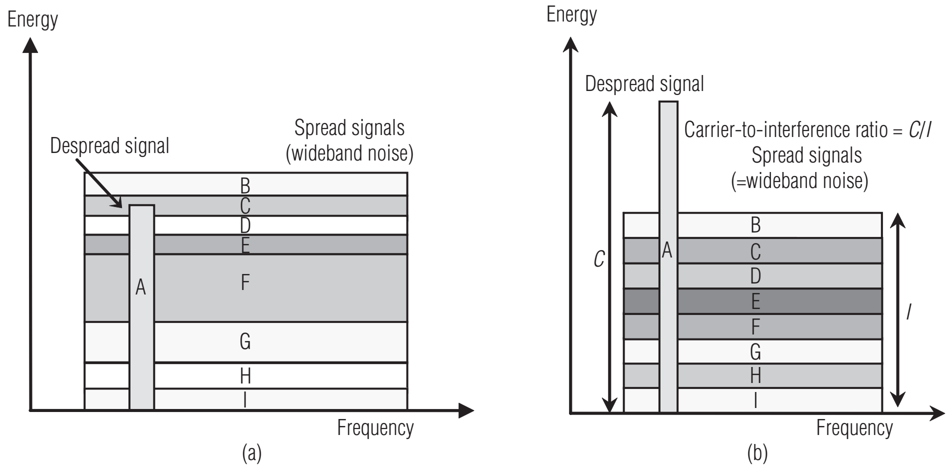


Figure 15.4 (a) Unrecoverable signal; (b) Recovery of despread signal

- Low probability of intercept
- Multiple access
- Multipath reception
- Diversity reception
- High resolution ranging
- Accurate universal timing

15.2.3 Advantages and disadvantages of spread spectrum technique

Using the Spread spectrum technology, several users can independently use the same bandwidth with very little interference. This property is used in cellular telephony applications, with a technique known as CDMA.

Advantages of spread spectrum

The advantages of spread spectrum are as follows:

1. Spread spectrum *improves the channel capacity* (i.e., the number of signals that can be transmitted at the same time over a given frequency band) as compared to the narrowband spectrum.
2. Spread spectrum *offers high resistance against narrowband interference* and tolerates narrowband interference because the signal is transmitted over a wide band.
   - Security against *tapping and jamming* is superior compared to narrowband spectrum techniques.
3. Signal gains *immunity from various kinds of noise and multipath distortion*. The earliest applications of spread spectrum were Global Positioning System (GPS), where it was used for its immunity to jamming.
4. It can also be used for *hiding and encrypting signals*. Signals of spread spectrum are indistinguishable from background noise to anyone who does not know the coding scheme.

Disadvantages of spread spectrum

The following are the disadvantages of spread spectrum:

1. Relatively high complexity of the coding mechanism is used in spread spectrum, which results in complex radio hardware designs and higher cost.
2. Bandwidth inefficiency.

15.3 Spread spectrum techniques

Depending on the way the frequency spectrum is used, four types of spread spectrum techniques are currently in use:

- Direct sequence spread spectrum (DSSS)
- Frequency-hopping spread spectrum (FHSS)
- Time-hopping spread spectrum (THSS)
- Multi-carrier spread spectrum (MCSS)

The primary spread spectrum techniques used in cellular system and GPS for multiple access are DSSS and FHSS. The DSSS is known as direct sequence code division multiple access (DS-CDMA; shown in Figure 15.5), and the FHSS is known as frequency-hopping code division multiple access

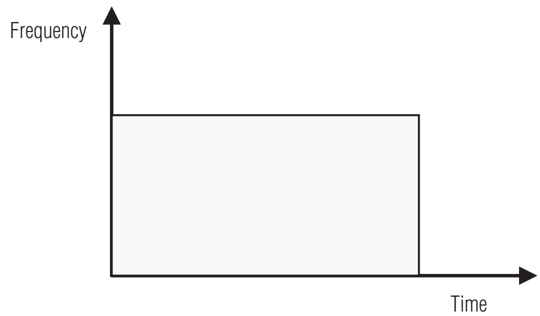


Figure 15.5 DS-CDMA principle

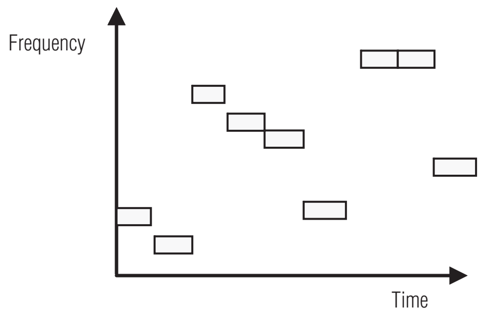


Figure 15.6 FH-CDMA principle

(FH-CDMA; shown in Figure 15.6). Example of cellular systems using MC-CDMA is CDMA2000 in the downlink and DS-CDMA in the uplink direction.

DSSS – The information signal is directly multiplied by a higher data rate spreading code. The resulting signal then modulates the digital wideband carrier. The chip rate of the spreading code must be much higher than the bit rate of the information signal. The spreading code is pseudogenerated randomly.

FHSS **–** The signal is rapidly switched between different frequencies according to a pseudorandom sequence. The pseudorandom sequence is a list of frequencies and the carrier hops through this list of frequencies.

THSS – Spread spectrum is achieved by carrying out the on–off keying in the time domain. The signal is transmitted in short bursts pseudorandomly, and the receiver knows beforehand when to expect the burst. The TH-CDMA principle is shown in Figure 15.7.

MCSS **–** As ***DSSS*** spreads the original data stream in the time (T)-domain with the aid of spreading sequences, MCSS spreads the signal in frequency domain. Each data symbol is transmitted simultaneously over *N* relatively narrowband subcarriers. Each subcarrier is encoded with a constant phase offset. Multiple access is achieved with different users transmitting at the same set of subcarriers, but with spreading codes that are orthogonal to the codes of the other users. The process is shown in Figure 15.8.

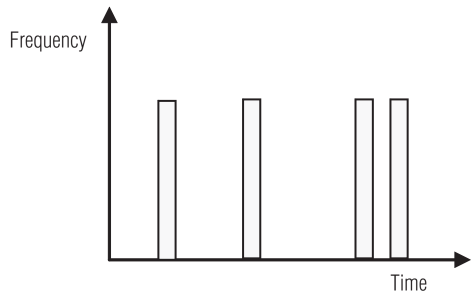


Figure 15.7 TH-CDMA principle

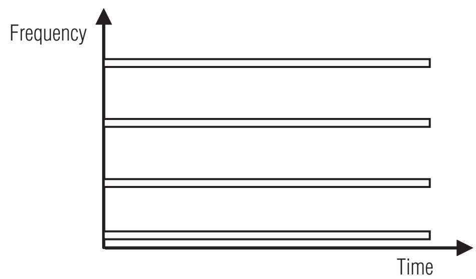


Figure 15.8 MC-CDMA principle

15.4 Direct sequence spread spectrum communications

The spreading modulation may be phase modulation or a rapid change of the carrier frequency, or it may be a combination of these two schemes. When spectrum spreading is performed by phase modulation, we call the resultant signal DSSS signal. In a DSSS system, the bandwidth of the baseband information carrying signals from a different user is spread by different codes with a bandwidth much larger than that of the baseband signals. The spreading codes used for different users are orthogonal or nearly orthogonal to each other. In the DSSS, the spectrum of the transmitted signal is much wider than the spectrum associated with the information rate. At the receiver, the same code is used for despreading to recover the baseband signal from the target user while suppressing the transmissions from all other users.

The principle behind the DSSS is that the information signal with bandwidth B is spread over a bandwidth B_s, where $B_s >> B$ (Figure 15.9). The processing gain (G_p) is defined as

$$G_p = \frac{B_s}{B}$$

The higher the processing gain, the lower the power density one needs to transmit the information. If the bandwidth is very large, the signal can be transmitted such that it appears like a noise. Typical processing gains of a DSSS system lie between 10 and 60 dB. With a DSSS system, the total noise level is determined both by thermal noise and by interference.

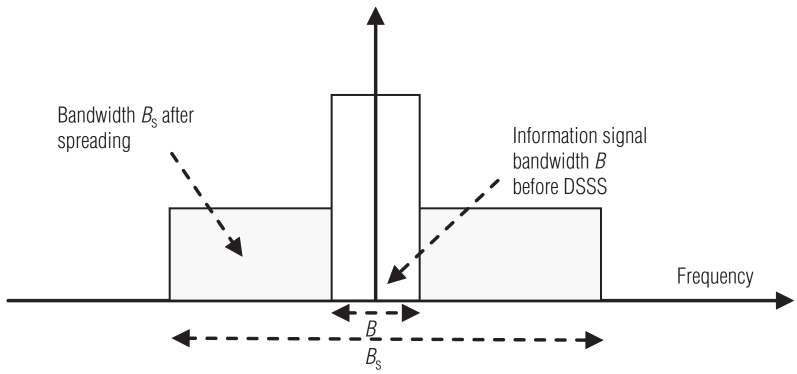


Figure 15.9 Principle of spread spectrum techniques

For a given user, the interference is processed as noise. The input and output *S*/ *N* ratios are related as

$$\left(\frac{S}{N}\right)_0 = G_p\left(\frac{S}{N}\right)_i \tag{15.1}$$

We express input $\left(\frac{S}{N}\right)_i$ ratio as

$$\left(\frac{S}{N}\right)_i = \left(\frac{E_b}{N_0}\times\frac{R_b}{B_w}\right) = \left(\frac{E_b}{N_0}\right)\times\frac{1}{G_P} \tag{15.2}$$

where E_b is the bit energy

N_0 is the noise power spectral density

Using Equation (15.2), we rewrite Equation (15.1) as

$$\left(\frac{S}{N}\right)_0 = G_p\left(\frac{S}{N}\right)_i = \left(\frac{E_b}{N_0}\right)_i \tag{15.3}$$

but

$$\left(\frac{S}{N}\right)_0 = \left(\frac{E_b}{N_0}\times\frac{R_b}{B_w}\right)_0 = \left(\frac{E_b}{N_0}\right)_0\times\frac{1}{G_p} \tag{15.4}$$

Therefore, using Equations (15.3) and (15.4) we get

$$\left(\frac{E_b}{N_0}\right)_0 = G_p\left(\frac{E_b}{N_0}\right)_i \tag{15.5}$$

15.4.1 Principle of direct sequence spread spectrum

For signal spreading, PRN codes with good cross- and auto-correlation properties are used. A PRN code is made up from a number of *chips* for mixing the data with the code (Figure 15.10 and 15.11). In order to recover the received signal, the code, which the signal was spread within the transmitter, is reproduced in the receiver and mixed with the spread signal. If the incoming signal and the locally generated PRN code are synchronized, the original signal after correlation can be recovered. In a multiuser environment, the user signals are distinguished by different PRN codes and the receiver needs only the knowledge of the user's PRN code and has to synchronize with it. This principle of user separation is referred to as DS-CDMA. The longer the PRN code, the more the noise-like signals appear. The drawback of the principle is that synchronization becomes more difficult unless synchronization information, such as pilot signals, is sent to aid acquisition.

One basic design problem with DS-CDMA is that, when multiple users access the same spectrum, it is possible that a single user could mask all other users at the receiver side if its power level is too high.

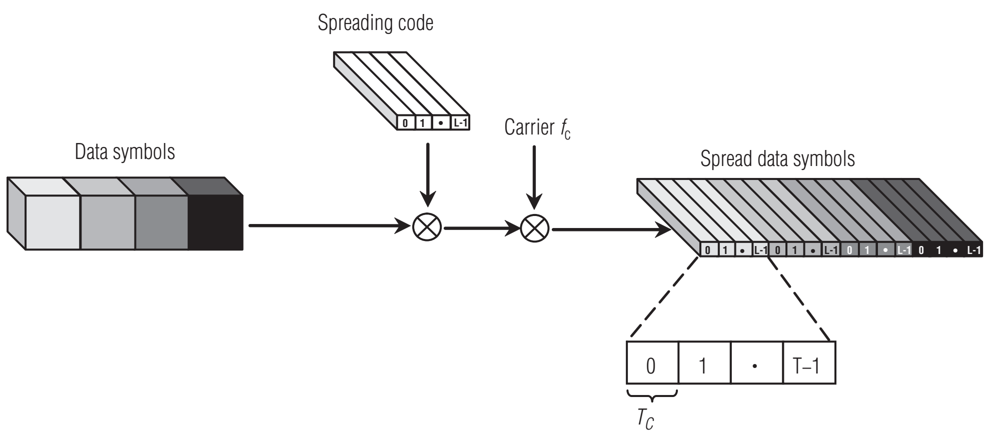


Figure 15.10 Principle of DS-CDMA

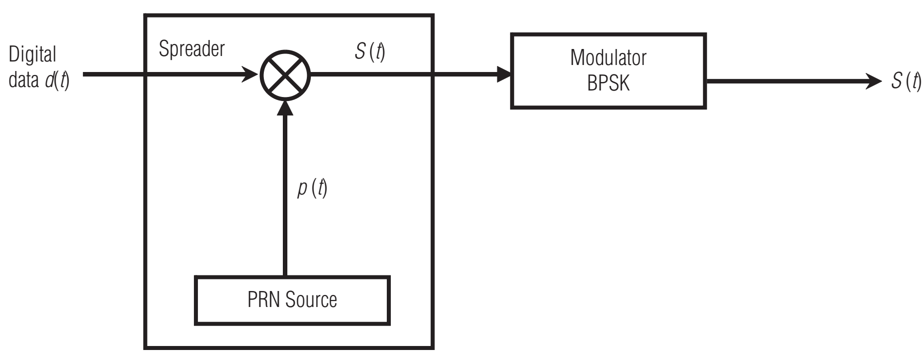


Figure 15.11 DSSS transmitter

15.4.2 DSSS with binary phase shift keying

We show how signal spreading and despreading is achieved with binary phase shift keying (BPSK). A DSSS signal is a spread-spectrum signal generated by the direct mixing of the data with a spreading waveform before the final carrier modulation. Ideally, a direct-sequence signal with BPSK or differential PSK (DPSK) data modulation can be represented by

$$S(t) = Ad(t)p(t)\cos(2\pi f_c t + \theta) \tag{15.6}$$

where

- A is the signal amplitude
- $d(t)$ is the data modulation
- $p(t)$ is the spreading waveform
- f_c is the carrier frequency
- θ is the phase at $t = 0$

The data modulation, shown in Figure 15.12(a), is a sequence of non-overlapping rectangular pulses of duration T_S each of which has an amplitude d_i = +1 if the associated data symbol is a 1, and d_i = –1 if it is a 0. The spreading code, $p(t)$, is shown in Figure 15.12(b). Spread spectrum signal, $S(t)$, is as shown in Figure 15.13.

The spreading wave form is given by

$$p(t) = p_i \psi(t - iT_c) \tag{15.7}$$

where each p_i equals +1 or –1 and represents one chip of the spreading sequence. The chip waveform $\omega(t)$ is ideally confined to the interval [0, T_c] to prevent inter-chip interference in the receiver. A rectangular chip waveform has $\psi(t) = w(t, T_c)$,

where
$$w(t, T_c) = \begin{cases} 1, & 0 \le t \le T_c \\ 0, & \text{otherwise} \end{cases} \tag{15.8}$$

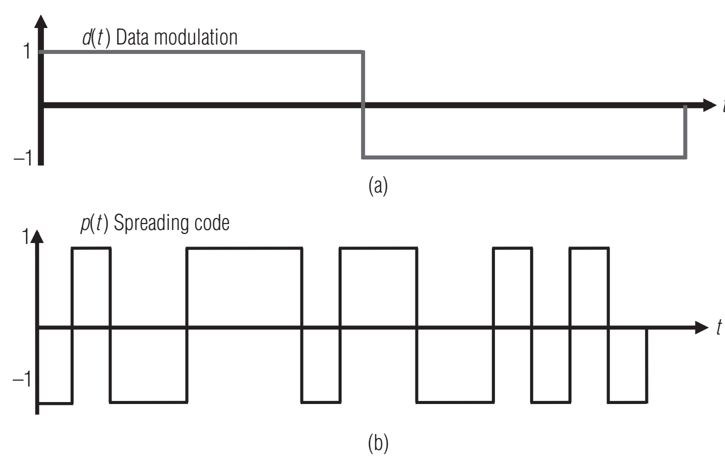


Figure 15.12 (a) Data modulation; (b) Spreading code (PRN)

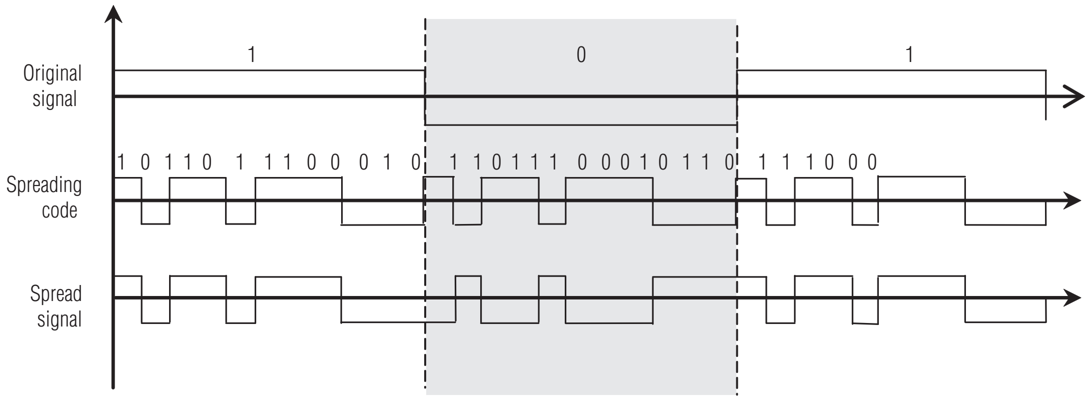


Figure 15.13 An example of data and rectangular chip waveform (spreading code)

If W_b is the bandwidth of $p(t)$ and B_d is the bandwidth of $d(t)$, the spreading due to $p(t)$ ensures that it has a bandwidth $W_b >> B_d$.

If T_b is the bit period of $s_i(t)$, then T_b may correspond to either a full period for $c_i(t)$, or to a fraction of a period. If T_b is less than one code period, then the data bits are modulating the polarity of a portion of a code period. The code $c_i(t)$ serves as a subcarrier for the source data. Since each mobile station (MS) uses the entire channel bandwidth and since the equation

$$m_i(t) = x_i(t)c_i(t) \tag{15.9}$$

has a code chip rate of $1/T_c$ chips per second, each BPSK carrier uses an RF bandwidth of $B_w = 1/T_c$. The available channel RF bandwidth determines the minimum chip width, and the code period determines its relation to the bit time. The number of code chips per bit is given by B_w/R_b which is the CDMA processing gain (G_p) or simply the spreading ratio of code modulation. This shows how much the RF bandwidth must be spread relative to the bit rate, R_b, to accommodate a given spreading code length. Each MS uses the same RF carrier frequency and RF bandwidth, but with its own spreading code $c_i(t)$.

The input signal equation

$$S_i(t) = x_i(t)c_i(t)\sqrt{2P}\cos(\omega_c t + \phi) \tag{15.10}$$

is transmitted using a distortion-less path with transmission. The signal is received together with some type of interference and/or Gaussian noise. Demodulation is performed in part by remodulating with the spreading code appropriately delayed as shown in Figure 15.14. This correlation of the received signal with the delayed spreading waveform is the despreading. This is a critical function in all spread spectrum systems. The signal component of the output of the despreading is

$$x_\mathrm{i}(t-\tau_\mathrm{d})\sqrt{2P}c_\mathrm{i}(t-\tau_\mathrm{d}) \times c_\mathrm{i}(t-\hat{\tau}_\mathrm{d})\cos(\omega_\mathrm{c}(t-\tau_\mathrm{d})+\phi) \tag{15.11}$$

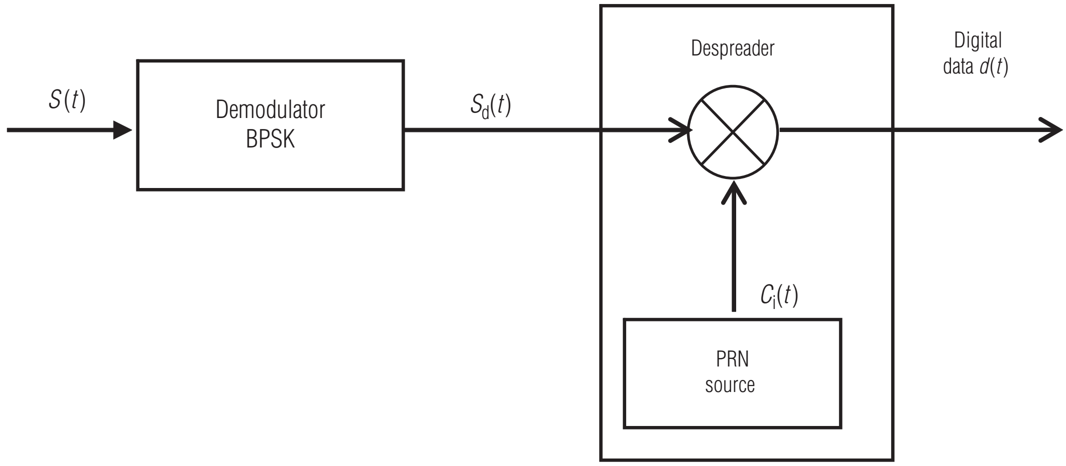


Figure 15.14 Block diagram of DSSS with BPSK

where

P is the carrier transmitted power

$\hat{\tau}_d$ is the receiver's best estimate of the transmission delay

Since $c_i(t) = \pm$, the product $c_i\,(t - \tau_d) \times c_i(t - \hat{\tau}_d)$ will be unity if $\tau_d = \hat{\tau}_d$.

That is, the code at the receiver is synchronized with the spreading code at the transmitter. When correctly synchronized, the signal component of the output of the receiver despreading is equal to

$$x_{\text{i}}(t-\tau_{\text{d}})\sqrt{2P}\cos(\omega_{\text{c}}(t-\tau_{\text{d}})+\phi) \tag{15.12}$$

which can be demodulated using a conventional coherent phase modulator. The bit error probability, P_e, associated with the coherent BPSK spread spectrum signal is the same as with the BPSK signal and is given as

$$P_e = \frac{1}{2}\text{erfc}\left[\sqrt{\left(\frac{E_b}{N_0}\right)}\right] = Q\left[\sqrt{2\left(\frac{E_b}{N_0}\right)_0}\right] = Q\left[\sqrt{2\left\{G_p\left(\frac{E_b}{N_0}\right)_i\right\}}\right] \tag{15.13}$$

The DSSS system with BPSK transmitter and receiver is shown in Figure 15.15.

$$x(t) = \sqrt{\frac{P}{2}C_I}(t-\tau_d)C_I(t-\hat{\tau}_d)\cos[\omega_c t+\phi] + \left(\sqrt{\frac{P}{2}}\right)C_Q(t-\tau_d)C_I(t-\hat{\tau}_d)\sin(\omega_c t+\phi) \tag{15.14}$$

$$y(t) = \sqrt{\frac{P}{2}}C_I(t-\tau_d)C_Q(t-\hat{\tau}_d)\sin[\omega_c t+\phi] + \left(\sqrt{\frac{P}{2}}\right)C_Q(t-\tau_d)C_Q(t-\hat{\tau}_d)\cos(\omega_c t+\phi) \tag{15.15}$$

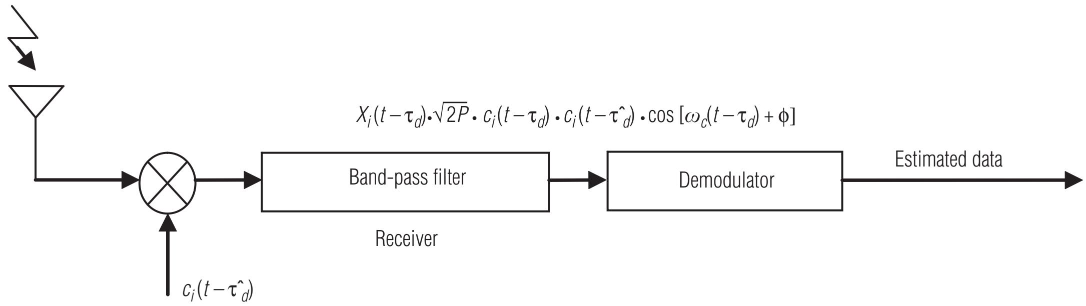


Figure 15.15 DSSS system with BPSK

If the receiver-generated replicas of spreading codes are correctly phased then

$$C_I(t-\tau_d)C_I(t-\hat{\tau}_d) = C_Q(t-\tau_d)C_Q(t-\hat{\tau}_d) = 1 \tag{15.16}$$

$$z(t) = x(t) + y(t) = \sqrt{2P}\cos[\omega_c t + \phi] \tag{15.17}$$

where $c_I(t)$ and $c_Q(t)$ are the in-phase and quadrature spreading codes, respectively. This condition is satisfied since $c_I(t)$ and $c_Q(t)$ are independent code waveforms. The receiver for the transmitted signal is shown in Figure 15.15. The band pass filter is centred at frequency ω_c and has a bandwidth sufficiently wide to pass the data-modulated carrier without distortion.

The signal $z(t)$ is the input to a conventional phase demodulator where data is recovered. When the spreading codes are staggered, one-half chip interval with respect to each other, the QPSK, is called offset-QPSK (OQPSK). In OQPSK, the phase changes every one-half chip interval, but it does not change more than ±90°. This limited phase change improves the uniformity of the signal envelope compared with BPSK and QPSK, since zero-crossings of the carrier envelope are avoided. Neither QPSK nor OQPSK modulation can be removed with a single stage of square-law detection. Two such detectors and the associated loss of signal-to-noise ratio are required. QPSK and OQPSK offer some low probability of detection advantages over the BPSK method.

15.4.3 Properties of direct sequence spread spectrum

In direct sequence modulation, the carrier frequency is fixed and the bandwidth of the transmitted signal is larger and independent of the bandwidth of the information signal.

Resistance to interference and anti-jamming effects

This characteristic is the real beauty of spread spectrum. Intentional or un-intentional interference and jamming signals are rejected because they do not contain the spread spectrum key. Only the desired signal, which has the key, will be seen at the receiver when the despreading operation is exercised. The demonstration of spread spectrum code used in a communication channel is shown in Figure 15.16(a). Figure 15.16(b) illustrates how the DSSS is immune to certain amount of noise, and Figure 15.16(c) illustrates how the DSSS is immune to interference.

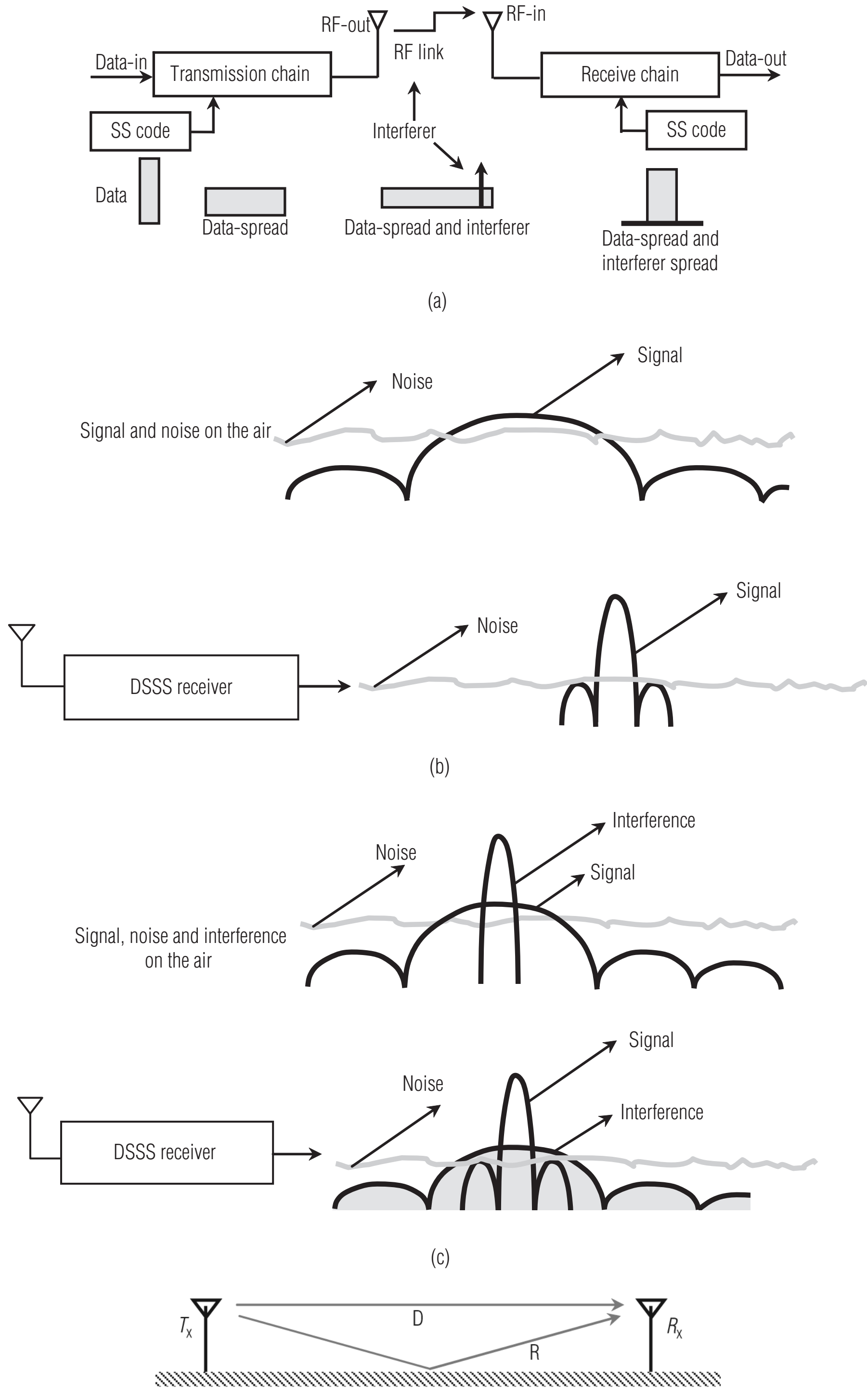


Figure 15.16 (a) Demonstration of SS code used in a communication channel; (b) DSSS is immune to certain amount of noise; (c) DSSS is immune to certain amount of interference; (d) Typical multipath scenario

Resistance to multipath effects

Wireless channels often include multiple-path propagation in which the signal has more than one path from the transmitter to the receiver. Such multipath can be caused by atmospheric reflection or refraction and by reflection from the ground or from objects such as buildings.

The reflected path (*R*) can interfere with the direct path (*D*) in a phenomenon called fading. Because the despreading process synchronizes to signal *D*, signal *R* is rejected even though it contains the same key. Methods are available to use the reflected-path signals by despreading them and adding the extracted results to the main one.

Jamming

Resistance to interception is the second advantage provided by spread spectrum techniques. Because non-authorized listeners do not have the key used to spread the original signal, they cannot decode it. Without the right key, the spread spectrum signal appears as noise or as an interferer (scanning methods can break the code, however, if the key is short). Even better, signal levels can be below the noise floor, because the spreading operation reduces the spectral density (total energy is the same, but it is widely spread in frequency). The message is thus made invisible, an effect that is particularly strong with the DSSS technique. Other receivers cannot "see" the transmission; they only register a slight increase in the overall noise level.

The receiver synchronizes to the code to recover the data. The use of an independent code and synchronous reception allows multiple users to access the same frequency band at the same time.

15.5 Frequency-hopping spread spectrum

FHSS is a method of transmitting radio signals by rapidly switching a carrier among many frequency channels, using a pseudorandom sequence known to both transmitter and receiver.

> *FHSS uses a frequency-hopping sequence to spread a user signal. FHSS shifts the carrier across a number of channels with a pseudorandom sequence that is known to both the transmitter and the receiver.*

User data is first modulated to narrowband signals, then a second modulation takes place–a signal with a hopping sequence of frequency is used as the radio carrier. The resulting spread signal is then sent to the receiver. The transmitted signal can be expressed as

$$S_k(t) = \sqrt{2P}\cos\{2\pi[f_c + f_m + f_k(t)]t + \Psi_m + \alpha_k(t)\} \tag{15.18}$$

MFSK signal $s_m(t)$ modulates a carrier $\cos(2\pi f_k(t)t)$ having a frequency $f_k(t)$, which is generated by a so-called frequency synthesizer generating the FH patterns required under the control of a PRN sequence, shown in Figure 15.17.

$f_k(t)$ represents the frequency-hopping pattern of user *k*, which is derived from a sequence $(f(k)j)$ of frequencies, where we have $f_k(t) = f(k)j$ for $jT_h \le t < (j+1)T_h$ and

$$f_k(t) = f_i^k \quad \text{for } jT_h \le t < (j+1)T_h \tag{15.19}$$

T_h represents the FH dwell time quantifying the amount of time the synthesizer camps at a given frequency. In Equation 15.18, $\alpha_k(t)$ represents the phase waveform of user *k* introduced

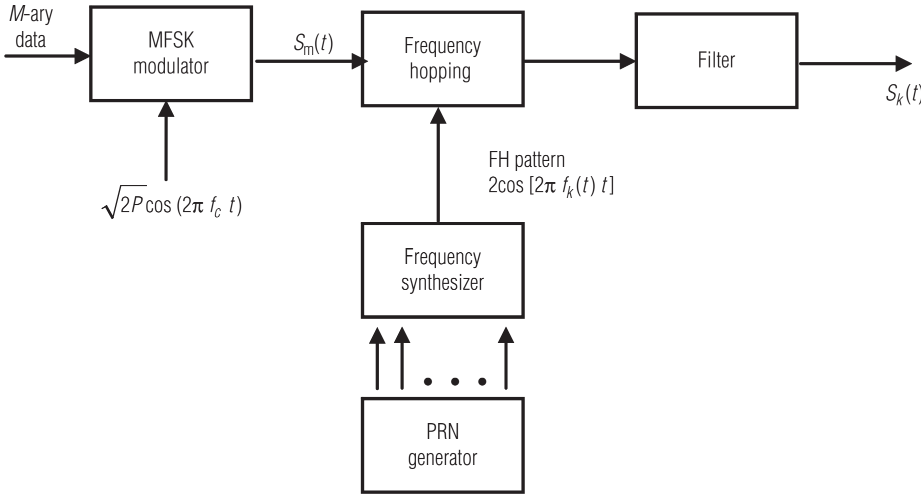


Figure 15.17 Transmitter schematic of FHSS systems using MFSK data modulation

by the k^{th} frequency hopper, which takes on the constant value of $\alpha k(j)$ during the j^{th} frequency-hopping dwell time having a duration of T_h seconds.

On the receiver side, two steps of modulations are required, (i) use the same frequency-hopping sequence to recover the narrowband signal, and (ii) demodulate the narrowband signal. In effect, the transmitter and the receiver follow the same pattern of synchronized frequency hopping. As a result, only if the hopping sequence is made known to the receiver can it recover the original user data bits; otherwise, the transmitted signals will appear as background noise. FHSS does not take up the entire allotted frequency band for transmission. Instead, at any given moment, only a portion of it is used for hopping. Two types of frequency-hopping systems are in use: (1) fast hopping systems and (2) slow hopping systems.

Fast hopping systems change frequency several times when transmitting a single bit, whereas in slow hopping systems each hop may transmit multiple bits.

The processing gain, G_p, of an FHSS system is given as

$$G_p = \frac{\text{Hopping bandwidth}}{\text{Minimum frequency spacing}} = \frac{M.\Delta f}{\Delta f} = M \qquad (15.20)$$

In the fast frequency-hop systems, there are L frequency hops during a symbol interval (T_s) (i.e., $T_s = LT_c$ or $R_c = LR_s$). Therefore,

$$\text{Hopping bandwidth} = (\text{KM}\Delta f)L \qquad (15.21)$$

where

K is the factor for frequency multiplication
$M = 2^b$ is the number of frequencies produced by frequency synthesizer
b is the bit in a symbol
L is the frequency hops per symbol

Therefore,

$$G_p = \frac{KM\Delta fL}{\Delta f} = MKL \tag{15.22}$$

The processing gain of a fast frequency-hop system is dependent on the number of frequencies used (M), the number of hops per symbol (L), and the frequency multiplication factor (K).

15.5.1 Advantages of FHSS

The advantages of FHSS are as follows:

- FHSS is resistant to narrow band interference since the spread signal causes the interfering signal to recede into the background.
- The signals are very difficult to intercept. FHSS signals make it seem like there has been an increase in background noise when a narrowband receiver detects them. In order to intercept the signal, the pseudorandom transmission hopping sequence has to be known.
- FHSS transmissions can share frequency bands with a number of other types of conventional transmissions without causing significant interference as shown in Figure 15.18. Each of these signals causes minimal interference and allows the bandwidth to be used more effectively.

15.5.2 Applications of FHSS

- Bluetooth uses FHSS technology.
- The protocol operates in the unlicensed ISM band at 2.4–2.4835 GHz.
- The Bluetooth protocol divides the band into 79 channels (each 1 MHz wide) and changes channels 1,600 times per second.

15.5.3 Properties of frequency hopping

The carrier frequency is varied, and the bandwidth of the transmitted signal is comparable to that of the information signal. Information is modulated on top of a rapidly changing carrier frequency. Some advantages and disadvantages of frequency-hopping systems are listed in Table 15.1.

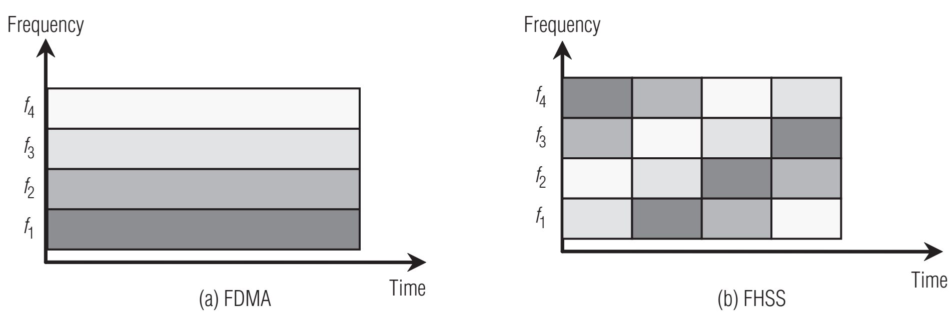


Figure 15.18 Multiple users using the same channel

Table 15.1 Advantages and disadvantages of FHSS

Advantages	Disadvantages
1. Carrier can be hopped over large portions of the spectrum	1. Complex frequency synthesizer
2. Can be programmed to avoid portions of the spectrum	2. Not useful for location and velocity measurements
3. Shorter acquisition time than direct sequence	3. Error correction required
4. Less affected by near-far problem than direct sequence	

15.6 Time-hopping spread-spectrum communications

In Section 15.5, the philosophy of FH spread-spectrum communications has been discussed, where the carrier frequency changes were controlled by a PRN sequence. Alternatively, the FH schemes concerned may also be viewed as a F-domain on–off modulation scheme, whose legitimate carrier frequencies are switched on or off, that is, activated or deactivated according to a PRN sequence. In a manner analogous to using F-domain on–off keying, time-domain on–off keying may also be employed for spread spectrum communications. The spectrum spreading is achieved by carrying out the on–off keying in the time-domain, referred to as *time-hopping* (TH). Explicitly, assume that in THSS systems, the time axis is segmented into T_f duration intervals representing the frames, and each frame is further segmented into a number of, say M, intervals referred to as time slots. Then in THSS communications, a user signal is transmitted employing a number of frames with each frame transmitting a single pulse, which is positioned in one of the M possible time slots of the frame based on the M-ary symbol transmitted. In TH systems, each spreading sequence corresponds to a specific arrangement of the time-slot locations. Hence, the spreading sequence appears in the form of time-slot locations, referred to as *TH patterns*.

Figure 5.19(a) shows THSS signals detection, where D_1 represents user 1's time frequency matrix after subtracting user 1's address of $a_1 = (4, 3, 7, 6, 2, 5)$ from the received matrix R, while D_2 represents user 2's time-frequency matrix after subtracting user 2's address of $a_2 = (2, 4, 6, 3, 1, 7)$ from the received matrix R. Observing that D_1 contains only one full row corresponding to the first user's transmitted 8-ary symbol of $X_1 = 5$, and that D_2 also contains only one row corresponding to the second user's transmitted 8-ary symbol of $X_2 = 3$.

Figure 15.19(b) shows the transmitter block diagram of a general THSS system. In the transmitter of Figure 15.19(b), the binary data are first temporarily stored in a buffer, awaiting their assignment to the appropriate time slots for transmission. As shown in Figure 15.19(b), time hopping is carried out by a gating circuit, which is switched on/off under the control of the TH pattern generated from a PRN sequence and the data symbol transmitted. When the "gate" is on, then a data symbol is passed to the modulator of Figure 15.19(b), where various baseband modulation schemes may be employed for the data modulation. Finally, as shown in Figure 15.19(b), the baseband-modulated signal modulates the carrier frequency, f_c, in order to translate the signal to the appropriate frequency band for transmission using the power P.

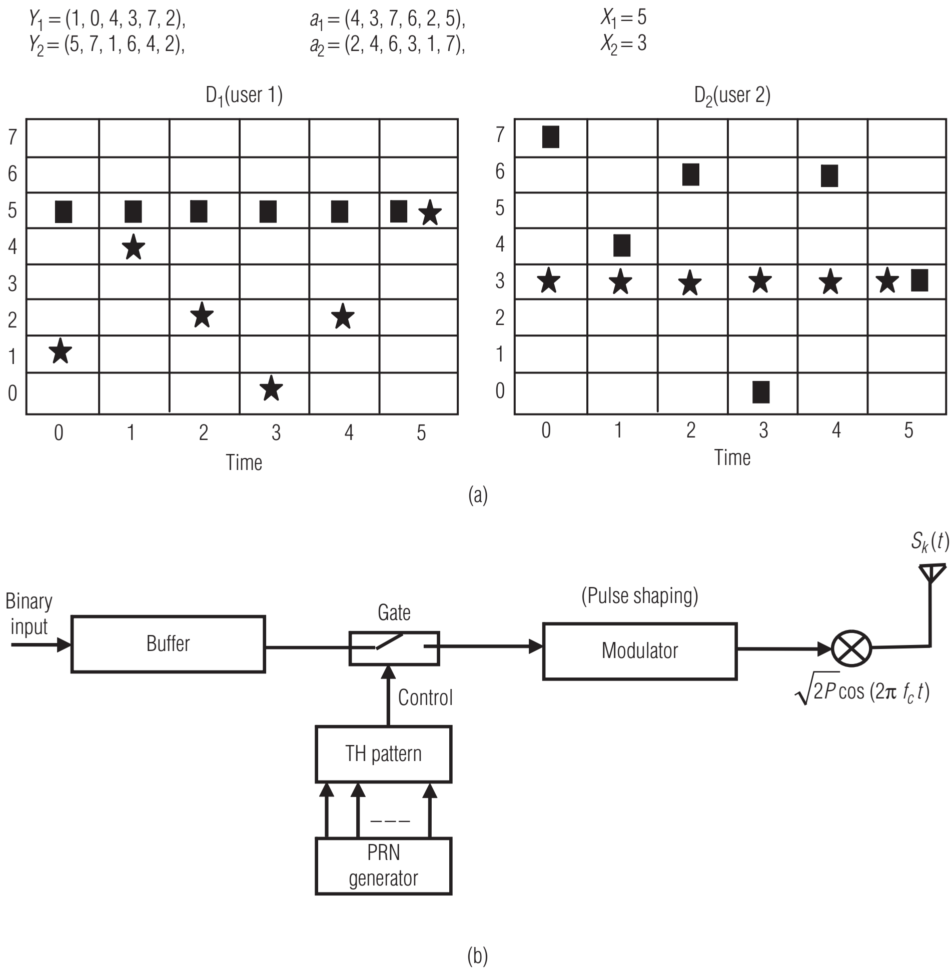


Figure 15.19 (a) Graphical representations of detections of THSS signals; (b) Transmitter schematic of a general THSS system

15.7 Multicarrier

In contrast to DSSS schemes, which spread the original data stream in the time (T)-domain with the aid of spreading sequences, MCSS schemes constitute a class of DSSS schemes that spread the original data stream in the frequency (F)-domain with the aid of a number of N subcarriers. Therefore, MCSS schemes can also be interpreted as frequency (F)-domain DSSS arrangements. However, as a consequence of using multiple subcarriers, the terminology of MCSS has established itself.

Figure 15.20 shows the transmitter block diagram of the MCSS scheme. In MCSS arrangements, the transmitter spreads the original data stream over N subcarriers using a given spreading code of $\{c_k[0], c_k[1], \ldots, c_k[N-1]\}$. Observe in Figure 15.20 that after multiplication with the corresponding chips of $n = 0, 1, \ldots, N-1$ of the N-chip spreading sequence c_k of the k^{th} user, the original data bits are mapped to N number of subcarriers. In the MCSS scheme, the data rate of each of the N subcarriers is the same as the input data rate. With reference to Figure 15.20, the k^{th} user's transmitted signal can be expressed as

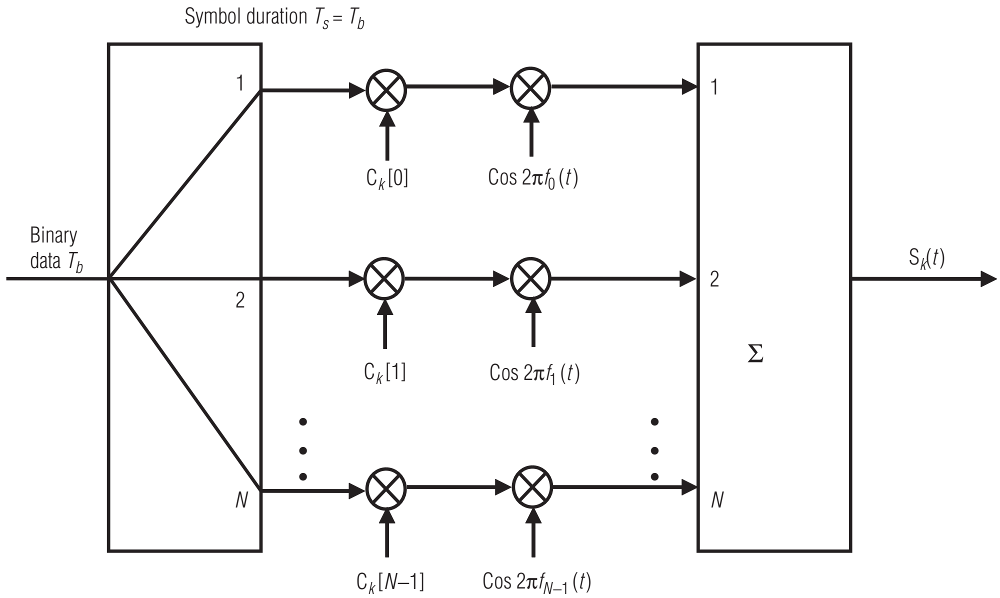


Figure 15.20 Transmitter schematic of the MCSS system

$$S_k(t) = \sqrt{\frac{2P}{N}} \sum_{n=0}^{N-1} b_k(t) c_k[n] \cos(2\pi f_n t + \psi_n^k) \tag{15.23}$$

where P represents the transmitted power of the MCSS signal, $C_k(n)$ represents the initial phase angle with respect to the k^{th} user and the n^{th} subcarrier, N is the number of subcarriers as well as the spreading gain, $\{c_k[0], c_k[1], \ldots, c_k[N-1]\}$ is the k^{th} user's spreading code, f_n $\{n = 0, 1, \ldots, N-1\}$ are the subcarrier frequencies, and finally, $b_k(t)$ represents the binary data sequence transmitted by the k^{th} user.

The baseband equivalent spectrum of the transmitted MCSS signal is shown in Figure 15.21, where we assumed that the MCSS system had eight subcarriers and the spacing between two adjacent subcarriers was $1/T_b$, where T_b represents the bit duration. Since each subcarrier signal in Equation (15.23) constitutes a conventional BPSK signal obeying Equation (15.6), the baseband equivalent PSD of the MCSS signal of Equation (15.23) in fact is given by the sum of N number of BPSK PSDs. In the MCSS system, the subcarrier waveforms are chosen to be orthogonal to each other, that is, the subcarrier waveforms satisfy the following condition

$$\int_0^{T_b} \cos(2\pi f_i t + \theta_i).\cos(2\pi f_j t + \theta_j) dt = 0, \text{ for } i \neq j \tag{15.24}$$

Therefore, the spacing between two adjacent subcarriers can take the value

$$\Delta = \frac{n}{T_b}$$

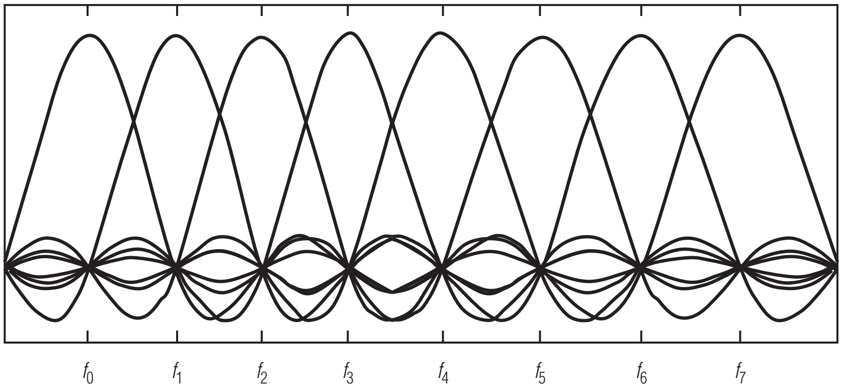


Figure 15.21 Stylized baseband equivalent spectrum of the transmitted MCSS signal

where n is an integer assuming values of $n = 1, 2, \ldots$. Usually, n assumes values of 1 or 2, so that the available frequency bandwidth can be efficiently exploited in MCSS.

If

$$\Delta = \frac{2}{T_b}$$

which is associated with allowing no spectral overlap amongst the F-domain main-lobes of the subcarrier signals' spectra, then the bandwidth required by the MCSS system for achieving a spreading gain of N becomes $2N/T_b$. It is worth noting, furthermore, that if each subcarrier's spectrum obeys the shape as shown in Figure 15.21, then the corresponding side-lobes have a zero at all legitimate subcarrier frequencies. This is advantageous in terms of minimizing the interference of the adjacent subcarrier signals.

In MC-CDMA, each data symbol is transmitted simultaneously over N relatively narrowband subcarriers. Each subcarrier is encoded with a constant phase offset. Multiple access is achieved with different users transmitting at the same set of subcarriers, but with spreading codes that are orthogonal to the codes of the other users. These codes are a set of frequency offsets in each subcarrier. It is unlikely that all of the subcarriers will be located in a deep fade and, consequently, frequency diversity is achieved.

Note that one of the IMT-2000 families of protocols is based on MC-CDMA technology. The IMT-MC protocol (CDMA2000) uses MC-CDMA spreading in the downlink, although in the uplink direction, the IMT-MC uses DS-CDMA, just like the UTRAN FDD mode. The first release of CDMA2000 will support only one downlink 1.2288-Mcps carrier ($1 \times$ RTT), so it cannot be regarded as an MC-CDMA system. However, in later releases, the IMT-MC downlink should support three parallel subcarriers ($3 \times$ RTT).

15.8 Comparison of DSSS and FHSS performances

To compare, FHSS is relatively simpler to implement than DSSS, but DSSS makes it much more difficult to recover the signal without knowing the chipping code and is more robust to signal

Table 15.2 Comparison between DSSS and FHSS

DSSS	FHSS
High cost	Low cost
High power consumption	Low power consumption
High data rates	Low data rates
Low aggregate capacity	High aggregate capacity
More range	Less range
Smaller number of geographically separate radio cells due to a limited number of channels	Most tolerant to signal interference

distortion and multipath effects. Both are widely used by a large array of wireless technologies operating on the unlicensed spectrum.

Examples of wireless communication applications of DSSS and FHSS techniques: IEEE 802.11b standard for wireless LAN employs DSSS over the 2.4-GHz free spectrums, whereas the Bluetooth standard uses FHSS for simplicity.

A brief comparison is made between DSSS and FHSS with respect to their performance for usage in a cellular system and is shown in Table 15.2.

15.9 Spreading codes

Spreading codes are also known as spreading sequences. Two types of spreading codes are used in the mobile communication:

- Orthogonal codes and
- PRN codes

In cellular phone communication, both codes are used together in the uplink and in the downlink transmissions. Examples of orthogonal spreading codes include

- Walsh codes
- Orthogonal variable spreading factor (OVSF) codes

Walsh codes provide a measure to uniquely identify each user on the forward link. Walsh codes have unique mathematical property that they are orthogonal. In OVSF, each data symbol is spread by an orthogonal code sequence.

Walsh codes are used in the IS95/CDMA2000 cellular phone systems. The spreading procedure in the mobile communication consists of two separate operations: channelization and scrambling.

Channelization uses orthogonal codes and scrambling uses PRN codes.

Channelization transforms each data symbol into several chips. In the scrambling procedure, the I- and Q-phases are further (after channelization) multiplied by a scrambling code. The same code is always used for both the spreading and despreading of a signal. This is possible because the spreading process is actually an XOR operation with the data stream and the spreading code.

15.9.1 Orthogonal codes

Channelization uses *orthogonal codes* and occurs before scrambling in the transmitter for both uplink and downlink operations. Channelization transforms each data symbol into multiple chips. This ratio (number of chips/symbol) is called the spreading factor (SF). Thus, this procedure expands the signal bandwidth. Data symbols on the I and Q branches are combined with the channelization code. Channelization codes are orthogonal codes (more precisely, OVSF codes), meaning that in an ideal environment they do not interfere with each other. However, orthogonality requires that the codes be time synchronized. Therefore, it can be used in the downlink to separate different users within one cell, but in the uplink only to separate the different services of one user. It cannot be used to separate different uplink users in a base station, as all mobiles are unsynchronized in time. Thus, their codes cannot be orthogonal (unless the system in question employs the TDD mode with uplink synchronization).

Also, orthogonal signals cannot be used as such between base stations in the downlink because there are only a limited number of orthogonal codes. The orthogonal codes must be reused in every cell, and therefore it is quite possible that user equipment (UE) in the cell boundary area receives the same orthogonal signal from two base stations, each directing their identical orthogonal codes to two different UEs. If only orthogonal spreading codes are used, these signals would interfere with each other very severely. However, in the uplink, the transmissions from one user are, of course, time synchronous; thus, orthogonal codes can be used to separate the different channels of a user. The generation method for channelization codes is defined in Section 15.9.1 and is illustrated in Figure 15.22.

This algorithm produces a tree of codes illustrated in Figure 15.23. This example shows only the root of the code tree. The UTRAN employs the SFs 4 through 512, where 4 to 256 appear in uplink, and SF 512 is added to the SF catalogue in the downlink direction. This code tree also illustrates how the codes can be allocated. If, for example, the code $C8,2$ is allocated, then no codes from its subtree can be used (i.e., $C16,4$, $C16,5$, $C32,8$). These subtree codes would not be orthogonal with their parent code. The use of orthogonal codes is depicted in Figure 15.23. A data sequence (1001) is combined with spreading code Cch,4,1 (1,1,–1,–1). This code has a SF of 4, which means that for each data signal there are four chips in the spreading code. The resulting signal bandwidth is four times wider than the bandwidth of the original signal. We then look to see what happens when this spread signal is despread with two codes, Cch,4,1 and Cch,4,2. Despreading with the correct code – the code which spread the signal – produces the original signal (1001) in the integrator, but if any of the wrong codes are used, then the result is noise. Note that an SF of 4 is a very low SF, the lowest possible in the UTRAN.

In the uplink direction, these orthogonal codes are assigned on as per UE basis, so the code management is quite straightforward. However, in the downlink direction, the same code tree is used by the base station for all mobiles in its cell area. Thus, careful code management is needed so that the base station does not run out of downlink channelization codes. Note that when the

$$\begin{bmatrix} C_{ch,2,0} \\ C_{ch,2,1} \end{bmatrix} = \begin{bmatrix} C_{ch,1,0} & C_{ch,1,0} \\ C_{ch,1,0} & -C_{ch,1,0} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$C_{ch,1,0} = 1 \qquad \begin{bmatrix} C_{ch,2^{n+1},0} \\ C_{ch,2^{n+1},1} \\ C_{ch,2^{n+1},2} \\ C_{ch,2^{n+1},3} \\ \vdots \\ C_{ch,2^{n+1},2^{n+1}-2} \\ C_{ch,2^{n+1},2^{n+1}-1} \end{bmatrix} = \begin{bmatrix} C_{ch,2^n,0} & C_{ch,2^n,0} \\ C_{ch,2^n,0} & -C_{ch,2^n,0} \\ C_{ch,2^n,1} & C_{ch,2^n,1} \\ C_{ch,2^n,1} & -C_{ch,2^n,1} \\ \vdots & \vdots \\ C_{ch,2^n,2^n-1} & C_{ch,2^n,2^n-1} \\ C_{ch,2^n,2^n-1} & -C_{ch,2^n,2^n-1} \end{bmatrix}$$

Figure 15.22 Generation of OVSF codes

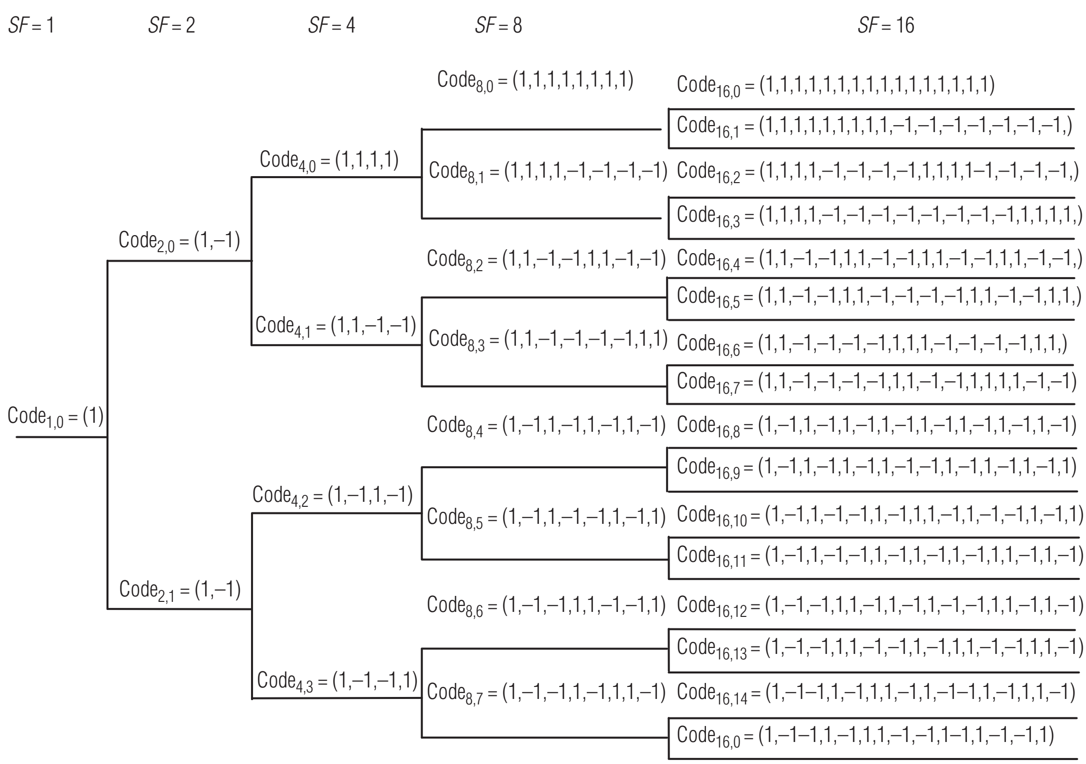


Figure 15.23 Spreading and despreading of orthogonal codes

signal was despread with a wrong code in Figure 15.23, the result in the integrator was exactly zero every time. This shows that in a fully orthogonal system, noise does not exist. Thus, in theory, power control would also be unnecessary in such a system. However, full orthogonality cannot be achieved in practice. There is always some noise in the system, and power control is needed to reduce it. The previous example demonstrated how spreading works with one user. Typical practice in the downlink has the same composite signal spread using several orthogonal codes (one for each user). This is depicted in Figure 15.23.

Note that the spreading codes must be time aligned, otherwise, the orthogonality is lost. This example shows how the original data streams can be resolved from the combined signal. The downlink transmissions from separate base stations are not orthogonal. A UE must first identify the right base station transmission according to the scrambling code, and then from that signal extract its own data using the orthogonal channelization code. Thus, in the real world, the downlink environment is never purely orthogonal and interference free. Intracell interference exists because of multipath reflections and intercell interference from asynchronous base stations. The inter-base-station non-orthogonality follows from the asynchronous nature of the system. However, the intercell interference is not as serious a problem as it might at first seem because power control and soft handovers (SHOs) should keep the other base stations from interfering too much with the downlink signal to a particular UE.

Disadvantages of orthogonal codes

The disadvantage of the orthogonal codes is that they do not fulfil the PRN properties. Therefore, the performance of such a system is not robust against non-synchronous data transmissions and multipath propagation. In addition, synchronization is often difficult to achieve because the size of the off-peak value relative to the peak value of the autocorrelation function is high.

15.9.2 Pseudorandom noise codes

A PRN code sequence is used in a DSSS system to achieve bandwidth spreading at the transmitter and the corresponding bandwidth reduction at the receiver. A pseudorandom code sequence has characteristics that are very close to those of purely random binary data streams, but is totally predictable and reproducible. In fact, a pseudorandom code sequence is a predetermined sequence of ones and zeroes that is periodic (repeats itself over and over).

Maximum length sequences or m-sequences

A pseudorandom code sequence can be produced using a linear feedback shift-register generator (LFSR generator). This type of generator consists of a shift register whose input is fed by a signal formed by combining the outputs of different stages of the shift register through modulo-2 adders. The combination of register output stages that is selected to produce the feedback signal determines the length of the pseudorandom code sequence and the arrangement of the ones and zeroes in the code sequence. PRN codes with the mathematical properties required for implementation of a DSSS radio are called maximum *length sequences (MLS) or m-sequences*. The *MLS* that can be obtained using an LFSR generator is determined by the following equation:

$$M = 2^n - 1 \tag{15.25}$$

where n is the number of stages in the shift register. These sequences are implemented by shift registers and exclusive-or (x-or) gates.

A DSSS transmitter is shown in Figure 15.24(a). The data is denoted by $d(t)$, the spreading signal is denoted by $c(t)$, and the spread waveform $q(t)$ is fed to the BPSK modulator operating at a carrier frequency f_c, and the transmitted signal is denoted by $x(t)$. $q(t)$ signal is shown in Figure 15.24(b).

The m-sequences are governed by primitive polynomials and possess good randomness properties including a two valued autocorrelation function.

For example, the 7-chip PRN sequence is governed by the primitive polynomial generator

$$C_7(x) = 1 + x^2 + x^3 \tag{15.26}$$

and the output chips are given by:

1 1 1 0 0 1 0 0 0 0 1 1 0 1 1 1 1 0 0 1 0. . .

Figure 15.25 depicts the 7-chip sequence and its autocorrelation function. Note that the autocorrelation also repeats every 7 chips, or once per bit of the actual data if each of the data bits is spread by the entire sequence. As another example, the 15-chip PRN sequence is governed by the primitive polynomial generator

$$C_{15}(X) = 1 + X^3 + X^4 \tag{15.27}$$

and the output chips are given by

1 1 1 1 1 1 1 1 1 1 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 1 1 0 0 1. . .

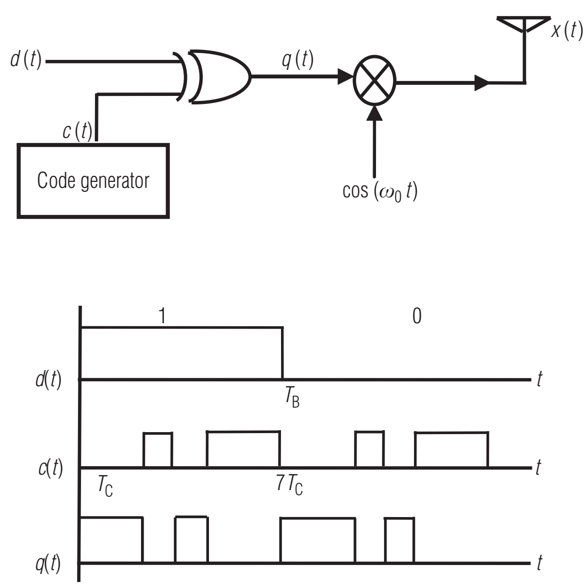


Figure 15.24 (a) DSSS transmitter; (b) $d(t)$, $c(t)$ with 7-sequences

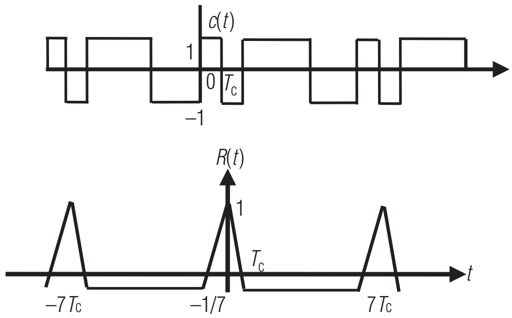


Figure 15.25 m-sequences with $q(t)$, $c(t)$

Barker codes

Barker codes, originally developed for radar, are short (13 bits or less) unique codes as compared to other spreading codes which run continually. They exhibit very good correlation properties and are very well suited for DSSS applications. The most notable property of Barker codes is that the minor peaks of their autocorrelation functions always consist of –1, 0, and +1. Barker sequences are not the natural product of linear feedback shift registers, but rather are hard-coded. A list of Barker codes is tabulated in Table 15.3.

The data stream is combined via an XOR function with a high-speed PRN as shown in Figure 15.26; the PRN specified is a 11-chip Barker code. The term “chip” is used instead of “bit” to denote the fact that the Barker code does not carry any binary information by and of itself. The result is an 11-Mbps digital stream, which is then modulated onto a carrier frequency using differential binary phase shift keying (DBPSK).

As shown in Figure 15.26, the effect of the Barker sequence is to spread the transmitted bandwidth of the resulting signal by a ratio of 11:1 (thus the term, “spread spectrum”). At the same time, the peak power of the signal is reduced by an identical ratio. Note, however, that the

Table 15.3 Barker and Willard codes

N	Barker sequences	Willard sequences
3	110	110
4	1110 or 1101	1100
5	11101	11010
7	1110010	1110100
11	11100010010	11101101000
13	1111100110101	1111100101000

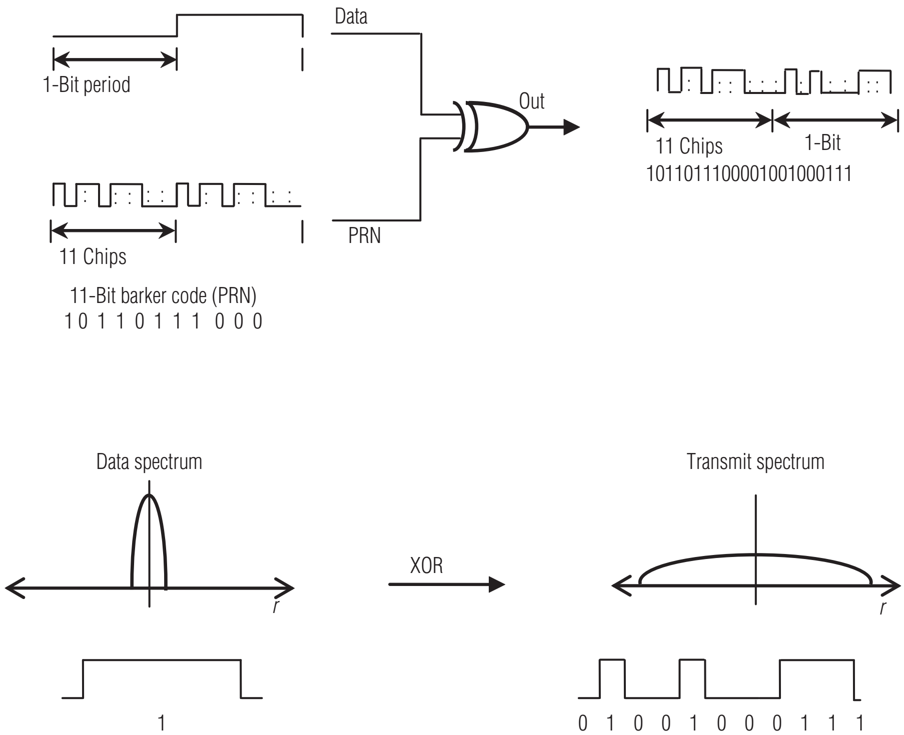


Figure 15.26 DSSS data and Barker sequence are combined via XOR function

total power is unchanged. Upon reception, the signal is demodulated and the 11 Mbcps binary stream is recovered.

Autocorrelation and cross-correlation functions of code sequences

Correlation is a measure of the degree of correspondence between two variables.

When applied to code sequences, correlation is the measure of the degree of correspondence between two code sequences when they are compared on a bit-per-bit basis. Figure 15.27 illustrates the correlation of two code sequences (Sequence 1 and Sequence 2). For each bit in the two sequences, the comparison results in an agreement (A) or a disagreement (D). The correlation value is calculated by subtracting the number of disagreements (D's) from the number of agreements (A's). The higher the correlation value, the higher the correspondence between the two sequences.

Auto-correlation

When PRN code sequences are used, it is often interesting to establish the correlation between a code sequence and a replica of this sequence that is delayed by a whole number of bit positions (chip positions). This is called autocorrelation. By varying the delay between the two sequences

SEQUENCE 1 : 1 1 1 1 0 0 0 1 0 0 1 1 0 1 0

SEQUENCE 2 : 1 1 0 1 0 0 1 1 0 0 0 1 0 0 0

COMPARISION RESULTS : A A **D** A A A **D** A A A **D** A A **D** A

CORRELATION VALUE = A's – **D's**
= 11 – **4**
= 7

Figure 15.27 Correlation of two-code sequences

over their complete length, an autocorrelation value is obtained for each delay value. Example is shown in Table 15.4.

From the autocorrelation values given for the various delays in Table 15.4, autocorrelation function plot related to the 7-bit m-sequences can be obtained, as shown in Figure 15.28.

Figure 15.28 shows that the autocorrelation is maximum and is equal to $2^n - 1$. When the delay is zero or equal to a multiple of the sequence length (0 bit, 7 bits, 14 bits, etc.), this corresponds to the situation when both sequences are aligned. For all other delays (1 to 6 bits, 8 to 13 bits, etc.), however, the two sequences are not aligned, and the autocorrelation is minimum and equal to –1. All m-sequences exhibit such an autospread spectrumcorrelation function. This

Table 15.4 Autocorrelation values for a 7-bit m-sequence

Reference m-sequence:0010111				
(Number of bit positions)	**Delayed sequence**	**Agreements (A's)**	**Disagreements (D's)**	**Autocorrelation value (A's–D's)**
0	0010111	7	0	7
1	1001011	3	4	–1
2	1100101	3	4	–1
3	1110010	3	4	–1
4	0111001	3	4	–1
5	1011100	3	4	–1
6	0101110	3	4	–1
7	0010111	7	0	7
8	1001011	3	4	–1
9	1100101	3	4	–1
10	1110010	3	4	–1

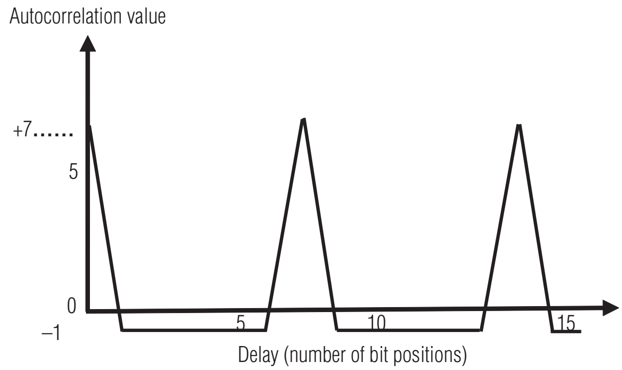


Figure 15.28 Autocorrelation function of a 7 bit (0010111) m-sequence

is an invaluable property of m-sequence since it provides a means of discriminating the alignment of the code sequence generated in a DSSS receiver with the received spread spectrum signal. Remember that correct alignment is crucial to achieve correct bandwidth reduction (correlation) of the received spread spectrum signal. The higher the length of the m-sequence, the higher the maximum autocorrelation value.

Cross-correlation

Two different pseudorandom code sequences can be compared to determine the degree of correlation between these two sequences. This is referred to as cross-correlation. Cross-correlation provides a measure of the degree of correspondence between two different code sequences. By varying the delay (a whole number of bit positions) between the two sequences over their complete length, a cross-correlation value is obtained for each delay value and a plot of the corresponding cross-correlation function can be drawn. Figure 15.29 shows the autocorrelation and cross-correlation functions for the two different 7-bit m-sequences that can be produced with a 3-stage LFSR generator. The cross-correlation function contains a few peaks where the correlation is fairly high. This indicates that for certain delay values, there is a relatively high degree of correspondence between the two sequences, even if both sequences are, in fact, different.

In DSSS applications, all PRN code sequences should ideally be m-sequences to have a good autocorrelation. Furthermore, when DSSS is used in a CDMA wireless communication system, the m-sequences used must be carefully selected to ensure that the cross-correlation functions, related to the selected sequences, are as low as possible. This is absolutely necessary to minimize the interference which any other spread spectrum signal received at the DSSS receiver input can produce when its code sequence momentarily matches that related to the spread spectrum signal desired (i.e., the one that is to be correlated). When designing a CDMA wireless communication system, the code sequence selection is a key issue that has a great impact on the system performance.

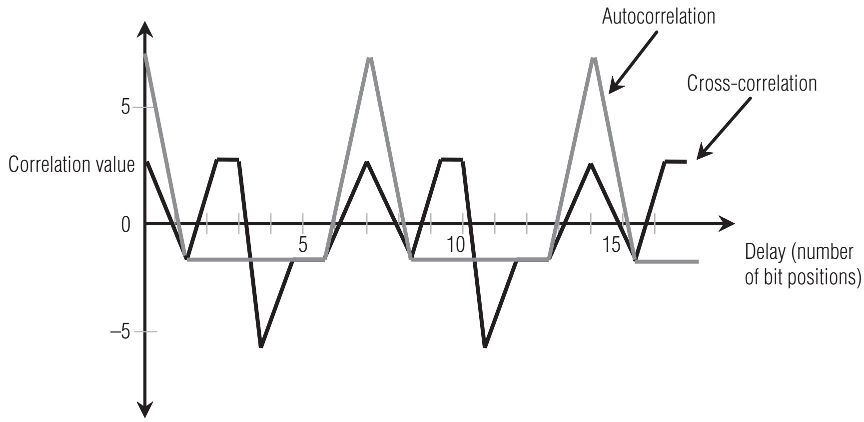


Figure 15.29 Auto- and cross-correlation functions for two different 7-bit m-sequences

15.10 Summary

- Spread spectrum techniques are used deliberately to spread the frequency domain of a signal from its narrowband domain. Various SS techniques are FHSS, DSSS, and Orthogonal Frequency Division Multiplexing (OFDM).
- In the spread spectrum technique, transmitted data occupy a larger bandwidth than required. Spreading of bandwidth is accomplished through the use of a code that is independent of the subscriber data.
- The two most common types of spread spectrum transmission are FHSS and DSSS.
- FHSS is a method of transmitting signals by rapidly switching channels, using a pseudorandom sequence known to both the transmitter and the receiver. Common examples of FHSS transmissions are wireless local area network (WLAN) cards and GSM mobile phone transmissions.
- FHSS offers three main advantages over a fixed frequency transmission:
  - Resistant to narrowband interference.
  - Difficult to intercept. An eavesdropper would only be able to intercept the transmission if they knew the pseudorandom sequence.
  - Can share a frequency band with many types of conventional transmissions with minimal interference.
- A direct sequence transmission increases the information bandwidth by mixing the information data signal with a much higher rate pseudorandom spreading (chip) sequence.
- Common examples of DSSS transmissions are WLAN cards and IS-95/3G mobile phone transmissions (CDMA).
- CDMA is a channel access method utilized by various radio communication technologies. CDMA employs spread spectrum technology and a special coding scheme (where each transmitter is assigned a code) to allow multiple users to be multiplexed over the same physical channel. By contrast, TDMA divides access by time, while FDMA divides it by frequency.

Example problem 15.1

1. If a 1 kHz signal is spread to 100 kHz, find the process gain.

Solution

The process gain expressed as a numerical ratio would be 100,000/1,000 = 100. Or, in decibels, $10\log_{10}(100) = 20$ dB.

Example problem 15.2

A DSSS system has a 1.2288 Mcps code clock rate and a 9.6 kbps information rate. Calculate the processing gain. How much improvement in information rate is achieved if the code generation rate is changed to 5 Mcps and the processing gain to 256?

Solution

$$G_p = \frac{R_c}{R_b}$$

$$\frac{1.2288\times 10^6}{9.6\times 10^3} = 128 = 10\log 128 = 21 \text{ dB}$$

$$R_b = \frac{R_c}{G_p}$$

$$R_b = \frac{5\times 10^6}{256} = 19.53 \text{ kbps}$$

Improvement in information rate = 19.53 – 9.6 = 9.93 kbps.

Example problem 15.3

In an FHSS system, a hopping bandwidth of 100 MHz and a frequency spacing of 10 kHz are used. What is the minimum number of PRN chips that are required for each frequency symbol?

Solution

Number of frequency tones in hopping bandwidth $= \dfrac{100\times 10^6}{10^4} = 10^4$

$$\text{Minimum number of chips} = \log_2(10^4) = 13$$

Example problem 15.4

A communication system transmits at 120 kbps and uses 32-FSK (frequency shift keying). A hop rate of 2,000 hops per second is used over an available spectrum of 10 MHz. Assuming a negligible synthesizer switching time between hops, calculate (a) data symbol transmitted per hop, and (b) the number of non-overlapping hop frequencies.

Solution

For 32-FSK, we have $32 = 2^b$ so here $b = 5$ bits per symbol

$$\text{Symbol rate} = \frac{120}{5} = 24 \text{ kbps}$$

Since the symbol rate is higher than the hope rate, the system is a slow FHSS system.

(a) Number of symbols per hop $= \dfrac{24{,}000}{2{,}000} = 12$

Minimum bandwidth B_w of an M-ary FSK~MRs:

$$B_w = 12 \times 24 = 288 \text{ kHz}$$

(b) Number of non-overlapping hop frequencies

$$n_{\text{FH}} = \frac{10 \text{ MHz}}{288 \text{ kHz}} \approx 35$$

Example problem 15.5

Consider an FHSS system in which the input data rate is 200 bits per second. The modulation scheme of 32-ary FSK is used to generate the modulation symbol. The frequency-hopping rate is 200 hops per second. Calculate (a) minimum separation between frequency tones; (b) number of frequency tones produced by a frequency synthesizer; (c) processing gain; and (d) hopping bandwidth. Assume a frequency multiplication factor $K = 1$.

Solution

With the 32-FSK modulation scheme, there are 5 chips per symbol.

$$\text{Symbol rate, } R_s = \frac{200}{5} = 40 \text{ symbols per sec}$$

Since the hop rate is higher than symbol rate, the system is a fast FHSS system.

$$\text{Symbol duration} = \frac{1}{40} = 25 \text{ ms}$$

$$L = \frac{200}{40} = 5 \text{ hops/symbol}$$

$$\text{Chip duration} = \frac{25}{5} = 5 \text{ ms}$$

$$\text{Minimum separation between tones} = \frac{1}{5 \times 10^{-3}} = 200 \text{ Hz}$$

$$M = 2^5 = 32 \text{ frequency tones}$$

$$\text{Frequency hopping band width} = KM\Delta fL = 1 \times 32 \times 200 \times 5 = 64 \text{ kHz}$$

$$G_p = MKL = 32 \times 1 \times 5 = 160$$

Review questions

1. Define spread spectrum. List its three benefits.
2. Compare and contrast FHSS and DSSS.
3. What are the basic differences between the DSSS and FHSS systems?
4. How does the spread spectrum technology work? Write its applications.
5. Define different types of spread spectrum technique. What are its advantages and disadvantages?
6. Explain DSSS with BPSK with neat diagrams and explain the properties of DSSS?
7. Differentiate DSSS, FHSS, THSS, and MCSS.
8. Write short notes on spreading codes.
9. Define autocorrelation and cross-correlation. Give an example for each.
10. Compare and contrast autocorrelation and cross-correlation.
11. Write short notes on (i) jamming and (ii) multipath fading.

Objective type questions and answers

1. Spread spectrum concept in wireless communication systems allows multiple users to occupy
   (a) same transmission band (b) multiple transmission bands
   (c) both a and b (d) none
2. The pulse in the spreading code is called
   (a) bit (b) chip (c) symbol (d) none
3. The rate at which the spread data (or PRN code) varies is called
   (a) chip rate (b) baud rate (c) bit rate (d) none
4. In DSSS, higher the processing gain
   (a) higher the power density (b) lower the power density
   (c) none
5. The applications of spread spectrum technology
   (a) anti-jamming (b) interference rejection
   (c) low probability of intercept (d) all of the above
6. The primary spread spectrum techniques used in cellular system and GPS for multiple access are
   (a) DSSS (b) FHSS (c) both (a) and (b) (d) none
7. Properties of DSSS are
   (a) resistance to interference (b) anti-jamming effects
   (c) resistance to multipath effects (d) all of the above

8. FHSS uses a ________ to spread a user signal.
9. The processing gain, G_p, of an FHSS system is given as ________.
10. Channelization uses ________ and scrambling uses ________.

Answers: 1. (c), 2. (b), 3. (a), 4. (b), 5. (d), 6. (c), 7. (d) 8. frequency hopping sequence, 9. $G_\text{p} = \dfrac{\text{Hopping bandwidth}}{\text{Minimum frequency spacing}} = \dfrac{M\Delta f}{\Delta f} = M$, 10. orthogonal codes and PRN codes.

Open book questions

1. How can spreading be achieved?
2. How can DSSS systems benefit from multipath propagation?
3. What is meant by DSSS modulation?
4. How is wideband frequency spectrum generated in a frequency-hopping (FH) technique? Distinguish between FDMA and FHMA.

Key equations

1. Spreading factor (SF) or processing gain (G_p).

   SF = Chip rate/Symbol rate or

   = Transmission signal bandwidth/Original signal bandwidth

2. The input and output S/N ratios are related as:

$$\left(\frac{S}{N}\right)_0 = G_P\left(\frac{S}{N}\right)_i$$

$$\left(\frac{E_b}{N_0}\right)_0 = G_p\left(\frac{E_b}{N_0}\right)_i$$

   where E_b is the bit energy and N_0 is noise power spectral density.

3. The bit error probability, P_e, associated with the coherent BPSK spread spectrum signal is given as

$$P_e = \frac{1}{2}\text{erfc}\left[\sqrt{\left(\frac{E_b}{N_0}\right)}\right] = \text{Q}\left[\sqrt{2\left(\frac{E_b}{N_0}\right)_0}\right] = \text{Q}\left[\sqrt{2\left\{G_p\left(\frac{E_b}{N_0}\right)_i\right\}}\right]$$

4. The processing gain, G_p, of an FHSS system is given as

$$G_\text{p} = \frac{\text{Hopping bandwidth}}{\text{Minimum frequency spacing}} = \frac{M\Delta f}{\Delta f} = M$$

Further reading

Belkerdid, M. A. Electrical and Computer Engineering Department, University of Central Florida, Orlando, FL.

Dixon, R. C. *Spread Spectrum Systems*. New York: Wiley-Interscience, 1984.

Garg, V. K. *CDMA IS-95 and cdma2000*. Upper Saddle River, NJ: Prentice Hall, 2000.

Garg, V. K., and Wilkes, J. E. *Wireless and Personal Communications Systems*, Prentice Hall, New Delhi, 1996.

Holma, H., and Toskala, A. (eds.). *WCDMA for UMTS*, New York: John Wiley & Sons, 2000.

Lee, W. C. Y. *Mobile Cellular Telecommunication Systems*, New York: McGraw-Hill, 1989.

Pahlwan, K., and Levesque, A. H. *Wireless Information Networks*, New York: John Wiley & Sons, 1995.

Peterson, R. L., Ziemer, R. E., and Borth, D. E. *Introduction to Spread Spectrum Communications*. Inglewood Cliffs, NJ: Prentice Hall, 1995.

Pickholtz, R. L., Schilling, D. L. and Milstein, L. B. "Theory of Spread-Spectrum Communications – A Tutorial," *IEEE Transactions on Communications*, Vol COM-30, May 1982.

Smith, C., and Collins, D. *3G Wireless Networks*, New York: McGraw-Hill, 2002.

Shannon, C. E. "Communications in the Presence of Noise," *Proceedings of the IRE*. No. 37, pp. 10–21, 1949.

Steele, R. *Mobile Radio Communications*. New York: IEEE Press, 1992.

Multiple Access Techniques and Advanced Transceiver Schemes 16

16.1 Introduction

Cellular systems divide a geographic region into cells, where a mobile phone in each cell communicates with a base station. The main objective of cellular systems design is to handle as many calls as possible (i.e., capacity) in a given bandwidth with reliability. The success of cellular network depends on the availability of the radio frequency spectrum. The radio frequency spectrum is a finite natural resource and has greater demands placed on it every day. This finite resource is usually defined in terms of *bandwidth*.

> *To allow many mobile users to share simultaneously a finite amount of radio spectrum, in a most efficient way, various technologies have been developed and the goal behind these methods is to handle as many calls as possible in a given bandwidth (i.e., call-handling capacity). This concept is called "multiple access" (Figure 16.1).*
>
> *The choice of multiple access (MA) technology is to share the available scarce bandwidth efficiently among a large number of users which could significantly enhance or lower the service quality delivered to end users.*

There are several different ways to allow access to the channel. The following are the four possible access methods: frequency division multiple access (FDMA), time division multiple access (TDMA), code division multiple access (CDMA), and space division multiple access (SDMA) (Figure 16.2). Since FDMA underlies all of the cellular systems, we first focus on the FDMA concept, and the later part of the chapter is devoted to describe the functionality of each access method (FDMA, TDMA, and CDMA), the advantages and disadvantages of each technology, and various forms of implementation for each technology. We also discuss the features of SDMA and basic principle of operation of orthogonal frequency division multiplexing (OFDM) technique.

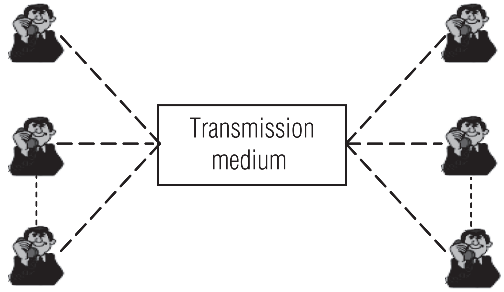


Figure 16.1 Multiple access scenario

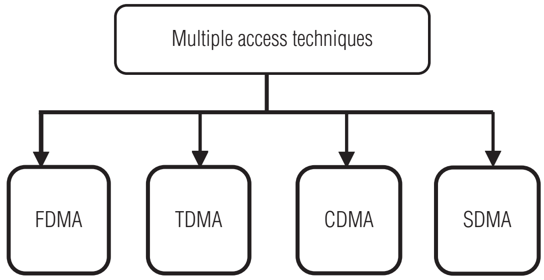


Figure 16.2 Multiple access techniques

16.2 Multiple access techniques

Multiplexing can create channels in frequency, time, code, etc., and the corresponding terms are frequency division multiplexing (FDM), time division multiplexing (TDM), code division multiplexing (CDM), etc.

> *Sharing the limited bandwidth efficiently among many users is one of the main objectives of multiple access schemes.*

The analogy of highway with several lanes (Figure 16.3) gives the simple example of multiple access methods such as SDMA, FDMA, and TDMA. In this figure, the medium is highway, the users are cars, and the interference is due to accidents.

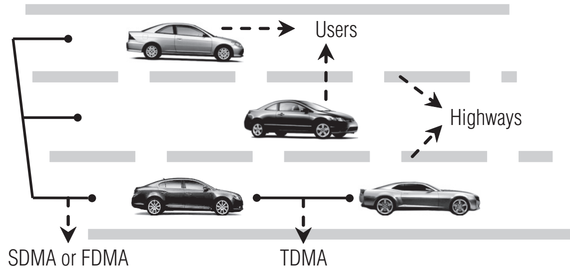


Figure 16.3 Multiple access methods

The best way to describe the differences among FDMA, TDMA, and CDMA technologies is with an example of how they work. One of the best examples shown in Figure 16.4 is explained below.

Imagine a large room with a group of people divided up into pairs. Conversation of each pair is independent of conversations of other pairs. For these conversations to take place without interruption from other conversations, it is necessary to define an isolated environment for each conversation. In this example, the room should be considered as a slice of the radio spectrum specifically allocated to this group of people. Imagine each pair communicating through cellular telephones or radios.

FDMA: Applying an FDMA system to this analogy, the single large room (slice of spectrum) would be partitioned with many dividing walls and creating a large number of smaller rooms (Figure 16.5). A single pair of people would enter each small room and hold their conversation. Each room is like a single frequency/channel. No one else could use the room (or frequency) until the conversation was complete, regardless of whether the parties were actually talking. When the conversation is completed, the first pair of people would leave and another pair would then be able to enter that small room.

The method of providing multiple access capability by transmitting the signals simultaneously in a non-overlapping frequency bands is called ***FDMA****.*

TDMA: In TDMA environment, each of the small rooms would be able to accommodate multiple conversations "simultaneously".

For example, with a three-slot TDMA system, each "room" would contain up to three pairs of people, with the different pairs taking turns talking. According to this system, each pair can speak for 20 s during each minute. Pair A would use 0:01 through 0:20 s, pair B would use 0:21 through 0:40 s, and pair C would use 0:41 through 0:60 s. However, even if there were fewer than three pairs in the small room, each pair would still be limited to 20 s/min (Figure 16.6).

The method of providing multiple access capability by transmitting the signals simultaneously in a non-overlapping time slot is called ***TDMA****.*

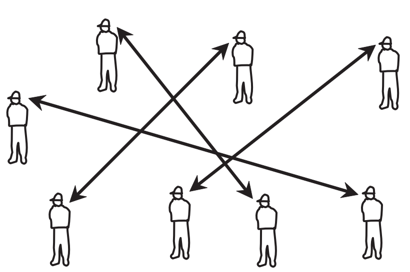


Figure 16.4 Example for multiplexing analogy

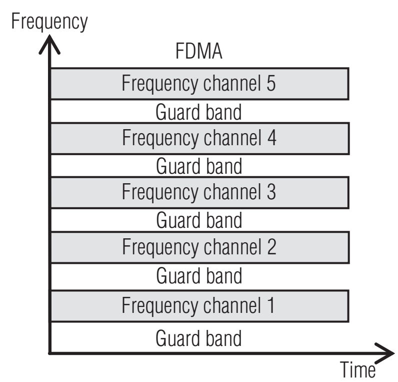


Figure 16.5 Channel allocation in FDMA

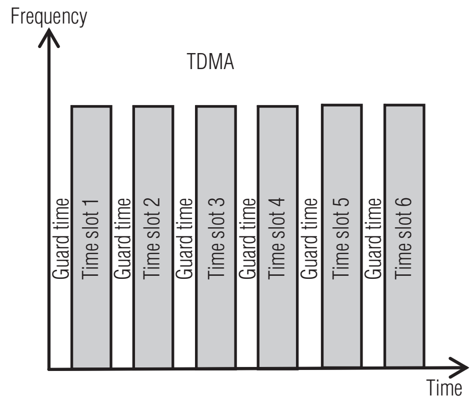


Figure 16.6 Time slots in TDMA

CDMA: Using the CDMA technology, all the little rooms would be eliminated. All pairs of people would enter the single large room (our spectrum space). Each pair would be holding their conversations in a different language and, therefore, they could use the air in the whole room to carry their voices while experiencing little interference from the other pairs (Figure 16.7).

1. The air in the room is analogous to a wideband "carrier" and the languages represent the "codes" assigned by the CDMA system. In addition, language "filters" would be incorporated, for example, people speaking French would hear virtually nothing from those speaking another language, say Russian.
2. Additional pairs could be added, each speaking a unique language (as defined by the unique code) until the overall "background noise" (interference from other users) made it too difficult to hold a clear conversation. By controlling the voice volume (signal strength) of all users to a minimum, the number of conversations that could take place in the room could be maximized (i.e., maximize the number of users per carrier).

The method of providing multiple access capability based on a spread-spectrum system is called CDMA. All users share same frequency and at the same time, but each user has own spreading code to encode data.

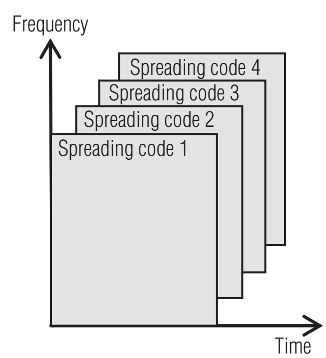


Figure 16.7 Illustration of CDMA

Table 16.1 Multiple access technologies used in 1G to 3G wireless systems

Cellular systems	MA technique	Modulated carrier bandwidth	Conversations/carrier	Modulation
1G cellular system				
Advanced mobile phone system (AMPS)	FDMA/FDD	30 kHz	1	Analogue FM (FSK for control signal)
2G cellular system				
Digital advanced mobile phone system (D-AMPS) or IS-136	TDMA/FDMA	30 kHz	3 or 6	Differential $\pi/4$ offset DQPSK
GSM	TDMA/FDD	30 kHz	8 or 6	Digital FM GMSK
IS-95	CDMA/FDD	1280 MHz	62 planned (12–18 achieved)	BPSK
3G cellular system				
IMT-2000 defined the 3G	WCDMA CDMA2000	5 MHz–wide carrier	Number of slots/frames: 15 Frame length: 10 msec	QPSK (forward) BPSK (reverse)

Spread-spectrum is a technique that increases signal bandwidth beyond the minimum necessary for data communication; the band spread is accomplished by means of a code that is independent of the data, and a synchronized reception with the code at the receiver is used for despreading and subsequent data recovery.

Various multiple access technologies and modulation techniques used in different wireless systems starting from 1G to 3G are presented in the Table 16.1.

All the aforementioned multiple access techniques are described in detail in the following sections.

16.3 Frequency division multiple access

Analogue transmission is considered as an "older" cellular phone technology. Analogue allows a cellular phone to transmit signals by sending voice, video, and data that are always changing.

> *The FDMA is the simplest scheme used to provide multiple access in analogue transmission. In FDMA systems, the radio frequency spectrum is divided into several frequency bands separated by a certain guard band. Each frequency band can be used simultaneously.*

In this technique, the bandwidth is divided into a number of channels and distributed among users with a finite portion of bandwidth for permanent use as illustrated in Figure 16.8. In addition, we can observe from Figure 16.8 that FDMA permits only one user per channel because it allows the user to use the channel 100 per cent of the time. Therefore, only the frequency

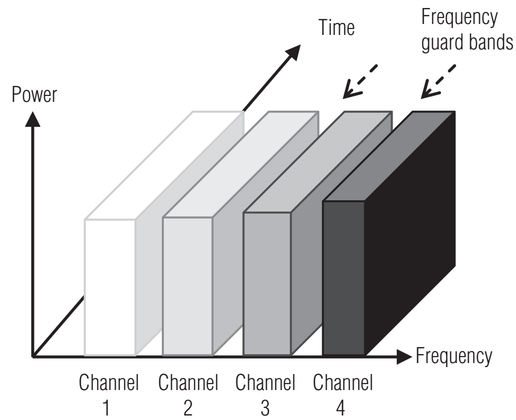


Figure 16.8 FDMA principle of operation

"dimension" is used to define channels. Each block represents a different user. Figure 16.9 shows how the three users occupy the time and the bandwidth with FDMA.

Frequency guard bands are provided between adjacent signal spectra to minimize crosstalk between adjacent channels.

In FDMA, the channel has two frequencies, namely forward channel and reverse channel. When the FDMA technique is employed, as long as the user is engaged in "conversation," no other user can access the same spectrum space.

Number of channels supported by FDMA system

The number of channels that can be simultaneously supported in FDMA system is given by (Figure 16.10)

$$\text{Number of channels } (N) = \frac{B_t - 2B_g}{B_c} \tag{16.1}$$

where

B_t is the allocated frequency spectrum
B_g is the guard band allocated at the edge of frequency spectrum
B_c is the channel bandwidth

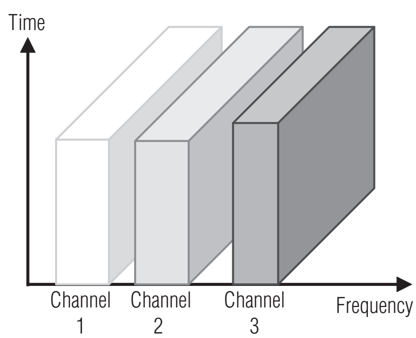


Figure 16.9 Time and bandwidth occupancy of three user signals with FDMA

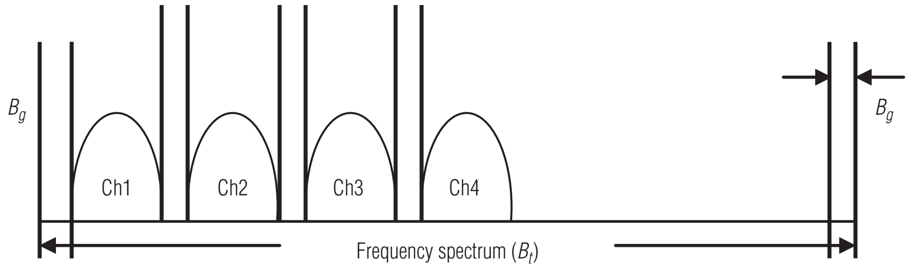


Figure 16.10 Frequency spectrum of FDMA

Example problem 16.1

If each cellular carrier is allocated N number of channels and *the allocated frequency spectrum is 12.5 MHz*, guard band is 10 KHz and the bandwidth of each channel is 30 KHz. Find the total number of channels provided by the cellular carrier.

Solution

From Equation (16.1)

$$\text{Number of channels } (N) = \frac{B_t - 2B_g}{B_c}$$

$$\text{Number of channels } (N) = \frac{[(12.5\times10^6) - 2(10\times10^3)]}{30\times10^3} = \mathbf{416}$$

Example problem 16.2

In the United States, the advanced mobile phone service (AMPS) cellular operator is allocated 15 MHz for each simplex band, and if B_t is 10 MHz, B_g is 10 KHz, and B_c is 25 KHz. Find the number of channels available in an FDMA system.

Solution

The number of channels available in the FDMA system is given as

$$N = \frac{B_t - 2B_g}{B_c}$$

$$N = \frac{10\ \times10^6 - 2(10\times10^3)}{25\times10^3} \approx 399$$

In the United States, each cellular carrier is allocated **399** channels.

16.3.1 Advantages, disadvantages, and applications of FDMA

The advantages and disadvantages of FDMA technology with respect to TDMA and CDMA are discussed in the following.

Advantages

1. A continuous transmission scheme, and therefore of lower complexity than TDMA scheme, for example, synchronization requirements are not severe
2. Simple to implement from a hardware standpoint, because multiple users are isolated by employing simple band pass filters
3. Fairly efficient with a small base population and when traffic is constant
4. No channel equalization required
5. Capacity can be increased by reducing the information bit rate and using an efficient digital speech coding scheme

Disadvantages

1. If an FDMA channel is not in use, then it sits idle and cannot be used by other users to increase or share the capacity. Therefore, FDMA implementation becomes *inefficient use of spectrum*.
2. FDMA requires tight RF filtering to minimize adjacent channel interference.
3. Network and spectrum planning are intensive.
4. Frequency planning is time consuming.
5. Even though no two users use the same frequency band at the same time, guard bands are introduced between frequency bands to minimize adjacent channel interference. Guard bands are unused frequency slots that separate neighbouring channels. This leads to a waste of bandwidth. When continuous transmission is not required, bandwidth goes wasted since it is not being utilized for a portion of the time.

Applications of FDMA

1. Walkie-talkies and mobile networks for closed user groups often use FDMA.
2. Another example of FDMA is AM or FM radio broadcasting, where each station has its own channel.
3. Early cellular telephony mostly used FDMA analogue transmission.

16.3.2 FDMA versus frequency division duplex

The pairing of channels is to provide a two-way communication in a direction, uplink or downlink (for instance, the traffic going back and forth between a mobile phone and a base-station), is termed as frequency division duplex (FDD) (Figure 16.11). The composite scheme, with separate frequency channels assigned for each direction of communication, is thus labelled FDMA/FDD (Figure 16.12).

FDD provides two simplex channels at the same time.

FDD is a physical layer technique that combines and transmits low-bandwidth channels through a high-bandwidth channel. FDMA, on the other hand, is an access method in the data link layer.

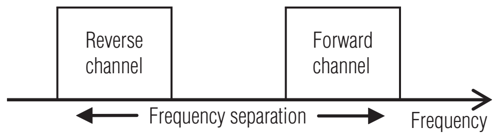


Figure 16.11 FDD provides two simplex channels

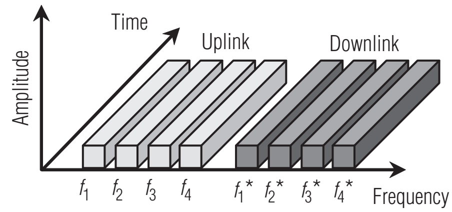


Figure 16.12 FDMA/FDD

16.3.3 Example and applications of FDMA in 1G cellular (or AMPS) and GSM 900

Applications of FDMA in 1G

1G cellular systems are analogue and they operate on the FDMA principle. The AMPS is 1G analogue cellular system used in North America and is an example of a pure FDMA. The AMPS system was allocated a frequency range of **869–894 MHz** for downlink or forward transmission from base station to mobile station and **824–849 MHz** band for the uplink or reverse communication from mobile station to base station. One channel in each direction, are assigned to one mobile user in these 25 MHz bands (894 MHz–869 MHz and 849 MHz–824 MHz). The assignment of frequency channels is sometimes referred to as a frequency-orthogonal assignment, since it enables multiple users to simultaneously use the system. Each user is accommodated a unique channel in AMPS. When a call is completed, or when a handoff occurs, the channel is vacated so that another mobile subscriber may use it.

Application of FDD/FDMA scheme in GSM 900 systems

In GSM 900, the frequencies for uplink and downlink communication are allotted by the base station as illustrated in Figure 16.13. The uplink frequency band is allotted between 890.2 and 915 MHz. The uplink channel frequency (f_u) allocation is done using

$$f_u = 890 \text{ MHz} + n \times 0.2 \text{ MHz} \tag{16.2}$$

where n is the channel number (starting from 1 to 124).

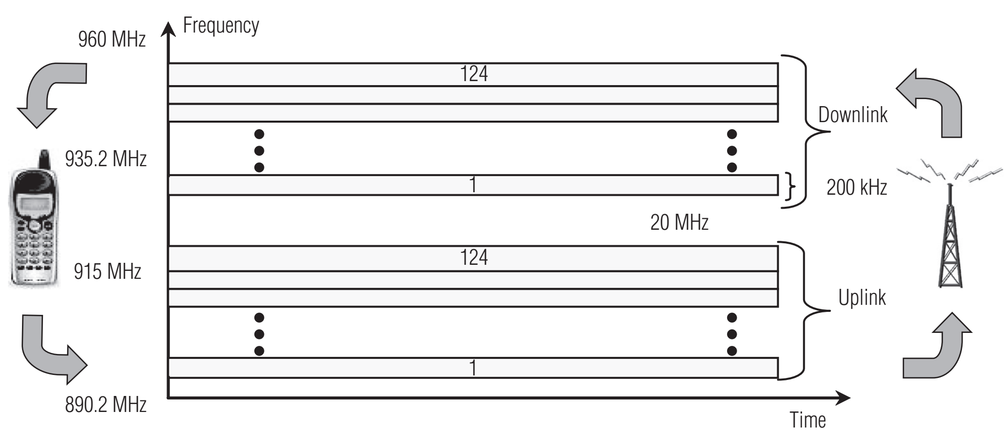


Figure 16.13 Illustration of frequencies for uplink and downlink communication

The downlink frequency band is allotted between 935 and 960 MHz and is allocated using

$$\text{Downlink channel frequency } (f_d) = f_u + 45 \text{ MHz} = 935 \text{ MHz} + n \times 0.2 \text{ MHz} \tag{16.3}$$

16.4 Time division multiple access

As the frequency spectrum experiences more traffic, spectrum efficiency becomes increasingly important. TDMA systems were developed as FDMA system spectrum efficiency became insufficient. In digital systems, continuous transmission is not required because users do not use the allotted bandwidth all the time. It allows several users to share the same frequency band by dividing the timescale into different time slots which are periodically allocated to each mobile user for the duration of a call.

TDMA systems have the capability to split users into time slots because they transfer digital data, instead of analogue data commonly used in legacy FDMA systems.

Digital 2G cellular systems that used the TDMA technology are GSM, IS-136, PDC, and DECT standard for portable phones

16.4.1 TDMA principle of operation

TDMA systems divide the radio spectrum into time slots and each user is allowed to either transmit or receive in each time slots (i.e., different users can use the *same frequency* in the same cell but at different times).

TDMA multiplexes three signals over a single channel. The current TDMA standard for cellular divides a single channel into six time slots, with each signal using two slots, providing a three to one gain in capacity over AMPS. Each caller occupies a cyclically repeating time slots.

In TDMA, when the caller depresses the push-to-talk (PTT) switch, a control channel registers the radio to the closest base station. During registration, the base station assigns the user an available pair of channels, one to transmit and the other to receive. However, unlike an FDMA system registration, a TDMA system registration also assigns an available time-slot within the channel. The user can only send or receive information at that time, regardless of the availability of other time-slots. Information flow is not continuous for any user, but rather is sent and received in bursts. The bursts are re-assembled at the receiving end and appear to provide continuous sound because the process is very fast.

Figure 16.14 shows the principle of operation. Each row in the Figure 16.14 represents a single channel and divided into three time-slots. Calls in a TDMA system start in analogue format and are sampled, transforming the call into a digital format. After the call is converted into digital format, the TDMA system places the call into an assigned time slot.

TDD: In cellular communications, when a user moves from one cell to another there is a chance that user could experience a call loss if there are no free time slots available. TDMA uses different time slots for transmission and reception. This type of duplexing is referred to as time division duplexing (TDD). TDD does not require duplexers. From Figure 16.15 TDD provides two simplex time slots on the same frequency.

Number of users supported by TDMA illustrates the improvement in the efficiency of TDMA system; let us compare Figures 16.8 and 16.13. It is obvious that the FDMA system (Figure 16.8)

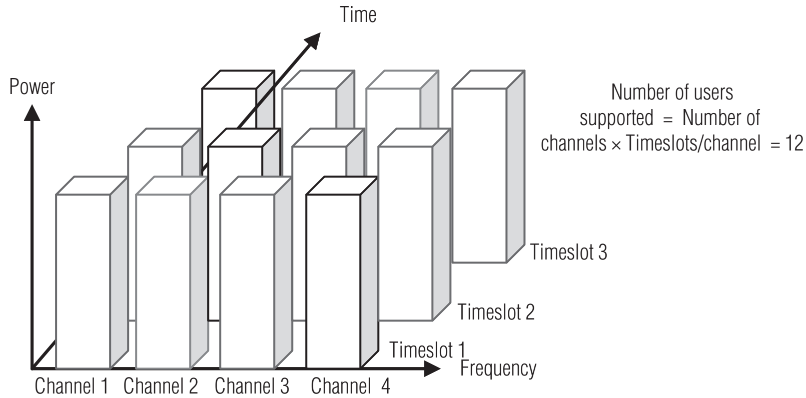


Figure 16.14 TDMA principle of operation

supports 4 users while the TDMA system (Figure 16.14) supports 12 users within the same bandwidth as the FDMA system (Equation (16.3)).

Number of users supported by the TDMA system

$$= \text{Number of channels in the frequency spectrum} \times \text{Time slots/channel} \qquad (16.4)$$
$$= 4 \times 3 = 12 \text{ (from Figure 16.14)}$$

Nowadays, there are systems in place that allow an increase of up to six times the capacity of FDMA alone. Because TDMA systems also split an allotted portion of the frequency spectrum into smaller slots (channels), they require the same level of frequency planning as FDMA systems. The same careful steps in frequency planning must be taken in both FDMA and TDMA systems.

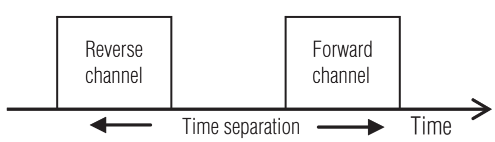


Figure 16.15 TDD two simplex time slots

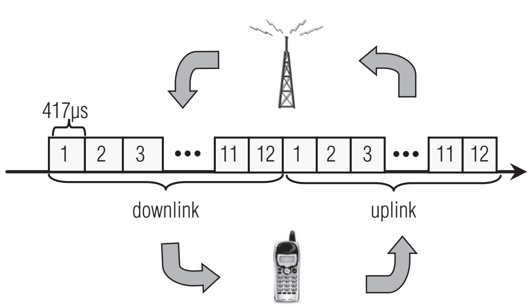


Figure 16.16 TDMA/TDD-example

16.4.2 TDMA frame structure

In a TDMA system, time is divided into equal time intervals called *slots*. User data is transmitted in the slots. Several slots make up a frame. Guard times are used between each user's transmissions to minimize crosstalk between channels (Figure 16.17).

Each user is assigned a frequency and a time slot to transmit data. The data is transmitted via a radio-carrier from a base station to several active mobiles in the downlink. In the reverse direction (uplink), transmission from mobiles to base stations is time-sequenced and synchronized on a common frequency for TDMA. The preamble carries the address and synchronization information that both the base station and the mobile stations use for identification.

16.4.3 Efficiency of TDMA

The efficiency of TDMA system is a measure of the percentage of transmitted data that contains information as opposed to providing overhead for the access scheme. The frame efficiency η is the percentage of bits per frame which contains transmitted data.

$$\text{Efficiency } (\eta) = \frac{\text{Number of bits per frame containing in the transmitted data}}{\text{Total number of bits per frame}} \tag{16.5}$$

$$= \frac{(b_T - b_{OH})}{b_T} \times 100 = (1 - b_{OH} / b_T) \times 100$$

where

b_T is the total number of bits per frame $= T_f \times R$ (16.6)

T_f is the frame duration

R is the channel bit rate

b_{OH} is the number of overhead bits per frame $= N_r \times b_r + N_t \times b_p + N_t \times b_g + N_r \times b_g$ (16.7)

where

N_r is the Number of reference bits per frame

N_t is the number of traffic bits per frame

b_r is the number of overhead bits per reference burst

b_P is the number of overhead bits per reference in each slot

b_g is the number of equivalent bits in each guard time interval

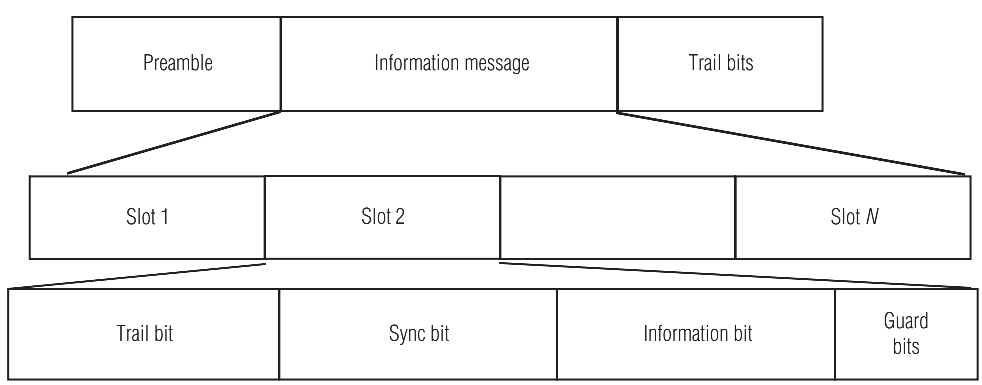


Figure 16.17 TDMA frame structure

Number of channels/time slots in TDMA system

$$N = \frac{M(B_{\text{total}} - 2B_{\text{guard}})}{B_C} \tag{16.8}$$

where

M is the number of time slots per carrier channel or maximum number of TDMA users supported on each radio channel

B_g is the guard band to present user at the edge of the band

16.4.4 Advantages and disadvantages of TDMA

Advantages

- Data Transmission is in discrete bursts
  - Extended battery life over FDMA and talktime
  - Handoff process is simpler, since it is able to listen for other base stations during idle time slots
- More efficient use of spectrum, compared to FDMA
  - Will accommodate more users in the same spectrum space than an FDMA system which improves capacity in high-traffic areas, such as large metropolitan areas
  - Efficient utilization of hierarchical cell structures – pico, micro, and macro cells
- Since different slots are used for transmission and reception, duplexers are not required

Disadvantages

- TDMA requires synchronization. If the time slot synchronization is lost, the channels may collide with each other.
- For mobiles and, particularly for handsets, TDMA on the uplink demands high-peak power in transmit mode that shortens battery life. If a TDMA interface consists of n channels, then the transmitted powers are 10 log n times higher than in an FDMA system.
- Network and spectrum planning are intensive.
- Dropped calls are possible when users switch in and out of different cells.
- Higher costs due to greater equipment sophistication.
- Equalization is required, since transmission rates are generally very high as compared to FDMA channels.

Example problem 16.3

In a digital cellular system such as GSM system, TDMA is used as multiple access system. The GSM system uses a frame structure where each frame consists of eight time slots, and each time slot contains 155.55 bits, and data is transmitted at 270.833 Kbps in the channel. Find (a) time duration of a bit, (b) time duration of a slot, (c) time duration of a frame, and (d) the duration for which a user occupying a single slot must wait between two simultaneous transmissions.

Solution

Time duration of a bit = T_b = 1/bit rate

$= 1/270.833 \times 10^3 = 3.692\ \mu\text{s}$

Time duration of a slot = $T_{\text{slot}} = 155.55 \times T_b = 0.574$ ms

Time duration of a frame $8 \times T_{\text{slot}} = 4.592$ ms
User has to wait 4.592 ms before next transmission

Example problem 16.4

If a normal GSM timeslot consists of 6 trailing bits, 8.25 guard bits, 26 training bits, and 2 traffic bursts of 61 bits of data, find the frame efficiency

Solution

Number of bits in a time slots = 6 + 8.25 + 26 + 2(61) = 162.25 bits.
Number of bits in Frame = 8 × 162.25 = 1,298 bits/frame.
The number of overhead bits per frame is given by

$$b_{OH} = 8(6) + 8(8.25) + 8(26) = 322 \text{ bits}$$

Frame efficiency, η = (1,298 – 322)/1,298 × 100 = 75.192 per cent

---

Example problem 16.5

Consider a GSM system, which is in TDMA/FDD system that uses the 30 MHz forward link, which is broken into number of channels of 240 kHz. If eight speech channels are supported on a single radio channel and if no guard band is assumed, find the number of simultaneous users that can be accommodated in GSM.

Solution

The number of simultaneous users that can be accommodated in GSM is

$$N = \frac{30\times10^6}{\left(240\times10^3/8\right)} = 1{,}000$$

Thus, GSM can accommodate 1,000 simultaneous users.

Example problem 16.6

If normal GSM time slot consists of 6 trailing bits, 8.50 guard bits, 28 training bits, and 2 traffic burst of 58 bits of data, find the frame efficiency.

Solution

A time slot has 6 + 8.50 + 28 + 2(58) = 158.5 bits
A frame has 8 × 158.50 = 1268 bits/frame
The number of overhead bits per frame is given by $b_{OH} = 8(6) + 8(8.5) + 8(28) = 340$ bits

The frame efficiency $\eta = \left(1 - \dfrac{340}{1268}\right) \times 100 = 73.19\%$

16.5 FDMA versus TDMA

- If an FDMA channel is not in use, it is idle and cannot be used by other users. After the assignment, the reverse and forward channels may transmit simultaneously and continuously.
- FDMA is implemented as single channel per carrier (SCPC), and is narrow band (NB) 30 kHz. The symbol time is large as compared to average delay spread. Inter-symbol interference (ISI) is low, and no equalization is required for NB-FDMA.
- FDMA mobile system is less complex as compared to TDMA. Due to advancement in DSP this is changing.
- FDMA is continuous transmission scheme, so fewer overhead bits are required as compared to TDMA.
- FDMA systems have higher cell site costs as compared to TDMA, because of SCPC in FDMA, and use of expensive pass band filters to eliminate the spurious radiations at base stations.
- The FDMA mobile unit uses duplexers, since both transmitter and receiver operate at the same time, resulting in an increase in the cost of mobile units and base stations.
- FDMA requires tight RF filtering to minimize adjacent channel interference.

16.6 Code division multiple access

CDMA is the third multiple access technique used in cellular systems. CDMA allows transmissions to occupy the entire bandwidth at the same time without interference.

> *CDMA uses spread-spectrum technique to increase spectrum efficiency over current FDMA and TDMA systems.*

A *spread-spectrum signal* is a signal that has an extra modulation that expands the signal bandwidth beyond what is required by the underlying data modulation. Spread-spectrum communication systems are useful for the following:

1. Suppressing interference
2. Making interception difficult
3. Accommodating fading and multipath channels
4. Providing a multiple-access capability

The most practical and dominant methods of spread-spectrum communications are direct-sequence modulation and frequency hopping of digital communications. These aspects, which are basis for the CDMA, are discussed in detail in under spread-spectrum techniques in Chapter 15.

CDMA cellular technology is originally known as *IS-95*, which competes with GSM technology for dominance in the cellular world. CDMA cellular systems operate in the 800 MHz and 1.9 GHz PCS bands. *QUALCOMM* is the developer of the CDMA air interface used in cellular systems. Compared to GSM cellular systems, CDMA requires fewer cell towers and provides up to five times the calling capacity. CDMA also provides more than 10 times the voice traffic of earlier analogue system (AMPS) and is the basis for 3G data transmission for GSM carriers.

16.6.1 CDMA principle of operation

CDMA uses unique spreading codes to spread the base band data before transmission.

> *CDMA assigns to each user a unique code sequence that is used to code data before transmission. If a receiver knows the code sequence related to a user, it is able to decode the received data.*

The codes are shared by the mobile phone and the base station. The codes are called *Pseudo-random code sequences*. All the users can access the entire spectrum allocation all of the time (Figure 16.18). A user's unique code separates the call from all other calls. The signal is transmitted in a channel, which is below noise level. The receiver then uses a correlator to despread the wanted signal, which is passed through a narrowband pass filter. Unwanted signals will not be despread and will not pass through the filter. Codes take the form of a carefully designed one/zero sequence produced at a much higher rate than that of the base band data. The rate of a spreading code is referred to as chip rate rather than bit rate.

CDMA channels can handle an unspecified number of users. The capacity of the system depends on the quality of current calls. *As more users are added, noise is added to the wideband frequency, therefore decreasing the quality of current calls*. Each user's transmission power increases the level of the frequency spectrum's "noise floor," and therefore decreases the overall call quality for all users. To eliminate the "noise floor," CDMA mobile phones and base stations use the minimum amount of power required to communicate with each other. They use precise power control to decrease users' transmission power. By decreasing a user's transmission power, the mobile phone has added battery life, increased talktime, and smaller batteries.

Types of codes used in CDMA: There are three types of codes generally used. They are Walsh code, Short PRN code, and Long PRN codes.

Walsh codes: These are orthogonal codes. The spreading on forward link is 1.2288 Mbps and on reverse link is 307.2 Kbps. 64-bit Walsh codes are used in IS 95A and IS 95B. 128-bit Walsh codes are used in CDMA2000.

Short PRN code: (16 bit) are used to identify the base station and the cell.

Long PRN code : (42-bit code) are used to identify mobile station on reverse link.

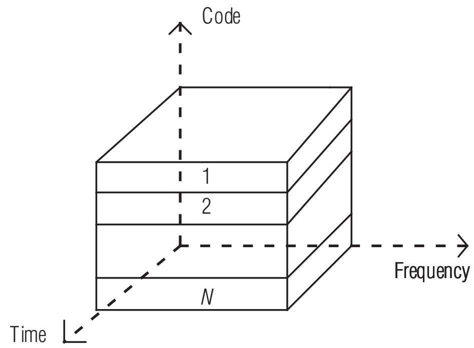


Figure 16.18 CDMA users access entire spectrum

16.6.2 Advantages and disadvantages of CDMA

Advantages

- Greatest spectrum efficiency: capacity increases about 8 to 10 times that of an analogue system and 4 to 5 times that of other digital systems, which makes it most useful in high traffic areas with a large number of users and limited spectrum.
- CDMA improves call quality by filtering out background noise, crosstalk, and interference.
- "Soft handoffs": because of the multiple diversities in use, handoffs between cells are undetected by the user.
- Simplified frequency planning: all users on a CDMA system use the same radio frequency spectrum.
  - Engineering detailed frequency plans are not necessary.
  - Frequency re-tunes for expansion are eliminated.
  - Fewer cells are required for quality coverage.
- Random Walsh codes enhance user privacy; a spread-spectrum advantage.
- Precise power control increases talktime and battery life for mobile phones.

Disadvantages

- Backwards compatibility techniques are costly.
- Currently, equipment is expensive.
- Difficult to optimize to maximize performance.
- Low traffic areas lead to inefficient use of spectrum and equipment resources.

16.6.3 Handoffs in CDMA mobile systems

The act of transferring a call of from one base station to another is termed as handoff. Handoff occurs when a call has to be handed off from one cell to another as the user moves between cells.

Hard handoff and soft handoff

In a traditional "hard" handoff, the connection to the current cell is broken and then the connection to the new cell is made. This is known as a "break-before-make" or hard handoff. Since all cells in CDMA use the same frequency, it is possible to make the connection to the new cell before leaving the current cell. This is known as a "make-before-break" or "soft handoff". Soft handoff requires less power, which reduces interference and increases capacity. The implementation of handoff is different in GSM and CDMA standards.

Differences between soft handoff and hard handoff

- Soft handoff is diverse from the traditional hard handoff process.
- Handoff is introduced and performed without the user attempting to have synchronized the traffic channel communications with the two base stations.
- With the soft handoff, a *conditional* decision is build whether to handoff or not. Depending on the changes in pilot signal strength from the two or more base stations included, a hard decision will ultimately be made to communicate with only one base station. This is normally happens after it is apparent that the signal from one base station is significantly stronger than those from the others. Temporarily, the user has simultaneous traffic channel communication with all candidate base stations.
- With hard handoff, a *definite* decision has to be made on whether to handoff or not.

Soft handoff example

A simple case of soft handoff is shown in Figure 16.19. There are only two base stations involved in the example. The same basic setup has two outcomes, one of which is shown in the left column as "scenario one," and the other in the right column as "scenario two". Scenario one shows a soft handoff from BS1 to BS2. Scenario two shows a case in which the user goes back to BS1 after a period of time in soft handoff. This might be the case when the mobile is temporarily obstructed from the line of sight of BS1. The equivalent scenario with hard handoff would be a hard handoff to BS2 and then another hard handoff back to BS1, wasting valuable network resources in the process of carrying out these unnecessary handoffs.

From Figure 16.19: (a) The left column shows a user at first having only BS1 in its active set; (b) Followed by a period time in which the pilot signal strengths of both BS1 and BS2 are strong, so the user has both BS1 and BS2 in its active set; (c) Eventually, the signal strength of BS1 declines and the signal strength of BS2 increase to the point where BS1 is removed from the active set. The column on the right shows the same situation, except that at the end BS2 is removed from the active set.

More detailed description on soft handoffs in CDMA mobile systems is given under CDMA mobile systems in Chapter 18.

16.6.4 Multipath error elimination in CDMA cellular system

One of the main advantages of CDMA systems is the capability of using signals that arrive in the receivers with different time delays. This phenomenon is called *multipath*.

FDMA and TDMA, which are narrowband systems, cannot discriminate between the multipath arrivals, and resort to equalization to mitigate the negative effects of multipath. Due to its wide bandwidth and rake receivers, CDMA uses the multipath signals and combines them to make a stronger signal at the receivers. One of the receivers (fingers) constantly search for different multipath and feeds the information to the other three receivers. Each receiver then

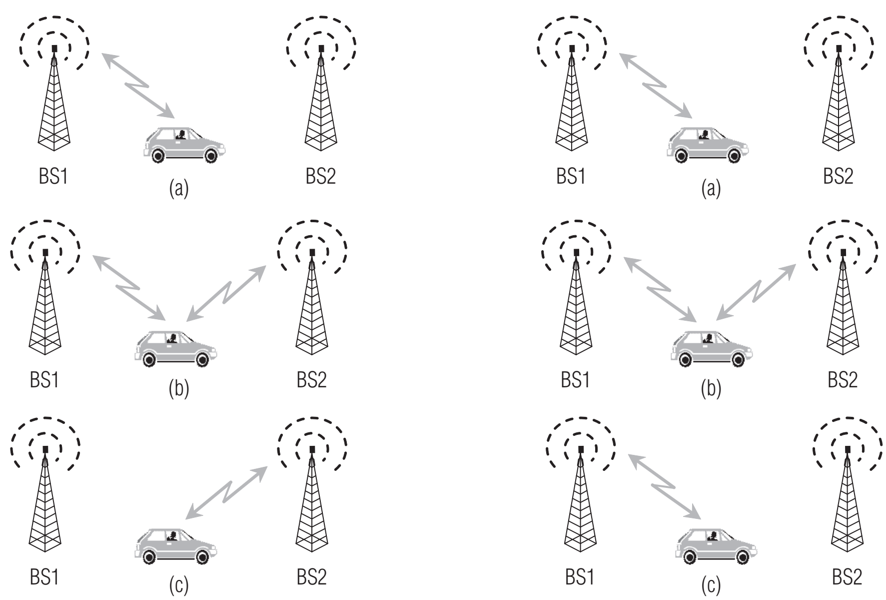


Figure 16.19 A simple soft handoff situation with two different outcomes

demodulates the signal corresponding to a strong multipath. The results are then combined together to make the signal stronger.

16.6.5 Comparison of CDMA- and TDMA/FDMA-based cellular systems

- In FDMA, all the users occupy disjoint frequency bands but are transmitted simultaneously in time.
- In TDMA, whereby all users occupy the same bandwidth but transmit in disjoint intervals of time.
- In CDMA, all the signals occupy the same bandwidth and are transmitted simultaneously in time. The different waveforms in CDMA are distinguished from one another at the receiver by the specific spreading codes they employ. It utilizes the most important use of spread-spectrum communications as a multiple accessing technique.
- In TDMA/FDMA, cell design requires more frequency planning which is tough job. Whereas in CDMA frequency planning is minimal.
- In TDMA, bandwidth available for transmission is small which leads to compromise in quality of transmission. Whereas in CDMA systems entire spectrum is used, this enhances voice quality.
- TDMA is band-limited system. CDMA is power-limited system.
- CDMA popularity is primarily due to the performance of the spread-spectrum waveforms display when transmitted over a multipath fading channel.

16.7 Near-far problem

A common difficulty encountered in CDMA wireless communication systems implemented using direct sequence spread spectrum (DSSS) technology is referred to as the near-far problem. The near-far problem occurs when two or more DSSS transmitters transmit the signals towards the same DSSS receiver as shown in Figure 16.20. In this figure, two DSSS transmitters transmit the signals towards a DSSS receiver, one transmitter being closer to the receiver than the other transmitter. The power of the two spread-spectrum signals transmitter is the same at the antenna of each DSSS transmitter.

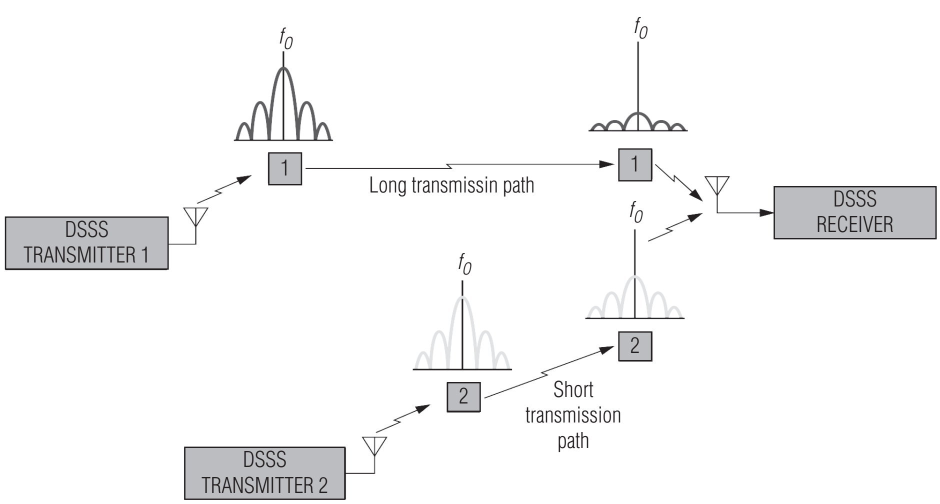


Figure 16.20 The near-far problem

However, the two spread-spectrum signals at the DSSS receiver antenna have different power levels because the paths between the two transmitters and the receiver are of different lengths. In Figure 16.20, the power of the spread-spectrum signal coming from DSSS transmitter 1 is lower than that coming from DSSS transmitter 2 because transmitter 1 is farther away from the receiver than transmitter 2. This can be a serious problem when the DSSS receiver is set to demodulate the spread-spectrum signal coming from DSSS transmitter 1, because the power level of this signal is lower than that of the spread-spectrum signal coming from DSSS transmitter 2. Since any spread-spectrum signal other than the desired one produces interference similar to that caused by noise, this results in a rather poor *S*/*N* ratio at the DSSS receiver input. Consequently, errors are likely to appear in the recovered data when the process gain of the system is not sufficient to overcome the *S*/*N* ratio deficit observed at the DSSS receiver input.

16.7.1 Near-far problem in the uplink of CDMA

The near-far problem described above is encountered in the CDMA cellular-telephone uplink networks. In such networks, each cellular telephone set is, in fact, a DSSS transmitter that transmits a spread-spectrum signal towards the base station (DSSS receiver). The power level of each spread-spectrum signal received depends on the distance that separates the corresponding cellular telephone set from the base station. The near-far problem can be greatly mitigated by having each cellular telephone set (DSSS transmitter) make its output signal power proportional to the distance that separates it from the base station. This causes the cellular telephone sets located close to the base station to reduce their output signal power. Conversely, the cellular telephone sets that are far from the base station boost their output signal power. The net effect is that the power levels of all spread-spectrum signals received at the base station are approximately the same. Power control in cellular telephone sets also offers another great benefit: it reduces the average energy consumption, thereby increasing the time during which the unit can be used before the battery needs to be recharged.

16.7.2 Near-far problem in the downlink of CDMA

The near-far problem does not happen in the downlink of CDMA cellular telephony networks. This is because the DSSS transmitters in the base station use special orthogonal codes (e.g., the Walsh codes) instead of pseudo-random code sequences to produce the various spread-spectrum signals and discriminate between the multiple users. These orthogonal codes provide perfect cross-correlation, that is, the cross-correlation values are null when the codes are compared to each other. This prevents interference between the multiple spread-spectrum signals transmitted by the base station. Note that for the special codes used in the DSSS transmitters of the base station to remain orthogonal, perfect synchronization between the codes is absolutely required. This is possible in the downlink because all codes are produced at the same location and transmitted from this same location (i.e., the base station). These severe conditions make the use of orthogonal codes in the uplink of CDMA cellular-telephone networks inapplicable.

16.8 Power control algorithm

The user capacity in direct-sequence CDMA (DS-CDMA) is limited by self-interference and is adversely affected by the near-far problem at the base station receiver. Thus, accurate power control of all the mobile radio transmitters in the system is essential and an added challenge for the transceiver design. The receiver includes an automatic gain control (AGC) loop to track the received power level, which varies because of large-scale path loss and small-scale fading. To compensate for those effects, CDMA (or IS95) employs two power control methods:

- Open-loop power control
- Closed-loop power control

Open-loop power control: The open loop method uses the power level at the mobile radio receiver (P_{Rx}) to estimate the forward-link path loss. It then specifies the transmit power (P_{Tx}) of the mobile radio as

$$P_{Tx} = -73\,\text{dBm} - P_{Rx} \tag{16.9}$$

Example problem 16.7

If the received power level is –85 dBm, find the transmitted power level.

Solution

Received power P_{Rx} = –85 dBm
Transmit power level P_{Tx} = (–73 dB) – (–85 dB) = +12 dBm

***Closed-loop power control*:** Adding a feedback signal completes the AGC loop and improves the accuracy of the open-loop method. The feedback signal is an error signal sent from the base station to the mobile radio that instructs the mobile radio to increase or decrease power by a set amount, generally 1 dB. It is sent once per power control group and is therefore updated at a rate of **800** Hz. As such, it is sufficient to support vehicle speeds up to 100 km/h. This second power control method is referred to as closed-loop power control.

16.9 Capacity of cellular CDMA with multiple cells

Spread spectrum modulation and CDMA techniques allow several users to share the radio interface; thus, the received waveform becomes the sum of ***k*** user signals and noise:

$$r(t) = \sum_{n=l}^{k} pn_n(t) A_n d_n(t) \cos(\omega t + \theta_n) + n(t) \tag{16.10}$$

where $d_n(t)$ is the data signal and $n(t)$ is the noise signal

The receiver retrieves the message signal by despreading the received signal. It does that by synchronizing its correlator to a specific spreading sequence, $pn_n(t)$ that is unique to the user and different from those of other users. As a result, the other user signals appear like noise.

The noise (N_t) seen by the correlator is the signal energy received from the k = 1 users and thermal noise, that is,

$$N_t = \sum_{n=l}^{k-l} S_n + WN_0 \tag{16.11}$$

where S_n is the received power from the n^{th} user, N_0 is the thermal noise power spectral density (PSD), and W is the channel bandwidth. If the received power from each user is assumed equal and k is large, such that $k - 1$ can be approximated by k, then

$$N_t = kS + WN_0 \tag{16.12}$$

Furthermore, the interference generally is much larger than the integrated thermal noise ($kS >> WN_0$), so that

$$N_t \approx I = kS \tag{16.13}$$

From this result, two important observations are made. First, the interference of the spread spectrum system increases linearly as the number of users is added. Second, the performance of the system suffers when any user transmits extra power, a problem known as the near-far effect.

The signal-to-noise ratio (SNR) is a key consideration in all communication systems. In digital communication systems, the SNR is characterized by a related figure of merit, the bit energy per noise density ratio (E_b/N_0). That parameter takes into account the processing gain of the communication system, a vital consideration in spread spectrum communications. The parameter normalizes the desired signal power to the bit rate R to determine the bit energy and the noise or interference signal power to the spreading bandwidth W to determine the noise spectral density. Recall that the correlator

- Despreads or integrates the desired signal to the narrow bandwidth of the original message signal (R)
- Spreads the interference to a wider bandwidth
- Leaves the uncorrelated noise unaltered

Therefore,

$$\frac{E_b}{N_0} = \frac{\mathrm{SIR}}{N/W} \approx \frac{s}{kS(R/W)} \tag{16.14}$$

Therefore amazingly, the interference from other users (i.e., self-interface) is reduced by the processing gain *(W/R)* of the system. A simple expression for the capacity of a CDMA system is developed from Equation (16.14) and is given by

$$k \approx \frac{W/R}{(E_b/N_0)_{\min}} \tag{16.15}$$

where $(E_b/N_0)_{\min}$ is the minimum value needed to achieve an acceptable level of receiver performance, typically measured as the bit error rate (BER). The expression shows that the capacity of CDMA communication systems depends heavily on the spreading factor and the receiver's performance. The capacity is tied to a flexible resource-power and is said to be *soft-limited*. In other words, if the required E_b/N_0 is lowered, the transmit signal power allocated to each user is reduced, and the number of users can be increased. In contrast, the capacity of systems that employ other multiple-access methods like FDMA and TDMA are hard-limited. That is because their capacity is fixed by system design.

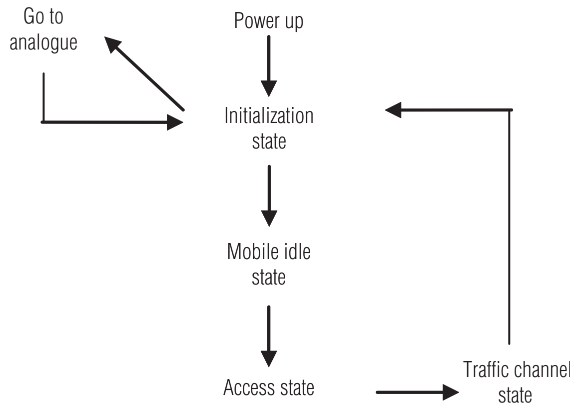


Figure 16.21 The four states of the CDMA mobile station: initialization, idle, access, and traffic channel states

16.10 Call processing in the CDMA mobile phones

Figure 16.21 shows the basic call-processing loop. After power up, the initialization state determines which system to use (whether analogue or CDMA). If it is CDMA, it goes into pilot and sync processing. Once the system is synchronized, the system goes into the mobile station idle state, where it monitors the paging channel. If a call is to be originated or the mobile is paged, the system goes into the access state.

Once a call is setup, the phone moves over to the traffic channel state, where the forward and reverse traffic channels are used to communicate voice and messaging. In the initialization state, after determining that there is a digital system, the handset monitors the paging channel. In determining the start and finish of the pilot channel, it can determine the timing of the sync channel. Once it can read the sync timing, it can further refine its timing (Figure 16.21). During the idle state, the mobile will monitor the paging channel. Various messages pertaining to setup and operation are on the paging channel.

Certain situations will trigger the mobile to drop out of the traffic state (drop the call on purpose):

- *Mobile ACK failure:* Certain messages require an ACK (Acknowledge signal); generally, a mobile will retransmit the message after 400 ms, but if no ACK comes after three tries, the mobile drops the call.
- *Base station ACK failure:* This is similar to the mobile ACK failure, but it is not standardized.
- *Mobile fade timer:* The timer is set to 5 s after receiving two consecutive good frames. If the timer gets to zero, the call drops.
- *Mobile bad frames:* If there are 12 consecutive bad frames, the mobile drops the call.
- *Base station bad frames:* This is similar to mobile bad frames, but not standardized (i.e., manufacturers can implement this, however, they choose).

16.11 Space division multiple access

In addition to frequency, time, and code domains, the spatial dimension can also be used for multiplexing of different data streams by transmitting the data streams over different, non-overlapping transmission channels. The use of space division multiplexing for multiple access is termed SDMA.

> *SDMA enables users to share simultaneously the same bandwidth in different geographical locations. SDMA solves capacity problem of wireless communication systems by* exploitation of the spatial dimension which makes it possible to identify the individual users, even when they are in the same time/frequency/code domains. *SDMA can be achieved using beam forming or sectorization.*

Omnidirectional antenna–based traditional cellular networks

In traditional mobile cellular network systems, the base stations have no information on the position of the mobile units within the cell, and base stations radiate the signal in all directions within the cell in order to provide radio coverage. These results in wasting power on retransmissions when there are no mobile units to reach, in addition to causing interference for adjacent cells using the same frequency, so called co-channel cells. In the same way, in reception, the antenna receives signals coming from all directions including noise and interference. These considerations have led to the development of the SDMA technique, which is based on deriving and exploiting information on the spatial position of mobile terminals.

Adaptive antenna or SDMA-based cellular network

By using adaptive antennas arrays, sometimes called smart antennas in mobile radio systems, signals can be received and sent only from and into a limited angular range, following the directional nature of multipath. This improves coverage or link quality in noise-limited situations and enhances capacity in interference-limited situations. The concept of SDMA is shown in Figure 16.22. Each user exploiting a single-transmitter-antenna-aided mobile station simultaneously communicates with the base station equipped with an array of receiver antennas.

16.11.1 Antenna arrays

By combining the outputs of the individual antennas in an array, a single effective antenna can be created with gain and directional characteristics that are very different from those of the individual elements comprising the array. For example, consider a row of, *m,* simple and

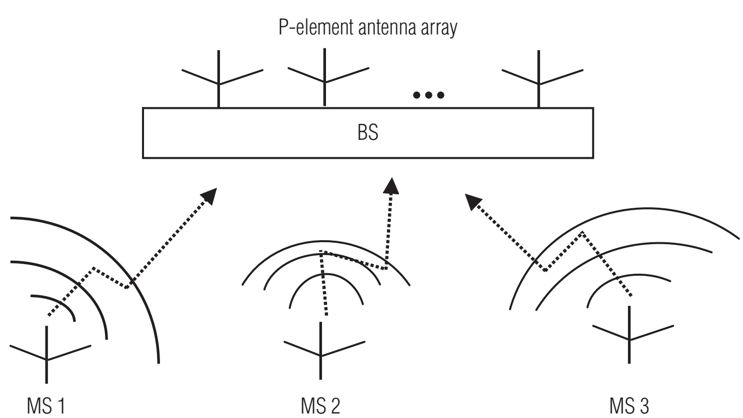


Figure 16.22 SDMA concept, employing a P-element receiver antenna array for supporting three mobile users

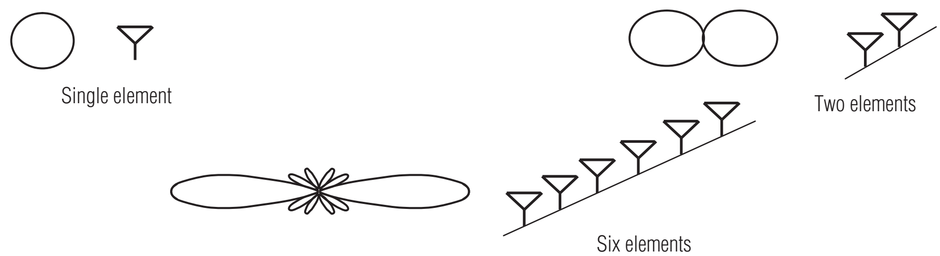


Figure 16.23 Antenna patterns for arrays of different number of elements

identical, antenna elements (called a linear array) with elements spaced one-half wavelength apart. Suppose that one simply adds the outputs of all the elements. Signals arriving at the array from broadside, or perpendicular to the axis of the array, arrive simultaneously at each element of the array, and their sum will therefore be m times as large (m^2 times as powerful) as the signal received by a single antenna. Simply adding the outputs of the m element array therefore results in an amplitude gain of m for signals arriving from broadside.

Figure 16.23 depicts the effective antenna pattern resulting from the above strategy for linear arrays consisting of a single antenna, two antennas, and six antennas (omnidirectional element patterns and half-wavelength spacing in all cases). The radius of the pattern in Figure 16.23 is proportional to the gain or strength of the signal at the output of the array. The increased gain in the broadside direction and the increases in gain with the number of array elements are evident in the Figure 16.23. It is also evident that there are certain directions in which the effective antenna has reduced sensitivity, or nulls.

Classical phased-array antenna systems have gain patterns or beam patterns that are shown in the figure, except that the direction of maximum gain might not be the broadside direction. More advanced systems have patterns that are optimized to enhance a particular user's signal while simultaneously rejecting interferers. Another differentiator of antenna array systems is whether or not the beam patterns can change with time. Antenna arrays can be built with combining strategies that are fixed. From an operational standpoint, this sort of array is no different from a conventional antenna with the same directional sensitivity. Alternatively, arrays can be equipped with combining hardware and software that make it possible for the pattern to be changed over time and adapted to the current operational scenario. Because radio reception and transmission are reciprocal, directive transmission with gain is also possible from an array. Any directivity pattern achievable for reception is also achievable for transmission.

SDMA controls the radiated energy for each user in space (Figure 16.24) by using adaptive antenna array patterns. These different areas covered by the antenna beam may be served by the same frequency (in a TDMA or CDMA system) or different frequencies (in an FDMA system). Sectorized antennas may be thought of as a primitive application of SDMA. Figure 16.25 shows spatially filtered base station antenna serving different users by using spot beams.

16.11.2 Advantages of SDMA technique

The advantages of SDMA system include the following:

1. ***Range extension:*** The coverage area of the antenna array is greater than that of any single element as a result of the gain provided by the array. When a system is constructed using SDMA, the number of cells required to cover a given area can be substantially reduced.

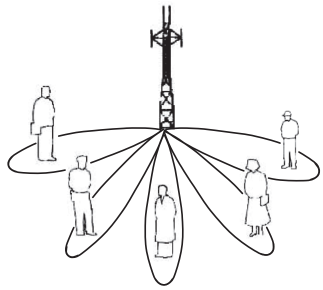


Figure 16.24 Multiple users belonging to the same cell use the same channel

A 10-element array offers a gain of 10, which typically doubles the range of the cell and thereby quadruples the coverage area.

2. ***Interference suppression:*** Interference from other systems and from users in other cells is significantly reduced by exploiting the desired user's unique channel impulse responses (CIRs). In "noisy" areas where range is limited by interference, spatially selective transmission and reception result in range extension.
3. ***Multipath effect elimination***: The copies of the desired signal that have arrived at the antenna after bouncing from objects between the signal source and the antenna can often be mitigated. In certain cases, the multipath can actually be used to reinforce the desired signal.
4. ***Capacity increase:*** Capacity increase can be done in two ways:
   - Channel reuse patterns in cellular systems can be significantly tighter because the average interference resulting from co-channel signals in other cells is markedly reduced (e.g., moving from a 7-cell to a 4-cell reuse pattern nearly doubles capacity).
   - Separate spatial channels can be created in each cell on the same conventional channel. In other words, intracellular reuse of conventional channels is possible **(**Figure 16.25).

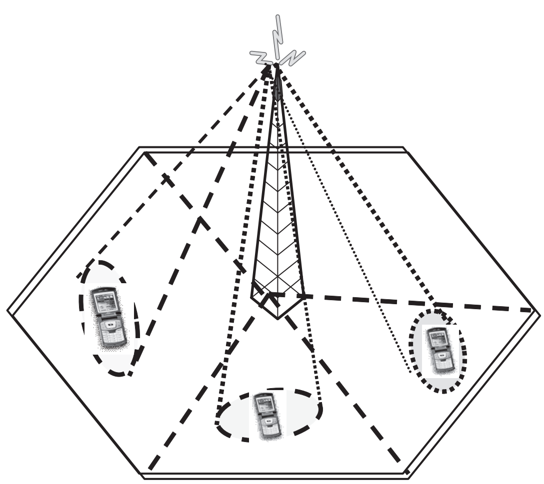


Figure 16.25 Intra-cell SDMA

5. ***Compatibility***: SDMA is compatible with most of the existing modulation schemes, carrier frequencies and other specifications. Furthermore, it can be readily implemented using various array geometries and antenna types.

16.11.3 Capacity of space division multiple access

For interference limited CDMA operating in an AWGN channel, with perfect power control with no interference from adjacent cells and with omnidirectional antennas used at the base stations, the average BER for a user can be found from the Gaussian approximation as

$$P_{\mathrm{b}} = Q\left(\sqrt{\frac{3N}{K-1}}\right) \tag{16.16}$$

where K is the number of users in a cell and N is the spreading factor. $Q(x)$ is the standard Q-function.

The directive antennas can improve the reverse link in a single-cell CDMA system. The omnidirectional receiver antenna will detect signals from all users in the system, and thus will receive the greatest amount of noise. The sectored antenna will divide the received noise into a smaller value and will increase the number of users in the CDMA system A spot beam for each user, and it is this implementation which is the most powerful form of SDMA. An ideal adaptive antenna is able to form a beam for each user in the cell of interest, and the base station tracks each user in the cell as it moves. Assume that a beam pattern, is formed such that the pattern has maximum gain in the direction of the desired user. Such a directive pattern can be formed at the base station using an N-element adaptive array antenna.

16.11.4 Comparison of SDMA, TDMA, FDMA, and CDMA

The performances of the various multiple access techniques are compared and shown in Table 16.2.

16.12 Orthogonal frequency division multiplexing

OFDM is an important technology for all next generation wireless systems (e.g., 4G system, HiperLAN) as it offers numerous advantages over other existing technologies, such as robust performance over *multipath fading channels* and the ability to achieve *high spectral efficiency*.

> *OFDM is a form of multi-carrier modulation where* high-rate data stream is split into a number of lower rate streams that are transmitted simultaneously over a number of subcarriers (Figure 16.26).

An OFDM signal consists of a sum of subcarriers that are modulated by using *phase shift keying* (PSK) *or quadrature amplitude modulation* (QAM). If d_i are the complex QAM symbol, N_s is the number of subcarriers, T the symbol duration and $f_i = f_0 + i/T$ the carrier frequency, then one OFDM symbol starting at $t = t_s$ *can* be written as:

$$\begin{aligned} s(t) &= \mathrm{Re}\left\{\sum_{i=0}^{N_s-1} d_i \exp(j2\pi f_i(t-t_s)\right\} \quad t_s \le t \le t_s + T \\ s(t) &= 0, \quad t < t_s \wedge t > t_s + T \end{aligned} \tag{16.17}$$

Table 16.2 Comparison of SDMA, TDMA, FDMA, and CDMA

Approach	SDMA	TDMA	FDMA	CDMA
Idea	Segment space into cells/sectors	Segment sending time into disjoint time slots, demand driven or fixed patterns	Segment the frequency band into disjoint sub-bands	Spread the spectrum using orthogonal codes
Terminals	Only one terminal can be active in one cell/one sector	All the terminals are active for short periods of time on the same frequency	Every terminal has its own frequency, uninterrupted	All terminals can be active at the same place at the same moment, uninterrupted
Signal separation	Cell structure directed antennas	Synchronization in the time domain	Filtering in the frequency domain	Code plus special receivers
Advantages	Very simple, increases capacity per km^2	Established fully digital, very flexible	Simple established, robust	Flexible, less planning needed, soft handover
Disadvantages	Inflexible antennas typically fixed	Guard space needed (multi-path propagation), synchronization difficult	Inflexible frequencies are as scarce resource	Complex receivers, needs more complicated power control for senders
Comment	Only combination with TDMA, FDMA, or CDMA useful	Standard in fixed networks, together with FDMA/SDMA used in many mobile networks	Typically combined with TDMA (frequency hopping patterns) and SDMA (frequency reuse)	Used in many 3G systems higher complexity, lowered expectations; integrated with TDMA/FDMA

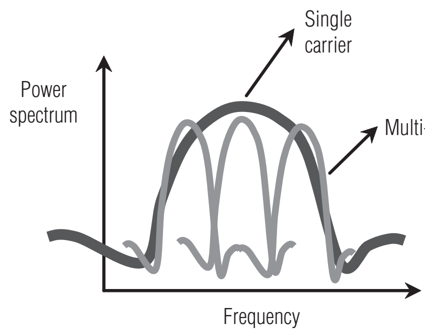


Figure 16.26 Power spectra of single- and multi-carrier systems

By using the equivalent complex notation, Equation (16.17) can be written as

$$s(t) = \sum_{i=0}^{N_S-1} d_i \exp(j2\pi f_i(t-t_s)) \quad t_s \le t \le t_s + T$$

$$s(t) = 0, \quad t < t_s \wedge t > t_s + T$$

In this representation, the real and imaginary parts correspond to the in-phase and quadrature parts of the OFDM signal, which have to be multiplied by a cosine and sine of the desired carrier frequency to produce the final OFDM signal. OFDM modulator (Figure 16.27) and its equivalent circuit is shown for transmission of QAM symbols on parallel subcarriers overlapping.

The biggest advantage of the OFDM technique is the mutual orthogonality of its carriers; this provides a high spectral efficiency. This is possible because there is no guard band, and carriers can be packed very close together.

We will discuss key features of OFDM technology in this section.

16.12.1 Basic principle of operation

The concept of parallel transmission of data is the basis for OFDM technique. The basic idea of OFDM is to transmit N symbols in parallel over N different subcarriers while enlarging the symbol duration N times.

We can describe OFDM principle through a rotation of the time frequency plane as shown in Figure 16.27. In a single-carrier system, each symbol occupies the full bandwidth. In a multi-carrier system the symbol duration is enlarged N times and simultaneously the bandwidth consumption of each symbol is reduced by the same factor N.

Figure 16.28 illustrates the difference between a single-carrier and a multi-carrier system through a rotation of the time-frequency plane. The transmitted data symbols are denoted $d[1]$ … $d[8]$. The multi-carrier system uses $N = 8$ subcarriers.

Figure 16.29 illustrates the subcarrier frequency-spectra in an OFDM system. The subcarrier has bandwidth cf. The centre frequency of subcarrier q is denoted by f_q.

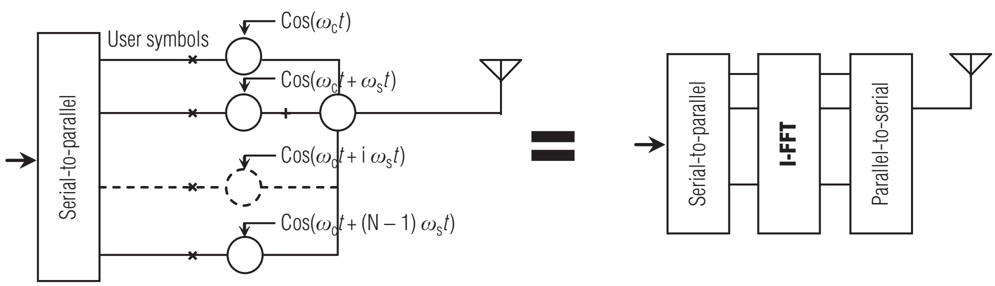


Figure 16.27 OFDM modulator

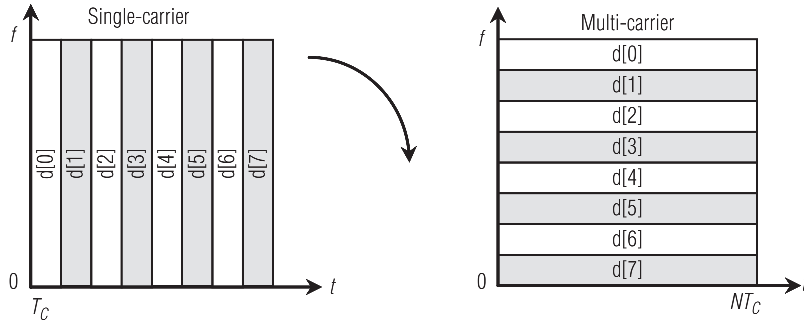


Figure 16.28 Single-carrier and multi-carrier systems

The overall data-rate and bandwidth consumption is kept constant through parallel transmission over N independent subcarriers. The subcarrier spectra overlap, as depicted in Figure 16.29. However, if the centre frequency of each subcarrier q is chosen as

$$f_q = q/(NT_C) \tag{16.18}$$

for $q \in \{0, \ldots ,N - 1\}$ the subcarriers are orthogonal despite their overlapping spectra. OFDM is a special case of a multi-carrier scheme with overlapping but orthogonal subcarriers.

Figure 16.30 shows all operations that are necessary for OFDM. Each subcarrier is modulated by a symbol (from a binary phase shift keying (BPSK) alphabet in this example) and the resulting signals are summed up. These operations are equivalent to an inverse discrete Fourier transform (DFT). The inverse DFT can be efficiently implemented by means of the inverse fast Fourier transform. The existence of such an efficient algorithm for the actual implementation is one major reason for the widespread application of OFDM.

OFDM needs the following processing steps (Figure 16.29):

- First, the subcarriers are multiplied by the individual data symbols, and then the resulting signals are added together.
- A copy of the signal tail is inserted at the beginning of each OFDM symbol (Figure 16.31).

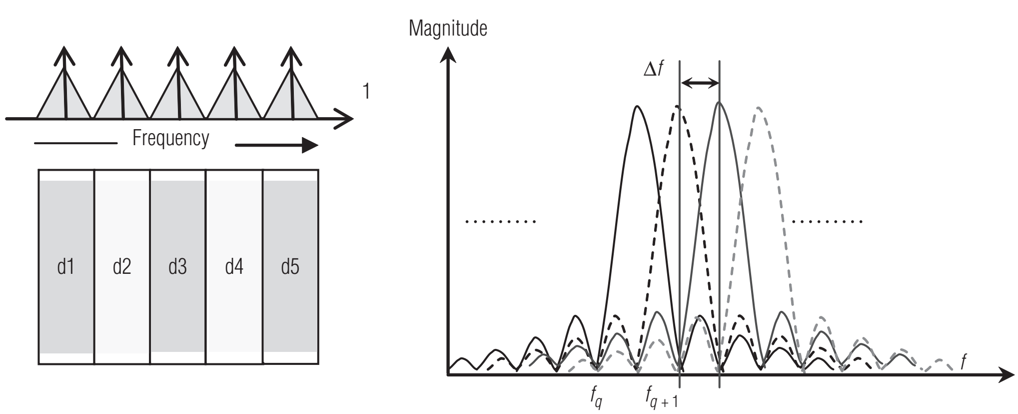


Figure 16.29 Subcarrier frequency-spectra in an OFDM system

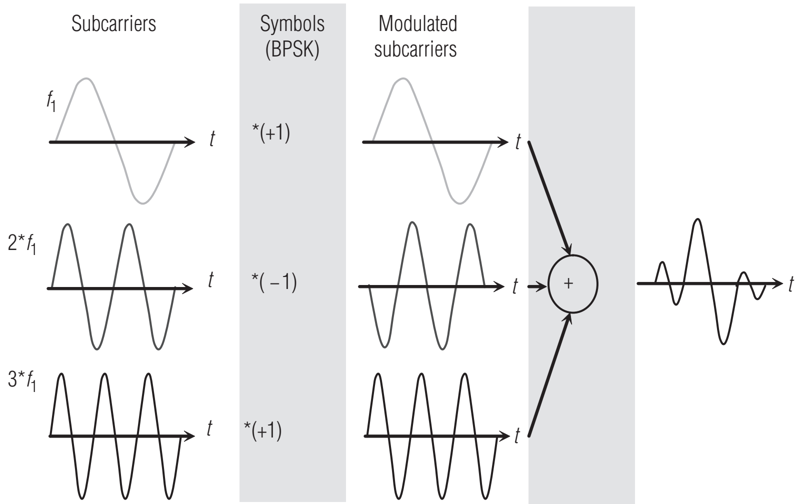


Figure 16.30 OFDM transmission

16.12.2 Cyclic prefix

In a fading multipath environment, inter-carrier interference (ICI) and ISI will be introduced. To solve this problem, every OFDM block is extended by a guard interval T_G, $T_G > T_d$

where T_d is the maximum multipath delay spread.

The most common guard interval is cyclic prefix. In order to completely remove the inter-symbol interference, a cyclic prefix is inserted in front of every OFDM symbol.

The cyclic prefix is a copy of the OFDM symbol tail.

Cyclic prefix refers to the prefixing of a symbol with a repetition of the end. Although the receiver is typically configured to discard the cyclic prefix samples, the cyclic prefix serves two purposes.

- As a guard interval, it eliminates the ISI from the previous symbol.
- As a repetition of the end of the symbol, it allows the linear convolution of a frequency-selective multipath channel to be modelled as circular convolution, which in turn may be transformed to the frequency domain using a discrete Fourier transform. This approach allows for simple frequency-domain processing, such as channel estimation and equalization.

We illustrate this operation in Figure 16.31. To eliminate ISI completely, the length of the cyclic prefix "G" must be longer than the essential support of the channel impulse response L,

$$G \geq L \tag{16.19}$$

The length of the OFDM symbol in chips after insertion of the cyclic prefix is denoted by

$$P = N + G$$

where N is number of symbols and G is length of cyclic prefix.

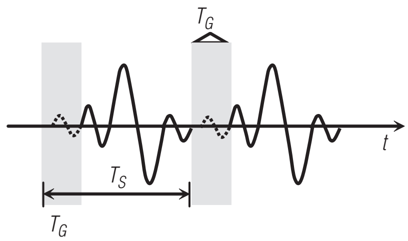


Figure 16.31 Cyclic prefix insertion

16.12.3 Multipath and ISI reduction in OFDM

Instead of using a single carrier for information transmission, the channel bandwidth "*B*" is divided into a number of equal bandwidth (Δf) creating *N* sub-channels ($N\Delta f/N$).

$$\text{Subcarrier interval } \Delta f = B/N \text{ where } B = \Delta f N$$

Consider a data stream operating at *R* bps and an available bandwidth of $N\Delta f$ centred at carrier frequency f_c. The entire bandwidth could be used to transmit a data stream, in which case the bit duration would be $1/R$. By splitting the data stream into *N* sub-streams using a serial-to-parallel converter, each sub-stream has a data rate of R/N and is transmitted on a separate subcarrier, with spacing between adjacent subcarriers of Δf (Figure 16.32). The bit duration is N/R.

The advantage of OFDM is that on a multiple channel the multipath is reduced relative to the symbol interval by a ratio of $1/N$ and thus imposes less distortion in each modulated symbol. OFDM overcomes ISI in a multipath environment. ISI has a greater impact at higher data rates because the distance between bits and symbols is smaller. With OFDM, the data rate is reduced by a factor of *N*, which increases the symbol duration by a factor of *N*. Thus, if the symbol duration is T_s for the source stream, the duration of OFDM signals is NT_s. This significantly reduces the effect of ISI.

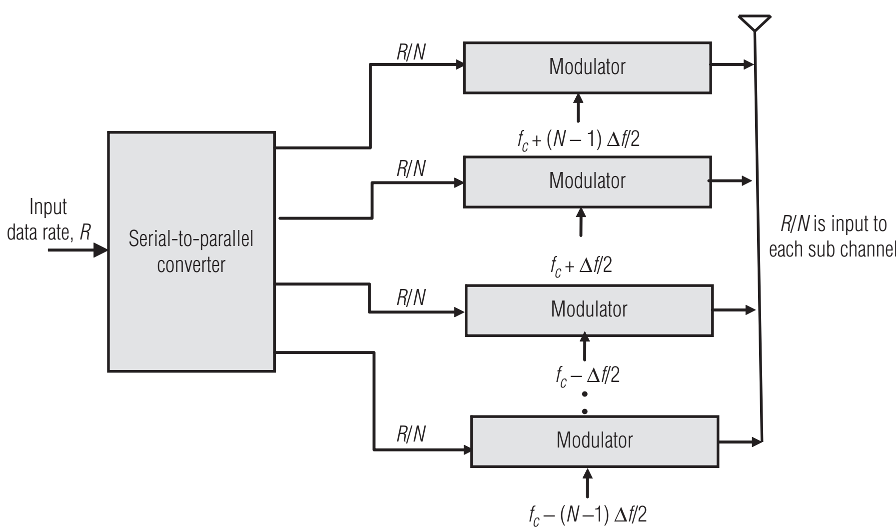


Figure 16.32 Orthogonal frequencies division multiplexing (OFDM)

16.12.4 Advantages and disadvantages of OFDM

Advantages

OFDM has several advantages over other widely used wireless access techniques such as TDMA, FDMA, and CDMA.

1. In OFDM, multiple symbols can be transmitted in parallel, while maintaining a high spectral efficiency because the radio channel is divided into many narrowband, low-rate, frequency-non-selective sub-channels or subcarriers
   i. Each subcarrier may deliver information for a different user, resulting in a simple multiple-access scheme known as orthogonal frequency-division multiple access (OFDMA). This enables different media such as video, graphics, speech, text, or other data to be transmitted within the same radio link, depending on the specific types of services and their quality-of-service (QoS) requirements.
   ii. Furthermore, in OFDM systems different modulation schemes can be employed for different subcarriers or even for different users. For example, the users close to the base station may have a relatively good channel quality, thus they can use high-order modulation schemes to increase their data rates. By contrast, for those users who are far from the base station or are serviced in highly loaded urban areas, where the subcarriers' quality is expected to be poor, low-order modulation schemes can be invoked.

Disadvantages

1. One problem is the associated increased peak-to-average power ratio (PAPR) in comparison with single-carrier systems, requiring a large linear range for the OFDM transmitter's output amplifier.
2. In addition, OFDM is sensitive to carrier frequency offset, resulting in ICI.
3. Not ideal for mobile phones, but fine for mobile computers with bigger batteries that are not sending data continuously.
4. Transmitter and receiver must be linear to preserve shape.

16.13 Packet radio

Early radio systems transmitted analogue signals. Today most radio systems transmit digital signals composed of binary bits, where the bits are obtained directly from a data signal or by digitizing an analogue signal. A *packet radio* is a digital radio that can transmit a continuous bit stream or it can group the bits into packets and is characterized by bursty transmissions. The radio is idle except when it transmits a packet. Packet radio can be simply defined as technology which enables the user to transmit packet data (digital data) via radio (wireless) interface.

The first network based on packet radio, ALOHANET, was developed at the University of Hawaii in 1971. This network enabled computer sites at seven campuses spread out over four islands to communicate with a central computer on Oahu via radio transmission. The network architecture used a star topology with the central computer at its hub. Any two computers could establish a bi-directional communications link between them by going through the central hub. ALOHANET incorporated the first set of protocols for channel access and routing in packet radio systems, and many of the underlying principles in these protocols are still in use today. The U.S. military was extremely interested in the combination of packet data and broadcast radio inherent to ALOHANET.

Throughout the 1970s and early 1980s the Defense Advanced Research Projects Agency (DARPA) invested significant resources to develop networks using packet radios for tactical communications in the battlefield. The nodes in these ad hoc wireless networks had the ability to self-configure (or reconfigure) into a network without the aid of any established infrastructure. DARPA's investment in ad hoc networks peaked in the mid-1980s, but the resulting networks fell far short of expectations in terms of speed and performance. These networks continue to be developed for military use. Packet radio networks also found commercial application in supporting wide-area wireless data services.

These services, first introduced in the early 1990s, enable wireless data access (including email, file transfer, and web browsing) at fairly low speeds, on the order of 20 Kbps. A strong market for these wide-area wireless data services never really materialized, due mainly to their low-data rates, high cost, and lack of "killer applications." These services mostly disappeared in the 1990s, supplanted by the wireless data capabilities of cellular telephones and wireless local area networks (LANs).

Multiple access techniques are primarily for continuous-time applications like voice and video, where a dedicated channel facilitates good performance. However, most data users do not require continuous transmission: data is generated at random time instances, so dedicated channel assignment can be extremely inefficient. Moreover, most systems have many more total users (active plus idle users) than can be accommodated simultaneously, so at any given time channels can only be allocated to users that need them. Random access strategies are used in such systems to efficiently assign channels to the active users. All random access techniques are based on the premise of packetized data or *packet radio*. In packet radio, user data is collected into packets of N bits, and once a packet is formed it is transmitted over the channel. Assuming a fixed channel data rate of R bps, the transmission time of a packet is

$$\tau = \frac{N}{R}\cdot$$

The transmission rate R is assumed to require the entire signal bandwidth, and all users transmit their packets over this bandwidth, with no additional coding that would allow separation of simultaneously transmitted packets. Thus, if packets from different users overlap in time a *collision* occurs, in which case neither packet may be decoded successfully.

Performance of random access techniques is typically characterized by the *throughput* of the system. The throughput, which is unit less, is defined as the ratio of the average number of packets successfully transmitted in any given time interval divided by the number of attempted transmissions in that interval. The throughput thus equals the offered load multiplied by the probability of successful packet reception (successful packet reception), where this probability is a function of the random access protocol in use as well as the channel characteristics, which can cause packet errors in the absence of collisions. The goal of a random access method is to make the throughput as large as possible in order to fully utilize the underlying link rates.

16.14 Packet radio protocols

ALOHA, slotted ALOHA, and carrier sense multiple access (CSMA) protocols are the packet radio protocols. Random access techniques were pioneered by Abramson with the ALOHA protocol, where data is packetized and users send packets whenever they have data to send. ALOHA is very inefficient due to collisions between users, which lead to very low throughput. The throughput can be doubled by slotting time and synchronizing the users, but even then, collisions lead to relatively low-throughput values. Modifications of ALOHA protocols are to avoid collisions and thereby increase the throughput include carrier sensing, collision detection, and collision avoidance. Long bursts of packets can be scheduled to avoid collisions, but this typically takes additional overhead. In this section, we will describe the various techniques for random access, their performance, and their comparison.

16.14.1 Pure ALOHA

Pure or unslotted ALOHA is the simplest protocol in which the user transmits data packets whenever it has data.

Whenever two frames from different users try to occupy the channel at the same time, there will be a collision and colliding frames will be destroyed. If the first bit of a new frame overlaps with just the last bit of a frame almost finished, both frames will be totally destroyed and both will have to be retransmitted later. This is shown in the Figure 16.33. Here all frames assumed to be same length because the throughput of ALOHA systems is maximized by having a uniform frame size rather than by allowing variable length frames. A frame will not suffer a collision if no other frames are sent within one frame time of its start.

Let t be the time required to send a frame. If any other user has generated a frame (Frame 1) between time t_0 and $t_0 + t$, the end of that frame will collide with the beginning of the already transmitted frame (Frame 3). Similarly, any other frame (Frame 2) started between $t_0 + t$ and $t_0 + 2t$ will bump into the end of the transmitted frame (Frame 3). The vulnerable period which is defined as the time interval during which the packets are susceptible to collisions with transmissions from other users. Each transmitted frame has a vulnerable period equal to two frame times. There will be no collision if and only if no other frame is transmitted within the vulnerable period (Figure 16.34).

The normalized throughput S of the pure ALOHA protocol in terms of the offered load is given by

$$S = Ge^{-2G} \tag{16.20}$$

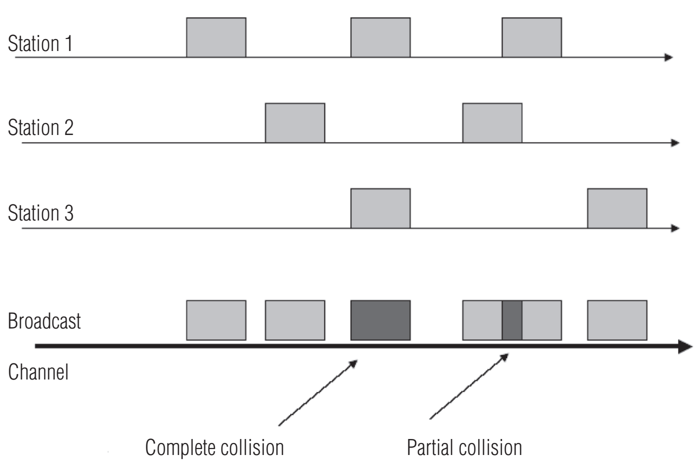


Figure 16.33 Collision in pure ALOHA

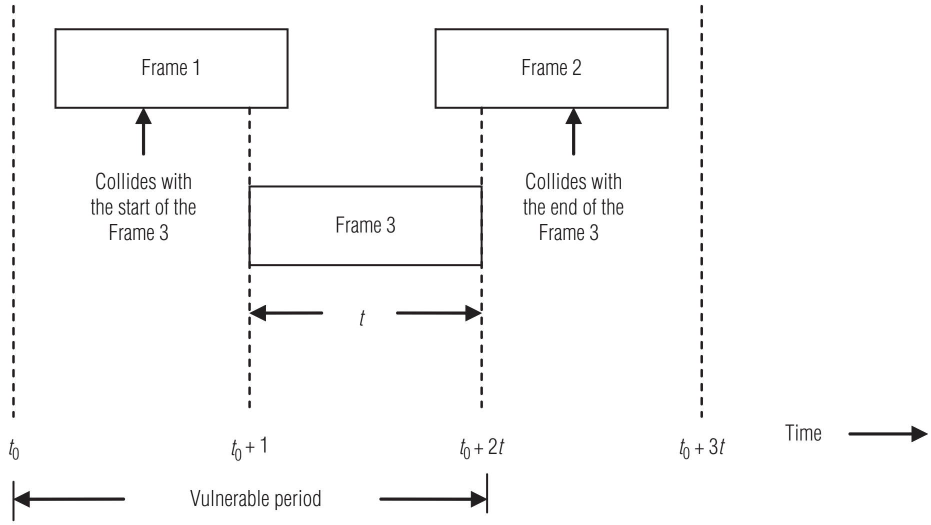


Figure 16.34 Vulnerable periods in pure ALOHA

From Equation (16.20), it should be noted that the maximum throughput occurs at traffic load G = 50 per cent and is $S = 1/2e$. This is about 0.184. Thus, the best channel utilization with the pure ALOHA protocol is only 18.4 per cent. At smaller offered load, channel capacity is under used and at higher offered load, too many collisions occur reducing the throughput.

Throughput of pure ALOHA

To evaluate the performance of pure ALOHA protocol technique in terms of throughput, it is assumed that all data packets transmitted by all subscribers in the system have a constant and a fixed packet length, L. Let the average transmission rate or mean arrival rate be λ (packets/s). This means that there are, on an average, λ data packets transmitted per second by all the subscribers. This type of situation is commonly modelled as a Poisson distribution process. The probability that there are k arrivals in the time period [0, t] is given by

Probability for k arrivals in the time period [0, t],

$$P_s = [(\lambda\, t)k / k!]e^{-\lambda t} \tag{16.21}$$

The Poisson distribution process has a memory less property. It implies that the probability that there are k arrivals in the period [0, t] is the same as the probability that there are k arrivals in the period [s, $t + s$] for some arbitrary value of s.

If two packets from two subscribers collide (overlap in time), it is assumed that the information in both the packets is lost. It is desired to determine the throughput S of this common channel (throughput is defined as the average number of packets that can be transmitted per time slot). If there is only one subscriber transmitting data at any time then the maximum throughput would be unity. However, in the case of a large number of subscribers, the probability that two or more packets will collide must be considered. It is obvious that for a packet transmitted at time t_0, any other packet transmitted in the interval $[t_0 - T, t_0 + T)$, where T is the packet duration, will cause a collision. With a Poisson distribution process for packet arrival times, if a packet is being transmitted, then the probability that no additional packets arrive during the period $[t_0 - T, t_0 + T]$, or $[t = 2T]$ is given by substituting $k = 0$ and $t = 2T$ in Equation (16.20). Probability that a packet is successfully received in time $2T$, $P_s = [(\lambda t)0 / 0!]\ e^{-2\lambda T}$

Or, probability that a packet is successfully received in time $2T$,

$$P_s = e^{-2\lambda T} \tag{16.22}$$

Consequently, the throughput S of an ALOHA system is given by the product of the packet arrival rate λ and the probability that a packet is successfully received in time $2T$, P_s, that is,

Throughput of an ALOHA system

$$S\,(\text{ALOHA}) = \lambda\, e^{-2\lambda T} \tag{16.23}$$

Normalizing this equation to packets received per packet time T, the normalized throughput, S (ALOHA) is given as Normalized throughput of an ALOHA system,

$$S\,(\text{ALOHA}) = \lambda\, Te^{-2\lambda T} \tag{16.24}$$

or,

$$S\,(\text{ALOHA}) = Ge^{-2G} \tag{16.25}$$

where $G = \lambda T$ is the traffic occupancy or the normalized channel traffic (measured in Erlangs)

The parameter G is defined as the normalized loading per packet period, that is, the average number of packets per slot time T, due to arriving and buffered packets. It is a relative measure of the channel utilization. If $G > 1$ then the packets generated by the subscribers exceed the maximum transmission rate of the channel. It implies that in order to obtain a reasonable acceptable throughput, the rate at which new packets are generated must lie within $0 < G < 1$. Under normal loading conditions, the throughput T is the same as the total offered load, L. The total offered load L is the sum of the newly generated packets and the retransmitted packets that suffered collisions in previous transmissions. The normalized throughput is always less than or equal to unity and may be considered as the fraction of time (or fraction of an Erlang) a channel is utilized. The normalized throughput is given as the total offered load times the probability of successful transmission. The peak normalized throughput S_0 *peak* occurs at $G = 0.5$.

Therefore, *peak throughput* $= 0.5\ e^{-2(0.5)} = 0.184$ packets per packet time

This means that with an ALOHA random-access channel, the maximum throughput is approximately 18.4 per cent or less than 19 per cent of the full channel capacity. In practice, the throughput is maintained at a much smaller value so as to ensure stability of the random-access approach.

Example problem 16.8

Calculate the maximum throughput of a pure ALOHA network with a large number of subscribers and a transmission rate of 2 Mbps. What is the throughput of the ALOHA network if only one subscriber is active?

Solution

For a large number of subscribers, each using a transmission rate of 1 Mbps to access a base station using ALOHA protocol, the maximum data rate that successfully passes through to the base station is given by

Maximum throughput of a pure ALOHA system, S_{max} (ALOHA),
S_{max} (ALOHA) = (Ge^{-2G}) × transmission rate, for $G = 0.5$
Therefore, S_{max} (ALOHA) = (0.5 $e^{-2(0.5)}$) × 2 Mbps
= 0.184 × 2 Mbps
= **368 Kbps**

If only one subscriber is active in the ALOHA network, there cannot be any collision and the only subscriber can transmit all the time. Therefore,

Throughput of the ALOHA network at 2 Mbps rate = **2 Mbps**

16.14.2 Slotted ALOHA

In 1972, Roberts published a method for doubling the capacity of an ALOHA system. In the slotted-ALOHA system, the transmission time is divided into time slot, each slot corresponding to one frame transmission time. Users are synchronized to the time slots and the user can transmit the frame only at the beginning of a slot. This reduces the vulnerable period from $2t$ to t ,that is, one frame duration and improves efficiency by reducing the probability of collision as shown in Figure 16.35.

The normalized throughput S of the slotted-ALOHA protocol in terms of the offered load is given by

$$S = Ge^{-G} \tag{16.26}$$

The maximum throughput occurs at traffic load $G = 1$ and equal to $1/e$. This is about 0.368. Thus, the best channel utilization with the slotted ALOHA protocol is only 36.8 per cent, which is twice than that of pure ALOHA protocol.

Throughput of slotted ALOHA

With slotted ALOHA, a collision occurs only if the two subscribers transmit during the same time slot T. Since the sequence of the newly generated and retransmitted packets follows Poisson distribution process, the probability of successful transmission is given by

$$P_sS = e^{-\lambda T} \tag{16.27}$$

Consequently, the throughput SS of a slotted ALOHA system is given by the product of the packet arrival rate and the probability that a packet is successfully received; that is, *throughput of a slotted ALOHA system*,

$$S = \lambda e^{-\lambda T} \tag{16.28}$$

Normalizing this equation to packets received per packet time T, and then the normalized throughput is given by normalized throughput of a slotted ALOHA system,

$$\mathrm{S} = \lambda\ T\ e^{-\lambda T}\ T \tag{16.29}$$

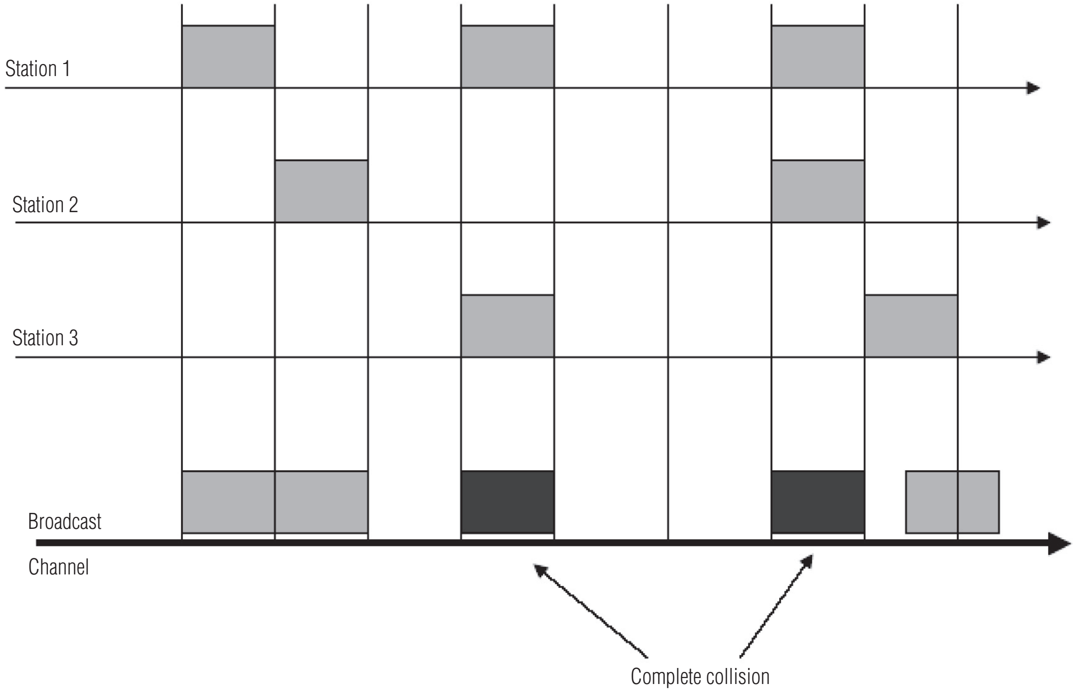


Figure 16.35 Probability of collision in slotted ALOHA

Using the definition of the normalized offered load $G = \lambda T$,

$$S\text{ (Slotted ALOHA)} = Ge^{-G} \tag{16.30}$$

The probability that no other data packets will be generated during the vulnerable period is e^{-G} . The peak throughput with slotted ALOHA occurs at $G = 1$, Therefore, S(Slotted ALOHA) = 1 $e^{-1} = 1/e \approx 0.368$ packets per packet time. That is, with a slotted ALOHA random-access protocol, the maximum throughput is 36.8 per cent of the full channel capacity. This is double that of pure or unslotted ALOHA. In other words, the slotted ALOHA protocol provides a maximum channel utilization of 0.368 Erlangs, double that of pure ALOHA protocol. Thus, the maximum throughputs of pure ALOHA and slotted ALOHA protocols are equal to 0.184 and 0.368, respectively.

16.14.3 Carrier sense multiple access protocols

The name carrier sense terms from the fact that the existence of the carrier wave on the communication channel is sensed by the user. In CSMA, a user having data to transmit first listens to the channel to check whether another transmission is in progress or not. The user starts sending only when the channel is idle, that is, there is no carrier.

There are several *classes* of CSMA protocols:

1-persistent CSMA

In 1-persistent CSMA, when a station has data to send, it first listens to the channel. If the channel is busy, the station waits until it becomes idle. When the station finds an idle channel, it transmits a packet. If a collision occurs, the station waits a random amount of time and then resumes listening to the channel. When the channel is again found to be idle, the packet is retransmitted immediately. The protocol is called 1-persistent because the station transmits with a probability of 1 when it finds the channel idle.

Non-persistent CSMA

In this type of CSMA, before transmitting the packet, a station senses the channel. If the channel is idle, the station begins its transmission. If the channel is busy, the station does not continually sense the channel. Instead, it waits a random period of time and then repeats the algorithm. This is popular for wireless LAN applications, where the packet transmission interval is much greater than the propagation delay to the farthermost user. This algorithm leads to better channel utilization but longer delays than 1-persistent CSMA.

P-persistent CSMA

P-persistent CSMA protocol is applied to slotted channels. If the channel is free, a station starts sending the packet. Otherwise, the station continues to monitor until the channel is free and then it sends the packet with probability p or in the next slot with probability $1 - p$.

CSMA/CD

CSMA with collision detection (CSMA/CD) is the modification of pure CSMA protocol. This protocol is used to improve the performance of CSMA by terminating transmission as soon as a collision is detected, thus reducing the probability of a second collision on retry. As the collided frames are terminated quickly, this protocol saves time and bandwidth. This is widely used on LANs in the medium access control (MAC) sub-layer.

16.14.4 Reservation protocols

With the pure ALOHA and slotted ALOHA protocols, users monitor the broadcast channel for the success or failure of their own transmission, which give rise to a maximum channel throughput of *1/(2e)* and *1/(e)* respectively. For broadcast networks with a very short propagation delay (relative to a packet transmission time), CSMA protocols were devised to significantly improve channel throughput by requiring the users to do carrier sensing before attempting a transmission.

However, the relatively large propagation delay (0.27 s) of a satellite channel causes carrier sensing impractical except in the case of an extremely long packet or low speed channel. The following reservation protocols were proposed to improve the throughput of a satellite channel beyond that of slotted ALOHA. Although it was invented for a satellite channel, the reservation protocols can be used for any of the other broadcast media.

Reservation ALOHA

Reservation ALOHA packet access scheme shown in Figure 16.36, is an improvement over slotted ALOHA and is a combination of slotted ALOHA and TDM. In this protocol, it is possible to reserve slots for the transmission of packets and the reservation can be permanent or can be reserved on request. In one type of reservation ALOHA, the user can reserve a slot permanently

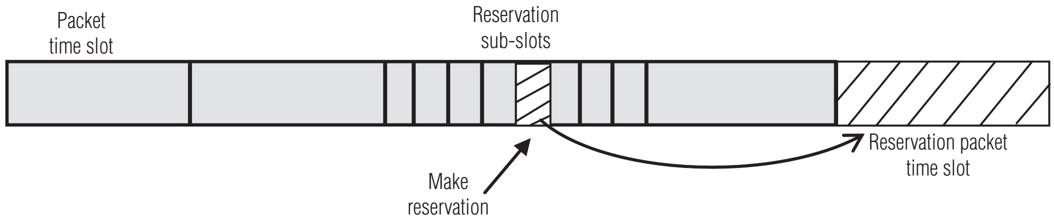


Figure 16.36 Protocol of reservation ALOHA

until its transmission is completed successfully. In another type of reservation ALOHA, it allows a user to transmit a request on a sub-slot which is reserved in each frame. If the transmission is successful, the user is allocated the next regular slot in the frame for data transmission

In the R-ALOHA protocol, time slots are divided into contention periods as well as contention-free periods. During the *contention period*, a subscriber uses very short packets to contend for the upcoming contention-free intervals that will be used for transmission of longer data packets. In this protocol, certain packet slots are assigned with priority. It is possible for subscribers to reserve some packet slots for the transmission of data packets either permanently or on request basis. The reservation period is followed by a transmission period. Usually, a very small data packet is sent to reserve transmission time slot. Reservations of time slots on request basis offers better throughput for high traffic conditions although this scheme causes a larger delay even under a light load.

Packet reservation multiple access

Packet reservation multiple access (PRMA) is a multiple access scheme with frames of a fixed number of slots. If a user has a series of data packets (or speech segments) to transmit, it competes for access in any free slot. If it successfully acquires the base station, the terminal achieves reservation in the corresponding slots of the next frames, until it delivers the reservation.

16.15 Capture effect in packet radio

When packet radio multiple access techniques are used with FM or spread-spectrum modulation, the strongest user may successfully capture the intended receiver, even when many other users are also transmitting. In addition, the closest transmitter may able to capture a receiver because of the small propagation path loss. The capture effect refers to the fact that a radio receiver may successfully decode a radio signal from one user despite the presence of interference from other users. The capture effect can be analysed using a parameter called capture ratio, which is defined as the ratio of the received signal power to the total interference power.

16.16 Summary

- The radio spectrum is a limited resource. Multiple access refers to techniques that enable multiple users to share the limited network resources efficiently. When there is more than one user to access such limited bandwidth, a multiple access scheme must be put in place to control the share of bandwidth among multiple users so that everyone can use services provided by the network and to make sure that no single user spends all available resources.

- Mobile cellular systems use various techniques to allow multiple users to access the same radio spectrum at the same time. They are: FDMA, TDMA, CDMA, and SDMA.
- *FDMA* is the first multiple access scheme and the signals from various users are assigned different frequencies, just as in an analogue system. Frequency guard bands are provided between adjacent signal spectra to minimize crosstalk between adjacent channels.
- In a *TDMA* system, each user uses the whole channel bandwidth. In a TDMA system, time is divided into equal time intervals, called *slots*. User data is transmitted in the slots. For example, 2G (GSM) and 2.5G (GPRS) adopted TDMA as their multiple access scheme.
- *In TDMA/FDMA, cell design requires more frequency planning which is tough. Whereas in CDMA, frequency planning is minimal.* In CDMA, all signals occupy the same bandwidth and are transmitted simultaneously in time. The different waveforms in CDMA are distinguished from one another at the receiver by the specific spreading codes they employ. It utilizes the most important use of spread-spectrum communications is as a multiple accessing technique.
- *CDMA assigns a unique spreading code to each user before the data transmission. If a receiver knows the correct code sequence related to the user, then only it is able to decode the received data.*
- Wideband CDMA (WCDMA) is one of the main technologies for the implementation of 3G cellular system. WCDMA can support mobile/portable voice, images, data, and video communications at up to 2 Mbps or 384 Kbps (wide area access). A 5-MHz wide carrier is used compared with 200 kHz wide carrier for narrowband CDMA. CDMA2000 is a family of 3G mobile technology standards, which use CDMA channel access, to send voice, data between mobile phones and base stations.
- SDMA system reuses the transmission frequency at suitable intervals of distance. If the distance between two base stations using the same frequency is large enough, the interference they inflict on each other is tolerable. The smaller this distance, the larger the system capacity. Therefore, various techniques have been developed to take advantage of this phenomenon. Sectorization divides a cell into smaller "sub-cells," some of which can reuse the same frequency.
- OFDM distributes the data over a large number of carriers that are spaced apart at precise frequencies. This spacing provides the "orthogonality" in this technique which prevents the demodulators from seeing frequencies other than their own.
- The reason for CDMA's popularity is primarily due to the performance that spread-spectrum waveforms display when transmitted over a multipath fading channel.
- 1G cellular system (AMPS) uses FDMA/FDD, 2G cellular systems (GSM uses TDMA/FDD and IS95 uses CDMA/FDD), 3G cellular system (WCDMA and cdma2000) uses CDMA/FDD and CDMA/TDD as their multiple access systems.
- Packet radio can be simply defined as technology which enables the user to transmit packet data (digital data) via radio (wireless) interface.
- All random access techniques are based on the *packet radio*. Performance of random access techniques is typically characterized by the throughput of the system.
- Pure or unslotted ALOHA is the simplest protocol in which the user transmits data packets whenever it has data.
- In the slotted-ALOHA system, the transmission time is divided into time slot, each slot corresponding to one frame transmission time.
- In CSMA user starts sending data only when the channel is idle, that is, there is no carrier,
- Reservation protocol is possible to reserve slots for the transmission of packets.

Review questions

1. What is the purpose of multiple access schemes?
2. What are the various multiple access techniques?
3. Define FDMA. What are the features of FDMA?
4. What are the advantages and drawbacks of FDMA?
5. Define TDMA.
6. Explain TDMA principle of operation.
7. What are the advantages of TDMA?
8. What are the drawbacks of TDMA?
9. Draw the TDMA frame structure.
10. Define the following terms: (a) Efficiency of TDMA (b) Number of channels in TDMA system.
11. What are the frame efficiency parameters in TDMA?
12. Explain the principle of CDMA.
13. What are the various types of CDMA?
14. Classify the existing cellular systems operation with respect to the multiple access method.
15. Describe the methods of CDMA.
16. What are hard handoff and soft handoff?
17. What is the difference between the hard handoff and the soft handoff?
18. Explain with a simple example how the soft handoff is implemented in CDMA mobile systems.
19. What are the advantages and drawbacks of CDMA?
20. Define "spread spectrum."
21. Compare the performance characteristics of FDMA, TDMA, CDMA, and SDMA multiple access methods.
22. What is SDMA system and how it is used as a multiple access system in cellular mobile communications?
23. What is cyclic prefix and explain how the ISI and CIR are removed in OFDM system?
24. Describe CSMA in detail.
25. Explain briefly about reservation protocols.
26. Compare the performance of slotted ALOHA with pure ALOHA.
27. Explain in detail about multiple access schemes. (Refer Section 16.2)
28. Explain about TDMA channels. (Refer Section 16.4.1)
29. Draw the TDMA frame structure and explain the significance of each slot. (Refer Section 16.4.2)

Exercise problems

1. GSM uses a frame structure in which each frame consists of 8 time slots, each time slot contains 156.25 bits, and data are transmitted at 270.833 Kbps over the channel. Find
   (a) The time duration of a bit
   (b) The time duration of a slot
   (c) The time duration of a frame
   (d) How long must a user occupying a single time slot wait between two consecutive transmissions?

2. If a total of 30 MHz of bandwidth is allocated to a particular cellular radio system that uses 30 kHz simplex channels to provide voice and control channels, compute the number of channels available per cell if a system uses
   (a) 4-cell reuse (b) 7-cell reuse (c) 12-cell reuse
   If 1 MHz of the allocated spectrum is dedicated to control channels, determine an equitable distribution of control and voice channels in each cell for each of the three systems.
3. An AMPS cellular operator is allocated 15 MHz for each simplex band, and if B_t is 15 MHz, B_g is 15 KHz, and B_c is 30 KHz, find the number of channels available in an FDMA system.
4. If the GSM uses a frame structure where each frame consists of 10 time slots and each time slot contains 157.25 bits, and data is transmitted at 270.833 Kbps in the channel, find (a) the time duration of a bit, (b) the time duration of a slot, (c) the time duration of a frame, and (d) how long must a user occupying a single time slot wait between two successive transmissions.
5. If a TDMA interface consists of 12 channels, then how many times the transmitted powers of TDMA is higher than FDMA system. (Hint: TDMA power is 10 log n times higher than in an FDMA system.)

Objective type questions and answers

1. Which type of multiple access method used in 3G cellular systems?
   (a) WCDMA/CDMA2000 (b) FDMA
   (c) CDMA (d) GSM
2. Which type of multiple access method preferred in GSM Cellular systems?
   (a) FDMA (b) FDMA/TDMA (c) CDMA (d) SDMA
3. The following communication channel can carry a voice conversation or, with digital service, carry digital data.
   (a) SDMA (b) FDMA/TDMA (c) CDMA (d) FDMA
4. The following are used to isolate channels from adjacent-channel interference.
   (a) guard bands (b) spectrum (c) adjacent channels (d) none of the above
5. When —— technique is employed, as long as the user is engaged in conversation, no other user can access the same spectrum space.
   (a) SDMA (b) FDMA/TDMA (c) CDMA (d) FDMA
6. ——technology has the longer handset battery life.
   (a) TDMA (b) FDMA/TDMA (c) CDMA (d) FDMA
7. —- allows handoff to/from an analogue AMPS channel.
   (a) TDMA (b) FDMA/TDMA (c) CDMA (d) FDMA
8. ——technology depends on sending a unique key to a receiver.
   (a) TDMA (b) FDMA/TDMA (c) CDMA (d) FDMA
9. Even though no two users use the same frequency band at the same time, guard bands are introduced between frequency bands to minimize
   (a) adjacent channel interference (b) channel equalization
   (c) spectrum inefficiency (d) (b) and (c)
10. Which multiple access method allows several users to share the same frequency channel by dividing the signal into different time slots?
    (a) FDMA (b) TDMA (c) FDD (d) (b) and (c)
11. CDMA uses —-technique to increase spectrum efficiency over current FDMA and TDMA systems.
    (a) spread spectrum (b) FDD
    (c) multipath mitigation (d) (a) and (c)

12. CDMA cellular systems operate in the
    (a) 900 MHz and 2.1 GHz (b) 800 MHz and 1.9 GHz
    (c) 1.6 GHz and 2. 1 GHz (d) 900 MHz and 1.8 GHz
13. CDMA code (i.e., pseudo-sequence) converts a ________ -signal to a noise-like ________ signal.
    (a) narrowband to wideband (b) wideband to narrowband
    (c) wideband to wideband (d) narrowband to narrowband
14. The act of transferring a call of a mobile from one base station to another is termed
    (a) mobile transfer (b) handoff (c) handover (d) takeoff
15. With ________ handoff, a definite decision is made on whether to handoff or not.
    (a) soft (b) hard (c) soft and hard (d) none of the above
16. One of the main advantages of CDMA systems is the capability of using signals that arrive in the receivers with different time delays. This phenomenon is called
    (a) fading (b) noise (c) efficiency (d) multipath
17. SDMA are realized by
    (a) adaptive arrays (b) fixed arrays (c) parabolic arrays (d) (a) and (b)
18. The best channel utilization with the pure ALOHA protocol is
    (a)18.4 per cent (b) 36.8 per cent (c) 73.6 per cent (d) none of the above
19. Maximum throughput in slotted ALOHA at traffic load G is equal to
    (a) 1 (b) 0.5 (c) 2 (d) none of the above

Answers: 1. (a), 2. (b), 3. (d), 4. (b), 5. (d), 6. (c), 7. (a), 8. (c), 9. (a), 10. (b), 11. (a), 12. (b), 13. (a), 14. (b), 15. (b), 16. (d), 17. (a), 18. (a), 19. (a).

Open book questions

1. Discuss near-far problem in cellular systems.
2. Discuss on which factor the performance of random access techniques depends.
3. Describe the SDMA technique.
4. What is the difference between TDMA and CDMA?
5. Mention the advantages and disadvantages of OFDMA.
6. What is rake receiver?
7. Explain the following terms of CDMA digital cellular systems:
   (i) Active set
   (ii) Long code
   (iii) Multiple sublayer
   (iv) Paging channel
8. State the differences between the architecture of TDMA and GSM and explain TDMA architecture.
9. Write short notes on
   (i) TDMA structure
   (ii) Frame length
   (iii) Frame offset
   (iv) Modulation timing
10. How many slots are present in TDMA frame and what is the length of the each slot?

Key equations

1. The number of channels that can be simultaneously supported in FDMA system is given by

$$\text{Number of channels } (N) = \frac{B_t - 2B_g}{B_c}$$

2. The efficiency of TDMA system

$$\text{Efficiency}(\eta) = \frac{\text{Number of bits per frame containing in the transmitted data}}{\text{Total number of bits per frame}}$$

3. Number of channels/time slots in TDMA system

$$N = \frac{M(B_{\text{total}} - 2B_{\text{guard}})}{B_c}$$

4. The received CDMA waveform becomes the sum of k user signals and noise:

$$r(t) = \sum_{n=l}^{k} pn_n(t)A_n d_n(t)\cos(\omega t + \theta_n) + n(t)$$

5. Capacity of a CDMA system is given by

$$k \approx \frac{W/R}{(E_b/N_0)_{\min}}$$

6. Normalized throughput S of the pure ALOHA protocol in terms of the offered load is given by

$$S = Ge^{-2G}$$

Further reading

Gilhousen, K. S., Jacobs, I. M., Padovani, R., Viterbi, A. J., Weaver, L. A.,Wheatly, C. E., “On the Capacity of a Cellular CDMA System,” *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, May 1991, pp. 303–312.

GPRS. Digital cellular communications system (Phase 2+): General Packet Radio Service Description, Stage 1, EN 301 113 v6.1.1 (1998–11) (GSM 02.60 version 6.1.1 Release 1997) European Telecommunications Institute (ETSI), Sophia Antipoli, Valbonne, France.

Gudsmundson, B. et al. 1992. “A Comparison of CDMA and TDMA Systems,” *Proceedings of IEEE Vehicular Technology Conference*, Denver, CO, May, pp. 732–735.

Maines, J. P., and Paige, E. G. S. “Surface Accoustic Wave Devices for Signal Processing Applications,” (Invited Paper), *Proceedings of the IEEE*, Vol. 64, No. 5, pp. 639–652, May 1976.

Mark, J. W., and Zhuang. W. *Wireless Communications and Networking*. Upper Saddle River, NJ: Pearson Education Inc., 2003.

Pickholtz R. L. Schilling, D. L., and Milstein, L. B. "Theory of Spread-Spectrum Communications – A tutorial," *IEEE Transactions on Communications*, Vol. 30, No. 5, May 1982, pp. 855–884.

Scholt, R. A. "The Origins of Spread-Spectrum Communications," *IEEE Transactions on Communications*, Vol. 30, No. 5, pp. 822–854, May 1982.

Simon, M. K., Omura, J. K., Scholtz, R. A. and Levitt, B. K., "*Spread Spectrum Communications*", Vol. 1, pp. 1–39, Rockville, Maryland: Computer Science Press, 1985.

Simon, M. K., Omura, J. K., Scholtz R. A., and Levitt, B. K., "*Spread Spectrum Communications*", Vol. 3, pp. 346–407, Rockville, Maryland: Computer Science Press, 1985.

Turin, G. L., "An Introduction to Digital Matched Filters," *Proceedings of the IEEE*, Vol. 64, No. 7, pp. 1092–1112, July 1976.

Wenk, B. "Aspects of Correlator Design for Industrial Applications," *Signal Processing II: Theories and Applications*.

Zeimer, R. E. *Digital Communications and Spread Spectrum Systems*. Macmillan Publishing Company, 1985.

17 Spectral Efficiency of FDMA, TDMA, and CDMA

17.1 Introduction

The spectral efficiency of a wireless network refers to the measure of the number of users that can be simultaneously supported by a limited radio frequency bandwidth in a defined geographic area. An increase in the number of users of the mobile communications system without a loss of performance increases the spectral efficiency.

Spectral efficiency of a mobile communications system shows how efficiently the spectrum is used by the system. Spectral efficiency of a mobile communications system depends on the choice of a multiple access scheme.

Spectral efficiency of a cellular system is also depends on the *frequency reuse factor* which is a measure of the efficiency of a radio plan or the modulation scheme. Therefore, the overall spectral efficiency of a multiple access technique is a combination of the efficiency of a radio plan or the modulation scheme and the one with respect to (i) time for TDMA, (ii) frequency for FDMA, (iii) modulation for both TDMA and FDMA.

This chapter briefly introduces the topics of power efficiency, bandwidth, and channel capacity for better understanding of the spectral efficiency computation in cellular system. Spectral efficiency computations for the various multiple access techniques and the cellular capacity of each multiple access technique are explained with example problems.

17.2 Parameters used to compare the efficiency of various cellular modulation techniques

Several parameters are used to compare the efficiency of the various modulation techniques. They are: power efficiency, bandwidth efficiency, and channel capacity. These are briefly described in these sections which are basis for the computation of spectral efficiency.

- Power efficiency is a measure of how much the signal power should be increased to achieve a particular bit error rate (BER) for a given modulation scheme. Signal energy per bit/noise power spectral density E_b/N_0 (see Section 17.2.1).
- Bandwidth efficiency is measured as data rate per hertz (rate/bandwidth – bits/s/Hz). This is also known as spectral efficiency (see Section 17.2.2).
- Channel capacity per bandwidth (see Section 17.2.3).

17.2.1 Power and energy of a signal

The energy (and power) of a signal represent the energy (or power) delivered by the signal when it is interpreted as voltage or current source feeding a 1 Ω resistor. The energy content of a signal *x(t)* is defined as the total work done and is represented as

$$E(x) = \underset{t\to\infty}{Lt} \int_{-\infty}^{\infty} |x(t)|^2 \, dt \tag{17.1}$$

The power content of a signal is defined as work done over time and is represented as

$$P(x) = \underset{t\to\infty}{Lt} \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 \, dt \tag{17.2}$$

Conventionally, power is defined as energy divided by time. A signal with finite energy is called an energy-type signal. A signal is energy type if $E(x) < \infty$ and a signal with positive and finite power is called a power-type signal and is power type if $0 < P < \infty$.

17.2.2 Bandwidth

The signal occupies a range of frequencies. This range of frequencies is called the bandwidth of the signal. In general, the bandwidth is expressed in terms of the difference between the highest and the lowest frequency components in the signal. A baseband signal or low pass signal bandwidth is a specification of only the highest frequency limit of a signal. A non-baseband bandwidth is the difference between the highest and lowest frequencies.

As the frequency of a signal is measured in Hz (hertz), so, the bandwidth is also expressed in Hz. In addition, we can say that the bandwidth of a signal is the frequency interval, where the main part of the power of the signal is located. The bandwidth is defined as the range of frequencies where the Fourier transform of the signal has a power above a certain amplitude threshold, commonly half the maximum value (half power = 3 dB) as

$10\log_{10}(P/P_{\text{half}})^{1/4}\ 10\log_{10}(1/2)^{1/4} - 3$ $(\approx -3$ dB, as $10\log_{10}(P/P_{\text{half}}) = 10\log_{10}(1/2) = -3)$. Power is halved $(P/2)$ at 3 dB points, for example, $P = P/2 = (V_0/\sqrt{2})(I_0/\sqrt{2})$, where V_0 and I_0 are the peak amplitude of voltage and current, respectively. (Figure 17.1)

However, in digital communication the meaning of "bandwidth" has been clouded by its metaphorical use. Technicians sometimes use it as slang for baud, which is the rate at which symbols may be transmitted through the system. It is also used more colloquially to describe channel capacity, the rate at which bits may be transmitted through the system.

Bit rate – This is the rate at which information bits (1 or 0) are transmitted. Normally, digital system requires greater bandwidth than analogue systems.

Baud – The baud (or signalling) rate defines the number of symbols transmitted per second. One symbol consists of one or several bits together, based on the modulation technique used. Generally, each symbol represents n bits, and has M signal states, where $M = 2^n$. This is called M-ary signalling.

17.2.3 Channel capacity

The maximum rate of communication via a channel without error is known as the capacity of the channel.

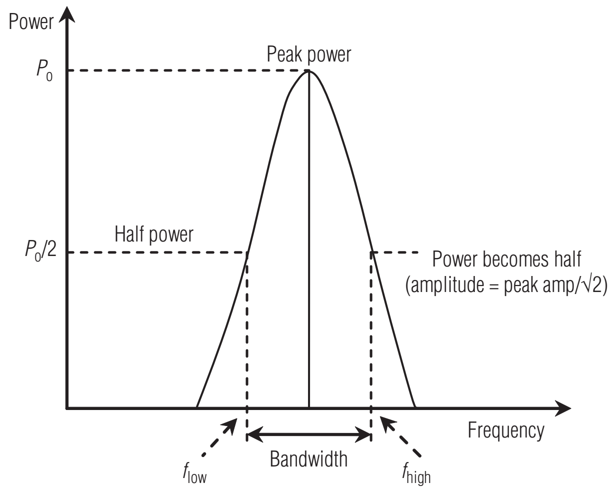


Figure 17.1 Bandwidth of a signal

Channel capacity refers to the maximum data rate a channel of a given bandwidth can sustain. An improved channel capacity leads to an ability to support more users of a specified data rate, implying a better spectral efficiency.

In a channel where noise is present, there is an absolute maximum limit for the bit rate of transmission.

This limit arises when the number of different signal levels is increased, as in such a case the difference between two adjacent sampled signal levels becomes comparable to the noise level. Claude Shannon extended Nyquist's work to a noisy channel.

Applying the classic sphere scenario, we can obtain an estimate of the maximum number of code words that can be packed in for a given power constant P, within a sphere of radius $\sqrt{N\sigma^2}$.

The noise sphere has a volume of $\sqrt{N\sigma^2}$. Thus, as shown in the Figure 17.2, the maximum number of code words that can be packed in with non-overlapping noise spheres is the

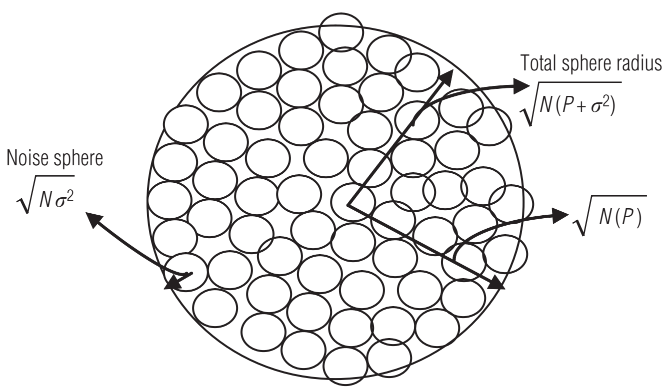


Figure 17.2 Number of noise spheres that can be packed into N-dimensional signal space

ratio of the volume of the sphere of radius $\sqrt{\{(N\sigma^2)+NP\}}$ to the volume of the noise sphere $[\sqrt{\{(N\sigma^2)+NP\}}]^N / [\sqrt{(N\sigma^2)}]^N$,

where N is the signal space dimension and σ^2 is the variance of the real Gaussian random variable.

With mean μ, this equation implies that the maximum number of bits per symbol that can be reliably communicated is

$$(1/N)\log\left[\sqrt{(N\sigma^2)+(NP)}\right]^N \Big/ \left[\sqrt{(N\sigma^2)}\right]^N = (1/2)\log\left(1+P/\sigma^2\right) \tag{17.3}$$

This is indeed the capacity of the additive white Gaussian noise (AWGN) channel. In a later chapter, we will see that for complex channels, the noise in I and Q components is independent. So it can be thought of as two independent uses of a real AWGN channel. The power constraint and noise per real symbol are represented as $P_{av}/2B$ and $N_0/2$, respectively. Hence, the capacity of the channel will be

$$\begin{aligned} C &= (1/2)\log\left(1+P_{av}/N_0.B\right) \text{ bits per real dimension} \\ &= \log\left(1+P_{av}/N_0.B\right) \text{ bits per complex dimension} \end{aligned} \tag{17.4}$$

This is the capacity in bits per complex dimension or degrees of freedom. As there are B complex samples per second, the capacity of the continuous time AWGN channel is

$$C_{\text{awgn}}(P_{av}, B) = B\log\left(1+P_{av}/N_0.B\right)\text{bits/s} \tag{17.5}$$

Now, the signal-to-noise ratio $(\text{SNR}) = \left(P_{av}/N_0.B\right)$, which is the SNR per degree of freedom. So, Equation (17.5) reduces to

$$C_{\text{awgn}} = \log\left(1+\text{SNR}\right) \text{ bits/s} \tag{17.6}$$

This equation measures the maximum achievable spectral efficiency through the AWGN channel as a function of the SNR.

17.3 Spectral efficiency of modulation

Radio spectrum is the one critical resource for all mobile radio communications. Without access to radio spectrum, there can be no mobile radio. Spectral efficiency is one of the crucial parameter for all wireless networks such as Wi-Fi, Wimax, and cellular networks and is defined as the maximum number of traffic channels per MHz per cell. *Spectral efficiency is calculated to figure out how efficiently* the spectrum allotted is being utilized and the main goal of service provider is to accommodate maximum number of users in the limited bandwidth. That is why spectral efficiency calculation plays very important role in wireless network planning.

The most desirable feature of a mobile communications system is efficient use of the spectrum. The measure of spectral efficiency allows us to estimate the capacity of a mobile communications system.

The spectral efficiency can be improved by following methods:

1. Reducing the channel bandwidth
2. Information compression (low-rate speech coding)

3. Variable bit-rate codec
4. Improved channel assignment algorithms
5. Dynamic channel assignment

The overall spectral efficiency of a mobile communications system can be estimated by knowing the *modulation* and the *multiple access* spectral efficiencies separately.

Spectral efficiency of modulation

In digital wireless networks, the units of system spectral efficiency or area spectral efficiency are bit/s/Hz/area.

$$\text{Spectral efficiency} = \frac{\text{Bit rate}}{\text{Transmission bandwidth}} \quad \text{(bps/Hz)}$$

For cellular mobile system, Spectral efficiency (η_m) is defined as

$$\eta_m = (\text{Traffic in Erlang})/(\text{Amount of spectrum in MHz} \times \text{Area in km}^2) \tag{17.7}$$

The system spectral efficiency of a cellular network can also be expressed as the maximum number of simultaneous phone calls per area unit over 1 MHz frequency spectrum in Erlangs per megahertz per cell (E/MHz)/cell). The other units of hm are (E/MHz)/sector, (E/MHz)/site, or (E/MHz)/km².

The spectral efficiency (η_m) with respect to the modulation used is defined as

$$\eta_m = \frac{\text{Total number of channels available in the system}}{\text{Bandwidth } \times \text{Total coverage area}} = \frac{\frac{B_w}{B_c} \times \frac{N_c}{N}}{B_w \times N_c \times A_c} = \frac{1}{B_c \times N \times A_c} \tag{17.8}$$

where

η_m is the modulation efficiency (channels/MHz/km²)
B_w is the bandwidth of the system (MHz)
B_c is the channel spacing (MHz)
N_c is the number of cells in a system
N is the frequency reuse factor of system or cluster size
A_c is the area covered by a cell (km²)

By introducing the radio trunking (sharing of a pool of resource) efficiency factor (η_t) *in* Equation (17.8), η_t is a function of the blocking probability and number of available channels per cell, the spectral efficiency of the system can be written as

$$\eta_m = \frac{\eta_t\left(\frac{B_w/B_c}{N}\right)}{B_w \times A_c} = \frac{\eta_t}{B_w \times N \times A_c} \text{channels/MHz/Km}^2 \tag{17.9}$$

where η_t is a function of the blocking probability and the total number of available channels per cell

$$\left[\frac{B_w/B_c}{N}\right]$$

From the above equation we can conclude the following:

- The voice quality will depend on the frequency reuse factor, N, which is a function of the signal-to-interference (S/I) ratio of the modulation scheme used in the mobile communications system (see Chapter 5).

- The relationship between system bandwidth, B_w, and the amount of traffic carried by the system is non-linear, that is, for a given percentage increase in B_w, the increase in the traffic carried by the system is more than the increase in B_w.
- From the average traffic per user (*Erlang/user*) during the busy hour and Erlang/MHz/km^2, the capacity of the system in terms of *users/km^2/MHz* can be obtained.
- The spectral efficiency of modulation depends on the blocking probability.

From Equation (17.8), we can observe that the spectral efficiency of modulation (η_m) does not depend on the bandwidth of the system. It only depends on the three important factors: channel spacing, the cell area, and the frequency reuse factor (N). By reducing the channel spacing, the spectral efficiency of modulation for the system is increased, provided the cell area (A_c) and reuse factor (N) remain unchanged. If a modulation scheme can be designed to reduce N then more channels are available in a cell and efficiency is improved.

17.4 Spectral efficiencies of multiple access techniques

Multiple access spectral efficiency is defined as the ratio of the total time or frequency dedicated for traffic transmission to the total time or frequency available to the system. Thus, the multiple access spectral efficiency is a dimensionless number with an upper limit of unity. In FDMA, users share the radio spectrum in the frequency domain. In FDMA, the multiple access efficiency is reduced because of guard bands between channels and also because of control/signalling channels. In TDMA, the efficiency is reduced because of guard time and synchronization sequence.

17.4.1 FDMA spectral efficiency

For FDMA, multiple access spectral efficiency is defined as

$$\eta_a = \frac{B_c N_t}{B_w} \leq 1 \tag{17.10}$$

where η_a is the multiple access spectral efficiency
N_t is the total number of traffic channels in the covered area
B_c is the channel spacing
B_w is the system bandwidth

17.4.2 TDMA spectral efficiency

TDMA can operate as wideband or narrowband. In the wideband TDMA, the entire spectrum is used by each individual user. For the wideband TDMA, multiple access spectral efficiency is given as

$$\eta_a = \frac{\tau M_t}{T_f} \tag{17.11}$$

where τ is the duration of a time slot
T_f is the frame duration
M_t is the number of time slots per frame

In Equation (17.11) it is assumed that the total available bandwidth is shared by all users. For the narrowband TDMA schemes, the total band is divided into a number of sub-bands, each using the TDMA technique. For the narrowband TDMA system, frequency domain efficiency is not unity as the individual user channel does not use the whole frequency band available to the system. The multiple access spectral efficiency of the narrowband TDMA system is given as:

$$\eta_a = \left(\frac{\tau M_t}{T_f}\right)\left(\frac{B_u N_u}{B_w}\right) \tag{17.12}$$

where B_u is the bandwidth of an individual user during his or her time slot

N_u is the number of users sharing the same time slot in the system, but having access to different frequency sub bands.

17.4.3 Overall spectral efficiency of FDMA and TDMA systems

The overall spectral efficiency (η) of a FDMA and TDMA mobile system is obtained by considering both the modulation and the multiple access spectral efficiencies

$$\eta = \eta_m \eta_a \tag{17.13}$$

17.5 Capacity of TDMA and CDMA systems

The cell capacity is defined as the maximum number of users that can be supported simultaneously in each cell.

17.5.1 Capacity and frame efficiency of a TDMA system

The capacity of a TDMA system is given by

$$N_u = \frac{\eta_b \mu}{vf} \times \frac{B_w}{RN} \tag{17.14}$$

where N_u is the number of channels (mobile users) per cell
η_b is the bandwidth efficiency factor (<1.0)
μ is the bit efficiency (=2 bit/symbol for QPSK, = 1 bit/symbol for GMSK as used in GSM)
vf is the voice activity factor (equal to one for TDMA)
B_w is the one-way bandwidth of the system
R is the information (bit rate plus overhead) per user
N is the frequency reuse factor

$$\text{Spectral efficiency } \eta = \frac{N_u \times R}{B_w} \text{ bit/s/Hz} \tag{17.15}$$

Efficiency of a TDMA frame

The number of overhead bits per frame is $b_0 = N_r b_r + N_t b_p + (N_t + N_r) b_g$ (17.16)

where N_r is the number of reference bursts per frame
N_t is the number of traffic bursts (slots) per frame
b_r is the number of overhead bits per reference burst
b_p is the number of overhead bits per preamble per slot
b_g is the number of equivalent bits in each guard time interval

The total number of bits per frame is

$$b_T = T_f \times R_{rf} \tag{17.17a}$$

where T_f is the frame duration
R_{rf} = bit rate of the RF channel

$$\text{Frame efficiency } \eta = \left(1 - b_0/b_T\right) \times 100 \text{ per cent} \tag{17.17b}$$

It is desirable to maintain the efficiency of the frame as high as possible. The number of bits per data channel (user) per frame is $b_c = RT_f$, where R is the bit rate of each channel (user).

$$\text{Number of channels/frame } N_{CF} = \frac{(total\ data\ bits)/(frame)}{(Bits\ per\ channel)/(frame)}$$

$$N_{CF} = \frac{\eta R_{rf} T_f}{RT_f} \tag{17.18a}$$

$$N_{CF} = \frac{\eta R_{rf}}{R} \tag{17.18b}$$

Equation (17.18b) indicates the number of times slot per frame.

17.5.2 Capacity of a DS-CDMA system

The capacity of a DS-CDMA system depends on the following parameters:

1. Processing gain, G_p (a ratio of spreading bandwidth, B_w, and information rate, R)
2. Bit energy-to-interference ratio, E_b/I_0.
3. Voice duty cycle, vf.
4. DS-CDMA omnidirectional frequency reuse efficiency, ηf.
5. Number of sectors, G in the cell-site antenna

The received signal power at the cell from a mobile is $S = R \times E_b$. The S/I ratio is

$$\frac{S}{I} = \frac{R}{B_w} \times \frac{E_b}{I_0} \tag{17.19}$$

where E_b is the energy per bit
I_0 is the interference density

In a cell with N_u mobile transmitters, the number of effective interferers is $N_u - 1$ because each mobile is an interferer to all other mobiles. This is valid regardless of how the mobiles are distributed within the cell since automatic power control (APC) is used in the mobiles. The APC operates such that the received power at the cell from each mobile is the same as for every other mobile in the cell, regardless of the distance from the centre of the cell. APC conserves battery power in the mobiles, minimizes interference to other users, and helps overcome fading. In a hexagonal cell structure, because of interference from each tier, the S/I ratio is given as (see Chapter 5):

$$\frac{S}{I} = \frac{1}{\left(N_u - 1\right)\left(1 + 6k_1 + 12k_2 + 18k_3 + \cdots\right)} \tag{17.20}$$

where N_u is number of mobile users in the band, B_w

k_i, $i = 1, 2, 3, \ldots$, is the interference contribution from all terminals in individual cells in tiers 1, 2, 3, etc., relative to the interference from the centre cell. This loss contribution is a function of both the path loss to the centre cell and the power reduction because of power control to an interfering mobile's own cell centre. If we define a frequency reuse efficiency, ηf, as in Equation (17.21a), then E_b/I_0 is given by Equation (17.20)

$$\eta f = \frac{1}{[1 + 6k_1 + 12k_2 + 18k_3 + \cdots]} \tag{17.21a}$$

$$\frac{S}{I} = \frac{\eta f}{(N_u - 1)} \tag{17.21b}$$

$$\frac{E_b}{I_0} = \frac{B_w}{R} \times \frac{\eta f}{(N_u - 1)} \tag{17.22}$$

This equation does not include the effect of background thermal and spurious noise (i.e., ρ) in the spreading bandwidth B_w. Including this as an additive degradation term in the denominator results in a bit energy-to-interference ratio of:

$$\frac{E_b}{I_0} = \frac{B_w}{R} \times \frac{\eta f}{(N_u - 1) + \rho/S} \tag{17.23}$$

Note that from Equation (17.23) the capacity of the DS-CDMA system is reduced by ρ/S which is the ratio of background thermal plus spurious noise to power level.

For a fixed $G_p = B_w/R$, one way to increase the capacity of the DS-CDMA system is to reduce the required E_b/I_0, which depends upon the modulation and coding scheme. By using a powerful coding scheme, the E_b/I_0 ratio can be reduced, but this increases system complexity. In addition, it is not possible to reduce the E_b/I_0, ratio indefinitely. The only other way to increase the system capacity is to reduce the interference. Two approaches are used: one is based on the natural behaviour of human speech and the other is based on the application of the satirized antennas. From experimental studies it has been found that typically in a full duplex 2-way voice conversation, the duty cycle of each voice is, on the average, less than 40 per cent. Thus, for the remaining period of time the interference induced by the speaker can be eliminated. Since the channel is shared among all the users, noise induced in the desired channel is reduced due to the silent interval of other interfering channels. It is not cost-effective to exploit the voice activity in the FDMA or TDMA system because of the time delay associated with reassigning the channel resource during the speech pauses. If we define vf as the voice activity factor (<1), then Equation (17.23) can be written as

$$\frac{E_b}{I_0} = \frac{\eta f}{vf} \times \frac{B_w}{R} \times \frac{1}{(N_u - 1) + \rho/S} \tag{17.24a}$$

$$(N_u - 1) + \frac{\rho}{S} = \left[\frac{\eta f}{vf}\right] \times \left[\frac{B_w}{R}\right] \times \left[\frac{I_0}{E_b}\right] \tag{17.24b}$$

The equation to determine the capacity of a DS-CDMA system should also include additional parameters to reflect the bandwidth efficiency factor, the capacity degradation factor due to imperfect power control, and the number of sectors in the cell-site antenna. Equation (17.24b) is augmented by these additional factors to provide the following equation for DS-CDMA capacity at one cell:

$$N_u = \frac{\eta f \eta_b c_d \lambda}{vf} \times \frac{B_w}{R \times (E_b/I_0)} + 1 - \frac{\rho}{S} \tag{17.25a}$$

Equation (17.25a) can be rewritten as Equation (17.25b) by neglecting the last two terms.

$$N_u = \frac{\eta f \eta_b c_d \lambda}{vf} \times \frac{B_w}{R \times (E_b/I_0)} \tag{17.25b}$$

where ηf is the frequency reuse efficiency <1
η_b is the bandwidth efficiency factor <1
c_d is the capacity degradation factor to account for imperfect APC <1
vf is the voice activity factor <1
B_w is the one-way bandwidth of the system
R is the information bit rate plus overhead
E_b is the energy per bit of the desired signal
E_b/I_0 is the desired energy-to-interference ratio (dependent on quality of service)
λ is the efficiency of sector-antenna in cell

For digital voice transmission, E_b/I_0 is the required value for a BER of about 10^{-3} or better, and ηf depends on the quality of the diversity. Under the most optimistic assumption, $\eta f < 0.5$. The voice activity factor, vf is usually assumed to be less than or equal to 0.6. E_b/I_0 for a BER of 10^{-3} can be as high as 63 (18 dB) if no coding is used and as low as 5 (7 dB) for a system using a powerful coding scheme. The capacity degradation factor, c_d will depend on the implementation but will always be less than 1.

17.6 Comparison of spectrum efficiencies of CDMA and TDMA

The spectrum efficiency relies upon the average BER that is adequate for a service quality.

TDMA system: A finite number of time slots are fixed to channel capacity and new users cannot be allowed when each of these slots are filled. The interference of the co-channel is reduced due to GSM's better error management and frequency hopping. Without any degradation in the quality of the service this allows the frequency to be reused. The efficiency of TDMA, with three different sets of frequency allotted to spread TDMA is 6.6.

CDMA system: CDMA handsets offer analogue capabilities where the GSM does not have. CDMA technology declares that its bandwidth is 13 times efficient than TDMA and 40 times efficient than analogue systems. CDMA also have higher data and voice transmission quality and better security because it uses the spread spectrum technology that has an increased resistance to multipath distortion. When compared to TDMA, CDMA has greater coverage area. Using very certain assumption CDMA gives an efficiency of 45 users per call per MHz and the pessimistic assumption the value is 12 which still gives CDMA a 2:1 advantage over TDMA.

17.6.1 Comparison of DS-CDMA versus TDMA system capacity

Using Equation (17.25b) with $vf = 1$ (no voice activity) for TDMA and $\lambda = 1.0$ (omnidirectional cell) for DS-CDMA the ratio of the cell capacity for the DS-CDMA and TDMA system is given as

$$\frac{N_{\text{CDMA}}}{N_{\text{TDMA}}} = \frac{c_d N \eta f}{E_b/I_0} \times \frac{1}{vf_{\text{odres}}} \times \frac{1}{\mu} \times \frac{R_{\text{TDMA}}}{R_{\text{CDMA}}} \tag{17.26}$$

17.7 Summary

- Bandwidth is the range of frequencies of a signal and is expressed in terms of the difference between the highest and the lowest frequency components in the signal.
- The maximum rate of communication via a channel without error is known as the capacity of the channel.
- Spectrum efficiency is a crucial factor for all wireless communications and defined as the *maximum number of traffic channels per MHz per cell*.
- The units of spectral efficiency are (E/MHz)/cell or (E/MHz)/sector, (E/MHz)/site, or (E/MHz)/km^2.
- The overall spectral efficiency of a multiple access technique is a combination of the efficiency of the modulation scheme and the one with respect to time (in case of TDMA) or to frequency (in case of FDMA).
- In FDMA, the multiple access efficiency is reduced because of guard bands between channels and also because of control/signalling channels. In TDMA, the efficiency is reduced because of guard time and synchronization sequence.
- In TDMA system, the channel capacity is fixed to a finite number of time slots and new users cannot be accommodated when each of these slots is filled.
- CDMA is said to have higher capacity than TDMA. The battery life is higher in TDMA compared to CDMA because CDMA handsets transmit data all the time and TDMA does not require constant transmission.
- When going for a cell phone, the user should choose the technology according to where they use it. For example, for users who travel abroad, it is better to go with GSM handsets. For the users in United States, CDMA is better than TDMA because of the coverage we can get at rural areas where digital signals cannot be transmitted.
- CDMA technology claims that its bandwidth is 13 times efficient than TDMA and 40 times efficient than analogue systems.

Example problem 17.1

Find spectral efficiency of modulation in the North American Narrowband FDMA cellular system (AMPS).

- System bandwidth: 12.5 MHz
- Channel spacing: 30 kHz
- Area of a cell: 8 km^2
- Total coverage area: 4,000 km^2
- Average number of calls per user during the busy hour: 1.2
- Average holding time of a call: 100 s
- Call blocking probably: 2 per cent

Frequency reuse factor: 7

Solution

Number of 30-kHZ channels = $(12.5 \times 1{,}000)/30 = 416$
Number of signalling channels = 21

Number of voice channels = 416 – 21 = 395
Number of channels per cell = 395/7 = 56
Number of cells = 4,000/8 = 500
With 2 per cent blocking for omnidirectional case, the total traffic carried by 56 channels (using Erlang-B formula) = 45.9 Erlangs/cell = 5.74 Erlangs/km^2
Number of calls per hour = (45.9 × 3,600)/100 = 1,652.4
Calls/hour/cell = 1,652.4/8 = 206.6 calls/hour/km^2
Number of users/cell = 1,652.4/1.2 = 1,377 users/hour/cell = 1,377/56 = 24.6 users/hour/channel
$\eta_m = (45.9 \times 500)/(4,000 \times 12.5) = 0.459$ Erlangs/MHz/km^2

Example problem 17.2

In the GSM800 digital channelized cellular system, the one-way bandwidth of the system is 12.5 MHz. The RF channel spacing is 200 kHz. Eight users share each RF channel and three channels per cell are used for control channels. Calculate the spectral efficiency of modulation (for a dense metropolitan area with small cells) using the following parameters:

Area of a cell = 8 km^2
Total coverage area = 4,000 km^2
Average number of calls per user during the busy hour = 1.2
Average holding time of a call = 100 s
Call blocking probability = 2 per cent
Frequency reuse factor = 4

Solution

$$\text{Number of 200-kHz RF channels} = \frac{12.5 \times 1000}{200} = 62$$

Number of traffic channels = 62 × 8 = 496
Number of signalling channels per cell = 3

$$\text{Number of traffic channels per cell} = \frac{496}{4} - 3 = 121$$

$$\text{Number of cells} = \frac{4000}{8} = 500$$

With 2 per cent blocking for an omnidirectional case, the total traffic carried by 121 channels (using Erlang-B tables) = 108.4 (1.0 – 0.02) = 106.2 Erlangs/cell or 13.28 Erlangs/km^2

$$\text{Number of calls per hour per cell} = \frac{106.2 \times 3600}{100} = 3,823$$

$$\text{Calls/hour/km}^2 = \frac{3823}{8} = 477.9$$

$$\text{Maximum number of users/cell/hour} = \frac{3823}{1.2} = 3186,$$

$$\text{user/hour/channel} = \frac{3186}{121} = 26.33$$

$$\eta_m = \frac{(\text{Erlangs per cell}) \times \text{Number of cells}}{B_w \times \text{Coverage area}} = \frac{106.2 \times 500}{12.5 \times 4000} = 1.06 \text{ Erlangs/MHz/km}^2.$$

Example problem 17.3

In a first-generation AMP system where there are 395 channels of 30 kHz each in a bandwidth of 12.5 MHz, what is the multiple access spectral efficiency for FDMA?

Solution

$$\eta_a = \frac{30 \times 395}{12.5 \times 1000} = 0.948.$$

Example problem 17.4 (see Section 10.4.3)

In the North American Narrowband TDMA cellular system, the one-way bandwidth of the system is 12.5 MHz. The channel spacing is 30 kHz and the total number of voice channels in the system is 395. The frame duration is 40 ms, with six time slots per frame. The system has an individual user data rate of 16.2 Kbps in which the speech with error protection has a rate of 13 Kbps. Calculate the multiple access spectral efficiency of the TDMA system.

Solution

The time slot duration that carries data: $\tau = \left(\frac{13}{16.2}\right)\left(\frac{40}{6}\right) = 5.35$ ms

$$T_f = 40 \text{ ms},\ M_t = 6,\ N_u = 395,\ B_u = 30 \text{ kHz, and } B_w = 12.5 \text{ MHz}$$

$$\eta_a = \frac{5.35 \times 6}{40} \times \frac{30 \times 395}{12500} = 0.76$$

The overhead portion of the frame = 1.0 – 0.76 = 24 per cent.

Example problem 17.5

Calculate the capacity and spectral efficiency of a TDMA system using the following parameters: bandwidth efficiency factor $b = 0.9$, bit efficiency (with QPSK) = 2, voice activity factor = $vf = 1.0$, one-way system bandwidth B_w = 12.5 MHz, information bit rate R = 16.2 Kbps, and frequency reuse factor $N = 19$.

Solution

$$N_u = \frac{0.9 \times 2}{1.0} \times \frac{12.5 \times 10^6}{16.2 \times 10^3 \times 19}$$

$N = 73.1$ (say 73 mobile users per cell)

Spectral efficiency $\eta = \dfrac{73 \times 16.2}{12.5 \times 1000} = 0.094$ bit/s/Hz

Example problem 17.6

Consider GSM TDMA system with the following parameters:

$N_r = 2$, $N_t = 24$ frames of 120 ms each with time slots per frame
$b_r = 148$ bits in each of 8 time slots
$b_p = 34$ bits in each of 8 time slots
$b_g = 8.25$ bits in each of 8 time slots
$T_f = 120$ ms
$R_{rf} = 270.8333333$ Kbps
$R = 22.8$ Kbps

Calculate the frame efficiency and the number of channels per frame.

Solution

$$b_0 = 2 \times (8 \times 148) + 24 \times (8 \times 34) + 8 \times 8.25 = 10{,}612 \text{ bits per frame}$$

$$b_T = 120 \times 10^{-3} \times 270.8333333 \times 10^3 = 32{,}500 \text{ bits per frame}$$

$$\eta = \left(1 - \frac{10612}{32500}\right) \times 100 = 67.35 \text{ per cent}$$

$$\text{Number of channels/frame} = \frac{0.6735 \times 270.8333333}{22.8} = 8$$

The last calculation, with an answer of 8 channels, confirms that our calculation of efficiency is correct.

Example problem 17.7

Calculate the capacity and spectral efficiency of the DS-CDMA system with an omnidirectional cell using the following data:

Bandwidth efficiency $\eta_b = 0.9$
Frequency reuse efficiency $\eta f = 0.45$
Capacity degradation factor $c_d = 0.8$
Voice activity factor $vf = 0.4$
Information bit rate $R = 16.2$ Kbps

$$E_b/I_0 = 7 \text{ dB}$$

One-way system bandwidth $B_w = 12.5$ MHz neglect other sources of interference

Solution

$E_b/I_0 = 5.02(7\text{ dB})$

$$N_u = \frac{0.45\times 0.9\times 0.8\times 1}{0.4}\times\frac{12.5\times 10^6}{16.2\times 10^3\times 5.02}$$

$$N_u = 124.5(say125)$$

$$\text{The spectral efficiency, } \eta = \frac{125\times 16.2}{12.5\times 10^3} = 0.162\text{ bits/s/Hz}$$

In these calculations, an omnidirectional antenna is assumed. If a three sector antenna (i.e., $G = 3$) is used at a cell site with $\lambda = 2.6$, the capacity will be increased to 325 mobile users per cell, and spectral efficiency will be 0.421 bits/s/Hz.

Example problem 17.8

Using data given in Example problems 17.5 and 17.6, compare the capacity of the DS-CDMA and TDMA omnidirectional cell.

Solution

$$\frac{N_{\text{CDMA}}}{N_{\text{TDMA}}} = \frac{0.8\times 19\times 0.45}{5.02}\times\frac{1}{0.4}\times\frac{1}{2}\times\frac{16.2}{16.2} = 1.703.$$

Review questions

1. What is spectral efficiency?
2. Mention the various methods to improve the spectral efficiency.
3. Write the equation for spectral efficiency of modulation. What are the various factors on which the spectral efficiency of modulation depends?
4. Write the multiple access spectral efficiencies of FDMA and TDMA cellular systems.
5. Write the various parameters on which the capacity of CDMA depends.
6. Compare the spectrum efficiencies of CDMA and TDMA.
7. What are the advantages of spectral efficiency of modulation?

Exercise problems

1. Calculate the capacity and spectral efficiency of a TDMA system using the following parameters: bandwidth efficiency factor $b = 0.8$, bit efficiency (with QPSK) = 1, voice activity factor = vf = 1.0, one-way system bandwidth B_w = 12.5 MHz, information bit rate R = 16.2 Kbps, and frequency reuse factor $N = 12$.
2. Calculate the capacity and spectral efficiency of the DS-CDMA system with an omnidirectional cell using the following data: bandwidth efficiency $\eta_b = 0.9$, frequency reuse efficiency $\eta f = 0.5$, capacity degradation factor $c_d = 0.9$, voice activity factor $vf = 1.0$, information

bit rate $R = 16.2$ Kbps, $E_b/I_0 = 5$ dB, one-way system bandwidth $B_w = 12.5$ MHz neglect other sources of interference.

Objective type questions and answers

1. The unit of spectral efficiency is
   (a) (E/MHz)/cell (b) (E/MHz)/site (c) (E/MHz)/km^2 (d) all of the above
2. CDMA technology is times efficient than TDMA
   (a) 30 (b) 40 (c) 50 (d) 90
3. FDMA technology efficiency reduced because of
   (a) guard bands (b) adjust channels (c) spectrum (d) none of the above
4. The efficiency of cellular modulation techniques depends up on
   (a) power signal (b) bandwidth (c) channel capacity (d) all of the above

Answers: 1. (d), 2. (a), 3. (a), 4. (d).

True/False

1. CDMA claims that its bandwidth is 13 times efficient than TDMA.
2. CDMA claims that its bandwidth is 40 times efficient than TDMA.
3. The overall spectral efficiency of a multiple access technique is a combination of the efficiency of a radio plan or the modulation scheme and the one with respect to (i) time for TDMA, (ii) frequency for FDMA, (iii) modulation for both TDMA and FDMA.
4. $E(x) = \underset{t\to\infty}{Lt} \int_{-\infty}^{\infty} |x(t)|^2 \, dt$
5. $P(x) = \underset{t\to\infty}{Lt} \frac{1}{T} \int_{-T/2}^{T/2} |x(t)|^2 \, dt$

Answers: 1. (T), 2. (T), 3. (T), 4. (T), 5. (T).

Open book questions

1. Explain on which factors do spectral efficiency of cellular system depend.
2. Write the various parameters on which the capacity of DS-CDMA depends.
3. What is frequency reuse?
4. State certain access technologies used in mobile satellite communication systems.
5. Define the term Erlang.

Key equations

1. The energy content of a signal $x(t)$ is defined as the total work done and is represented as

$$E(x) = \underset{t\to\infty}{Lt} \int_{-\infty}^{\infty} |x(t)|^2 \, dt$$

2. The power content of a signal is defined as work done over time and is represented as

$$P(x) = \underset{t\to\infty}{Lt}\frac{1}{T}\int_{-T/2}^{T/2}\left|x(t)\right|^2 dt$$

3. The capacity of the continuous time AWGN channel is

$$C_{\text{awgn}}(P_{av}, B) = B\log\left(1+P_{av}/N_0.B\right) bits/\text{s}$$

4. The total number of bits per frame is

$$b_T = T_f \times R_{rf}$$

5. The signal-to-interference ratio is

$$\frac{S}{I} = \frac{R}{B_w}\times\frac{E_b}{I_0}$$

Further reading

Calhoun, G. *Digital Cellular Radio*. Boston: Artech House, 1988.

Chen, Q., Sousa, E. S., and Pasupathy, S. "Performance of a Coded Multi-Carrier DS-CDMA in Multipath Fading Channels," *Kluwer Journal of Wireless Personal Communications*, Vol. 2, 1995, 167–183.

Fazel, K. "Performance of CDMA/OFDM for Mobile Communication System," *IEEE Proceedings of ICUPC*, Vol. 2, October 1993, 975–979.

Garg, V. K., and Wilkes, J. E. *Wireless and Personal Communications*. Upper Saddle River, NJ: Prentice Hall, 1996.

Gilhousen, S., Jacods, I. M., Padovani, R.; Weaver, L. A., "Increased Capacity Using CDMA for Mobile Satellite Communication," *IEEE Journal on Selected Areas in Communications*, Vol. 8, No. 4, pp. 503–514, May 1990.

Kondo, S., and Milstein, L. "Performance of Multi-Carrier DS-CDMA Systems," *IEEE Transactions on Communications*, Vol. 44, No. 2, February 1996, 238–246.

Lee, W. C. Y. "Spectrum Efficiency in Cellular," *IEEE Transactions on Vehicular Technology*, Vol. 38, May 1989, 69–75.

Lee, W. C. Y. "Overview of Cellular CDMA," *IEEE Transactions on Vehicular Technology*, Vol. 40, No. 2, May 1991, 291–302.

Rappaport, S. T. *Wireless Communications*. New Jersey: Prentice Hall, 1996.

Viterbi, A. J. "When Not to Spread Spectrum-A Sequel," *IEEE Communications Magazine*, Vol. 23, April 1985, 12–17.

Viterbi, A. J., and Roberto, T. "Implications of Mobile Cellular CDMA," *IEEE Communications Magazine*, Vol. 30, No. 12, 1992, 38–41.

Walrand, J. *Communications Networks: A First Course*. Homewood, IL: Irwin, 1991.

Woo, 111, and Cho, Dong-Ho. "A Study on the Performance Improvements of Error Control Schemes in Digital Cellular DS/CDMA Systems," *IEICE Transactions Communications,* Vol. E77-B, No. 7, July 1994.

Yee, N., Linnartz, J -P., and Feltweis, G. "Multi-Carrier CDMA in Indoor Wireless Radio Networks," *IEEE Proceedings of PIMRC*, Vol. 1, September 1993, 109–113.

Handoff Technologies 18

18.1 Introduction

In a cellular environment, a large geographical area is divided into small areas each covered by a cell site or base station (BS). When a mobile originates a call, a dedicated circuit has to be established between the mobile and the called party. The first link of the circuit is a wireless link between the mobile and the closest BS. The second link is established between the BS and mobile switching centre (MSC), which can be through a wireless or a wired media. When a mobile phone user moves from one cell (one base station) to another cell (another base station) continuation of an active call is one of the most important quality measurements in the cellular systems. Call-handoff or handover is the process where a *call-in-progress* is seamlessly transferred from one cell BS to another (via a channel change) while maintaining the call's connection to the cellular system via the MSC. Generally, a handoff is performed when the quality of the link between the BS and the mobile terminal on the move is decreasing.

> *The switching of an ongoing call to a different channel or cell (base station) is known as handover or handoff.*

Handoff is divided into three main categories: hard, soft, and softer handoffs. Hard handoff is characterized by *break-before-make* and soft handoff is characterized by *make-before-break*. In hard handoffs, current resources are released before new resources are used. In soft handoffs, both resources exist and new resources are used during the handoff process. Softer handoff refers to a mobile that communicates with two sectors within the cell.

All the handoff forms enable the cellular phone to be connected to a different cell or different cell sector. Handoffs used in GSM phones which use time division multiple access (TDMA) technology are different from code division multiple access (CDMA) system. These are performed in slightly different ways and are available under different conditions. Various handoff types, exact implementation, and specific handoff procedures are discussed in this chapter.

18.2 Handoff

Mobility is the most important feature of a wireless cellular communication system. Even though the mobile phone users are moving inside the cells, they are able to communicate with other users or callers. Each cell has a BS which provides frequency channels to the mobile phones. The BSs are also known as cell sites. All the BSs are linked to the MSCs which are responsible for controlling the calls and acting as a gateway to other networks. When an active user (say mobile phone using a frequency channel) reaches the boundary of the cell, it needs to change its current frequency channel to another belonging to the neighbour cell without any interruption to the call.

> *The network procedure, which switches a connection of a call made by a mobile phone, from one BS or cell to another, is called a handover or handoff.*

The handoff transfers the *ongoing calls* through the communication channels in terms of *time slot* (TDMA), *frequency band* (FDMA), *code word* (CDMA), or a hybrid scheme to a new BS.

Why handoff

The handoff procedure is a very important parameter in cellular networks that affects connection quality and also the phone call continuity.

The purposes of handoff are as follows:

- If the quality of a communication has become worse than a threshold, a decision of handoff is made for rescuing connection.
- It keeps a continuous communication with a moving mobile, that is it avoids call termination.
- It improves the cellular network performance by reducing the call drop rate and the congestion rate.
- It frees up some capacity for other users.

18.2.1 Handoff parameters

The following basic parameters are needed to determine whether a handoff is required or not.

1. Signal strength of the BS with which communication is being made.
2. Signal strengths of the surrounding BSs.
3. Availability of channels.

The handoff parameters are measured in the following way:

1. Signal strengths of BSs are measured by the mobile devices.
2. Channel availability status is known at the cellular network.
3. Cellular network makes the decision about when the hand over is to take place in which channel of which cell.

18.2.2 Handoff process in cellular mobile communication

The mobile phone continuously monitors the signal strength of the surrounding BSs, including the one which is currently using, and it feeds this information back. When the strength of the signal from the current BS starts to fall to minimum acceptable level, the cellular network looks at the reported strength of the signals from other BSs reported by the mobile. It then checks for channel availability, if it finds a channel, it informs the new BS to reserve a channel for the incoming mobile. When ready, the current BS passes the information for the new channel to the mobile. When the mobile on the new channel sends a message, it informs to the network that it has arrived. If this message is successfully sent and received then the network shuts down communication with the mobile on the old channel, freeing it up for other users, and all communication takes place on the new channel. Handoff process in GSM is illustrated with an example in Figure18.1.

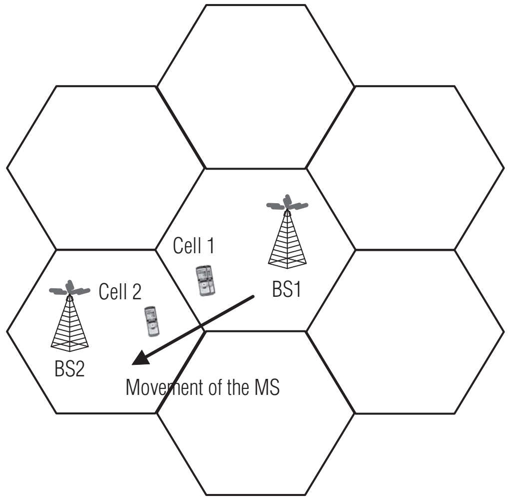


Figure 18.1 Handoff between BS1 and BS2 (Cell 1 to Cell 2)

In Figure 18.1, handoff takes place when the mobile is moving from BS1 to BS2. When the mobile is moving from BS1 towards BS2 the signal strength of BS1 reduces to minimum acceptable level or threshold and the signal strength of BS2 is dominated. Then the call automatically connects to BS2, that is call is transferred without any interruption.

Handoff situation will become worsen if the number of mobile users per cell increases. With the increased rate of mobile users per cell, a big cell can be divided into more number of smaller cells. Hence, more number of handoff situations occurs.

18.2.3 Simple handoff algorithm

The simple handoff algorithm is based on the received signal strength (RSS) and the threshold. In this algorithm a threshold "*T*" is to be defined at a set level, that is above the minimum acceptable signal level.

$$T \text{ (dBm)} = \text{Minimum acceptable signal strength level (dBm)} + \delta_{th} \text{ (dB)} \tag{18.1}$$

When the RSS falls below the threshold "T", a handoff is initiated. The following are several design tradeoffs in this algorithm:

1. The additive term "δ_{th}" cannot be too large as it will result in too many dropped calls due to the signal dropping below an acceptable level before the handoff is completed.
2. At the same time the "δ_{th}" cannot be too small as it will result in too many handoffs, by putting burden on MSC.
3. Averaging of the received signal level must be performed, to ensure the dropping of the call below the threshold is not just too momentary fading. However, averaging introduces a delay in processing.

This trade-off must be used in the design of the system. In 1G cellular network the handoff time is often on the order of 10 s and "δ_{th}" is typically set between 6 and 12 dB. With 2G TDMA

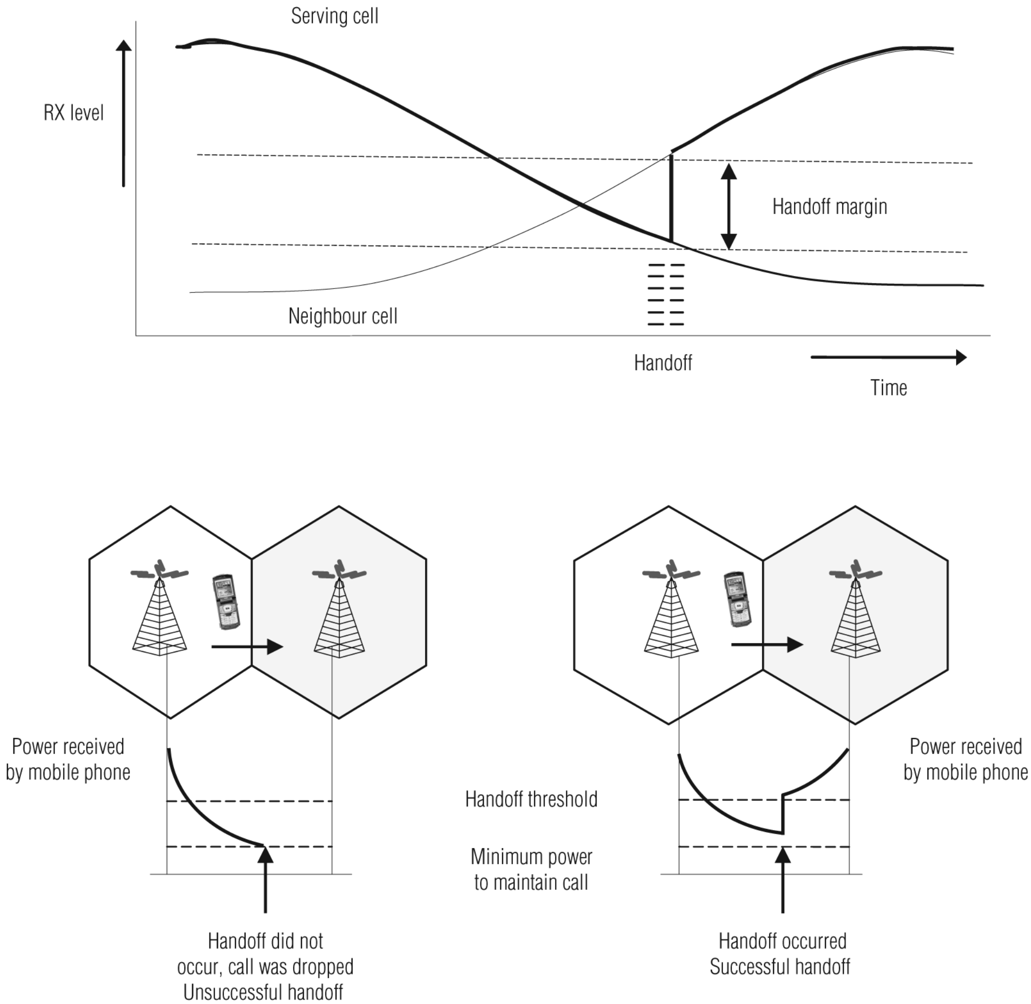


Figure 18.2 Handoff process between serving cell and neighbour cell

systems, handoff time is often on the order of 1 to 2 s and "δ_{th}" is typically set between 0 and 6 dB. Figure 18.2 illustrates the handoff process between serving cell and neighbour cell.

It is observed from the figure that the handoff process between the serving cell and the neighbour cell is based on the RSS of the BSs which follows Equation (18.1). The handoff does not occur, if the RSS of the neighbouring cell is more than the serving cell. If the RSS of serving cell falls below the threshold or handoff margin, then handoff takes place.

18.2.4 Handoff scenarios

Depending on the BS and MSC arrangement of the cellular network the handoff may occur in the following scenarios, based on the movement of a mobile station (MS).

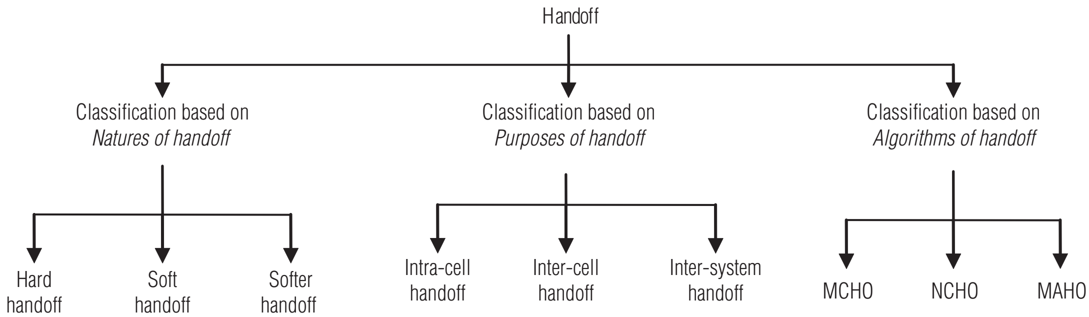


Figure 18.3 Classifications of handoffs

1. *Inter-carrier handoff*: Handoff between one channel of a BS to another (within the same BS) when there is too much traffic in the first channel.
2. *Intra-BSC handoff:* Handoff between cells (base transceiver stations) under the control of the same base station controller (BSC).
3. *Inter-BSC handoff:* Handoff between cells under the control of different BSCs, but belonging to the same MSC.
4. *Inter-MSC handoff:* Handoff between cells under the control of different MSCs.
5. *Intra-network handoff:* The handoff carried out when the MS moves *across the border of cells managed by two* MSCs, which are relatively free, in order to prevent interruption in communication.
6. *Inter-network handoff:* The handoff due to roaming across GSM networks of *two wireless operators* in which the real-time communication will be interrupted and the connection will have to be re-established.

18.2.5 Types of handoffs

Handoffs can be classified based on three decisive factors, the natures of handoff, the purposes of handoff, and the algorithms of handoff. These classifications are further classified into various handoffs as shown in Figure 18.3 and each one is explained in the following sections.

18.3 Classification based on natures of handoff

In this classification, the handoff mechanism is usually categorized as follows:

- Hard handoff
- Soft handoff
- Softer handoff

18.3.1 Hard handoff

A hard handoff is also known as break-before-make handoff.

Usually, a mobile phone communicates with one BS in a given cell. In hard handoff, as the mobile crosses the border to another cell, the communication between mobile and BS is first broken before the communication is started between mobile and the other BS. As a consequence, the transition is not smooth. Although there is generally a short break in transmission, this is normally short enough and not to be noticed by the user.

Figure 18.4 illustrates the *hard handoff* between BS1 and BS2. In Figure 18.4(a) the dark (solid) line indicates connection between BS1 and mobile. The dotted line indicates no connection and request for establishing new connection with BS2, which indicates the mobile before handoff. In Figure 18.4(b) the dark line indicates connection between BS2 and mobile, similarly the dotted line indicates no connection with BS1, which indicates the mobile after handoff.

Hard handoff can be seamless or non-seamless. Seamless handoff means the handoff that is not perceptible to the user. In general, a handoff that requires a change of the carrier frequency (inter-frequency handoff) is always performed as hard handoff.

A hard handoff can be practically employed with more efficiency in frequency-division multiple access (FDMA) and TDMA network access schemes. In FDMA, different frequency ranges are used for adjacent channels in order to minimize the channel interference. So when the mobile moves from one BS to another BS, it becomes impossible for it to communicate with both BSs (since different frequencies are used). Analogue AMPS, GSM, and D-AMPS are the examples of the cellular mobile systems where hard handoff is implemented. Hard handoff implementation is also fairly simple. The MS performs a handoff when the signal strength of a neighbouring cell exceeds the signal strength of the current cell with a given threshold.

Hard handoff is less expensive to implement and is more spectral (network bandwidth) efficient, particularly in heavy data traffic environments.

Hard handoff process in GSM systems

Based on location and usage there are four different types of handoff used in the GSM system:

1. Handoff between channels in the same cell – inter-carrier handoff. In this case, the mobile is diverted to a different traffic channel within the same cell. This channel is generated with a different frequency or time slot. The decision about the handoff is made by the BSC that controls the cell.

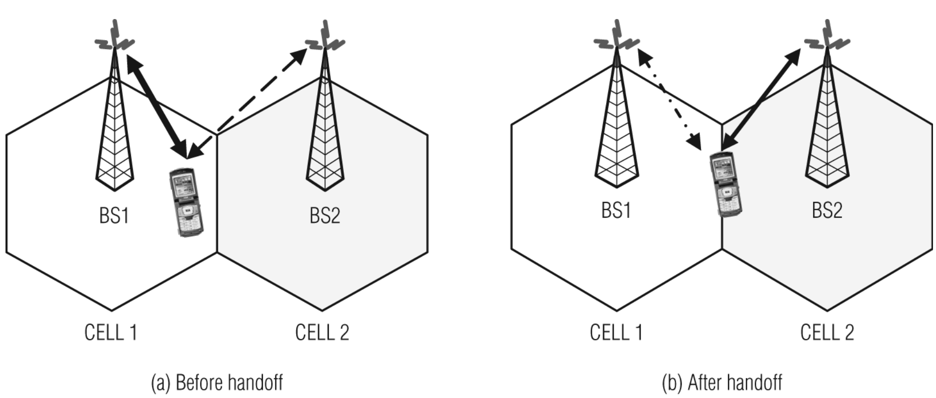


Figure 18.4 Hard handoff between BS1 and BS2

2. Handoff between cells (base transceiver stations) under the control of the same BSC – intra-BSC handoff. This handoff takes place when a mobile user moves from one cell into a neighbouring cell both are controlled by the same BSC. The communication with the old cell is discontinued as soon as the connection is established with the new cell. This process is controlled by the BSC.
3. Handoff between cells under the control of different BSCs, but belongs to the same MSC – inter-BSC handoff.
4. Handoff between cells under the control of different MSCs – inter-MSC handoff.

The first two types of handoff involve only one BSC without involving the MSC, except for notifying the completion of the handoff. To save the signalling bandwidth, they are managed by the BSC. The last two types of handoff are managed by the MSCs. An important aspect of GSM is that the original MSC is solely responsible for the most call-related functions, with the exception of subsequent inter-BSC handoffs under the control of the new MSC.

In GSM systems, the handoff can be initiated by either the mobile or the MSC. During its idle time slots, the mobile scans the broadcast control channel up to 16 neighbouring cells and forms a list of the six best cells for possible handoff based on the RSS. This information is passed to the BSC and MSC, at least once per second and this is used by the handoff algorithm.

18.3.2 Soft handoff

The soft handoff is also known as make-before-break handoff.

In soft handoff, the handoff from one BS to another occurs in a smooth manner. In this the MS keeps its communication with the original BS, until it establishes a connection with the other BSs. The excess connections are given up only when the new link has the sufficient quality.

In soft handoff, unlike hard handoff, for smooth transition of a call from one cell to another, the mobile continues to talk to both the cells. The soft handoff between two BSs (BS1 and BS2) is illustrated in Figure 18.5. As the mobile moves from cell A (BS1) to cell B (BS2), at some point the communication is broken with the old BS of cell A. During this handoff process, the MS

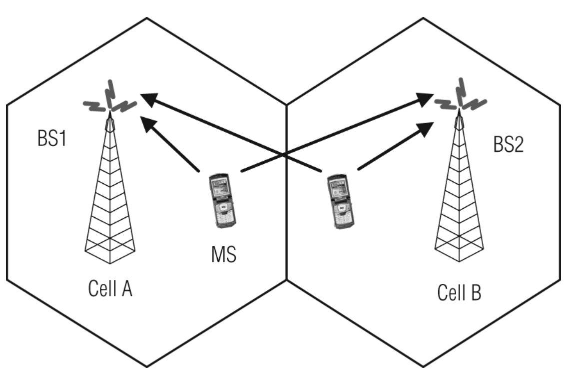


Figure 18.5 Soft handoff between BS1 and BS2

remains in communication with the original cell as well as with the new cell. The handoff is completed when the mobile selects the best BS as the target. Soft handoff is more careful in selecting the target BS, because the target BS needs to provide the strongest signal among the available BSs.

Normally soft handoff can be used when cells operated on the same frequency are changed.

The new 3G technologies use CDMA where it is possible to have neighbouring cells on the same frequency. This opens the possibility of having a form of handoff where it is not necessary to break the connection. In UMTS most of the handoffs that are performed are intra-frequency soft handoffs.

Soft handoff process in CDMA systems

Soft call-handoffs are different from hard call-handoffs. The soft handoff allows both the original cell and one to two new cells to service a call during the handoff transition. This is achieved because all the cell sites are transmitting the same frequency, unlike an analogue system where each cell site has a unique set of frequencies. In CDMA wireless systems, 35% to 45% of the capacity of each BS can be reserved for soft handoff processing.

CDMA systems require a GPS receiver and antenna at every BS. The GPS antennas synchronize all cell sites to one timing source. This is an absolute necessity for soft handoffs because timing is critical among the multiple sites that may simultaneously handle a call during the soft handoff process.

The sequence of events in a soft handoff is as follows:

1. After a mobile call is initiated, the MS continues to scan the neighbouring cells to determine the strongest signals than that of the original cell.
2. When this happens, the MS knows that the call has entered a new cell's coverage area such that a handoff can be initiated.
3. The MS transmits a control message to the MSC, which states that the mobile is receiving a stronger signal from the new cell site, and the mobile identifies that new cell site.
4. The MSC initiates the handoff by establishing a link to the MS through the new cell while maintaining the link to the old cell that was managing the call.
5. Although the MS is located in the transition region between the two cell sites, the call is supported by communication through both cells. This eliminates the ping-pong effect of repeated requests to handle the call back and forth between two cell sites.
6. The original cell site will discontinue handling the call only when the MS is firmly established in the new cell.

Comparison of hard handoff and soft handoff

The comparison of hard handoff and soft handoff is illustrated in the Figure 18.6.

It is shown that in soft handoff, the connection with new cell is established before the connection with old cell is disconnected. Whereas, in the hard handoff before making the connection with new cell the connection with old cell becomes weaker and will break soon. CDMA cellular systems such as IS-95 use *soft handoff*, while TDMA cellular systems such as GSM and D-AMPS typically use *hard handoff*.

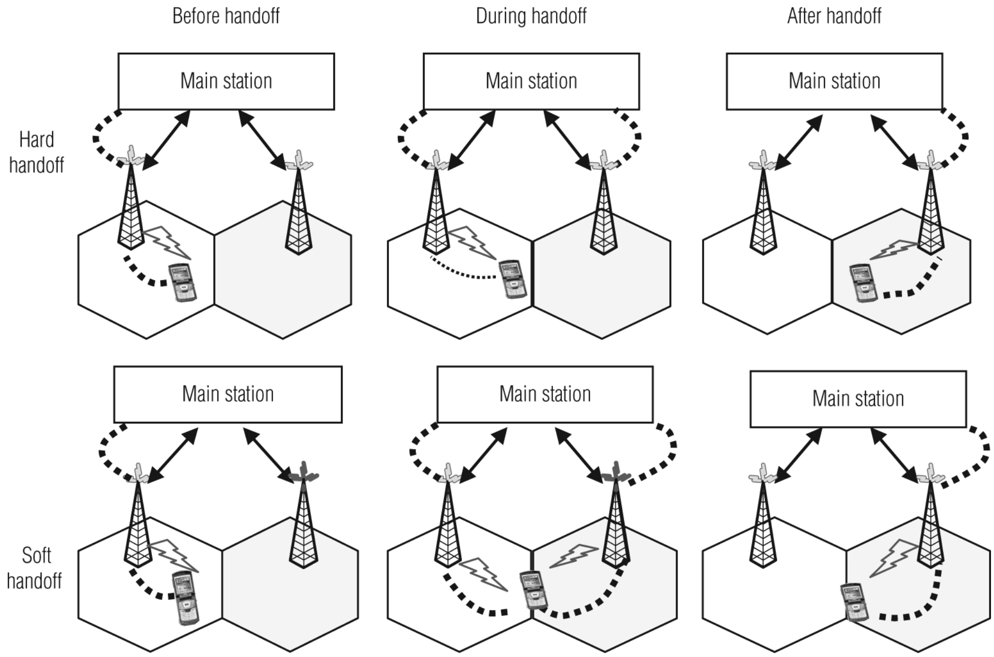


Figure 18.6 Comparison of hard handoff and soft handoff

Advantages of soft handoff

1. Soft handoff reduces/eliminates the "ping-pong" effect which results in
   i. fewer load on the network from handoff signalling and overhead
   ii. smoother user communications without the "clicks" when speech transmissions are stopped momentarily during handoffs
2. The soft handoff results in less delay and equivalent to "instantaneous" macroscopic selection diversity. This is accomplished by instantaneous switching to the best BS signal during a soft handoff (uplink), and avoids the additional interference associated with handoffs. Hence,
   i. Keeping BS separations and transmitter powers (BS and user) fixed, the overall uplink interference is reduced, leading to
      a. better communication quality for a given number of users
      b. more users (i.e. greater capacity) for the same required E_C/I_0 (ratio of received energy per chip to the total received spectral density)
      c. smaller required uplink transmitter powers, further reducing uplink interference
   ii. Keeping required outage probability and BS separation fixed, the system fade margins are reduced. This leads to smaller required downlink transmitter powers and downlink interference.
   iii. Keeping the same required outage probability and fade margins, BS separation increases.

3. Soft handoff imposes fewer time constraints on the network. There is a longer mean queuing time to get a new channel from the target BS, so this helps to reduce the blocking probability or probability of dropped calls.

Disadvantages of soft handoff

1. Additional network resources are used during soft handoffs.
2. Soft handoff is more complex and expensive to implement.
3. Downlink interference (to other users) increases when soft handoff is in progress.
4. Forward interference increases with soft handoff since several BSs can connect to the same MS. This increase in forward interference can become a problem if the handoff region is large, such that there are many MSs in soft handoff mode.

18.3.3 Softer handoff

A softer handoff occurs when the mobile is communicating with two sectors of the same BS in a given cell simultaneously.

Cell sectorization

One approach to increase the subscriber capacity of a cellular network is replacing the omnidirectional antenna at each BS by three or six directional antennas with a 120° or 60° opening of the corresponding receivers to illuminate a fraction of a cell. Each sector can be recognized as a new cell, with its own (set of) frequency channels.

By using directional antenna on a BS, each pointing in different directions, it is possible to sectorize the BS so that several different cells are served from the same location. This increases the traffic capacity of the BS.

An example of a cell with three sectors is shown in Figure 18.7. In this case, each sectoral antenna uses a 120° beamwidth directional antenna pattern in the horizontal plane. Assuming that the antenna radiation pattern is such that the gain is virtually zero outside the beam, the interference outside of the 120° sector radiation angle is essentially suppressed, which reduces the relative other cell interference by a factor of 3 and improves CDMA capacity by the same factor. Three-sector cells are widely deployed in CDMA cellular systems, except for indoor or picocell systems where distributed antenna systems may be better suited.

The use of directional sector antennas substantially reduces the interference among co-channel cells. This allows denser frequency reuse.

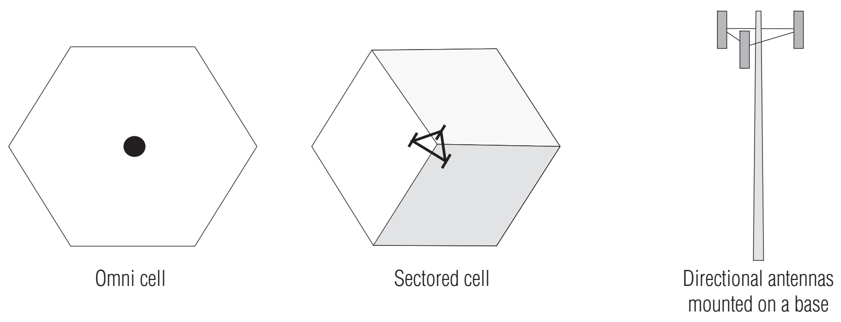


Figure 18.7 Example of sectorization with three sectors per cell

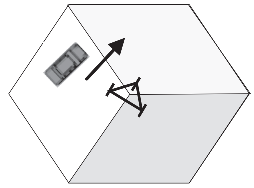


Figure 18.8 Example of softer handoff

For Example, if the cluster size is 7, in a three-sector antenna, each sector uses one-third of the allocated channels and the mobile is interfered by two BSs instead of six. The signal-to-interference (S/I) ratio in a three-sector (120° each sector) antenna is given by

$$(S/I)_{120^\circ} = (S/I)_{360^\circ \text{ or omni}} + 10\log 3 = (S/I)_{360^\circ \text{ or omni}} + 4.8\text{dB}$$

Sectorization is less expensive than cell-splitting, as it does not require the acquisition of new BS sites.

Softer handoff may also occur when a signal is replaced by a stronger signal from a different sector under the same BS. This type of handoff is available within UMTS as well as CDMA2000.

Softer handoff refers to the handoff between the sectors of a cell. Soft handoff explained in the above section refers to a feature used by the CDMA standards, where a cell phone is simultaneously connected to two or more cells (or cell sectors) during a call. If the sectors are from the same physical cell site (a sectorized site as shown in Figure 18.8), it is referred to as *softer handoff*. The forward link is the same as soft handoff. In reverse link there is no need to communicate with other BS. Thus, it is faster.

Advantages of softer handoffs

- Softer handoff is handled by BS internally
- Softer handoff probability is about 5%–15%
- No extra transmissions between BSs are needed

CDMA system uses soft and softer handoff technique to improve receptions when MSs move between cells or sectors (on cell or sector boundaries).

18.4 Classification based on purposes of handoff

In this classification the handoff can be of three types: intra-cell handoff, inter-cell handoff, and inter-system handoff. The classification is based on network view and is shown in Figure 18.9.

1. If the mobile unit is assigned a new channel within the same BS or cell is referred to as intra-cell handoff. This is done to avoid interference or channel quality reasons.
2. If the connection to a mobile unit is transferred over the cell boundary to a new cell or BS, the handoff is known as inter-cell handoff. Inter-cell handoff is mainly due to weak signal or bad channel quality or traffic loading balancing reasons.

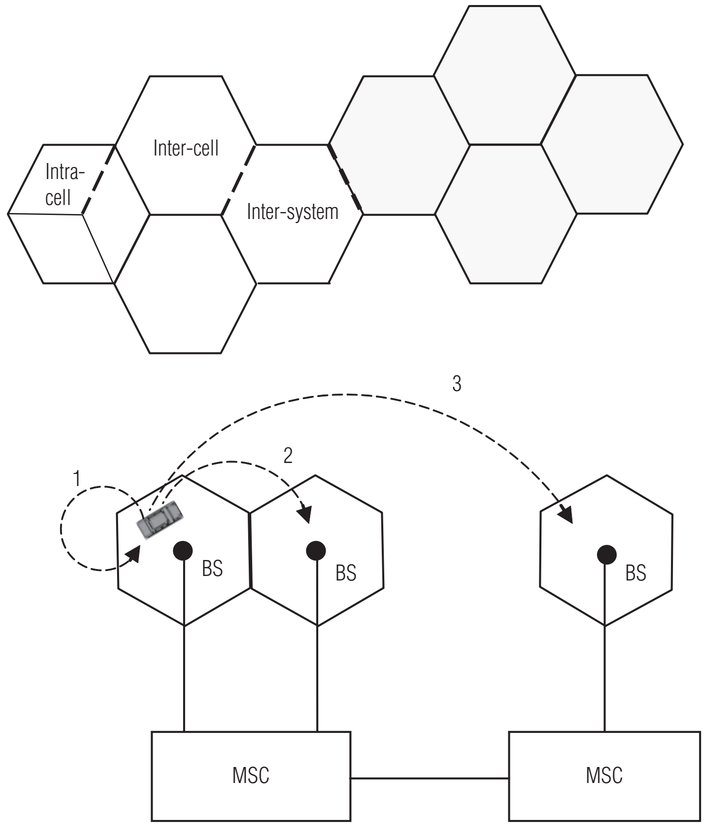


Figure 18.9 Classification based on purposes of handoff

3. If the handoff takes place between two MSCs of different cellular systems is known as inter-system handoff.

These three handoff types are explained in detail in the following sections.

18.4.1 Intra-cell handoff

The demand for wireless service is increasing day by day and the number of frequencies assigned to a cell became insufficient to support the required number of mobile subscribers. Thus, a cellular design technique is referred to as *cell sectoring*, where a single omni-directional antenna is replaced with several directional antennas at the BS. In general, a cell is sectorized into "n" number of sectors. A cell with three sectors is shown in Figure 18.10. In this case, the total 360° of cell area is sectored into three 120° individual areas and the handoff in between sector-to-sector is called as *intra-cell handoff*.

The handoff between two sectors of a cell is known as intra-cell handoff.

18.4.2 Inter-cell or inter – BS handoff

The most basic form of handoff is when a phone call is in progress is redirected from its current (source) cell channel to a new (target) cell channel. In terrestrial networks the source and the

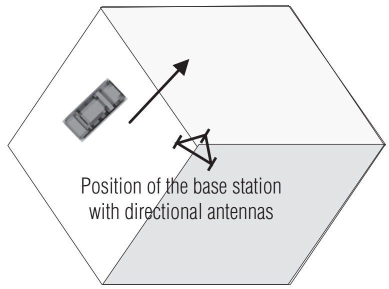


Figure 18.10 Intra-cell handoff

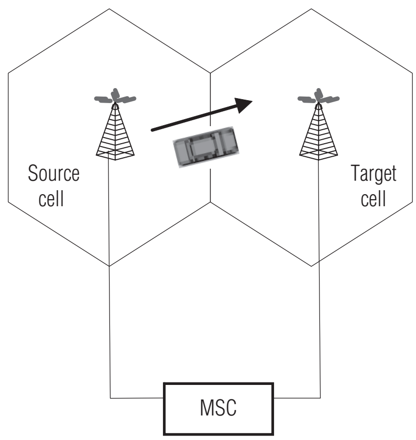


Figure 18.11 Inter-cell handoff

target cells may be served from two different cell sites or from one and the same cell site. Such a handoff, in which the source and the target are different cells is called *inter-cell handoff* as shown in Figure 18.11.

The purpose of inter-cell handoff is to maintain the call as the subscriber is moving out of the area covered by the source cell and entering the area of the target cell.

The inter-cell handoff switches a call in progress from one cell to another cell, and the intra-cell handoff switches a call in progress from one physical channel of one cell to another physical channel of the same cell.

Figure 18.12 illustrates the combination of intra-cell and inter-cell handoffs.

18.4.3 Inter-system handoff

The handoff which takes place when a mobile moves away from one system controlled by an MSC and enters into another system controlled by a different MSC is called as inter-system handoff. Inter-system handoff may be between two same type systems or between two different systems.

As shown in Figure 18.13, the mobile initially in a system (say CDMA) controlled by MSC-A is moving from one place to another place. While it is moving, it enters into another region where

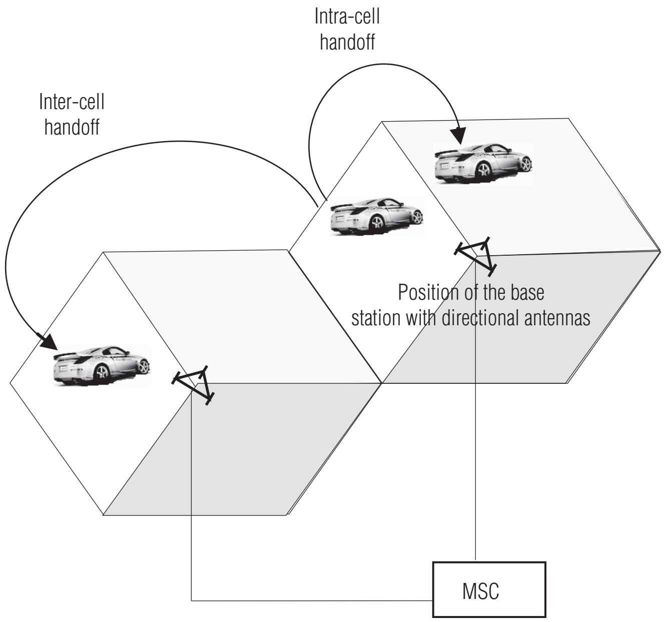


Figure 18.12 Intra-cell and inter-cell handoff

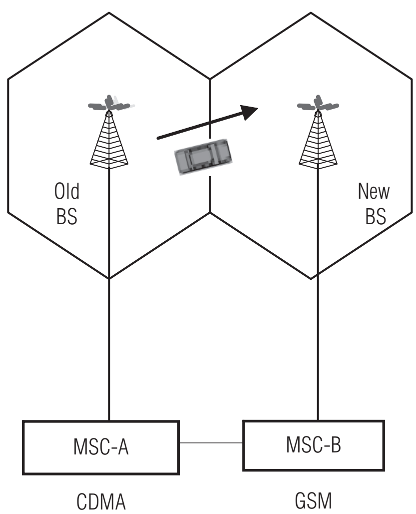


Figure 18.13 Inter-system handoff or Inter-MSC handoff
(The new and the old BSs are connected to different MSCs)

there is no availability of its own system service and in that region another system (say GSM) is working which is controlled by MSC-B. In this case, when the signal strength is decreasing MSC-A searches for new MSC of its own system and if there is no MSC of its own, then it makes a handoff request to another system MSC-B to provide handoff so as to avoid the dropped call.

To perform inter-system handoff, the MSCs need to have special kind of software since the handoff is between different systems and every system will have its own procedure for communication and the software will have to provide required interface between two systems. Inter-system handoff is a very essential technology to perform smooth handoffs between various systems so as to increase the quality of communication.

18.5 Handoff schemes based on algorithms of handoff (handoff protocols)

The mobile unit and the BS are connected via radio links which carry data as well as signalling information. There are three different handoff strategies based on algorithms of handoff, which have been proposed for transferring the connection to a new BS.

- MCHO (mobile-controlled handoff)
- NCHO (network-controlled handoff)
- MAHO (mobile-assisted handoff)

Since the number of handoffs increases with decreasing cell size, it will be an almost impossible task to make a handoff decision for every mobile by one central switch (centralized). Moreover, in microcells the connection between MS and BS can deteriorate very quickly. Fast handoff decisions required in such situations can be achieved more readily by decentralizing the handoff decision process.

18.5.1 Mobile-controlled handoff (MCHO)

In this case, the mobile phone is the only entity which measures the handoff criteria and makes a decision based on them. The MSC is not involved in the handoff process resulting in reduced burden on the MSC. The mobile has to choose the optimum BS based on the measurements. Since the handoff process is implemented in the mobile itself, the delay is usually smaller with a typical value of 0.1 s and is suitable for microcellular systems.

In this strategy, the mobile continuously monitors the radio signal strengths and quality of surrounding BSs. A handoff can be initiated if the signal strength of the serving BS is lower than that of another BS by a certain threshold. Then the mobile requests the target BS for a channel with the lowest interference and handoff mechanism will take place. In such a case, the MS does not have any information about the signal quality of other users, but handoff must not cause interference to other users. MCHO is the highest degree of handoff decentralization. Some of the advantages of handoff decentralization are as follows:

- Handoff decisions can be made fast.
- MSC does not have to make handoff decisions for every mobile, which is a very difficult task for the MSC of high-capacity microcellular systems (radius < 1 km).

An example of a MCHO-based handoff control network is the standard for cordless phones in Europe – digital European cordless telephone (DECT).

18.5.2 Network-controlled handoff (NCHO)

NCHO is a centralized handoff protocol. In this type of handoff the network (*surrounding BSs, the MSC or both*) makes a handoff decision based on measurements of the RSSs of mobile and the interferences from different BSs. The signal-to-interference ratio (SIR) is measured by means of a supervisory audio tone (SAT). If the mobile is measured to have a weaker signal in its old cell, while a stronger signal in a neighbouring cell, then a handoff decision could be made by the network to switch BS from the old cell to the new cell. Such a type of handoff in general takes 100–200 ms and produces a noticeable "interruption" in the conversation. However, overall delay of such a type of handoff is in general in the range of 5–10 s. Thus, this type of handoff is not suitable to a rapidly changing environment and to a high density of users due to the associated delay.

The NCHO is widely used in the first-generation cellular systems, such as AMPS, Total Accesses Communications System (TACS) and Nordic Mobile Telephone (NMT). In NCHO, the MSC is solely in charge of the handoff process and the MSs are completely passive.

18.5.3 Mobile-assisted handoff (MAHO)

MAHO is a variant of NCHO strategy. To improve the handoff reaction time and to reduce the handoff administration load of the MSC, the handoff decisions should be distributed towards the mobile phones. One way to achieve this could be to let the mobile phones make the measurements and the MSC make the decisions.

In the MAHO strategy, the network (BS and/or MSC) directs the mobile to measure the signal strengths from the surrounding BSs and to report those measurements back to the network. The network then uses these measurements to determine where a handoff is required with which channel. The delay in this protocol starting from the handoff initiation till the handoff execution is around 1 s. This time may still be too long to avoid dropping a call due to street corner effect.

Some examples of present cellular networks which implement MAHO are the GSM system and the IS-95 system.

Comparison of MAHO and MCHO performances

There would not be a major difference in the performance between MAHO and MCHO, if all signalling on the air interface were error free. The critical difference is that in MAHO a handoff request is transmitted from the BS to the MS. If that message is not received correctly the call may be dropped. Also if new BSs are not identified or recent measurement reports are missing, the handoff request might be delayed causing a call dropout.

18.6 Handoff initiation techniques

Handoff initiation means when a handoff should be initialized. It is the process of deciding when to request a handoff. Handoff decision is based on RSS from current BS and neighbouring BSs. Initiation of the handoff may begin when the signal strength at the mobile received from BS2 is greater than that of BS1.

When the mobile is moving from one BS (BS1) to another (BS2), the mean signal strength of BS1 decreases as the mobile moves away from it. Similarly, the mean signal strength of BS2 increases as the mobile approaches it. The signal strength measures the signal levels averaged over a chosen amount of time. The following are different methods for handoff initiation:

- Relative signal strength
- Relative signal strength with threshold

- Relative signal strength with hysteresis
- Relative signal strength with hysteresis and threshold
- Prediction techniques

18.6.1 Relative signal strength

This method prefers the strongest signal received by BS at all times. The selection depends upon a mean measurement of the received signal. In Figure 18.14, the handoff would occur at position A. The RSSs of current BS (BS1) and one neighbouring BS (BS2) are shown in Figure 18.15. The RSS gets weaker as mobile unit goes away from BS1 and gets stronger as it gets closer to the BS2 as a result of signal propagation.

A major problem with this approach is that the received signals of both BSs often fluctuate if the mobile is moving and cause it to rapidly switch links with either BS. The BSs bounce the link with the mobile back and forth. This phenomenon is called *ping-ponging*. Due to this ping-pong effect, several handoffs can be requested while BS1's RSS is still sufficient to serve mobile unit to cause unnecessary handoffs.

To avoid unnecessary handoffs, a better method is to use the averaged signal levels relative to a threshold and hysteresis margin for handoff initiation is to be used.

Ping-pong handoff

Ping-pong handoff is an undesirable effect that frequently occurs during handoffs. It is a handoff made to neighboring cell, but it returns to the original cell after a short time (<10 s) due to power budget criterion. A handoff is mainly executed on the basis of this criterion in cells with good radio coverage and only minimal disruption due to interference. To avoid such "ping-pong"

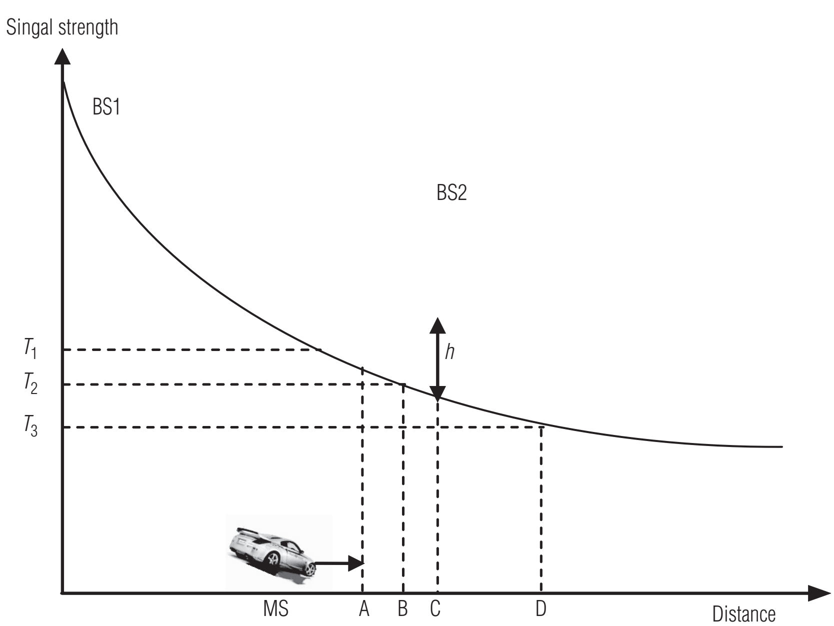


Figure 18.14 Illustration of handoff initiation techniques

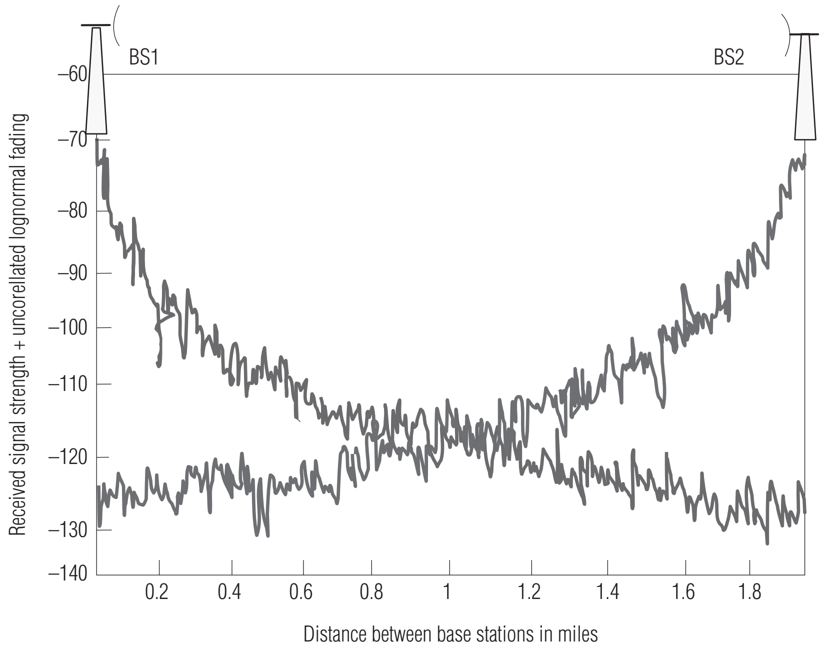


Figure 18.15 Handoff mechanism based on signal strength measurement between two base stations

effect the MS is allowed to continue maintaining a radio link with the current BS, until the signal strength from BS2 exceed that of BS1 by some pre-specified threshold (handoff margin) value *H*. Besides transmitting power, the handoff also depends on the mobility of the MS. In normal operation, a threshold (*H*) of 5–10 dB is used to prevent minor variation in signal level of different BSs from causing a handoff. Strong shadow fading caused by large obstacle can cause up to 30 dB. If such an obstacle is found in the line-of-sight (LOS) of the serving BS but not of the neighbouring station, it is possible that a handoff may be triggered. As soon as the MS moves out of the shadowing area the level again become normal and handoff takes place to the original cell. A medium and high mobility MS results in handoff to a neighbouring station back to the original BS within a short period of time (<10 s) (ping-pong handoff).

Handoff is only possible theoretically between points A and C, but is recommended at point B because that is where the level of BS2 has first fallen below the allowable hysteresis value *h*, in other words sufficiently above that of BS. Due to propagation-related signal level fluctuations at the receiver, the curved characteristics shown only apply to the statistical average as shown in Figure 18.14. Therefore, the location of a handoff is randomly distributed in the area around B.

18.6.2 Relative signal strength with threshold

This method introduces a threshold value, T_1,(Figure 18.14) to overcome the ping-pong effect. The handoff is initiated if BS1's RSS is lower than the threshold value and at the same time BS2's RSS is stronger than BS1's. So according to this method the handoff request is issued at point B as shown in Figure 18.14.

Effect of threshold value

The effect of the threshold depends on its relative value compared to the signal strengths of point A. If the threshold is higher than this value, say T_1, this scheme performs exactly like the relative signal strength scheme, so the handoff occurs at position A. If the threshold is lower than this value, say T_2, the mobile unit would delay handoff until the current signal level crosses the threshold at position B. In the case of T_3, the delay may be so long that the mobile unit drifts too far into the new cell.

18.6.3 Relative signal strength with hysteresis

This scheme allows a user to handoff only if the new BS is sufficiently stronger (by a hysteresis margin, "h" in Figure 18.14) than the current one. In this case, the handoff would occur at point C, which results in ping-pong effect (the repeated handoff between two BSs caused by rapid fluctuations in the RSSs from both BSs).

18.6.4 Relative signal strength with hysteresis and threshold

This scheme hands a MS over to a new BS only if the current signal level drops below a threshold and the target BS is stronger than the current one by a given hysteresis margin. In Figure 18.14, the handoff would occur at point D if the threshold is T_3.

In this method, we combine both the threshold and hysteresis value so as to reduce number of unnecessary handoffs. The handoff is requested when the BS1's RSS is below T_1 (Figure 18.16) and BS2's RSS is stronger than BS1's by the hysteresis value (point C in Figure 18.16). If we would

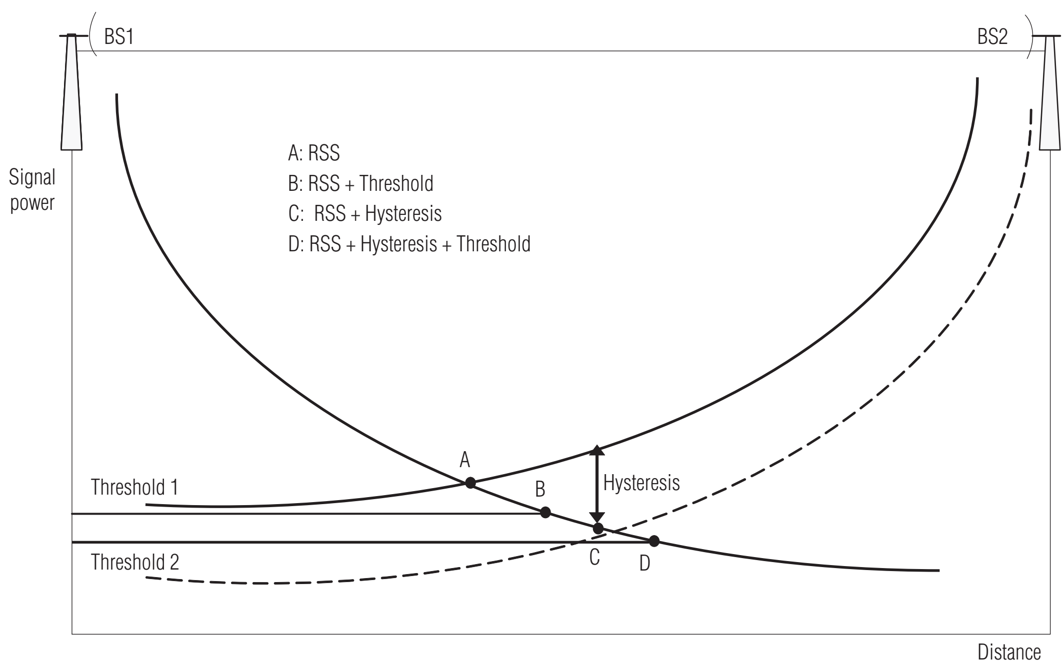


Figure 18.16 Traditional handoff algorithms using RSS, threshold and hysteresis

choose a lower threshold level than T_1 (but higher than T_2), then the handoff initiation would be somewhere at the right of point C (left of point D).

18.6.5 Prediction techniques

Prediction techniques base the handoff decision on the expected future value of the RSS. These techniques have been proposed to indicate a better result, which means reduction of unnecessary handoffs than previous approaches.

The salient features of the above five traditional handoff algorithms which are based on the RSS or received power "*P*" are given below:

- RSS (choose a BS (BS_{new}) if $P_{new} > P_{old}$)
- RSS plus threshold (if $P_{new} > P_{old}$ and $P_{old} < T$)
- RSS plus hysteresis ($P_{new} > P_{old} + H$)
- RSS, hysteresis, and threshold ($P_{new} > P_{old} + H$ and $P_{old} < T$)

18.7 Basic cellular structures

At present, there are different types of cellular network structures and that the related handoff procedures are also changing with these cellular structures. A handoff algorithm with fixed parameters cannot perform well in different system environments. Specific characteristics of the communication systems should be taken into account while designing handoff algorithms. Various types of cell structures based on the cell coverage area are shown in Figure 18.17. They are given in Table 18.1.

Advantages of decreasing cell size:

- Increased user capacity.
- Increased number of handoffs per call.
- Increased complexity in locating the subscriber.
- Lower power consumption in mobile terminal, so it gives longer talk time and safer operation.

The two important basic cellular structures, macrocells and microcells, along with handoff mechanisms are described in the following sections.

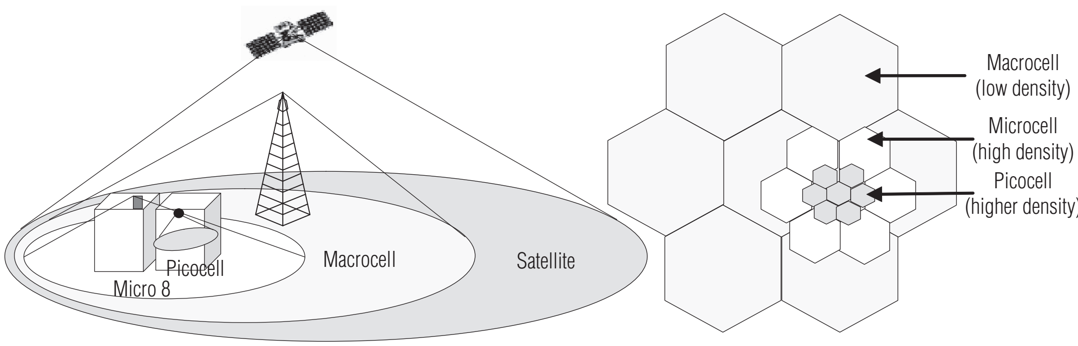


Figure 18.17 Basic cellular structures based on coverage area

Table 18.1 Different cell structures

S.No	Type of cell	Coverage area (metres)
1.	Macrocell	1–30 km
2.	Microcell	200–2000 m
3.	Picocell	4–200 m

18.7.1 Macrocell

Macrocells are big cells and generally their radii are in several kilometres. Since they are big in size there is low cell crossing rate. Due to the low cell crossing rate, centralized handoff is possible despite the large number of MSs the MSC has to manage. The signal quality in the uplink and downlink is approximately the same. The transition region between the BSs is large.

Handoff schemes should allow some delay to avoid flip-flopping. However, the delay should be short enough to preserve the signal quality because the interference increases as the MS penetrates the new cell. This cell penetration is called "cell dragging". Macrocells have relatively gentle path loss characteristics. The averaging interval (i.e. the time period used to average the signal strength variations) should be long enough to get rid of fading fluctuations.

Figure 18.18 shows three clusters of seven cells in a macrocellular system. A cluster consists of a group of cells marked cell1 through cell7 represented from A to G. First- and second-generation cellular systems provide wide-area coverage even in cities using macrocells. Typically, a BS transceiver in a macrocell transmits high output power with the antenna mounted several metres high upon a tower to illuminate a large area.

18.7.2 Microcells

The microcells are cells with small radii and employed in highly populated areas such as city buildings and streets to meet high system capacity by frequency reuse. By cell splitting the big cell is subdivided into smaller cells called as microcells. A microcell is shown in Figure 18.19. Since some capacity improvement techniques such as larger bandwidths, channel coding, and

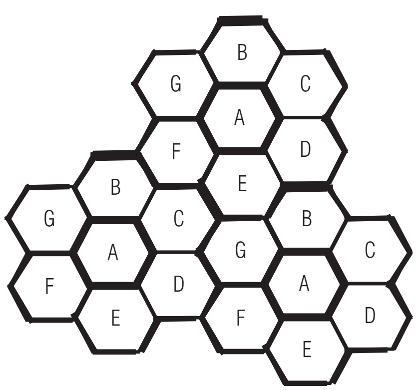


Figure 18.18 Three clusters of seven cells each in a macrocellular system

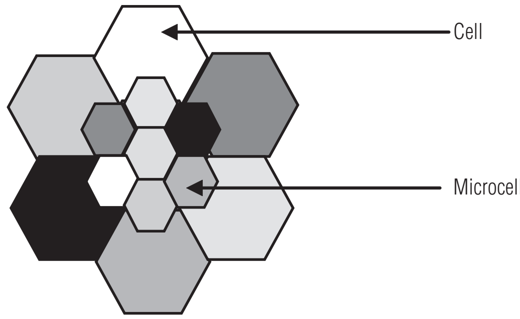


Figure 18.19 A cell spitted into microcells

modulation were not enough to satisfy the required service demand, hence the use of microcells is considered as the most effective means of increasing the capacity of cellular systems.

In microcellular systems the BSs are placed densely to provide for high capacity. The BS antennas are mounted on building walls and lampposts to obtain radio coverage basically along streets, that is LOS coverage. Since the signal strength drops very rapidly as a MS turns around a corner (street corner effect) the call must be handed off to a BS serving the crossing street very quickly in order not to drop the call.

Microcells increase capacity of the channel, but radio resource management becomes more difficult, depending on whether they are along a road or a highway, covering an area such as a number of adjacent roads, or located in multi-level buildings.

Advantages of microcells are as follows:

- In this cell structure, the BS transceiver transmits low output power with the antenna mounted approximately 5 m above ground.
- The MS also transmits low power, which leads to longer battery life.
- Since BS antennas have lower heights compared to the surrounding buildings, RF signals propagate mostly along the streets. The antenna may cover 100–200 m in each street direction, serving a few city blocks.

As a comparison with macrocells, microcells are more sensitive due to short-term variations such as traffic and interference with medium- or long-term alterations (e.g. new buildings).

The main disadvantage with microcell structure is the number of handoffs per cell is increased by an order of magnitude, and at the same time the time available to make a handoff is decreased. Due to the increase in the microcell boundary crossings and expected high-traffic loads, a higher degree of decentralization of the handoff process becomes necessary.

In a microcellular system, there are two types of handoff procedures, and they are as follows:

1. Line-of-sight (LOS) handoff
2. Non-line-of-sight (NLOS) handoff

Line-of-sight handoff

LOS handoff is a handoff from one LOS BS to another LOS BS. In Figure 18.20 we have two streets intersecting with three BSs, BS1 and BS3 have LOS with each other. The handoff between BS1 and BS3 is called LOS handoff.

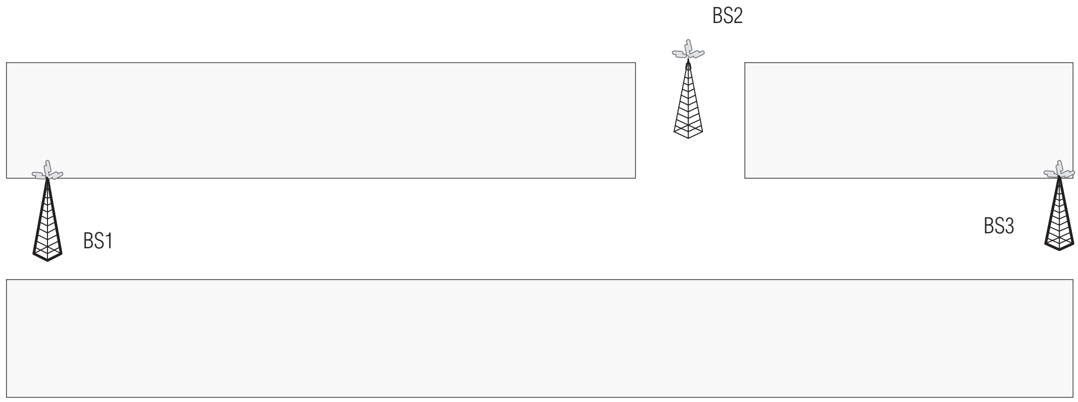


Figure 18.20 A cellular structure with three base stations

Similarly, when a cell is sectorized into more number of sectors as shown in Figure 18.7, if the handoff from a microcell in a sector to its adjacent microcell in the same sector is present, then in such situations LOS handoff can be implemented. In this handoff, premature handoff requests should be prevented.

Non-line-of-sight handoff

A NLOS handoff is a handoff from a NLOS BS to a LOS BS. As shown in Figure 18.20, the handoff between BS1 and BS2 is a NLOS handoff since they don't have LOS. Similarly, when a cell is sectorized into more number of sectors as shown in Figure 18.7, if the handoff from a microcell in one sector to its adjacent microcell in other sector (even though both the microcells are adjacent they become NLOS cells due to sectoring) is required, in such situations NLOS handoff can be implemented. In a NLOS handoff, the handoff must be done as fast as possible as the user turns the corner.

18.8 Delaying handoff

A handoff could be delayed if no available cell could take the call. One of the advantages of a delayed handoff is to make the handoff occur at the proper location. The situations needed to delay the handoff procedure instead of starting are the following:

- When call traffic is heavy, the switching processor is loaded heavily such that a lower number of handoffs would help the processor to handle call processing more efficiently.
- When the mobile unit is located at a signal-strength hole within a cell but not at the boundary.
- When the mobile unit approaches a cell boundary but no channels in the new cell are available (neighbouring cells are busy).

The following are the two approaches for delaying handoffs:

1. Implementing a two-level handoff
2. Queuing the handoff calls

Queuing of handoff approach is more effective than a two-level handoff.

18.8.1 Implementing a two-level handoff procedure

In this approach, based on the RSS, two handoff request levels and a threshold level are defined first. Then average signal strength can be taken as a function of time. Let us consider L1 and L2 are two handoff levels and "x" is the threshold level as shown in Figure 18.21.

When the RSS drops below the first handoff level L1 (Figure 18.21(a)), a handoff request is initiated but does not take place immediately. At this stage for every 5 s handoff request is made and handoff takes place if the new signal from adjacent cell is stronger. The basic advantage with this is when the mobile unit is in low signal area (e.g. may be a hole), where actually handoff is not needed, after few seconds the RSS may increase above L1.

Similarly, if a neighbouring cell is busy, the handoff is needed to delay. In both cases waiting for some time is good instead of starting the handoff immediately. If the handoff reached level L2 (as shown in Figure 18.21(b)), handoff takes place immediately without considering any condition.

Even after second level (L2) if no neighbouring cells (BSs) are available, then the call continues until the RSS drops below the threshold level "*X*"(shown in Figure 18.21(c)) and after the threshold level the call is dropped since no handoff will take place.

18.8.2 Queuing the handoff calls

Queuing is another way of delaying handoffs. The queuing handoff scheme queues the handoff calls. If the BS finds all the available channels occupied in the target cell, a handoff request is put in the queue. If a channel is released when the queue for handoff requests is not empty, the channel is assigned to request on the top of the queue. If the RSS from the current BS falls below the receiver threshold level prior to the mobile being assigned a channel in the target cell, the call is forced to terminate.

Queuing new calls results in increased handoff blocking probability. The probability of a successful handoff can be improved by queuing handoff requests at the cost of increased new call

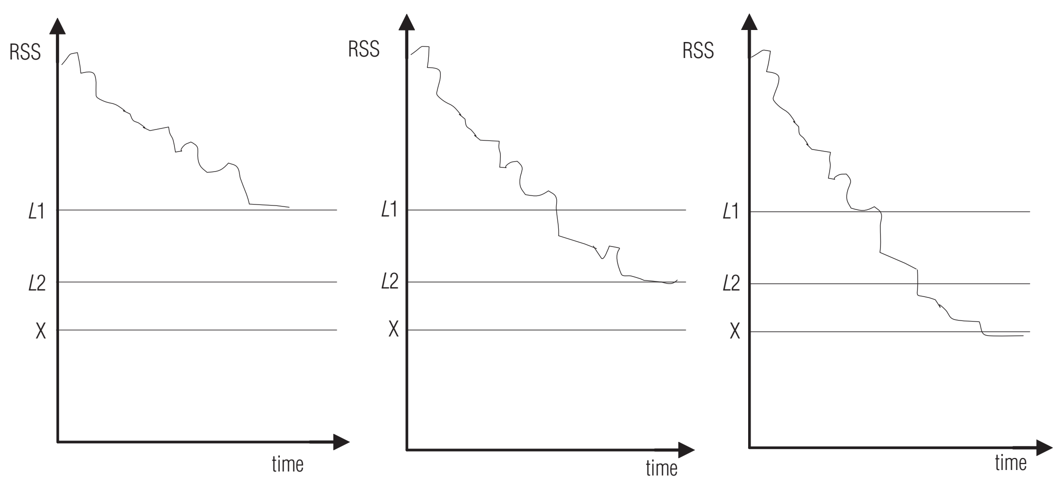


Figure 18.21 Implementation of two-level handoff scheme. (a) Handoff at first threshold L1 (b) Handoff at second threshold L2 (c) Dropped call

blocking probability. This results in a decrease in the ratio of carried-to-admitted traffic since new calls are not assigned a channel until all the handoff requests in the queue are served.

Queuing is possible due to the overlap region between the adjacent cells in which MS can communicate with more than one BS. If handoff requests occur uniformly, queuing is not needed. Queuing is effective only when handoff requests arrive in large number at a time.

Queuing is more effective in macrocell structures since the large area of macrocells allow the MS to wait some time for handoff before signal quality drops to an unacceptable level. However, the effectiveness of queuing decreases for microcells due to stricter time requirements. The combination of queuing and channel reservation can be employed to obtain better performance.

The general model for handoff queuing as shown in Figure 18.22 where

T_m = mean rate of channel holding time

A_H = arrival rate of handoff calls (calls per second)

A_O = arrival rate of originating calls (calls per second)

N = number of channels

In this model N channels are allotted for both originating calls (calls in the cell which do not needed any kind of handoff means they remain in the same cell) and handoff calls. Arrival rate information of both originating and handoff calls can be given to MSC which can allow blocking probability of handoff calls as well as originating calls. Here, we discuss only handoff calls.

The MSC will queue the requests of handoff calls instead of rejecting them if the new cell sites are busy. A queuing scheme becomes effective only when the requests for handoffs arrive at the MSC in batches or bundles. Here, there are three different situations considered and blocking probability for both handoff calls and originating calls obtained are given below:

1. Without queuing
2. Queuing originating calls only
3. Queuing handoff calls only

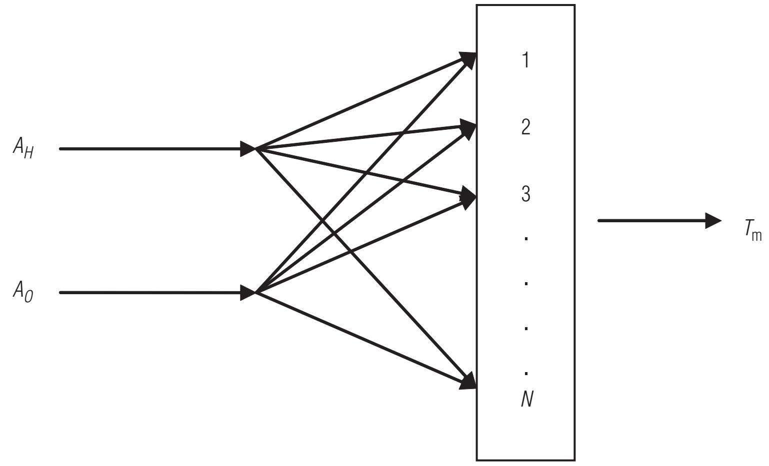


Figure 18.22 A general system model for handoff queuing

Let us consider the following notations which are used in all three cases.

$a = (A_O + A_H)$ T

$b_1 = A_O\ T$

$b_2 = A_H\ T$

T = average calling time in seconds, including new calls and handoff calls in each cell

M_O = size of queue for originating calls

M_H = size of queue for handoff calls

Without queuing

If there is no queuing procedure used to control originating calls as well as handoff calls then the blocking probability (B_0) for either an originating call or a handoff call is

$$B_O = \frac{a^N}{N!} P(0) \tag{18.2}$$

where

$$P(0) = \left(\sum_{n=0}^{N} \frac{a^N}{n!} \right)^{-1} \tag{18.3}$$

Queuing origination calls only

In this case, queuing is done for originating calls but not for handoff calls. The blocking probability for originating calls is

$$B_{OO} = \left(\frac{b_1}{N} \right)^{M_0} P_q(0) \tag{18.4}$$

The blocking probability for handoff calls is

$$B_{OH} = \frac{1-(b_1/N)^{M_0+1}}{1-(b_1/N)} p_q(0) \tag{18.5}$$

where

$$P_q(0) = \left[N! \sum_{n=0}^{N-1} \frac{a^{n-N}}{n!} + \frac{1-(b_1/N)^{M_0+1}}{1-(b_1/N)} \right]^{-1} \tag{18.6}$$

Queuing handoff calls only

In this case, queuing is done for handoff calls and no queuing procedure for originating calls. The blocking probability for originating calls is

$$B_{OO} = \frac{1-(b_2/N)^{M_H+1}}{1-(b_2/N)} P_q(0) \tag{18.7}$$

The blocking probability for handoff calls is

$$B_{OH} = \left(\frac{b_2}{N} \right)^{M_H} P_q(0) \tag{18.8}$$

Here,

$$P_q(0) = \left[N! \sum_{n=0}^{N-1} \frac{a^{n-N}}{n!} + \frac{1-(b_2/N)^{M_H+1}}{1-(b_2/N)} \right]^{-1} \tag{18.9}$$

In general, it is better to queue only the handoff calls so that it will give a decreased blocking probability of handoff calls and it should not affect the originating calls. Similarly, queuing of originating calls will give an increased blocking probability on handoff calls.

18.9 Forced handoff

A forced handoff is defined as a handoff which would normally occur but is prevented from happening or a handoff that should not occur is forced to happen.

18.9.1 Controlling handoff

The cell site can assign a low handoff threshold in a cell to keep a mobile unit in a cell longer or assign a high handoff threshold level to request a handoff earlier. The MTSO also can control a handoff by making either a handoff earlier or later after receiving a handoff request from a cell site.

18.9.2 Creating a handoff

In this case the cell site does not request a handoff but the MTSO finds that some cells are too congested while others are not. Therefore, the MTSO can request cell sites to create early handoffs for those congested cells. In other words, a cell site has to follow the MTSO's order and increase the handoff threshold to push the mobile units at the new boundary and to handoff earlier.

18.10 Dropped calls and dropped call rate

In mobile communication, it is always needed to reduce the number of dropped calls to increase the quality of service. The possibility that a call will drop due to the poor signal of the assigned voice channel is called as "dropped call".

If the phone cannot find an alternative cell to move in order to take over the call, the call is lost. Co-channel and adjacent-channel interference can also be responsible for dropped calls in a wireless network. Neighbouring cells with the same frequencies interfere with each other, deteriorating the quality of service and producing dropped calls.

One of major reason of dropped calls is improper handoff, a proper timely handoff is one of the procedures to reduce dropped calls. During handoff between two cells due to an imbalance of traffic between the two cell site areas, it cannot accept the additional traffic of the call then there is a chance of call dropping. The dropping probability is defined as the percentage of handoff attempts that are denied because of insufficient resources in the cell into which the mobile is moving.

The number of dropped calls in cellular system is dependent on the dropped call rate. The dropped call rate is dependent on the following factors:

- The channel capacity
- Level of traffic in the system (highly populated areas such as metro cities and business areas have more chances of handoffs and so the dropped call rate increases).
- Voice quality
- Probability that the signal below the receiver threshold (δ)
- Probability that the signal below the specified co-channel interference level (μ)

18.10.1 Channel capacity

Channel capacity is directly proportional to bandwidth of the system. If bandwidth is more, then more number of channels (users) can be allotted. With the increase in channels, adjacent-channel interference also increases and so signal-to-interference ratio decreases. This leads to poor signal quality and increased dropped call rate.

There is a relation between channel capacity, the number of voice channels, and the signal-to-noise ratio as given below:

The radio capacity

$$\mathrm{RC} = \frac{N}{\sqrt{2/3(S/I)}} \tag{18.10}$$

where N = total number of channels

S/I = required SIR ratio for designing a system.

The Equation (18.10) is obtained based on six co-channel interferers which occur in busy traffic (i.e. a worst case). From the equation if the channel capacity is increased the voice or information (signal) quality will decrease since (S/I) will decrease. As a consequence dropped call rate will increase.

18.10.2 Level of traffic in the system

Highly populated areas such as metro cities, business areas, railway stations, and so on have more chances of handoffs since high rate of mobility from one cell to another. So in such regions dropped call rate will become high and number of dropped calls will be more.

Traffic intensity is the measurement of traffic generated by a user during the busy hour (BH). The total number of voice calls originated or terminated in a mobile during the BH is called voice traffic arrival rate and voice traffic is generally represented by the unit called Erlang. Erlang is defined as a voice call of one hour duration. Each voice call is held for certain duration. The average duration of all voice calls is called holding time of a call. Similarly, departure rate can be considered as "$1/T$".

If R represents the arrival rate of voice calls during a BH (call/s)and T represents average holding time of a call (in seconds), the total BH voice traffic is given by RT. Then,

$$\text{the total traffic (in Erlangs)} = \frac{RT}{3600}\,\text{Erlang.} \tag{18.11}$$

18.10.3 Receiver threshold (δ) and co-channel interference level (μ)

If we consider a whole cellular system, the general formula for dropped call rate D will be given as

$$D = 1 - \left[\sum_{n=0}^{N} \alpha_n X^n\right] = \sum_{n=0}^{N} \alpha_n D_n \tag{18.12}$$

and

$$D_n = 1 - X^n \tag{18.13}$$

where

D_n = the probability of a dropped call when the call has gone through n handoffs and

$$X = (1 - \delta)\,(1 - \mu)\,(1 - \theta\zeta)\,(1 - \beta)^{2} \tag{18.14}$$

δ = probability that the signal is below the specified receiver threshold (in a noise limited case)

μ = probability that the signal is below the specified co-channel interference level (in interference-limited case)

ζ = probability that no channel is available for handoff when moving into a new cell

θ = probability that the call will return to the original cell

β = probability of blocking circuits between BSC and MSC during handoff

α_n = the weighted value for those calls having "n" handoffs, and $\sum_{n=0}^{N}\alpha_n = 1$

N = the highest number of handoffs for those calls

In general, the values of ζ, θ, and β are assumed to be very small and can be neglected. Hence, we can take

$$X = (1 - \delta)\,(1 - \mu) \tag{18.15}$$

Now we are able to deduce the expressions of dropped call rate for the following two cases:

1. Noise-limited system, $\mu \to 0$.
2. Interference-limited system, $\delta \to 0$.

Noise-limited system, $\mu \to 0$

Here, we are considering only noise limited system, so the effect of receiver threshold signal can be considered and also assumed that there will not be any co-channel interference. In such a case, since $\mu \to 0$ the expression for dropped call rate is

$$D = \sum_{n=0}^{N}\alpha_n D_n = \sum_{n=0}^{N}\alpha_n[1-(1-\delta)^n] \tag{18.16}$$

Interference-limited system, $\delta \to 0$

Here, we consider only interference-limited system, so the effect of co-channel interference can be considered and also assumed that there will not be any kind of noise which is introducing in the system. In such a case, since $\delta \to 0$ and the expression for dropped call rate is

$$D = \sum_{n=0}^{N}\alpha_n D_n = \sum_{n=0}^{N}\alpha_n[1-(1-\mu)^n] \tag{18.17}$$

18.11 Vehicle locating methods

Vehicle locating systems are the devices used for tracking location of vehicles in real time. Finding the location of vehicle is done with mobile tracking. The movement of vehicle is observed for its signal strength as it moves away from BS. If the vehicle enters in a new cell then inter-cellular tracking is being used and the mobility is analysed.

The vehicle locating methods can be classified into two ways:

1. Installation of equipments in the vehicles
2. Installation of equipments at the cell sites

18.11.1 Installing equipment in the vehicles

There are two methods of locating the vehicles by installing the equipment in the vehicle. They are triangulation and GPS method.

Triangulation In the triangulation method, three or more transmitting antennas are used at different cell sites. Each cell site forms a circle. The centre of the circle is the cell site transmitter location and the radius of the circle is the distance between the cell site transmitter and the vehicle location. Since the locations of the sites are known, the location of the vehicles can be obtained on calculating the distance from the cell sites. The distance is determined in the vehicle by measuring the travel time of signal from cell site to the vehicle. However, the accuracy is limited by the multipath phenomenon.

Global Positioning System (GPS) The GPS is a satellite-based navigation system that was developed by the United States in the early 1970s. GPS provides continuous positioning and timing information, anywhere in the world under all weather conditions. The GPS system consists of 24 satellites orbiting at an altitude of 20,200 km over earth surface. The 24 satellites are positioned in six orbital planes with four satellites in each plane. The orbits are equally spaced above the equator at a 60° separation with an inclination angle of 55°. The GPS satellites travel at a velocity of 3.9 km/s. Satellite vehicles (SVs) are arranged such that observers anywhere on the earth's surface will always have at least four satellites in view. The nominal orbital period of a GPS satellite is 11 h 58 min. The satellites transmit two pseudo-random noise (PRN) radio signals. The signals consist of a coarse acquisition (C/A) code at 1.023 MHz and a precision (P) code at 10.23 MHz bandwidths. The signals are transmitted at two frequencies, L_1 (1,575.42 MHz) and L_2 (1,227.60 MHz).

The basic principle behind GPS is the measurement of distance between satellites and the vehicle. It utilizes the concept of time-of-arrival (TOA) ranging to determine vehicle position. Each GPS satellite continuously transmits two frequencies. Timing information is embedded within the satellite ranging signal that enables the receiver to calculate when the signal has left the satellite. From this information, along with the received signal time, receiver measures the time taken by the satellite signal to reach the GPS receiver. This time interval is referred to as the signal propagation time. The propagation time is then multiplied by the speed of the signal to obtain satellite to receiver distance. By making TOA measurements to multiple satellites, three-dimensional positioning of the vehicle is achieved.

The ranges measured from satellites are called pseudoranges since biases in the receiver clock prevent the precise measurement of true ranges. The location of the vehicle can be found by the "triangulation method". The pseudorange (R_i) from each SV defines a sphere on which user may be located in three-dimensional space. Pseudoranges from three satellites define three spheres, the intersection of which defines the vehicle location (Figure 18.23).

For example, let the vehicle be at x_u, y_u, and z_u in earth-fixed, earth-centred coordinate system and the satellites be at x_i, y_i, z_i, (where $i = 1, 2, 3$) in the same coordinate system as the vehicle. The observation equation of the pseudorange measured between the vehicle (x_u, y_u, z_u) and known satellite (x_i, y_i, z_i) in space can be given in Cartesian coordinate system as

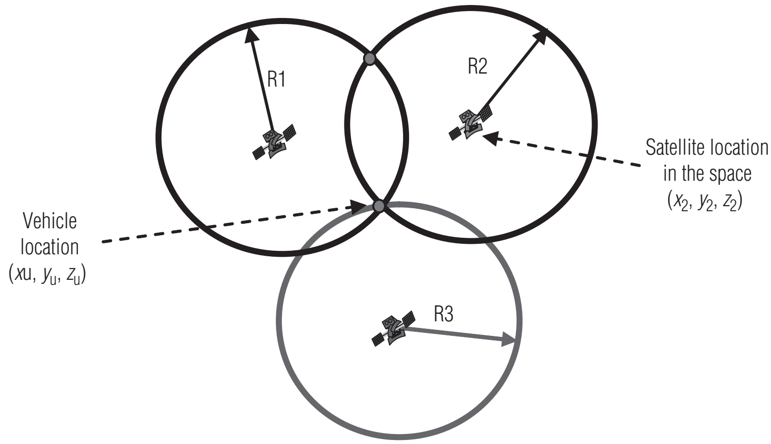


Figure 18.23 GPS positioning using three satellites

$$R_i = \sqrt{(x_u - x_i)^2 + (y_u - y_i)^2 + (z_u - z_i)^2} \tag{18.18}$$

Vehicle location can be determined by solving the above equation using either least squares approximation or Kalman-filter method.

18.11.2 Installing equipment at the cell site

There are three methods of locating the vehicles based on installing the equipment at the cell site. They are as follows:

- Triangulation based on signal strength
- Triangulation based on angular arrival
- Triangulation based on response-time arrival

Triangulation based on signal strength: In this triangulation method the location of the vehicle is determined by measuring the signal strength received from the mobile unit at each cell site. The degree of accuracy is very poor with this method because of the multipath phenomenon in signal propagation.

Triangulation based on angular arrival: In this triangulation method the location of the vehicle is determined by finding the direction of signal arrival at each cell site.

Triangulation based on response time arrival: In this method the cell site transmits a signal towards the mobile unit. The time of arrival or a change in phase angle of the signal is measured at each cell site. The location of the vehicle is determined by calculating the distance of the mobile unit at each cell site, with the measured time delay or phase difference of the signal.

18.12 Summary

- Handoff refers to the mechanism of transferring a call carrying voice, data, or video in a communication session from one BS to another geographically adjacent BS in order to maintain an uninterrupted communication for mobile user.

- Handoffs can be classified based on nature of handoffs, purposes of handoffs, and algorithms of handoffs. MCHO, NCHO, and MAHO are three important algorithms and can also be treated as handoff protocols.
- Two main actions in handoff are handoff initiation and handoff implementation.
- Relative signal strength, relative signal strength with threshold, relative signal strength with hysteresis, and relative signal strength with hysteresis and threshold are the four ways to initiate a handoff.
- Implementation of handoff can be done based on signal strength, carrier-to-information ratio, and power difference.
- The cellular structure is an important thing to choose particular handoff mechanism. Due to small radius, compared to macrocellular structures, microcellular structures will have more number of handoffs per second.
- There are some situations where it is required to delay the handoff procedure. For delaying the handoff, queuing and two-level handoff procedures can be used.
- While queuing off the handoff calls, it is always required to consider the originating calls also. Queuing new calls results in increased handoff blocking probability.
- The probability of a successful handoff can be improved by queuing handoff requests at the cost of increased new call blocking probability and a decrease in the ratio of carried-to-admitted traffic. If handoff requests occur uniformly, queuing is not needed.
- Untimely, handoffs can lead to dropped calls with an increase in dropped call rate. The dropped call rate depends upon channel capacity, level of traffic in the system, receiver threshold, and co-channel interference.

Example problem 18.1

In a cellular system the number of channels at the cell site is 60. The call holding time is 0.024 h. The number of originating calls attempted per hour is 1,834, and the number of handoff calls attempted per hour is 62. Calculate the blocking probability of the system when queuing is done for originating calls only, if the queue size is 5.

Solution

Given parameters are:
Number of channels, $N = 60$
Call holding time, $T = 0.024$ h
Mean rate of originating calls, $A_O = 1{,}834$
Mean rate of handoff calls, $A_H = 62$
Queue size of originating calls, $M_O = 5$
Then, $a = (A_O + A_H)\,T = (1{,}834 + 62) \times 0.024 = 45.504$
And $b_1 = A_O \times T = 1{,}834 \times 0.024 = 44.016$

$$P_q(0) = \left[N! \sum_{n=0}^{N-1} \frac{a^{n-N}}{n!} + \frac{1-(b_1/N)^{M_0+1}}{1-(b_1/N)} \right]^{-1}$$

$$= \left[60! \sum_{n=0}^{60-1} \frac{45.504^{n-60}}{n!} + \frac{1-(44.016/60)^{5+1}}{1-(44.016/60)} \right]^{-1}$$

$$= 0.006352$$

When queuing of originating calls is done then blocking probability of originating calls can be given calculated as

$$B_{OO} = \left(\frac{b_1}{N} \right)^{M_0} P_q(0)$$

$$= \left(\frac{44.016}{60} \right)^5 \times 0.006352$$

$$= 1.35 \times 10^{-3}$$

$$= 0.00135$$

Example problem 18.2

In a system the number of channels at the cell site is 50. The call holding time is 0.022 h. The number of originating calls attempted per hour is 2,035, and the number of handoff calls attempted per hour is 32.

Calculate the blocking probability of

(a) originating calls

(b) handoff calls when queuing is done for handoff calls only and the queue size is 4.

Solution

Given parameters are:

Number of channels, $N = 50$

Call holding time, $T = 0.022$ h

Mean rate of originating calls, $A_O = 2{,}035$

Mean rate of handoff calls, $A_H = 32$

Queue size of originating calls, $M_H = 4$

$a = (A_O + A_H)\, T$

$= (2{,}035 + 32) \times 0.022 = 45.474$

and $b_2 = A_H \times T$

$= 32 \times 0.022 = 0.704$

When queuing only the handoff calls,

$$P_q(0) = \left[N! \sum_{n=0}^{N-1} \frac{a^{n-N}}{n!} + \frac{1-(b_2/N)^{M_H+1}}{1-(b_2/N)} \right]^{-1}$$

$$= \left[50! \sum_{n=0}^{50-1} \frac{45.474^{n-50}}{n!} + \frac{1-(0.704/50)^{4+1}}{1-(0.704/50)} \right]^{-1}$$

$$= 0.0584$$

(a) Blocking probability of originating calls

$$B_{OO} = \frac{1-(b_2/N)^{M_H+1}}{1-(b_2/N)} p_q(0) = \frac{1-(0.704/50)^{4+1}}{1-(0.704/50)} \times 0.0584 = 0.0592$$

(b) Blocking probability of handoff calls

$$B_{OH} = \left(\frac{b_2}{N}\right)^{M_H} p_q(0) = \left(\frac{0.704}{50}\right)^4 \times 0.0584 = 2.29 \times 10^{-9}$$

Example problem 18.3

In a cellular system, the measurements of arrival rate of data during busy hour (BH) is found to be 150 calls per second. The average holding time of the call is measured at 120 s. Find the estimated traffic in Erlang and the departure rate per second for the system?

Solution

Given parameters are

Call holding time, $T = 120$ s

Arrival rate of data, $R = 150$ s

Then traffic in Erlangs, $= \dfrac{RT}{3600} = \dfrac{120 \times 150}{3600} = 5$

and departure rate, $= 1/\mathrm{T}$

$= 1/120 = 0.00833$

Example problem 18.4

In a system the probability that the signal below the specified receiver threshold is 20 per cent and the probability that the signal above the specified co-channel interference level is 15 per cent. Then what is the probability of a dropped call when the call has gone through three handoffs?

Solution

Given parameters are:

$\delta = 20\% = 0.2$

$\mu = 15\% = 0.15$

Number of handoffs, $n = 3$

From the given data the constant "X" can be evaluated as

$$\begin{aligned} X &= (1 - \delta)\,(1 - \mu) \\ &= (1 - 0.2)\,(1 - 0.15) \\ &= 0.8 \times 0.85 = 0.68 \end{aligned}$$

Then the probability of a dropped call when call has gone through three handoffs can be given as

$$\begin{aligned} D_n &= 1 - X^n \\ &= 1 - 0.68^3 = 1 - 0.3144 = 0.68 \end{aligned}$$

Review questions

1. What is meant by handoff?
2. Describe the classification of handoff processes.
3. Differentiate the soft, softer, and hard handoffs.
4. What is the difference between intra-cell handoff and inter-cell handoff methods?
5. Differentiate the mobile-controlled handoff and mobile-assisted handoff algorithms.
6. How can handoff be initiated at the boundary of two cells, based upon threshold point considering signal strengths at two BSs?
7. What are different categories of handoff procedures in GSM?
8. How can handoff be implemented based on signal strength?
9. Why do the microcellular structures have more number of handoffs per second compared to macrocell structures?
10. Explain the LOS handoff and non-LOS handoff procedures in a microcellular system.
11. What are the different methods of delaying the handoff? Explain briefly.
12. What is meant by a dropped call? And what are the factors that influence the dropped call rate?
13. What are the different vehicle locating methods? Explain GPS method.
14. The number of channels at the cell site is 70. The call holding time is 101 s. The number of originating calls attempted per hour is 2,270. The number of handoff calls attempted per hour is expressed as 80. Calculate the blocking probability for originating calls only if the queue size is 5. (Ans: 0.0474)
15. For a system the probability above the specified co-channel interference is given as 18 per cent and the probability below the specified receiver threshold is 23 per cent. If the call goes through four handoffs, find the probability of a dropped call? (Ans: 0.841)
16. Explain how handoff is initiated. (Refer Section 18.6)
17. Why handoff is necessary for cellular systems? Determine the two types of handoffs based on signal strength and C/I ratio. (Refer Section 18.2)
18. Explain the concept of delayed handoff. (Refer Section 18.8)
19. Plot the signal strength for a two-level handoff scheme and explain it. (Refer Sections 18.2.3 and 18.8.1)
20. Derive the blocking probabilities for handoff calls and the blocking probability of originating calls. (Refer Sections 18.8.2.1 to 18.8.2.3)
21. What are the different types of handoffs? Explain their implementation. (Refer Sections 18.3.1 and 18.3.2)
22. Discuss the method of queuing of handoffs. (Refer Section 18.8.2)
23. Define the dropped call rate. How dropped calls are considered? (Refer Section 18.10)
24. What type of handoff is used when a call initiated in one cellular system enters another system before terminating? Explain how it works. (Refer Section 18.4.3)
25. Classify different handoff mechanisms and define each technique. (Refer Sections 18.3.1 and 18.3.2)
26. What is forced handoff? Explain. (Refer Section 18.9)
27. Explain the necessity of power difference handoff. Also explain the different conditions based on the power difference handoff. (Refer Section 18.2.3)

Objective type questions and answers

1. In normal handoff procedure the handoff request is based on ________
   (a) power level (b) signal strength (c) peak current (d) time delay
2. Generally, the soft handoff procedure involves ________ number of cell-site(s).
   (a) 1 (b) 3 (c) 6 (d) several
3. In ________, a call communication link uses only one channel at any moment.
   (a) softer handoff (b) soft handoff (c) hard handoff (d) intra-cell handoff
4. Creating handoff is requested by ________
   (a) PSTN (b) MTSO (c) BSC (d) cell site
5. The radio capacity for the lower voice quality ________
   (a) remains unaffected (b) decreases
   (c) approaches to zero (d) increases
6. The decentralized handoff protocol that distributes the handoff decision process.
   (a) Network-controlled handoff (b) Mobile-assisted handoff
   (c) Soft handoff (d) Mobile-controlled handoff
7. The cellular networks that uses the NCHO protocol is ________
   (a) GSM (b) ETACS (c) USDC (d) AMPS
8. The overall handoff delay in mobile-assisted handoff algorithm is typically ________
   (a) 5–10 s (b) 1 s (c) 2–3 s (d) less than 1 s
9. The relation between channel capacity and bandwidth of the system is ________
   (a) directly proportional (b) inversely proportional
   (c) both a & b (d) none
10. One of the main probable reasons of call drop is ________
    (a) co-channel interference (b) handoff (c) external noise
    (d) channel assignment

Answers: 1. (b), 2. (d), 3. (c), 4. (b), 5. (d), 6. (b), 7. (d), 8. (b), 9. (a), 10. (b).

Open book questions

1. What is meant by handoff and handoff algorithm?
2. What is the purpose of handoff?
3. What are the advantages of handoff process?
4. Describe the handoff criteria to make a handoff decision.
5. The handoff is always implemented on a voice channel, but not on a control channel. Justify?
6. What are the advantages of soft handoff?
7. Differentiate the inter-cell handoff and intra-cell handoff.
8. What are the various handoff schemes based on handoff algorithms? Discuss briefly.
9. Name the various handoff initiation techniques?
10. How queuing is important for the handoff procedure?
11. Explain how the coverage of a noise limited system is increased by the parameters of the system.
12. Discuss the advantages of delayed handoffs.
13. Explain how to calculate the number of handoffs per call.

14. What are the circumstances where handoffs are necessary but cannot be made?
15. If the number of channels at the cell site $N = 45$, the call holding time is 1.76 min, the number of originated calls per hour expressed as λ_1 is 2,270, and the number of handoff calls attempted per hour expressed as λ_2 is 80, find the probability of queuing the originated calls but not the handoff calls.

Key equations

1. The signal-to-interference (S/I) ratio in a three-sector (120° each sector) antenna is given by

$$(S/I)_{120^\circ} = (S/I)_{360^\circ \text{ or omni}} + 10\log 3 = (S/I)_{360^\circ \text{ or omni}} + 4.8\text{dB}$$

2. The blocking probability (B_0) for either an originating call or a handoff call is

$$B_O = \frac{a^N}{N!}P(0)$$

3. The radio capacity is

$$\text{RC} = \frac{N}{\sqrt{2/3(S/I)}}$$

4. The total traffic (in Erlangs) is

$$\frac{RT}{3600}\text{Erlang.}$$

5. The observation equation of the pseudorange measured between the vehicle (x_u, y_u, z_u) and known satellite (x_i, y_i, z_i) in space can be given in Cartesian coordinate system as

$$R_i = \sqrt{(x_u - x_i)^2 + (y_u - y_i)^2 + (z_u - z_i)^2}$$

Further reading

G. Liodakis and P. Stavroulakis, "A novel approach in handover initiation for microcellular systems," Proc. 44th IEEE VTC, 1994, pp. 1820–23.

P. Marichamy, S. Chakrabati and S. L. Maskara, "Overview of handoff schemes in cellular mobile networks and their comparative performance evaluation", IEEE VTC'99, vol. 3, 1999, pp. 1486–1490.

Nishint D. Tripathi, Jeffrey H. Reed and Hugh F. VanLandinoham, "Handoff in Cellular Systems", IEEE Personal Communications, vol. 5, December 1998, pp. 26–37.

G. P. Pollini, "Trends in handover design," IEEE Communication Magazine, March 1996, pp. 82–90.

S. Tekinay and B. Jabbari, "Handover and channel assignment in mobile cellular networks", IEEE Communications Magazine, pp. 42–46, November 1991.

Global System for Mobile Communications 19

19.1 Introduction

Digital cellular systems have many features such as improved communication quality due to the use of various digital signal processing technologies, new services (e.g. non-telephony services), improved ciphering, greater conformity with digital networks, and efficient utilization of the radio spectrum.

The development of digital cellular system was triggered by standardization efforts in Europe, which was home to many competing analogue systems. In Europe, analogue cellular systems were using different frequency bands and schemes which made interconnection impossible across national borders. In 1982, the European Conference of Postal and Telecommunications Administrations (CEPT) established the global system for mobile (GSM) and developments were carried out under the leadership of the European Telecommunications Standards Institute (ETSI). The GSM-based services were launched in 1992.

In the United States, the IS-54 was developed by the Electronic Industries Association (EIA) and the Telecommunications Industry Association (TIA). IS-54 services launched in 1993 were required to satisfy dual-mode (both analogue and digital cellular) operations and adopted time division multiple access (TDMA). Studies on CDMA inclusive of field tests had been carried out in a vigorous manner from 1989 onwards and consequently the IS-95 standard based on CDMA technology was adopted in 1993.

All multiple access techniques depend on the adoption of digital technology. Digital technology is a standard for digital cellular systems where all analogue calls are converted to digital form for transmission over the backbone.

Digital transmission has a number of advantages over analogue transmission:

- It economizes bandwidth.
- It allows easy integration with personal communication system (PCS) devices.
- It maintains superior quality of voice transmission over long distances.
- It is difficult to decode.
- It can use lower average transmitted power.

The worldwide market figures for digital cellular networks as follows: The most popular digital system is GSM with approximately 70 per cent market share. The analogue advanced mobile phone system (AMPS) holds 3 per cent and the Japanese PDC holds 5 per cent (60 million users). The remainder is split between CDMA (12 per cent), TDMA (10 per cent) systems, and other technologies. In Europe, almost all users use digital GSM (over 370 million) with no analogue systems left. The situation is different in the United States and some other countries that have adopted US technology (e.g. South Korea, Canada). Here, the digital market is split among TDMA, CDMA, and GSM systems with 107 million TDMA, 135 million CDMA, and only 16 million GSM users (North America only).

This chapter deals with GSM and its architecture, GSM specifications, GSM operation, GSM channels, protocol stack configuration of GSM, and its basic call flow.

19.1.1 Brief history of GSM

During the early 1980s, analogue cellular telephone systems were experiencing rapid growth not only in Europe (particularly in Scandinavia and the United Kingdom) but also in France and Germany. Each country developed its own system, which was incompatible with everyone else's in equipment and operation. This was an undesirable situation, because not only the ME was limited to operation within the national boundaries, which in a unified Europe were increasingly unimportant, but there was a very limited market for each type of equipment, so economies of scale, and the subsequent savings, could not be realized.

The Europeans realized this early on, and in 1982, the Conference of European Posts and Telegraphs (CEPT) formed a study group called the Groupie Special Mobile (GSM) to study and develop a pan European public land mobile system. The proposed system had to meet certain criteria:

- Good subjective speech quality
- Low terminal and service cost
- Support for international roaming
- Ability to support handheld terminals
- Support for range of new services and facilities
- Spectral efficiency
- ISDN compatibility

In 1989, the responsibility of GSM was transferred to the European Telecommunication Standards Institute (ETSI) and phase I of the GSM specifications were published in 1990. Commercial service was started in mid-1991 and by 1993, there were 36 GSM networks in 22 countries with 25 additional countries having already selected or considering GSM.

This is not only a European standard – South Africa, Australia, and many Middle and Far East countries have chosen GSM. By the beginning of 1994, there were 1.3 million subscribers worldwide. The acronym GSM now stands for Global System for Mobile telecommunications.

The developers of GSM chose an unproven (at the time) digital system, as opposed to the standard analogue cellular systems like AMPS in the United States and TACS in the United Kingdom. They had faith that advancements in compression algorithms and digital signal processors would allow the fulfilment of the original criteria and the continual improvement of the system in terms of quality and cost. This is done in part by providing descriptions of the interfaces and functions of each of the functional entities defined in the system.

19.2 Global system for mobile

GSM is most widely used and globally implemented digital cellular technology. It is used for transmitting data and mobile voice services. As was mentioned earlier it is the name of a standardization group formed by CEPT. It is basically a circuit-switched system in which every

200 kHz channels are divided into 25 kHz time slots. In Europe it operates in the band 900 MHz and 1.8 GHz, whereas in United States the operation band is 1.9 GHz and 850 MHz. It is estimated that GSM technology serves more than 80 per cent of world's digital subscribers that is more than one billion subscribers throughout the world.

In GSM, time division multiple access (TDMA) technique is used for transmitting voice and data through air interface. TDMA is a digital technology and support data rates in the range between 64 kbps and 120 Mbps. As mentioned in the standard it supports roaming service, which makes it possible to use one GSM mobile phone number in another GSM network.

A few important features of GSM are discussed in the following sections.

19.2.1 Flexibility and increased capacity

With an analogue air interface, every connection between a mobile station (MS) and a cell site requires a separate radio frequency (RF) carrier, which in turn requires separate RF hardware set. In order to expand the capacity of a cell site by a given number of channels, an equivalent quantity of hardware must be added. This makes system expansion time consuming, expensive, and labour intensive. Re-configuration of an analogue site suffers similar problems since much of the equipment requires manual re-tuning and this makes the system inflexible.

GSM equipment is fully controlled by its software. Network re-configurations can be made quickly and easily with minimum manual intervention. In addition, since one carrier can support eight users, expansion can be made with less equipment.

An enhancement soon to be realized is the half-rate speech channel where mobiles will use new speech algorithms requiring half as much data to be sent over the air interface. By implementing the half-rate speech channel, one carrier will be able to support 16 users, effectively doubling the capacity of the network. However, this is optimum since the mobile as well as the base transceiver station (BTS) will need to be modified to support half rate.

GSM networks also offer the flexibility of *international roaming*. This allows the mobile user to travel to foreign countries and still use their mobiles on the foreign network. If necessary, the user may leave their ME at home and carry only the subscriber identity module (SIM) card, making use of a hired mobile or any available equipment.

GSM's use of a digital air interface makes it more resilient to interference from users on the same or nearby frequencies and so the cells can be packed closer together, which means more carriers in a given area to give better frequency reuse.

Multi-band networks and mobiles are available where a user can make use of both the 900 MHz network and the 1800/1900 networks. The mobile must be capable of operating in dual frequency bands. This enables network operators to add in capacity and reduce network interference by using cells operating in different frequency bands. The operator will be required to show that they have made efficient use of their existing frequencies before they will be granted access to frequencies in another band. This means using techniques like sectorization, microcells, and frequency hopping.

GSM is highly software dependent and, although this makes it very complex, it also provides for a high degree of flexibility.

19.2.2 Frequency, channel spacing, and transmission rate

In this section, a few important characteristics of GSM are described in detail. This includes a discussion of the frequency, channel spacing, and the transmission rate parameters.

Frequency

An MS communicates with a BTS by transmitting or receiving radio waves which consist of electromagnetic energy. The frequency of a radio wave is the number of times that the wave oscillates per second. Frequency is measured in hertz (Hz), where 1 Hz indicates one oscillation per second. RFs are used for many applications in the world today. Some common uses are as follows:

- Television: 300 MHz approx.
- FM Radio: 100 MHz approx.
- Police radios: Country dependent
- Mobile networks: 300–2,000 MHz approx.

The frequency used by mobile networks varies according to the standard being used. An operator applies for the available frequencies or, as in the United States, the operator bids for frequency bands at an auction. Figure 19.1 displays the frequencies used by the major mobile standards:

Table 19.1 summarizes the frequency-related specifications of each of the GSM systems. The terms used in the table are explained in the remainder of this section.

Channel spacing (Carrier separation)

In addition to the duplex distance, every mobile system includes carrier separation. This is the distance on the frequency band between channels being transmitted in the same direction. This is required in order to avoid the overlapping of information in one channel into an adjacent channel. The length of separation between two channels is dependent on the amount of information which is to be transmitted within the channel. The greater the amount of information to transmit the greater the amount of separation required.

From Figure 19.2, it can be seen that the information to be sent is modulated around the carrier frequency of 895.4 MHz. The same is true to the information to be sent on 895.6 MHz. To avoid interference between the two sets of information, a separation distance of 200 kHz is required. If less separation were used, they would interfere and a caller on 895.4 MHz may experience crosstalk or noise from the caller on 895.6 MHz.

Transmission rate

The amount of information transmitted over a radio channel over a period of time is known as the transmission rate. Transmission rate is expressed in bits per second or bps. In GSM, the net bit rate over the air interface is 270 kbps.

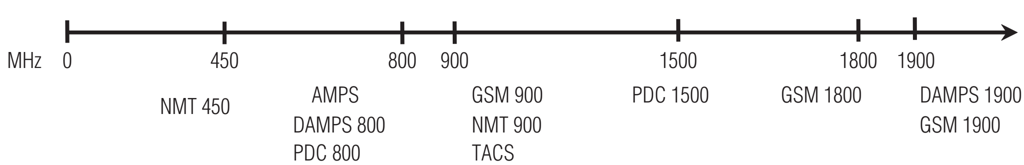


Figure 19.1 Frequencies of major mobile standards

Table 19.1 Frequencies of major mobile standards

System	P-GSM 900	E-GSM 900	GSM 1800	GSM 1900
Frequencies				
• Uplink	890–915 MHz	880–915 MHz	1,710–1,785 MHz	1,850–1,910 MHz
• Downlink	935–960 MHz	925–960 MHz	1,805–1,880 MHz	1,930–1,990 MHz
Wavelength	~ 33 cm	~ 33 cm	~ 17 cm	~ 16 cm
Bandwidth	25 MHz	35 MHz	75 MHz	60 MHz
Duplex distance	45 MHz	45 MHz	95 MHz	80 MHz
Carrier separation	200 kHz	200 kHz	200 kHz	200 kHz
Radio channels	125	175	375	300
Transmission rate	270 kbps	270 kbps	270 kbps	270 kbps

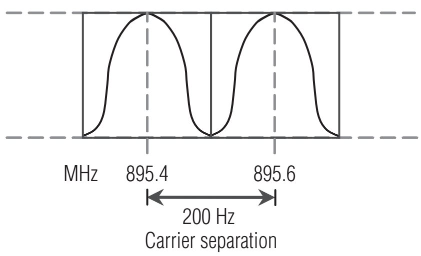


Figure 19.2 Carrier separation

19.2.3 Improved security and confidentiality

Security figures high on the list of problems encountered by some operators of analogue systems. In some systems, it is virtually non-existent and the unscrupulous were quick to recognize this. With some of the "first generation" systems, it has been estimated that up to 20 per cent of cellular phone calls were stolen.

Extensive measures have been taken when specifying the GSM system to substantially increase security with regard to both call theft and equipment theft.

With GSM, both the mobile equipment (ME) and mobile subscriber are identified. The ME has a unique number coded into it when it is manufactured. This can be checked against a database every time the mobile makes a call to validate the actual equipment. The subscriber is authenticated by use of a smart card known as a SIM. Again, this allows the network to check a MS subscriber against a database for authentication.

GSM also offers the capability to encrypt all signals over the air interface. Different levels of encryption are available to meet different subscriber/country requirements.

With authentication processes for both the ME and the subscriber, together with the encryption and digital encoding of the air interface signals, it makes it very difficult for the casual "hacker" to listen-in to personal calls.

In addition to this, the GSM air interface supports frequency hopping. This entails each "burst" of information being transmitted to/from the MS/base site on a different frequency, again making it very difficult for an observer (hacker) to follow/listen to a specific call. However, it should be noted that frequency hopping is employed to optimize network performance by overcoming interference problems in busy areas, to increase call quality and capacity.

19.2.4 Flexible handover processes

Handovers take place as the MS moves between cells, gradually losing the RF signal of one and gaining that of the other.

The MS switches from channel-to-channel and cell-to-cell as it moves to maintain call continuity. With analogue systems, handovers are frequently a problem area and the subscriber is often aware that a handover has occurred.

When GSM was specified, a great deal of thought went into the design and implementation of handovers. Although the GSM system is more complicated than analogue in this area, the flexibility of the GSM handover processes offer significant improvements which provide a much better quality of service to the subscriber as shown in Figure 19.3.

GSM provides handover processes for the following:

- Quality (uplink/downlink)
- Interference (uplink/downlink)
- RF level (uplink/downlink)
- MS distance
- Power budget

More handover algorithms have been developed for specific applications, such as microcellular, and are currently being implemented.

19.2.5 Switching and control

Having established radio coverage through the use of cells, both omni-directional and directional (sectored sites), now consider what happens when the MS is in motion (as MSs tend to be). At some point the MS will have to move from one cell's coverage area to another cell's coverage area. Handovers from one cell to another could be for a number of reasons (e.g. the signal strength of the "serving cell" is less than the signal strength of a "neighbour cell" or the MS is suffering a

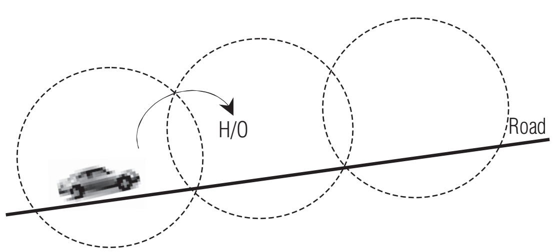


Figure 19.3 Handover process

quality problem in the serving cell) and by handing over to one of its neighbours this may stop the quality problem.

Regardless of the reason for a "handover", it has to be controlled by some entity, and in GSM, that entity is the mobile services switching centre (MSC).

To perform a handover, the network must know which neighbour cell to hand the MS over to. To ensure that we handover to the best possible candidate, the MS performs measurements of its surrounding neighbour cells and reports its findings to the network. These are then analysed together with the measurements that the network performs and a decision is made on a regular basis as to the need for a handover. If a handover is required, then the relevant signal protocols are established and the handover is controlled by the MSC as shown in Figure 19.4.

Handovers must be transparent to the MS subscriber. That is, the subscriber should be unaware that a handover has occurred. As we will see later in this chapter, handovers are just one of the functions of the MSC. Many more are performed by the MSC and its associated entities (e.g. such as authentication of MS, ciphering control, location updating, gateway to PSTN).

Note: Some networks may allow certain handovers to be performed at the BSS level. This would depend on the manufacturer's equipment.

19.2.6 Noise robust

In cellular telephone systems, such as AMPS, total access communication system (TACS), or nordic mobile telephone (NMT), the MS communicates with the cell site by means of analogue radio signals. Although this technique can provide an excellent audio quality (e.g. it is widely used for stereo radio broadcasting), it is vulnerable to noise, as anyone who has tried to receive broadcast stereo with a poor aerial will testify.

The noise which interferes with the current system may be produced by any of the following sources:

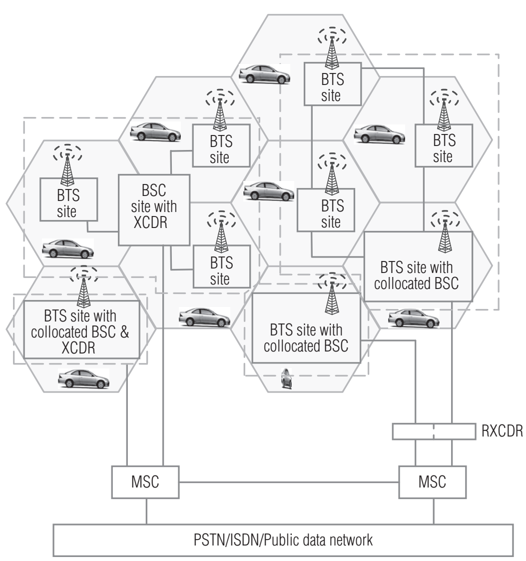


Figure 19.4 Switching and control

- A powerful or nearby external source (a vehicle ignition system or a lightning bolt, perhaps)
- Another transmission on the same frequency (co-channel interference)
- Another transmission "breaking through" from a nearby frequency (adjacent-channel interference)
- Background radio noise intruding because the required signal is too weak to exclude it

In order to combat the problems caused by noise, GSM uses digital technology instead of analogue. By using digital signals, we can manipulate the data and include sophisticated error protection, detection, and correction software. The overall result is that the signals passed across the GSM air interface withstand more errors (i.e. we can locate and correct more errors than current analogue systems). Due to this feature, the GSM air interface in harsh RF environments can produce a usable signal, where analogue systems would be unable to. This leads to better frequency reuse patterns and more capacity.

19.2.7 User services

GSM has much more to offer than voice telephony. Additional services allow you greater flexibility in where and when you use your phone. There are three basic types of services offered through GSM which you can ask for:

- Telephony (also referred to as teleservices) services
- Data (also referred to as bearer services) services
- Supplementary services

Teleservices or telephony services

A teleservice utilizes the capabilities of a bearer service to transport data, defining which capabilities are required and how they should be set up.

Voice calls: The most basic teleservice supported by GSM is telephony. This includes full-rate (FR) speech at 13 kbps and emergency calls, where the nearest emergency service provider is notified by dialling three digits. A very basic example of emergency service is 911 services available in the United States.

Videotext and facsimile: Another group of teleservices includes videotext access, teletex transmission, facsimile alternate speech, and facsimile group 3, automatic facsimile group 3, and so on.

Bearer services or data services

Using your GSM phone to receive and send data is the essential building block leading to widespread mobile Internet access and mobile data transfer. GSM currently has a data transfer rate of 9.6 kbps. New developments that will push up data transfer rates for GSM users are high-speed circuit-switched data (HSCSD) and general packet radio service (GPRS) are now available.

Data can be sent over the air using some of the present systems, but this requires specially designed "add-ons" to protect the data content in the harsh environment of the air interface.

Special provision is made in the GSM technical specifications for data transmission. Therefore, like ISDN, GSM is "specially designed" for data transmission. GSM can be considered as an extension of ISDN into the wireless environment.

Text files, images, messages, and fax may all be sent over the GSM network. The data rates available are 2.4 kbps, 4.8 kbps, and 9.6 kbps. In addition to supporting data transmission, GSM also provides for group 3 fax transmission.

Supplementary services

Supplementary services are provided on top of teleservices or bearer services, and include features such as caller identification, call forwarding, call waiting, multi-party conversations, and barring of outgoing (international) calls, among others. A brief description of supplementary services is given below:

- *Short text messages* (SMS)*:* SMS service is a text messaging which allows you to send and receive text messages on your GSM mobile phone. Services available from many of the world's GSM networks today — in addition to simple user generated text message services — include news, sport, financial, language, and location-based services, as well as many early examples of mobile commerce such as stocks and share prices, mobile banking facilities, and leisure booking services.
- *Multiparty service or conferencing:* The multiparty service allows a mobile subscriber to establish a multiparty conversation. That is, a simultaneous conversation between three or more subscribers to setup a conference call. This service is only applicable to normal telephony.
- *Call waiting:* This service allows a mobile subscriber to be notified of an incoming call during a conversation. The subscriber can answer, reject, or ignore the incoming call. Call waiting is applicable to all GSM telecommunications services using a circuit-switched connection.
- *Call hold:* This service allows a subscriber to put an incoming call on hold and then resume this call. The call hold service is only applicable to normal telephony.
- *Call forwarding:* The call forwarding supplementary service is used to divert calls from the original recipient to another number and is normally set up by the subscriber himself. It can be used by the subscriber to divert calls from the MS when the subscriber is not available, and so to ensure that calls are not lost. A typical scenario would be when a salesperson turns off his mobile phone during a meeting with customers, but does not wish to lose potential sales leads while he is unavailable.
- *Call barring:* The concept of barring certain types of calls might seem to be a supplementary disservice rather than service. However, there are times when the subscriber is not the actual user of the MS, and as a consequence may wish to limit its functionality, so as to limit the charges incurred. Alternatively, if the subscriber and user are one and the same, the call barring may be useful to stop calls being routed to international destinations when they are routed. The reason for this is because it is expected that the roaming subscriber will pay the charges incurred for international re-routing of calls. Therefore, GSM devised some flexible services that enable the subscriber to conditionally bar calls.
- *Number identification:* There are following supplementary services related to number identification:
    - **Calling Line Identification Presentation:** This service deals with the presentation of the calling party's telephone number. The concept is for this number to be presented, at the start of the phone ringing, so that the called person can determine who is ringing prior to answering. The person subscribing to the service receives the telephone number of the calling party.
    - **Calling Line Identification Restriction:** A person not wishing their number to be presented to others subscribes to this service. In the normal course of event, the restriction service overrides the presentation service.
    - **Connected Line Identification Presentation:** This service is provided to give the calling party the telephone number of the person to whom they are connected. This may

seem strange since the person making the call should know the number they dialled, but there are situations (such as forwardings) where the number connected is not the number dialled. The person subscribing to the service is the calling party.

- ◦ **Connected Line Identification Restriction:** There are times when the person called does not wish to have their number presented and so they would subscribe to this person. Normally, this overrides the presentation service.
- ◦ **Malicious Call Identification:** The malicious call identification service was provided to combat the spread of obscene or annoying calls. The victim should subscribe to this service, and then they could cause known malicious calls to be identified in the GSM network, using a simple command. This identified number could then be passed to the appropriate authority for action. The definition for this service is not stable.

- **Advice of Charge (AoC):** This service was designed to give the subscriber an indication of the cost of the services as they are used. Furthermore, those service providers who wish to offer rental services to subscribers without their own SIM can also utilize this service in a slightly different form. AoC for data calls is provided on the basis of time measurements.
- **Closed User Groups (CUGs):** This service is provided on GSM to enable groups of subscribers to only call each other. These types of services are being offered with special discount and are limited only to those members who wish to talk to each other.
- **Unstructured supplementary services data (USSD):** This allows operator-defined individual services.

19.2.8 ISDN compatibility in GSM

Integrated services digital network (ISDN) is a standard that most developed countries are committed to implement.

This is a new and advanced telecommunications network designed to carry voice and user data over standard telephone lines. Major telephone companies in Europe, North America, Hong Kong, Australia, and Japan are committed to commercial enterprises using ISDN.

The GSM network has been designed to operate with the ISDN system and provides features which are compatible with it as shown in Figure 19.5. GSM can provide a maximum data rate of 9.6 kbps while ISDN provides much higher data rates than this (standard rate 64 kbps and primary rate 2.048 Mbps).

2B + **D** refer to the signals and information which may be carried on an ISDN line. There are effectively three connections, one for signalling (“*D*”) and the other two for data or speech (“2*B*”).

Figure 19.5 ISDN compatibility in GSM

19.3 GSM Network architecture

Overview

GSM networks are made up of mobile service switching centres (MSC), base station systems (BSS), and MS. These three entities can be broken down further into smaller entities. Within the BSS, we have Base Station Controllers (BSCs), BTSs, and Transcoders as shown in Figure 19.6. Now we will use the three major entities.

With the MSC, BSS, and MS, we can make calls, receive calls, perform billing, and so on, as any normal PSTN network would be able to do. The only problem for the MS is that all the calls made or received are from other MSs. Therefore, it is also necessary to connect the GSM network to the PSTN.

MSs within the cellular network are located in "cells". These cells are provided by the BSSs. Each BSS can provide one or more cells depending on the manufacturers' equipment.

GSM network consists of several functional entities whose functions and interfaces are defined. Figure 19.7 shows the different parts of GSM network.

- The MS
- The base station subsystem (BSS)
- The network switching subsystem (NSS)
- The operation and maintenance/support system (OSS)

The GSM network adds some other components as shown in Figure 19.8. These added components of the GSM architecture include the functions of the databases and messaging systems.

Each network component is designed to communicate over an interface specified by the GSM standards. This provides flexibility and enables a network provider to utilize system components from different manufacturers. For example, Motorola BSS equipment may be coupled with an Ericsson NSS.

The following are the principle component groups of a GSM network:

- *Mobile station (MS)*: This consists of the mobile telephone, fax machine, and so on. This is the part of the network that the subscriber will see.
- *Base station system (BSS)*: This is the part of the network which provides the radio interconnection from the MS to the land-based switching equipment.

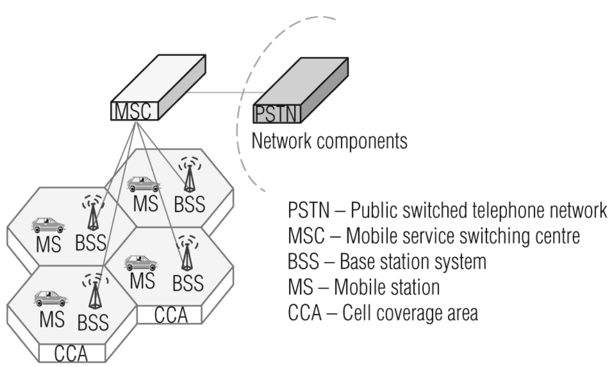


Figure 19.6 GSM network

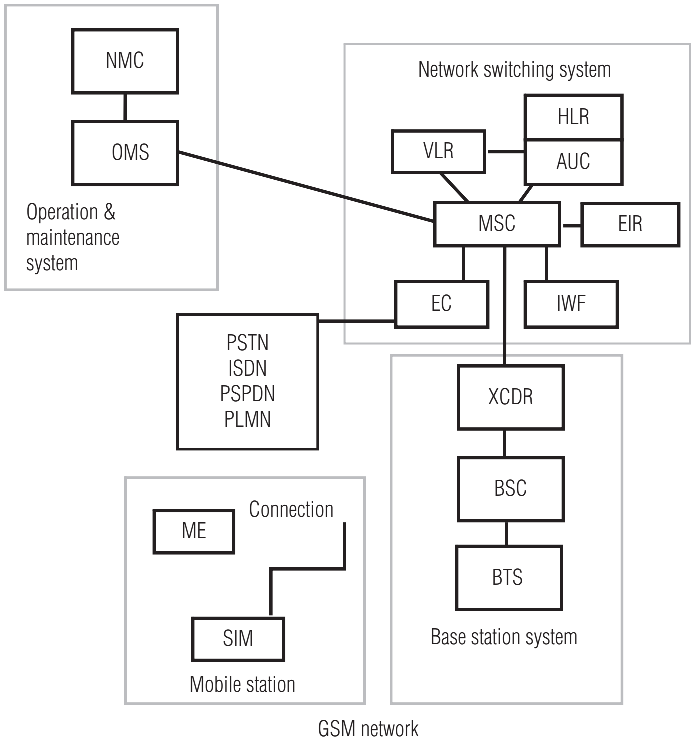


Figure 19.7 Simple GSM network architecture

- *Network switching system (NSS)*: This consists of the MSC and its associated system-control databases and processors together with the required interfaces. This is the part which provides for interconnection between the GSM network and the public switched telephone network (PSTN).
- *Operations and maintenance system (OMS)*: This enables the network provider to configure and maintain the network from a central location.

19.3.1 The mobile station

The MS consists of two components as shown in Figure 19.9: the mobile equipment (ME) and an electronic "smart card" called a subscriber identity module (SIM). The ME is the hardware used by the subscriber to access the network. The hardware has an identity number associated with it, which is unique for that particular device and permanently stored in it. This identity number is called the international mobile equipment identity (IMEI) and enables the network operator to identify ME which may be causing problems on the system.

The SIM is a card which plugs into the ME. This card identifies the MS subscriber and also provides other information regarding the service that the subscriber should receive. The subscriber is identified by an identity number called the international mobile subscriber identity (IMSI). ME may be purchased from any store but the SIM must be obtained from the GSM network provider.

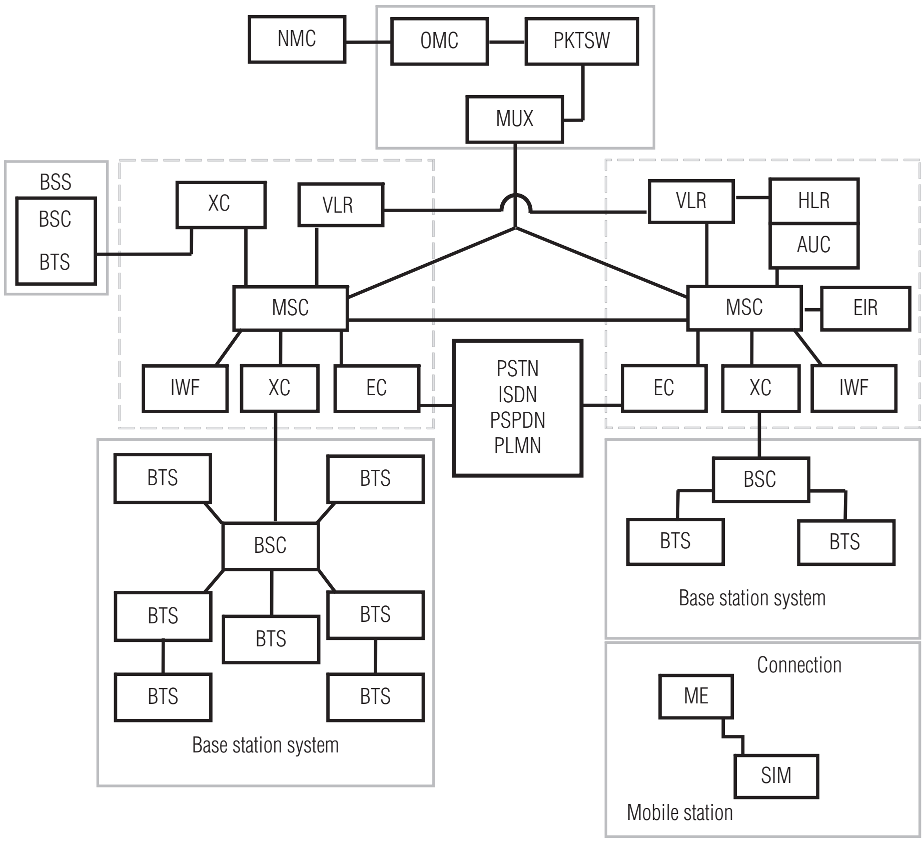


Figure 19.8 GSM network along with added elements

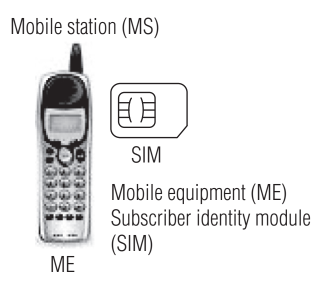


Figure 19.9 The mobile station

The SIM may be contained within the MS, or it may be a removable unit that can be inserted by the user. It provides personal mobility so that the user can have access to all subscribed services irrespective of both the location of the terminal and the use of a specific terminal. Without the SIM inserted, the ME will only be able to make emergency calls. By making a distinction between the subscriber identity and the ME identity, GSM can route calls and perform billing based on the identity of the "subscriber" rather than the equipment or its location.

The MS also acts as a receptor for SMS messages, enabling the user to toggle between the voice and data use. Moreover, the mobile facilitates access to voice messaging systems. The MS also provides access to the various data services available in a GSM network. These data services include the following:

- X.25 packet switching through a synchronous or asynchronous dial-up connection to the PAD at speeds typically at 9.6 kbps.
- GPRS using either an X.25 or IP-based data transfer method at speeds up to 115 kbps.
- High-speed circuit-switched data at speeds up to 64 kbps.

Mobile equipment (ME)

The ME is the only part of the GSM network which the subscriber will really see. There are three main types of ME. These are listed below:

- *Vehicle mounted*: These devices are mounted in a vehicle and the antenna is physically mounted on the outside of the vehicle.
- *Portable mobile unit*: This equipment can be handheld when in operation, but the antenna is not connected to the handset of the unit.
- *Hand portable unit*: This equipment comprises of a small telephone handset not much bigger than a calculator. The antenna is connected to the handset.

The ME is capable of operating at a certain maximum power output dependent on its type and use. These mobile types have distinct features which must be known by the network.

Subscriber identity module (SIM)

The SIM as mentioned previously is a "smart card" which plugs into the ME and contains information about the MS subscriber, hence the name Subscriber Identity Module (Figure 19.10).

The SIM contains several pieces of information:

- *International mobile subscriber identity* (IMSI): This number identifies the MS subscriber. It is transmitted over the air only during initialization.
- *Temporary mobile subscriber identity* (TMSI): This number identifies the subscriber; it is periodically changed by the system management to protect the subscriber from being identified by someone attempting to monitor the radio interface.
- *Location area identity* (LAI): Identifies the current location of the subscriber. The fields of LAI are (1) country code (CC): three decimal places; (2) mobile network code (MNC): two decimal places; (3) location area code (LAC): maximum five decimal places or maximum twice the 8 bits coded in hexadecimal (LAC < FFFF).
- *Subscriber authentication key* (Ki): This is used to authenticate the SIM card.

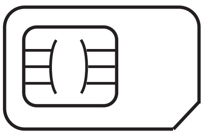


Figure 19.10 Subscriber identity module

- *Mobile station international services digital network* (MSISDN): This is the telephone number of the mobile subscriber. It is comprised of a country code, a network code, and a subscriber number. The MSISDN categories follow the international ISDN number plan and therefore have the following structure.

1. Country Code (CC): Up to three decimal places.
2. National Destination Code (NDC): Typically two to three decimal places.
3. Subscriber Number (SN): Maximum 10 decimal places.

Most of the data contained within the SIM is protected against reading (Ki) or alterations (IMSI). Some of the parameters (LAI) will be continuously updated to reflect the current location of the subscriber.

The SIM card and the high degree of inbuilt system security provide protection of the subscriber's information and networks against fraudulent access. SIM cards are designed to be difficult to duplicate. The SIM can be protected by use of a personal identity number (PIN) password, similar to bank/credit charge cards, to prevent unauthorized use of the card.

The SIM is capable of storing additional information such as accumulated call charges. This information will be accessible to the customer via handset/keyboard key entry.

19.3.2 Base station subsystem (BSS)

The GSM BSS is the equipment located at a cell site. It comprises of a combination of digital and RF equipment. The BSS provides the link between the MS and the MSC. Generally, the BSS communicates with the MS over the digital air interface and with the MSC via 2 Mbps links.

The BSS contains a BSC and one or more sub-standing BTSs. The BSS is responsible for all functions related to the radio resources channel management. This includes the management of radio channel configuration with respect to use as speech, data, or signalling channels, allocation and release of channels for call set-up and release, control of frequency hopping, and transmitted power at the MS.

The BSS is composed of the following:

- Base transceiver station (BTS)
- Base station controller (BSC)
- Transcoder (XCDR)

The BTS and BSC communicate over an interface known as Abis interface. This interface establishes compatibility between components from different suppliers and allows operations between them.

The BSS consists of four to seven or nine cells and one or more base stations. A high-speed line (T1 or E1) is established between BSS and MSC as shown in Figure 19.11.

Base transceiver station

The BTS contains a radio transceiver and antennas. Power transmitted from the BTS defines a cell in the network. As shown in Figure 19.12, BTS is placed in the centre of cell and handles the radio link protocols with MS. A large number of BTSs are deployed in a large urban area. Number of transceivers with each BTS depends on density of users in the cell. Each BTS may contain 1–16 transceivers.

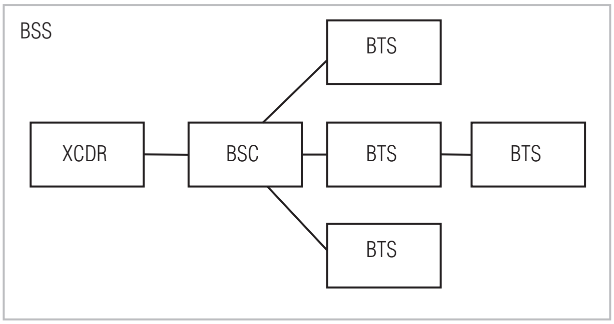


Figure 19.11 GSM base subsystem station

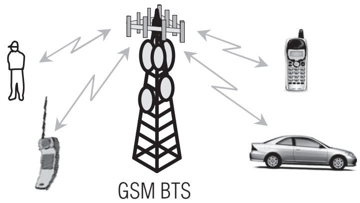


Figure 19.12 GSM base transceiver station

The BTS does the following functions:

- Frequency and time synchronization
- Encoding, encrypting, multiplexing, modulating, and feeding RF signals to the antenna.
- Voice through full- or half-rate services
- Timing advances
- Transcoding and rate adaptation
- Decoding, decrypting, and equalizing received signals
- Random access detection
- Uplink channel measurements

Base station controller

BSC is a switching device. It acts as a radio resource manager of BTSs as it performs the functions of frequency hopping, radio channel setup, and handovers (both intercell and intracell). It also does the following functions:

- Establishes connection between MSC and mobile
- Manages the time slots and frequencies for the MS
- Translates the voice channel (13 kps) over the radio link to standard channel (64 kbps) used by PSTN or ISDN
- Allocates necessary time slots between the BTS and MS in its area
- Reallocates frequencies among BTSs
- Controls frequency hopping

- Provides an interface to the operations and maintenance centre (OMC) for the BSS
- Performs traffic concentration to reduce the number of lines from the MSC
- Synchronizes frequency and time
- Does time-delay measurements of received signals from the MS
- Power management

The transcoder

The transcoder is used to compact the signals from the MS so that they are more efficiently sent over the terrestrial interfaces. Although the transcoder is considered to be a part of the BSS, it is very often located closer to the MSC. The transcoder is used to reduce the rate at which the traffic (voice/data) is transmitted over the air interface. Although the transcoder is part of the BSS, it is often found physically closer to the NSS to allow for more efficient use of the terrestrial links.

The transcoder converts the voice or data output from MSC into a form mentioned in GSM specification, that is it converts signals from 64 kbps to 16 kbps and vice versa.

The 64 kbps pulse code modulation (PCM) circuits from the MSC, if transmitted on the air interface without modification, would occupy an excessive amount of radio bandwidth. This would use the available radio spectrum inefficiently. The required bandwidth is therefore reduced by processing the 64 kbps circuits so that the amount of information required to transmit digitized voice falls to a gross rate of 16 kbps. The transcoding function may be located at the MSC, BSC, or BTS (Figure 19.13).

The content of the 16 kbps data depends on the coding algorithm used. There are two speech coding algorithms available and selecting which one to use depends on the capabilities of the ME and the network configuration.

The full rate (FR) speech algorithm is supported by all mobiles and networks. It produces 13 kbps of coded speech data plus 3 kbps of control data which is commonly referred to as transcoder

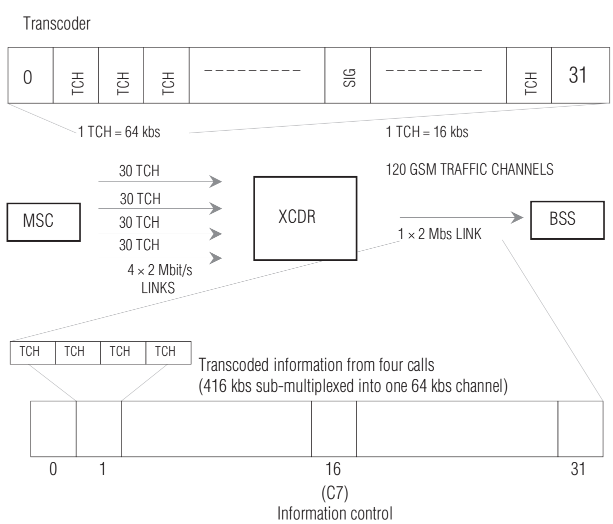


Figure 19.13 The transcoder (XCDR)

rate adaptation unit (TRAU) data. The TRAU data on the downlink will be used by the BTS and therefore removed from the 13 kbps of speech data before transmission on the air interface.

The 13 kbps of speech data is processed at the BTS to form a gross rate of 22.8 kbps on the air interface which includes forward error correction. In the uplink direction, the BTS adds in TRAU data which will be used by the transcoder.

Enhanced FR (EFR) is an improved speech coding algorithm and is only supported by Phase 2+ mobiles and is optional in the network. It produces 12.2 kbps from each 64 kbps PCM channel. The TRAU data in this case is made up to 3.8 kbps to keep the channel rate of the BTS at 16 kbps as for FR. As with FR, the TRAU data is used at the BTS and the transcoder.

For data transmissions, the data is not transcoded but data rate is adapted from 9.6 kbps (4.8 kbps or 2.4 kbps may also be used) up to a gross rate of 16 kbps for transmission over the terrestrial interfaces. Again, this 16 kbps contains a 3 kbps TRAU.

As can be seen from the Figure 19.13, although the reason for transcoding was to reduce the data rate over the air interface, the number of terrestrial links is also reduced approximately on a 4:1 ratio.

BSS configurations

As we have mentioned, a BSC may control several BTSs. The maximum number of BTSs which may be controlled by one BSC is not specified by GSM. Individual manufacturer's specifications may vary greatly. The BTSs and BSC may either be located at the same cell site ("co-located"), or located at different sites ("remote"). In reality, most BTSs will be remote, as there are many more BTSs than BSCs in a network (Figure 19.14).

Another BSS configuration is the daisy chain. A BTS need not communicate directly with the BSC which controls it and it can be connected to the BSC via a chain of BTSs. Daisy chaining reduces the amount of cabling required to set up a network as a BTS can be connected to its

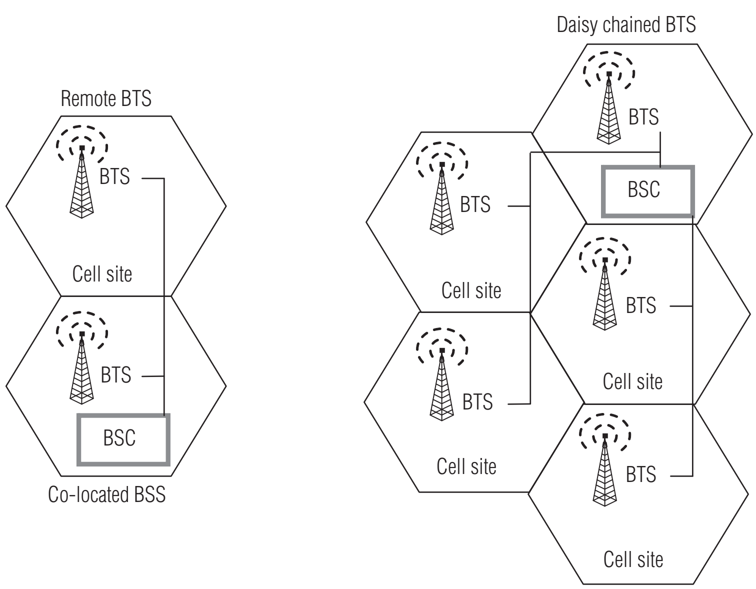


Figure 19.14 BSS configurations

nearest BTS rather than all the way to the BSC. Problems may arise when chaining BTSs, due to the transmission delay through the chain. The length of the chain must, therefore, be kept sufficiently short to prevent the round trip speech delay becoming too long.

Other topologies are also permitted, including stars and loops. Loops are used to introduce redundancy into the network. For example, if a BTS connection is lost, the BTS may still be able to communicate with the BSC if a second connection is available.

19.3.3 Network switching subsystem

The NSS includes the main switching functions of the GSM network. The main part of NSS is the mobile switching centre (MSC) which performs the switching of calls between the mobile and other fixed or mobile network users, as well as the management of mobile services such as authentication (Figure 19.15). It also contains the databases required for subscriber data and mobility management. Its main function is to manage communications between the GSM network and other telecommunications networks.

The components of the NSS are listed below:

- Mobile service switching centre (MSC)
- Home location register (HLR)
- Visitor location register (VLR)
- Equipment identity register (EIR)
- Authentication centre (AUC)
- Interworking function (IWF)
- Echo canceller (EC)

In addition to the more traditional elements of a cellular telephone system, GSM has location register network entities. These entities are HLR, VLR, and EIR. The location registers are database-oriented processing nodes which address the problems of managing subscriber data and keeping track of a MS's location as it roams around the network.

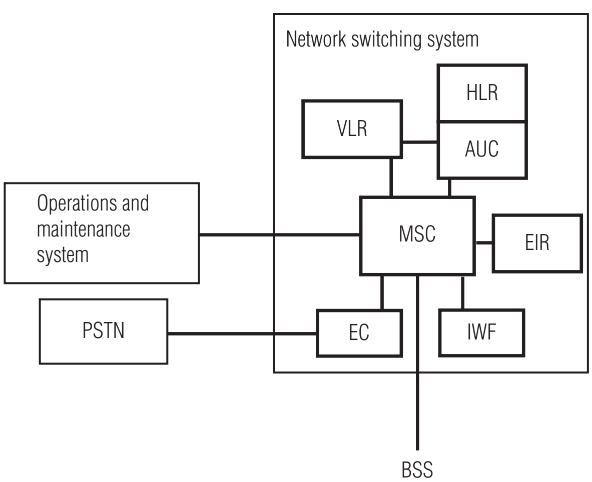


Figure 19.15 The elements of network switching subsystem

Functionally, the IWF and the ECs may be considered as parts of the MSC, since their activities are inextricably linked with those of the switch as it connects speech and data calls to and from the MSs.

Mobile service switching centre (MSC)

The MSC is included in the GSM system for call-switching. Its overall purpose is the same as that of any telephone exchange. However, because of the additional complications involved in the control and security aspects of the GSM cellular system and the wide range of subscriber facilities that it offers, the MSC has to be capable of fulfilling many additional functions.

The central component of the network subsystem is the MSC. The MSC performs the switching of calls between the mobile and other fixed or mobile network users, as well as the management of mobile services such as registration, authentication, location updating, handovers, and call routing to a roaming subscriber.

The MSC will carry out several different functions depending upon its position in the network. When the MSC provides the interface between the PSTN and the BSSs in the GSM network, it is known as a gateway MSC. In this position, it will provide the switching required for all MS originated or terminated traffic.

Each MSC provides service to MSs located within a defined geographic coverage area. The network typically contains more than one MSC. One MSC is capable of supporting a regional capital with approximately one million inhabitants. An MSC of this size will be contained in about half a dozen racks.

The functions carried out by the MSC are listed below:

- *Call processing:* Includes control of data/voice call setup, supervision and release, inter-BSS, and inter-MSC handovers, and control of mobility management (subscriber validation and location).
- *Operations and maintenance support:* Includes database management, management of radio resources, registration, location updating and traffic metering and measurement, and a man–machine interface.
- *Internetwork interworking:* Manages the interface between the GSM network and the PSTN.
- *Billing:* Collects call billing data.

Home location register (HLR)

The HLR is the reference database for subscriber parameters used for storage and management of new and old subscriptions. It stores permanent data about subscribers which include service profile, activity status, and location information. It is therefore considered as most important database. When a person gets a SIM, then all his information mentioned in subscription form is registered in HLR.

Usually, one HLR is developed for each GSM network for administration of subscriber configuration and services. Besides the up-to-date location information for each subscriber, which is dynamic, the HLR maintains the following subscriber data on a permanent basis:

- International mobile subscriber identity (IMSI)
- Service subscription information
- Service restrictions
- Supplementary services
- Mobile terminal characteristics
- Billing/accounting information

Visitor location register

The VLR is a database that lessens the burden on HLR. It is integrated with MSC and contains temporary information about subscribers. This information is required by the MSC in order to service visiting subscribers. The VLR will get the data about a MS from the HLR, whenever the MS roams into a new MSC area. Later this information contained in VLR will be used for call setup, whenever that MS makes a call without interrogating HLR. The VLR also contains information about locally activated features such as call forward on busy.

The additional data stored in the VLR is listed below:

- Mobile status (busy/free/no answer, etc.)
- Location area identity (LAI)
- Temporary mobile subscriber identity (TMSI)
- Mobile Station Roaming Number (MSRN)

Location area identity (LAI)

Cells within the public land mobile network (PLMN) are grouped together into geographical areas. Each area is assigned a LAI. A location area may typically contain 30 cells. Each VLR controls several LAIs and as a subscriber moves from one LAI to another, the LAI is updated in the VLR. As the subscriber moves from one VLR to another, the VLR address is updated at the HLR.

Temporary mobile subscriber identity

The VLR controls the allocation of new TMSI numbers and notifies them to the HLR. The TMSI will be updated frequently. This makes it very difficult for the call to be traced and therefore provides a high degree of security for the subscriber.

The TMSI may be updated in any of the following situations:

- Call setup
- On entry to a new LAI
- On entry to a new VLR

Mobile subscriber roaming number

As a subscriber may wish to operate outside its "home" system at some time, the VLR can also allocate a MSRN. This number is assigned from a list of numbers held at the VLR (MSC). The MSRN is then used to route the call to the MSC which controls the base station in the MSs current location.

Equipment identity register

The EIR maintains information to authenticate terminal equipment so that fraudulent, stolen, or non-type-approved terminals can be identified and service can be denied. The information is in the form of white, gray, and black lists that may be consulted by the network when it wishes to confirm the authenticity of the terminal requesting service.

The EIR database consists of lists of IMEIs (or ranges of IMEIs) organized as follows:

White list: Contains those IMEIs which are known to have been assigned to valid MS equipment.

Black list: Contains IMEIs of MS which have been reported stolen or which are to be denied service for some other reason.

Gray list: Contains IMEIs of MS which have problems (e.g., faulty software). These are not, however, sufficiently significant to warrant a "black listing".

The EIR database is remotely accessed by the MSCs in the network and can also be accessed by an MSC in a different PLMN.

As in the case of the HLR, a network may well contain more than one EIR with each EIR controlling certain blocks of IMEI numbers. The MSC contains a translation facility, which when given an IMEI returns the address of the EIR controlling the appropriate section of the equipment database.

Authentication centre (AUC)

The AUC is a processor system. It performs the "authentication" function. It will normally be co-located with the HLR as it will be required to continuously access and update, as necessary, the system subscriber records. The AUC/HLR centre can be co-located with the MSC or located remote from the MSC.

The authentication process will usually take place each time the subscriber "initializes" on the system.

Process of authentication

The process of authentication is shown in Figure 19.16.

To discuss the authentication process, we will assume that the VLR has all the information required to perform that authentication process (Kc, SRES, and RAND). If this information is unavailable, then the VLR would request it from the HLR/AUC.

Steps:

1. Triples (Kc, SRES, and RAND) are stored at the VLR.
2. The VLR sends RAND via the MSC and BSS to the MS (unencrypted).
3. The MS, using the A3 and A8 algorithms and the parameter Ki stored on the MS SIM card, together with the received RAND from the VLR, calculates the values of SRES and Kc.
4. The MS sends SRES unencrypted to the VLR.
5. Within the VLR, the value of SRES is compared with the SRES received from the mobile. If the two values match, then the authentication is successful.
6. If ciphering is to be used, Kc from the assigned triple is passed to the BTS.
7. The mobile calculates Kc from the RAND and A8 and Ki on the SIM.
8. Using Kc, A5 and the GSM hyper frame number, encryption between the MS and the BSS can now occur over the air interface.

   Note: The triples are generated at the AUC by the following:
   RAND = randomly generated number
   SRES = derived from A3 (RAND, Ki)
   Kc = derived from A8 (RAND, Ki)
   A3 = From 1 of 16 possible algorithms defined on allocation of IMSI and creation of SIM card
   A8 = From 1 of 16 possible algorithms defined on allocation of IMSI and creation of SIM card
   Ki = Authentication key, assigned at random together with the versions of A3 and A8

The first time a subscriber attempts to make a call, the full authentication process takes place. However, for subsequent calls attempted within a given system control time period, or within a single system provider's network, authentication may not be necessary, as the data generated during the first authentication will still be available.

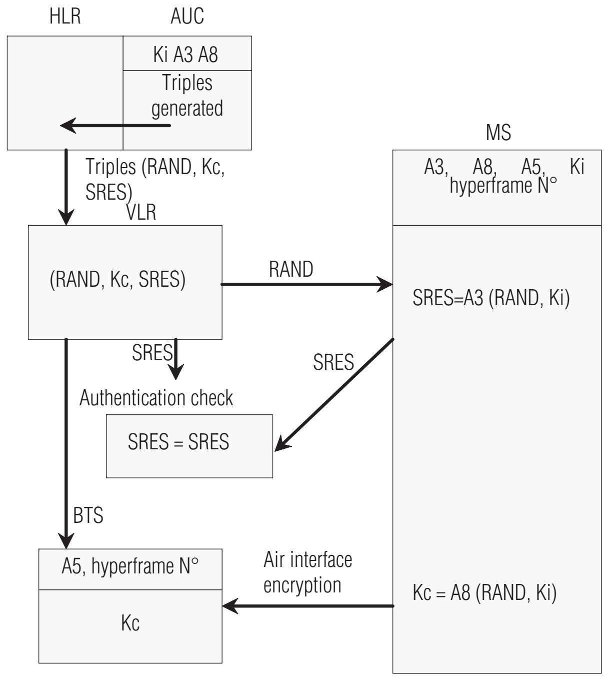


Figure 19.16 Process of authentication

Interworking function

The IWF provides the function to enable the GSM system to interface with the various forms of public and private data networks currently available. The basic features of the IWF are listed below:

- Data rate adaption
- Protocol conversion

Some systems require more IWF capability than others; this depends upon the network to which it is being connected.

The IWF also incorporates a “modem bank”, which may be used when, for example, the GSM data terminal equipment (DTE) exchanges data with a land DTE connected via an analogue modem as shown in Figure 19.17.

Echo canceller

An EC is used on the PSTN side of the MSC for all voice circuits. Echo control is required at the switch because the inherent GSM system delay can cause an unacceptable echo condition, even on short distance PSTN circuit connections (Figure 19.18).

The total round trip delay introduced by the GSM system (e.g. the cumulative delay caused by call processing, speech encoding, and decoding) is approximately 180 ms. This would not be apparent to the MS subscriber, but the inclusion of a two-wire to four-wire hybrid

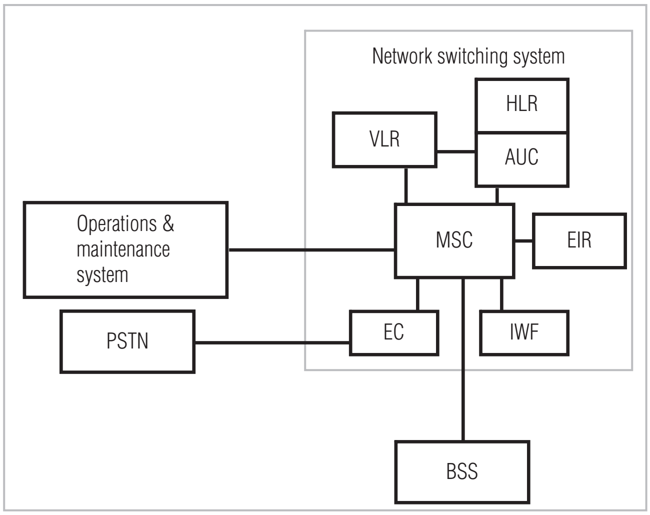


Figure 19.17 Interworking functional unit

transformer in the circuit is required at the land party's local switch because the standard telephone connection is two-wire. The transformer causes the echo. This does not affect the land subscriber.

During a normal PSTN land-to-land call, no echo is apparent because the delay is too short and the user is unable to distinguish between the echo and the normal telephone "side tone". However, without the EC and with the GSM round trip delay added, the effect would be very irritating to the MS subscriber, disrupting speech and concentration.

The standard EC will provide cancellation of up to 68 ms on the "tail circuit" (the tail circuit is the connection between the output of the EC and the land telephone).

19.4 GSM—specifications

The specifications for different personal communication services (PCS) systems vary among the different PCS networks. The GSM specification is listed below with important characteristics.

19.4.1 Modulation

Modulation is a form of change process where we change the input information into a suitable format for the transmission medium. We also changed the information by demodulating the signal at the receiving end.

The GSM uses Gaussian minimum shift keying (GMSK) modulation method.

19.4.2 Access methods

Because radio spectrum is a limited resource shared by all users, a method must be devised to divide up the bandwidth among as many users as possible.

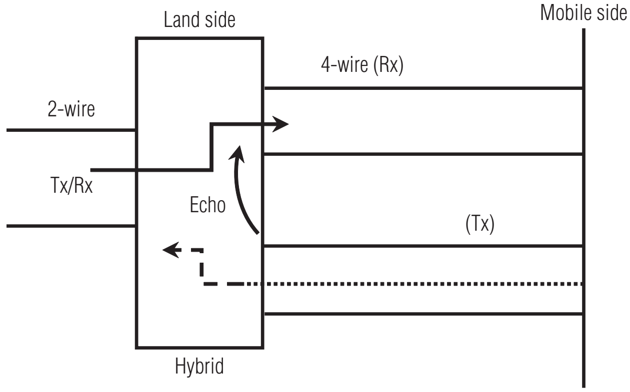


Figure 19.18 Generation of echoes at two-wire to four-wire interface

GSM chose a combination of TDMA/FDMA as its method. The FDMA part involves the division by frequency of the total 25 MHz bandwidth into 124 carrier frequencies of 200 kHz bandwidth.

One or more carrier frequencies are then assigned to each BS. Each of these carrier frequencies is then divided in time, using a TDMA scheme, into eight time slots. One time slot is used for transmission by the mobile and one for reception. They are separated in time so that the mobile unit does not receive and transmit at the same time.

19.4.3 Transmission rate

The total symbol rate for GSM at 1 bit per symbol in GMSK produces 270.833 K symbols/second. The gross transmission rate of the time slot is 22.8 kbps.

GSM is a digital system with an over-the-air bit rate of 270 kbps.

19.5 GSM—addresses and identifier

GSM distinguishes explicitly between user and equipment and deals with them separately. Besides phone numbers and subscriber and equipment identifiers, several other identifiers have been defined; they are needed for the management of subscriber mobility and for addressing of all the remaining network elements. The most important addresses and identifiers are presented in the following:

19.5.1 International mobile station equipment identity

The IMEI uniquely identifies a MS internationally. It is a kind of serial number. The IMEI is allocated by the equipment manufacturer and registered by the network operator who stores it in the EIR. By means of IMEI, one recognizes obsolete, stolen, or non-functional equipment.

The parts of an IMEI are as follows:

- *Type approval code* (TAC): six decimal places, centrally assigned.
- *Final assembly code* (FAC): six decimal places, assigned by the manufacturer.
- *Serial number* (SNR): six decimal places, assigned by the manufacturer.
- *Spare* (SP): one decimal place.

Thus, IMEI = TAC + FAC + SNR + SP. It uniquely characterizes a MS and gives clues about the manufacturer and the date of manufacturing.

19.5.2 International mobile subscriber identity

Each registered user is uniquely identified by its IMSI. It is stored in the SIM. A MS can only be operated if a SIM with a valid IMSI is inserted into an equipment with a valid IMEI.

The parts of an IMSI are as follows:

- *Mobile country code (MCC)*: three decimal places, internationally standardized.
- *Mobile network code (MNC)*: two decimal places, for unique identification of mobile network within the country.
- *Mobile subscriber identification number (MSIN)*: Maximum 10 decimal places, identification number of the subscriber in the home mobile network.

19.5.3 Mobile subscriber ISDN number

The real telephone number of a MS is the MSISDN. It is assigned to the subscriber (his or her SIM, respectively) such that a MS set can have several MSISDNs depending on the SIM.

The MSISDN categories follow the international ISDN number plan and therefore have the following structure.

- *Country code (CC)*: Up to three decimal places.
- *National destination code (NDC)*: Typically two to three decimal places.
- *Subscriber number (SN)*: Maximum 10 decimal places.

19.5.4 Mobile station roaming number (MSRN)

The MSRN is a temporary location dependent ISDN number. It is assigned by the locally responsible VLR to each MS in its area. Calls are also routed to the MS by using the MSRN.

The MSRN has same structure as the MSISDN.

- *Country code* (CC) of the visited network.
- *National destination code* (NDC) of the visited network.
- *Subscriber number* (SN) in the current mobile network.

19.5.5 Location area identity

Each location area of an PLMN has its own identifier. The LAI is also structured hierarchically and is internationally unique as follows:

- *Country Code* (CC): three decimal places.
- *Mobile Network Code* (MNC): two decimal places.
- *Location Area Code* (LAC): maximum five decimal places or maximum twice 8 bits coded in hexadecimal (LAC < FFFF).

19.5.6 Temporary mobile subscriber identity

The VLR, which is responsible for the current location of a subscriber, can assign a TMSI which has only local significance in the area handled by the VLR. It is stored on the network side only in the VLR and is not passed to the HLR.

Together with the current location area, TMSI allows a subscriber to be identified uniquely and it can consist of up to 4×8 bits.

19.5.7 Local mobile subscriber identity

The VLR can assign an additional searching key to each MS within its area to accelerate database access. This unique key is called the local mobile subscriber identity (LMSI). The LMSI is assigned when the MS registers with the VLR and is also sent to the HLR.

An LIMSI consists of four octets (4×8 bits).

19.5.8 Cell identifier

Within the location area, the individual cells are uniquely identified with a cell identifier (CI) of maximum 2×8 bits. Together with the global cell identity (LAI + CI), calls are thus also internationally defined in a unique way.

19.6 GSM operation

The operation of the GSM system can be understood by studying the sequence of events that takes place when a call is initiated from the MS.

19.6.1 Call from mobile phone to PSTN

When a mobile subscriber makes a call to a PSTN telephone subscriber, the following sequence of events takes place:

1. The MSC/VLR receives the message of a call request.
2. The MSC/VLR checks if the MS is authorized to access the network. If so, the MS is activated. If the MS is not authorized, service will be denied.
3. MSC/VLR analyses the number and initiates a call setup with the PSTN.
4. MSC/VLR asks the corresponding BSC to allocate a traffic channel (a radio channel and a time slot).
5. The BSC allocates the traffic channel and passes the information to the MS .
6. The called party answers the call and the conversation takes place.
7. The MS keeps on taking measurements of the radio channels in the present cell and neighbouring cells and passes the information to the BSC. The BSC decides if a handover is required. If so, a new traffic channel is allocated to the MS and the handover is performed. If handover is not required, the MS continues to transmit in the same frequency.

19.6.2 Call from PSTN to mobile phone

When a PSTN subscriber calls a MS, the sequence of events is as follows:

- The gateway MSC receives the call and queries the HLR for the information needed to route the call to the serving MSC/VLR.
- The GMSC routes the call to the MSC/VLR.
- The MSC checks the VLR for the location area of the MS.
- The MSC contacts the MS via the BSC through a broadcast message, that is through a paging request.
- The MS responds to the page request.
- The BSC allocates a traffic channel and sends a message to the MS to tune to the channel. The MS generates a ringing signal and, after the subscriber answers, the speech connection is established.
- Handover, if required, takes place, as discussed earlier.

The MS codes the speech at 13 kbps for transmission over the radio channel in the given time slot. The BSC converts (or transcodes) the speech to 64 kbps and sends it over a land link or radio link to the MSC. The MSC then forwards the speech data to the PSTN. In the reverse direction, the speech is received at 64 kbps rate at the BSC and the BSC does the transcoding to 13 kbps for radio transmission.

In its original form, GSM supports 9.6 kbps data, which can be transmitted in one TDMA time slot. Over the last few years, many enhancements were done to the GSM standards (GSM Phase 2 and GSM Phase 2+) to provide higher data rates for data applications.

19.7 Operations and maintenance system

Overview

The operations and maintenance system provides the capability to manage the GSM network remotely.

This area of the GSM network is not currently tightly specified by the GSM specifications. It is left to the network provider to decide what capabilities they wish it to have. As shown in Figure 19.19, the operations and maintenance system comprises of two parts:

- Network management centre (NMC)
- Operations and maintenance centre (OMC)

Network management centre (NMC): The NMC has a view of the entire PLMN and is responsible for the management of the network as a whole. The NMC resides at the top of the hierarchy and provides global network management.

Operations and maintenance centre (OMC): The OMC is a centralized facility that supports the day-to-day management of a cellular network as well as providing a database for long-term network engineering and planning tools. An OMC manages a certain area of the PLMN, thus giving regionalized network management.

19.7.1 Network management centre

The NMC offers the ability to provide hierarchical regionalized network management of a complete GSM system. It is responsible for operations and maintenance at the network level, supported by the OMCs which are responsible for regional network management.

The NMC is therefore a single logical facility at the top of the network management hierarchy.

The NMC, on the one hand, has a high-level view of the network, as a series of network nodes and interconnecting communications facilities. The OMC, on the other hand, is used to filter information from the network equipment for forwarding to the NMC, thus allowing it to

OMC (Regional)	NMC (Global)
Multiple OMCs per network *Regionalized* network management Employed in *daily operations* Used by network *operators*	*single* NMC per network *Global* network management Employed in *long-term planning* 24 hours supervision

Figure 19.19 Operations and maintenance system

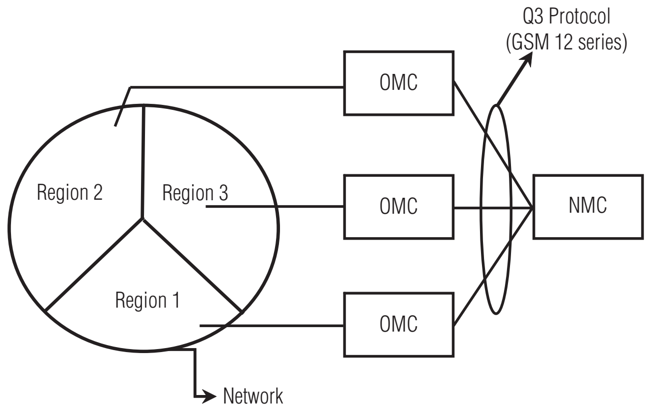


Figure 19.20 Functional units of network management centre

focus on issues requiring national co-ordination. The NMC can also co-ordinate issues regarding interconnection to other networks, for example the PSTN (Figure 19.20).

The NMC can take regional responsibility when an OMC is not manned, with the OMC acting as a transit point between the NMC and the network equipment. The NMC provides operators with functions equivalent to those available at the OMC.

Functionality of the NMC

- Monitors nodes on the network
- Monitors GSM network element statistics
- Monitors OMC regions and provides information to OMC staff
- Passes on statistical information from one OMC region to another to improve problem solving strategies
- Enables long-term planning for the entire network

19.7.2 Operations and maintenance centre (OMC)

The OMC provides a central point from which to control and monitor the other network entities (i.e. base stations, switches, database, etc.) as well as monitor the quality of service being provided by the network.

At present, equipment manufacturers have their own OMCs which are not compatible in every aspect with those of other manufacturers. This is particularly the case between radio BS equipment suppliers, where in some cases the OMC is a separate item, and digital switching equipment suppliers, where the OMC is an integral, but functionally separate part of the hardware and OS is the operating system in the OMC (Figure 19.21).

There are two types of OMC. They are as follows:

OMC (R) L: OMC controls specifically the BSS.
OMC (S): OMC controls specifically the NSS.

The OMC should support the following functions as per ITS–TS recommendations:

- Event/alarm management
- Fault management

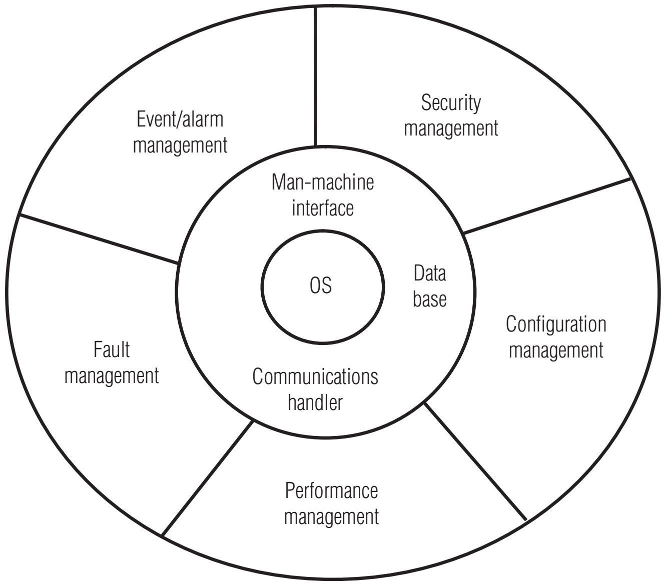


Figure 19.21 The OMC functional architecture

- Performance management
- Configuration management
- Security management

19.8 Channels on the air interface

One of the key elements of the development of the GSM was the development of the GSM air interface. There were many requirements that were placed on the system, and many of these had a direct impact on the air interface. Elements including the modulation, GSM slot structure, burst structure, and the like were all devised to provide the optimum performance.

GSM uses a variety of *channels* in which the data is carried. In GSM, these channels are separated into *physical channels* and *logical channels*. The physical channels are determined by the timeslot, whereas the logical channels are determined by the information carried within the physical channel. It can be further summarized by saying that several recurring timeslots on a carrier constitute a physical channel. These are then used by different logical channels to transfer information. These channels may either be used for user data (payload) or signalling to enable the system to operate correctly.

19.8.1 Transmission of analogue and digital signals

The main reasons why GSM uses a digital air interface are as follows:

- It is "noise robust" enabling the use of tighter frequency reuse patterns and minimizing interference problems.
- It incorporates error correction, thus protecting the traffic that it carries.

- It offers greatly enhanced privacy to subscribers and security to network providers.
- It is ISDN compatible, uses open standardized interfaces and offers an enhanced range of services to its subscribers.

Modulation techniques: There are three methods of modulating a signal so that it may be transmitted over the air:

- *Amplitude modulation (AM)*: Amplitude modulation is very simple to implement for analogue signals but it is prone to noise.
- *Frequency modulation (FM)*: Frequency modulation is more complicated to implement but provides a better tolerance to noise.
- *Phase modulation (PM)*: Phase modulation provides the best tolerance to noise but it is very complex to implement for analogue signals and therefore is rarely used.

Digital signals can use any of the modulation methods, but phase modulation provides the best noise tolerance. Since phase modulation can be implemented easily for digital signals, this is the method which is used for the GSM air interface. Phase modulation is known as phase shift keying (PSK) when applied to digital signals.

Transmission of digital signals:Phase modulation provides a high degree of noise tolerance. However, there is a problem with this form of modulation. When the signal changes phase abruptly, high-frequency components are produced and thus, a wide bandwidth would be required for transmission.

GSM has to be as efficient as possible with the available bandwidth. Therefore, it is not this technique but a more efficient development of phase modulation that is actually used by the GSM air interface. It is called Gaussian minimum shift keying (GMSK).

Gaussian minimum shift keying (GMSK): Gaussian minimum shift keying or Gaussian filtered minimum shift keying, is a form of modulation used in a variety of digital radio communications systems. It has advantages of being able to carry digital modulation while still using the spectrum efficiently.

GMSK modulation is based on MSK, which is itself a form of phase shift keying. One of the problems with standard forms of PSK is that sidebands extend out from the carrier. To overcome this, MSK and its derivative GMSK can be used.

With GMSK, the phase change which represents the change from a digital '1' or a '0' does not occur instantaneously as it does with binary phase shift keying (BPSK). Instead it occurs over a period of time and therefore the addition of high-frequency components to the spectrum is reduced as shown in Figure 19.22.

With GMSK, first the digital signal is filtered through a Gaussian filter. This filter causes distortion to the signal and the corners are rounded off. This distorted signal is then used to phase shift the carrier signal. The phase change therefore is no longer instantaneous but spread out.

Example problem 19.1

The GSM system uses the GMSK modulation scheme. Show that the bandwidth efficiency of the standard GSM system is 1.35 bps/Hz.

Solution

The channel bandwidth = 200 kHz (standard)
The channel data rate = 270.833 kbps (standard)

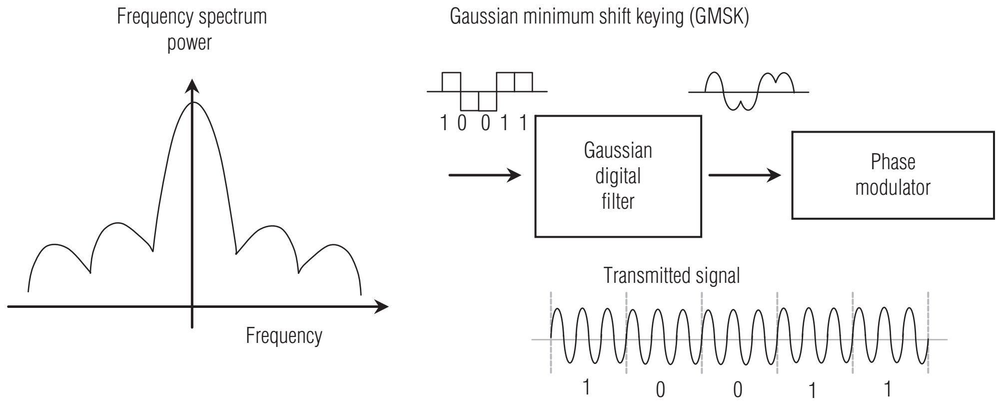


Figure 19.22 Process of Gaussian minimum shift keying

Bandwidth efficiency = channel data rate/channel bandwidth
Therefore, bandwidth efficiency = 270.833 kbps/200 kHz
Hence, bandwidth efficiency = 270.833 kbps/200 kHz = **1.35 bps/Hz**

Example problem 19.2

Show that the 3 dB bandwidth for a Gaussian LPF used to produce $B \times T_b = 0.3$ GMSK modulation in GSM standard is 81.3 kHz. The channel data rate is 270.833 kbps.

Solution

Channel data rate, R_b = 270.833 kbps (given)
Baseband symbol duration, $T_b = 1/R_b$
Baseband symbol duration, T_b = 1/270.833 kbps = 3.69 μs
Product of 3 dB bandwidth and baseband symbol duration, $B \times T_b = 0.3$ (given)
Therefore, 3 dB bandwidth, $B = 0.3/T_b$
Hence, 3 dB bandwidth, B = 0.3/3.69 μs = 81.3 kHz

Hence, the 3 dB bandwidth for a Gaussian LPF used to produce $B \times T_b = 0.3$ GMSK modulation in GSM standard is 81.3 kHz.

19.8.2 Physical and logical channels

The physical channel is the medium over which the information is carried. In the case of a terrestrial interface, this would be a cable. The logical channels consist of the information carried over the physical channel.

GSM physical channels

A single GSM RF carrier can support up to eight MS subscribers simultaneously. The diagram opposite shows how this is accomplished. Each channel occupies the carrier for one-eighth of the time. This is a technique called time division multiple access (TDMA).

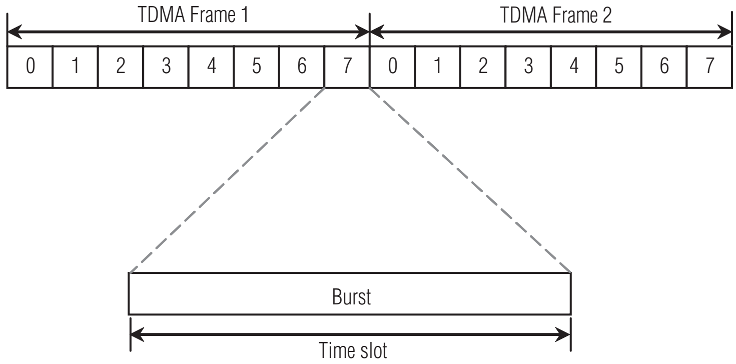


Figure 19.23 Timeslots and TDMA frame

Time is divided into discrete periods called "timeslots" shown in Figure 19.23. The timeslots are arranged in sequence and are conventionally numbered 0 to 7. Each repetition of this sequence is called a "TDMA frame".

Each MS telephone call occupies one timeslot (0–7) within the frame until the call is terminated, or a handover occurs. The TDMA frames are then built into further frame structures according to the type of channel. We shall later examine how the information carried by the air interface builds into frames and multi-frames and discuss the associated timing.

For such a system to work correctly, the timing of the transmissions to and from the mobiles is critical. The MS or BS must transmit the information related to one call at exactly the right moment, or the timeslot will be missed. The information carried in one timeslot is called a "burst". Each data burst, occupying its allocated timeslot within successive TDMA frames, provides a single GSM physical channel carrying a varying number of logical channels between the MS and the BTS.

GSM logical channels

There are two main groups of logical channels: traffic channels and control channels.
GSM traffic channels (TCH): The traffic channel carries speech or data information. The different types of traffic channels are listed below:

Full rate

TCH/FS : Speech (13 kbps net, 22.8 kbps gross)
TCH/EFR : Speech (12.2 kbps net, 22.8 kbps gross)
TCH/F9.6 : 9.6 kbps of data
TCH/F4.8 : 4.8 kbps of data
TCH/F2.4 : 2.4 kbps of data

Half rate

CH/HS : Speech (6.5 kbps net, 11.4 kbps gross)
TCH/H4.8 : 4.8 kbps of data
TCH/H2.4 : 2.4 kbps of data

Acronyms

TCH : Traffic channel
TCH/FS : Full rate speech channel

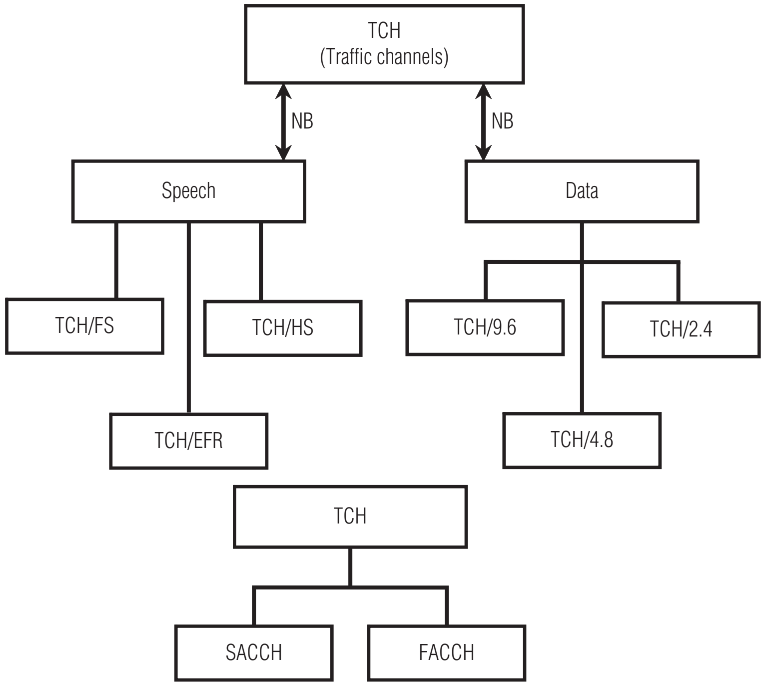


Figure 19.24 Subtypes of traffic channels

TCH/EFR: Enhanced full rate speech
TCH/HS: Half rate speech channel
TCH/9.6: Data channel 9.6 kbps
TCH/4.8 : Data channel 4.8 kbps
TCH/2.4: Data channel 2.4 kbps
SACCH: Slow associated control channel
FACCH: Fast associated control channel
NB: Normal burst

Speech channels: Speech channels are supported by two different methods of coding known as FR and EFR. EFR coding provides a speech service that has improved voice quality from the original FR speech coding, whilst using the same air interface bandwidth. EFR employs a new speech coding algorithm in addition to the FR channel coding algorithm to accomplish this improved speech service. However, it will only be supported by Phase 2+ mobiles (Figure 19.24).

19.8.3 GSM control channel groups

The control channel groups may be divided into broadcast control channel (BCCH), common control channel (CCCH), and the dedicated control channel (DCCH) groups as shown in Figure 19.25.
BCCH group: The BCCH are downlink only (BSS to MS) and comprise the following:

- BCCH carries information about the network, a MS's present cell and the surrounding cells. It is transmitted continuously as its signal strength is measured by all MSs on surrounding cells.
- The synchronizing channel (SCH) carries information for frame synchronization.

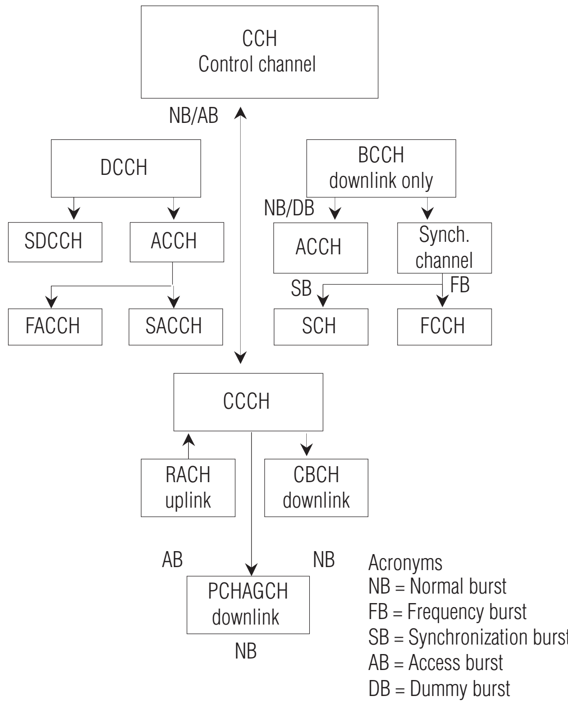


Figure 19.25 Subtype of CCH

CCCH group: The CCCH group works in both uplink and downlink directions:

- Random access channel (RACH) is used by MSs to gain access to the system.
- Paging channel (PCH) and access granted channel (AGCH) operate in the "downlink" direction. The AGCH is used to assign resources such as a stand-alone dedicated control channel (SDCCH) to the MS. The PCH is used by the system to call a MS. The PCH and AGCH are never used at the same time.
- Cell broadcast channel (CBCH) is used to transmit messages (e.g. road traffic information, sporting results, and so on) to be broadcast to all MSs within a cell.

DCCH group: DCCHs are assigned to a single MS for call setup and subscriber validation. DCCH comprises of the following:

- SDCCH which supports the transfer of data to and from the MS during call setup and validation.
- ACCH consists of slow ACCH which is used for radio link measurement and power control messages. Fast ACCH is used to pass "event" type messages, for example handover messages. Both FACCH and SACCH operate in uplink and downlink directions.

Broadcast control channel

The BCCH is transmitted by the BTS at all times. The RF carrier used to transmit the BCCH is referred to as the BCCH carrier. The information carried on the BCCH is monitored by the MS periodically (at least every 30 s), when it is switched on and not in a call. BCCH carries the following information (this is only a partial list):

- Location area identity (LAI)
- List of neighbouring cells which should be monitored by the MS
- List of frequencies used in the cell
- Cell identity
- Power control indicator
- DTX permitted
- Access control (e.g., emergency calls, call barring)
- CBCH description

The BCCH is transmitted at constant power at all times, and its signal strength is measured by all MS which may seek to use it. “Dummy” bursts are transmitted to ensure continuity when there is no BCCH carrier traffic.

1. *Frequency correction channel (FCCH)*: This is transmitted frequently on the BCCH timeslot and allows the mobile to synchronize its own frequency to that of the transmitting base site. The FCCH may only be sent during timeslot 0 on the BCCH carrier frequency and therefore it acts as a flag to the mobile to identify timeslot 0.
2. *Synchronization channel (SCH)*: The SCH carries the information to enable the MS to synchronize to the TDMA frame structure and know the timing of the individual timeslots.

The following parameters are sent:

- Frame number
- Base site identity code (BSIC)

The MS will monitor BCCH information from surrounding cells and store the information from the best six cells. The SCH information on these cells is also stored so that the MS may quickly resynchronize when it enters a new cell as seen in Figure 19.26.

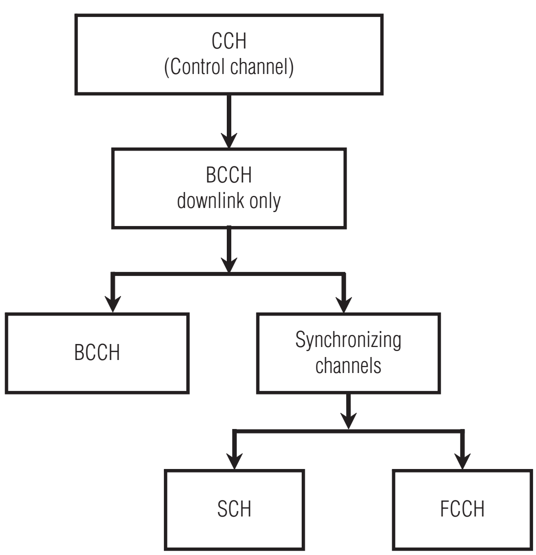


Figure 19.26 Subtype of BCCH

Common control channels (CCCH)

The CCCH is responsible for transferring control information between all mobiles and the BTS. This is necessary for the implementation of "call origination" and "call paging" functions (Figure 19.27).

It consists of the following:

Random access channel (RACH): Used by the mobile when it requires to gain access to the system. This occurs when the mobile initiates a call or responds to a page.
Paging channel (PCH): Used by the BTS to page MS (paging can be performed by an IMSI, TMSI, or IMEI).
Access grant control channel (AGCH): Used by the BTS to assign a DCCH to a MS in response to an access message received on the RACH. The MS will move to the dedicated channel in order to proceed with either a call setup, response to a paging message, location area update, or short message service.
Cell broadcast channel (CBCH): This channel is used to transmit messages to be broadcast to all MSs within a cell. The CBCH uses a DCCH to send its messages. However, it is considered a common channel because the messages can be received by all mobiles in the cell.

Active MSs must frequently monitor both BCCH and CCCH. The CCCH will be transmitted on the RF carrier with the BCCH.

Dedicated control channels

The DCCH is a single timeslot on an RF carrier which is used to convey eight SDCCHs. A SDCCH is used by a single MS for call setup, authentication, location updating, and point-to-point SMS (Figure. 19.28).

As we will see later, SDCCH can also be found on a BCCH/CCCH timeslot. This configuration allows only four SDCCHs.

Associated control channels (ACCH): These channels can be associated with either an SDCCH or a TCH. They are used for carrying information associated with the process being carried out on either the SDCCH or the TCH.

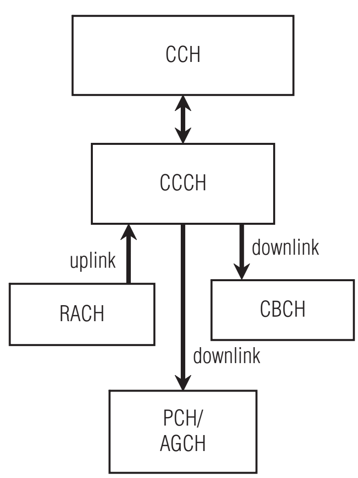


Figure 19.27 Subtype of CCCH

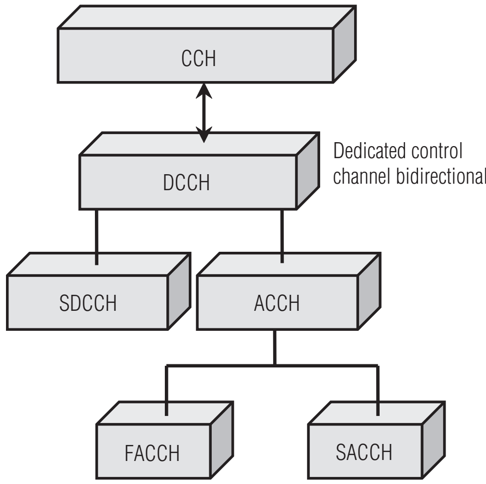


Figure 19.28 Subtype of DCCH

(i) SACCH: Conveys power control and timing information in the downlink direction (towards the MS) and receive signal strength indicator (RSSI) and link quality reports in the uplink direction.

(*FACCH:* The FACCH is transmitted instead of a TCH. The FACCH "steals" the TCH burst and inserts its own information. The FACCH is used to carry out user authentication, handovers, and immediate assignment.

All control channels are required for system operation. However, in the same way that we allow different users to share the radio channel by using different timeslots to carry the conversation data, the control channels share timeslots on the radio channel at different times. This allows efficient passing of control information without wasting capacity which could be used for call traffic. To do this, we must organize the timeslots between those which will be used for traffic and those which will carry control signalling.

19.8.4 Effective usage of channels

In this section, the effective usage of channels is discussed in detail.

Channel combinations

The different logical channel types mentioned are grouped into what are called channel combinations. The four most common channel combinations are listed below:

- Full-rate traffic channel combination: TCH8/FACCH + SACCH
- Broadcast channel combination: BCCH + CCCH
- Dedicated channel combination: SDCCH8 + SACCH8
- Combined channel combination: BCCH + CCCH + SDCCH4 + SACCH4

The half-rate channel combination (when introduced) will be very similar to the FR traffic combination.

- Half-rate traffic channel combination: TCH16/FACCH + SACCH

Table 19.2 Channel combination

Channel combination	Timeslots
Traffic	Any timeslot
Broadcast	0, 2, 4, 6 (0 must be used first)*
Dedicated	Any timeslot
Combined	0 only

*If broadcast is assigned to timeslots 2, 4, or 6 then FCCH and SCH will be replaced with dummy bursts since these control channels may only occur on timeslot 0.

Note: Only one BCCH/CCCH timeslot is required per cell (not RF carrier).

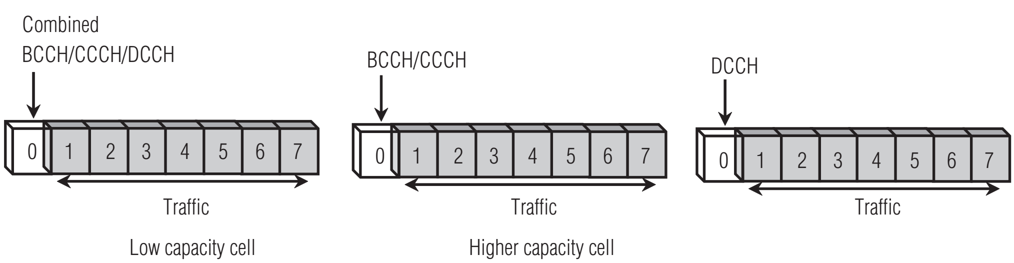


Figure 19.29 Channel combinations and timeslots

Timeslots

The channel combinations we have identified are sent over the air interface in a selected timeslot. Some channel combinations may be sent on any timeslot, but others must be sent on specific timeslots. Table 19.2 shows mapping the channel combinations to their respective timeslots (also shown in Figure 19.29).

Figure 19.29 illustrates how these different channel combinations may be mapped onto the TDMA frame structure.

19.9 Frame structure for GSM

Transmission in any TDMA-based wireless communication system is in the form of a repetitive sequence of frames. Each TDMA frame is divided into a number of uniform time slots. Each time slot position across the sequence of frames forms a separate logical channel. It is very critical to determine the length and composition of the logical channel time slot that will provide effective speech and data transmission with efficient use of the available frequency spectrum.

Each user transmits a burst of data during the time slot assigned to it. These data bursts may have specific format. Figure 19.30 illustrate the data burst used for TCH and DCCH transmissions on both the forward and the reverse link.

3 start bits	58 bits of encrypted data	26 training bits	58 bits of encrypted data	3 stop bits	8.25 bits guard period

Figure 19.30 Time slot data bursts in GSM

Figure 19.31 illustrates the frame structure of GSM. It consists of 148 bits which are transmitted at a rate of 270.833333 kbps (an unused guard time of 8.25 bits is provided at the end of each burst). Out of the total 148 bits per time slot, 114 are information-bearing bits which are transmitted as two 57-bit sequences close to the beginning and end of the burst. The mid-amble

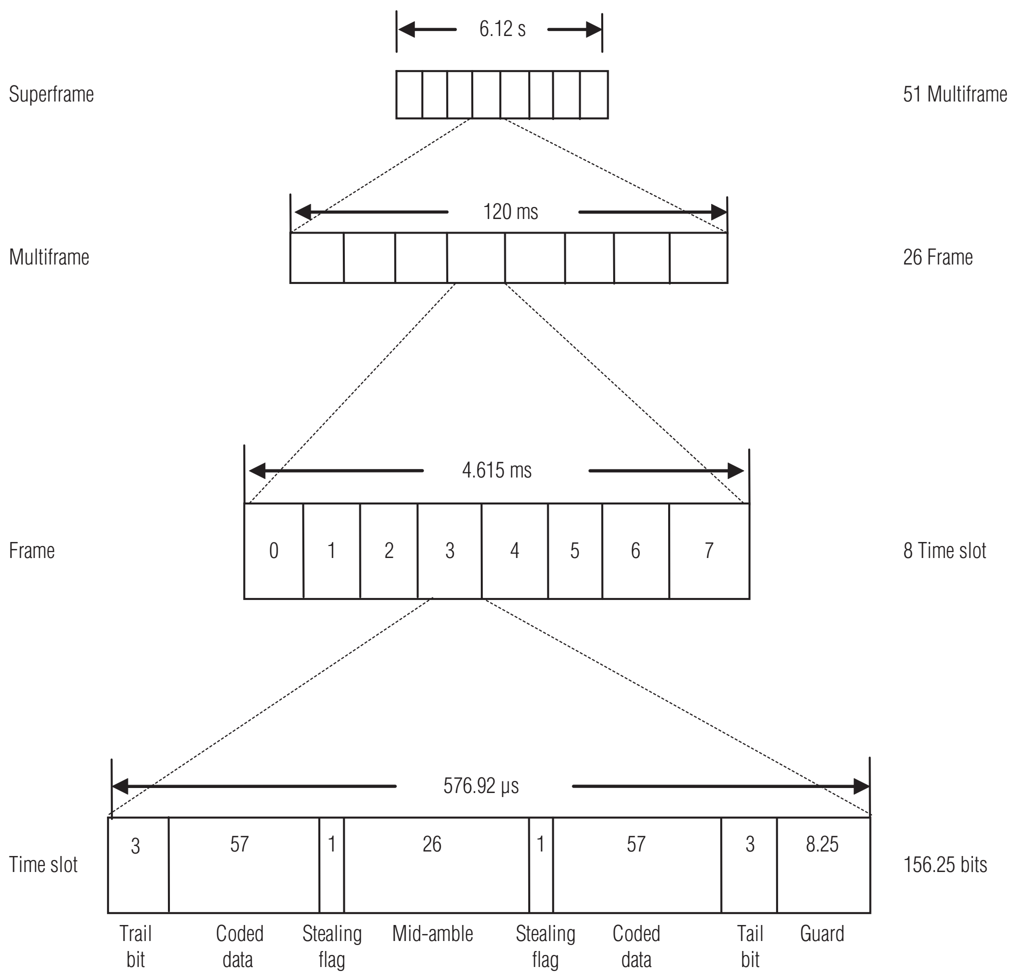


Figure 19.31 GSM frame structure

consists of a 28-bit training sequence which allows the adaptive equalizer in the mobile or base station receiver to analyse the radio channel characteristics before decoding the user data. On either side of the mid-amble there are control bits called stealing flags. These two flags are used to distinguish whether the time slot contains voice (TCH) or control (FACCH) data, both which share the same physical channel. During a frame, a GSM subscriber unit uses one time slot to transmit, one time slot to receive and may use the six spare time slots to measure signal strength on five adjacent base stations as well as its own base station.

As shown in Figure 19.31, there are eight timeslots per TDMA frame and the frame period is 4.615 ms. A frame contains $8 \times 156.25 = 1{,}250$ bits, although some bit periods are not used. The frame rate is 270.833 kbps/1,250 bits/frame, or 216.66 frames/s. The 13th or 26th frame is not used for traffic, but for control purposes. Each of the normal speech frames are grouped into larger structures called multiframes which in turn are grouped into superframes and hyperframes (hyperframes are not shown in Figure 19.31). One multiframe contains 26 TDMA frames, and one superframe contains 51 multiframes, or 1,326 TDMA frames. A hyperframe contains 2,048 superframes, or 2,715,648 TDMA frames. A complete hyperframe is sent about every 3 h, 28 min, and 54 s, and is important to GSM since the encryption algorithms rely on the particular frame number, and sufficient security can only be obtained by using a large number of frames as provided by the hyperframe.

Example problem 19.3

GSM uses a frame structure where each frame consists of eight time slots, and each time slot contains 156.25 bits and data is transmitted over a channel at 270.833 kbps. Find (i) time duration of a bit, (ii) time duration of a time slot, (iii) time duration of a TDMA frame, and (iv) how long must a user wait when occupying a single time slot between two successive transmissions.

Solution

(i) To find time duration of a bit, T_b
Channel data rate = 270.833 kbps (given)
Time duration of a bit, T_b = 1/data rate
Hence, time duration of a bit, $T_b = 1/270.833 \text{ kbps} = 3.69\ \mu\text{s}$

(ii) To find time duration of a time slot, T_{slot}
Number of bits per time slot = 156.25 bits (given)
Time duration of a time slot, $T_{\text{slot}} = 156.25 \text{ bits} \times T_b$
Time duration of a time slot, $T_{\text{slot}} = 156.25 \text{ bits} \times 3.69\ \mu\text{s} = 577\ \mu\text{s}$

(iii) To find time duration of a TDMA frame, T_f
Number of time slots per TDMA frame = 8 (given)
Time duration of a frame, T_f = number of time slots $\times T_{\text{slot}}$
Time duration of a frame, $T_f = 8 \times 577\ \mu\text{s} = 4.616 \text{ ms}$

(iv) To find time duration for a user occupying a single time slot between two successive transmissions has to wait for the time duration of a frame. Hence, a user has to wait for 4.616 ms between two successive transmissions.

19.10 GSM protocol stack configuration

In any telecommunication system, signalling is required to coordinate the necessarily distributed functional entities of the network. The transfer of Signalling information in GSM follows the layered OSI model.

The GSM protocol architecture used for the exchange of signalling messages pertaining to mobility, radio resource, and connection management functions is shown in Figure 19.32. The protocol layering consists of the physical layer, the data link layer (DLL), and the Layer 3.

It is noted to the OSI-minded reader to be careful in not confusing the Layer 3 protocol functions defined by GSM with what is normally defined to be the Layer 3 functions in the OSI model. The GSM Layer 3 protocols are used for the communication of network resource, mobility, code format, and call-related management messages between the various network entities involved. Since, in the OSI model, some of these functions are actually provided by the higher layers, the term "message layer" may be a more appropriate term for referring to the Layer 3 in GSM.

General description of each layer: Table 19.3 indicates the various entities and their functionalities of each sub-layer presented in the GSM Protocol Stack.

Overview of Interfaces in GSM Protocol Stack and Description of the Layers

The following specification indicates the interfaces:

- *Um interface*: The "air" or radio interface standard that is used for exchanges between a mobile (ME) and a base station (BTS/BSC). For signalling, a modified version of the ISDN LAPD, known as link access procedure on the Dm channel (LAPDm) is used.
- *Abis interface*: This is a BSS internal interface linking the BSC and a BTS, and it has not been totally standardized. The Abis interface allows control of the radio equipment and RF allocation in the BTS.

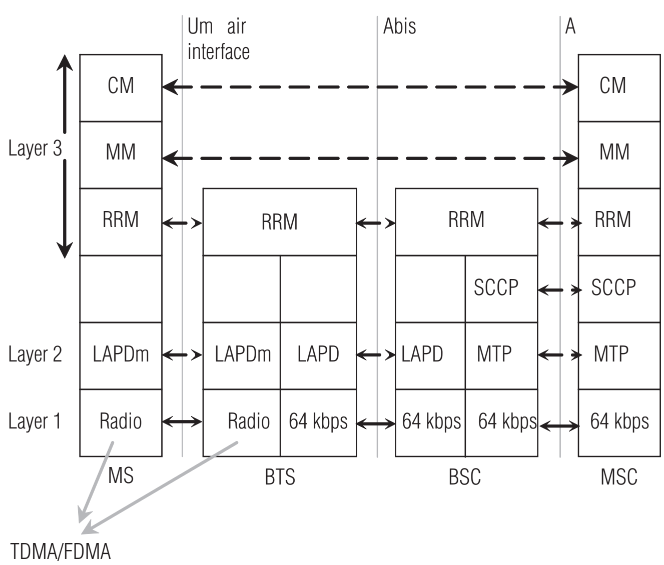


Figure 19.32 GSM protocol architecture

Table 19.3 Functionalities of each sublayer presented in the GSM protocol stack

Layer name	General description
PHY	It specifies physical layer (= wireless signal format) between UE and BS for GSM/ GPRS. GSM ciphering is performed by PHY
Data Link	It establishes link to exchange GSM Layer-3 message between UE and BTS
RLC/MAC	It abbreviates as radio link control/medium access control, which establishes for GPRS data communication between UE and BS. Also manages multiple UEs to share wireless resource (MAC)
LLC	It is logical link control which establishes link between UE and core network and GPRS ciphering is performed
SNDCP	It is sub-network dependent convergence protocol which converts high-layer data (IP data) to GPRS data. It also performs reverse conversion
RR	It is radio resource management which manages GSM wireless resource Typical message: ASSIGNMENT COMMAND
CC	It is call control to manage GSM call procedure Typical message: CALL SETUP
MM	It is mobility management which manages the GSM UE information location Typical procedure: GSM location updating
GRR	It is GPRS radio resource management which manages wireless resources for GPRS data communication Typical procedure: Downlink TBF establishment
GMM	It is GPRS mobility management to manage UE location information for GPRS data communications Typical procedure: GPRS Routing area update
SM	It is session management which manages session for GPRS data communications. Typical procedure: PDP context activation

- *A interface*: The A interface is used to provide communication between the BSS and the MSC. The interface carries information to enable the channels, timeslots, and the like to be allocated to the MEs being serviced by the BSSs. The messaging required within the network to enable handover to be undertaken is carried over the interface.
- *B interface*: The B interface exists between the MSC and the VLR. It uses a protocol known as the MAP/B protocol. As most VLRs are collocated with an MSC, this makes the interface purely an “internal” interface. The interface is used whenever the MSC needs access to data regarding a MS located in its area.
- *C interface*: The C interface is located between the HLR and a GMSC or a SMS-G. When a call originates from outside the network, that is from the PSTN or another mobile network it has to pass through the gateway so that routing information required to complete the call may

be gained. The protocol used for communication is MAP/C, the letter "C" indicating that the protocol is used for the "C" interface. In addition to this, the MSC may optionally forward billing information to the HLR after the call is completed and cleared down.

- *D interface*: The D interface is situated between the VLR and HLR. It uses the MAP/D protocol to exchange the data related to the location of the ME and to the management of the subscriber.
- *E interface*: The E interface provides communication between two MSCs. The E interface exchanges data related to handover between the anchor and relay MSCs using the MAP/E protocol.
- *F interface*: The F interface is used between an MSC and EIR. It uses the MAP/F protocol. The communications along this interface are used to confirm the status of the IMEI of the ME gaining access to the network.
- *G interface*: The G interface interconnects two VLRs of different MSCs and uses the MAP/G protocol to transfer subscriber information, for example during a location update procedure.
- *H interface*: The H interface exists between the MSC the SMS-G. It transfers short messages and uses the MAP/H protocol.
- *I interface*: The I interface can be found between the MSC and the ME. Messages exchanged over the I interface are relayed transparently through the BSS.

Although the interfaces for the GSM cellular system may not be as rigorously defined as many might like, they do at least provide a large element of the definition required, enabling the functionality of GSM network entities to be defined sufficiently.

19.10.1 MS protocols

The signalling protocol in GSM is structured into three general layers depending on the interface.
Layer 1: Physical layer: Used for radio transmission.
Layer 2: Data link layer (DLL): Provides error-free transmission between adjacent entities, based on the ISDN's LAPD protocol for the Um and Abis interfaces, and on SS7's Message Transfer Protocol (MTP) for the other layer interfaces.
Layer 3: Networking or Messaging Layer: Responsible for the communication of network resources, mobility, code format, and call-related management messages between various network entities.

Layer 1: Physical layer

The physical layer on the radio link was discussed in the section on radio channel structure. The traffic channels on the landside are formed from TDM slots implemented on 2.048 Mbps links. The signalling channels are basically logically multiplexed on an aggregate of the TDM slots.

The following are the functions of physical layer:

- Modulation techniques – GMSK
- Channel coding: Block code and convolutional code
- Interleaving: To distribute burst error
- Power control methodology – to minimize co-channel interference
- Time synchronization approaches

Layer 2:Data link layer

Provides error-free transmission between adjacent entities and the corresponding functionalities are listed below:

- Connection-based network: for traffic, signalling, and control.
- Signalling and control data are conveyed through Layer 2 and Layer 3 messages in GSM.
- Purpose of Layer 2 is to check the flow of packets for Layer 3.
- DLL checks the address and sequence (for Layer 3).
- Also manages acknowledgements for transmission of the packets.
- Allows two SAPs for signalling and SMS.
- SMS traffic is carried through a fake signalling packet that carries user information over signalling channels.
- DLL allows SMS data to be multiplexed into signalling streams.
- Signalling packet delivered to the physical layer is 184 bits which conforms with the length of the DLL packets in the LAPD protocol used in ISDN network.
- The LAPD protocol is used for A and Abis interface.
- The DLL for the Um interface is LAPDm

LAPDm:
LAPDm is the protocol for use by the DLL on the radio interface.

Functions of LAPDm:

- Organization of Layer 3 information into frames
- Peer-to-peer transmission of signalling data in defined frame formats
- Recognition of frame formats
- Establishment, maintenance, and termination of one or more (parallel) data links on signalling channels

Layer 3: Networking or message layer
The Layer 3 protocols are used for the communication of network resources, mobility, code format, and call-related management messages between various network entities.

This third layer of the GSM signalling protocol is divided into three sub-layers:

- Radio resource management (RR)
- Mobility management (MM)
- Connection management (CM)

A number of mechanisms are needed to establish, maintain, and terminate a mobile communication session. Layer 3 implements the protocols needed to support these mechanisms (Figure 19.33).

A signalling protocol is composed of a sequence of communication events or messages. Layer 3 defines the details of implementation of messages.

Transaction identifier (TI): Allows multiple protocols to operate in parallel to identify a protocol that consists of a sequence of messages.

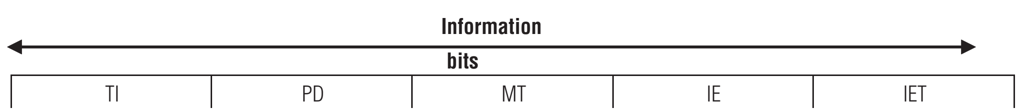


Figure 19.33 Layer 3 message format

Protocol discriminator (PD): Identifies the category of the operation (such as management, supplementary services, and call control)

Message type (MT): Identifies the type of messages for a given PD

Information elements (IE): Identifies the information elements.

Radio resource (RR) management sublayer

The RR management sublayer terminates at the BSS and performs the functions of establishing physical connections over the radio for the purpose of transmitting call-related signalling information such as the establishment of signalling and traffic channels between a specific mobile user and the BSS. The RR management functions are basically implemented in the BSS.

Mobility management sublayer (MM)

The MM sublayer is terminated at the MSC and the related messages from or to the MS is relayed transparently in the BSS using the DTAP process. The MM sublayer provides functions that can be classified into three types of procedures. These are called the MM specific procedures, the MM common procedures, and the MM connection-related procedures. These procedures are discussed in the following.

The protocols in the MM layer involve the SIM, MSC, VLR, and the HLR, as well as the authentic centre (AuC) (which is closely tied with the HLR).

Connection management sublayer (CM)

The CM functional layer is divided into three sublayers:

- Call control (CC)
- Supplementary services
- Short message service

Call control (CC) sublayer: Manages call routing, establishment, maintenance, and release, and is closely related to ISDN call control.

Supplementary services sublayer: Manages the implementation of the various supplementary services (call forwarding/waiting/hold) and also allows users to make a call.

Short message service sublayer: Handles the routing and delivery of short messages, both from and to the mobile subscriber.

19.10.2 MS to BTS protocols

The RR layer oversees the establishment of a link, both radio and fixed, between the MS and the MSC. The main functional components involved are the MS, the BSS, and the MSC. The RR layer is concerned with the management of an RR-session, which is the time that a mobile is in dedicated mode, as well as the configuration of radio channels, including the allocation of dedicated channels.

The MM layer is built on top of the RR layer and handles the functions that arise from the mobility of the subscriber, as well as the authentication and security aspects. Location management is concerned with the procedures that enable the system to know the current location of a powered-on MS so that incoming call routing can be completed.

The CM layer is responsible for CC, supplementary service management, and short message service (SMS) management. Each of these may be considered as a separate sublayer within the CM layer. Other functions of the CC sublayer include call establishment, selection of the type of service (including alternating between services during a call), and call release.

19.10.3 BSC protocols

After the information is passed from the BTS to the BSC, a different set of interfaces is used. The Abis interface is used between the BTS and BSC. At this level, the radio resources at the lower portion of Layer 3 are changed from the RR to the base transceiver station management (BTSM). The BTS management layer is a relay function at the BTS to the BSC.

The RR protocols are responsible for the allocation and reallocation of traffic channels between the MS and the BTS. These services include controlling the initial access to the system, paging for MT calls, the handover of calls between cell sites, power control, and call termination. The RR protocols provide the procedures for the use, allocation, reallocation, and release of the GSM channels. The BSC still has some radio resource management in place for the frequency coordination, frequency allocation, and the management of the overall network layer for the Layer 2 interfaces.

From the BSC, the relay is using SS7 protocols so the MTP 1–3 is used as the underlying architecture, and the BSS mobile application part or the direct application part is used to communicate from the BSC to the MSC.

19.10.4 MSC protocols

In MSC, the information is mapped across the A interface to the MTP Layers 1 through 3 from the BSC. Here, the equivalent set of radio resources is called the BSS MAP. The BSS MAP/DTAP and the MM and CM are at the upper layers of Layer 3 protocols. This completes the relay process. Through the control-signalling network, the MSCs interact to locate and connect to users throughout the network. Location registers are included in the MSC databases to assist in the role of determining how and whether connections are to be made to roaming users.

Each user of a GSM MS is assigned a HLR that is used to contain the user's location and subscribed services. A separate register, the VLR, is used to track the location of a user. As the users roam out of the area covered by the HLR, the MS notifies a new VLR of its whereabouts. The VLR in turn uses the control network (which happens to be based on SS7) to signal the HLR of the MS's new location. Through this information, MT calls can be routed to the user by the location information contained in the user's HLR.

19.11 GSM basic call flow

Figure 19.34 shows the basic components and processes involved in setting up a call between a GSM MS and an ordinary "land" telephone.

19.11.1 Call flow from the MS to land (PSTN)

The BTS receives a data message from the MS which passes it to the BSC. The BSC relays the message to the MSC via C7 signalling links, and the MSC then sets up the call to the land subscriber via the PSTN. The MSC connects the PSTN to the GSM network, and allocates a terrestrial circuit to the BSS serving the MS's location. The BSC of that BSS sets up the air interface channel to the MS and then connects that channel to the allocated terrestrial circuit, completing the connection between the two subscribers.

When a mobile subscriber makes a call to a PSTN telephone subscriber, the following sequence of events takes place:

- The MSC/VLR receives the message of a call request.
- The MSC/VLR checks if the MS is authorized to access the network. If so, the MS is activated. If the MS is not authorized, service will be denied.

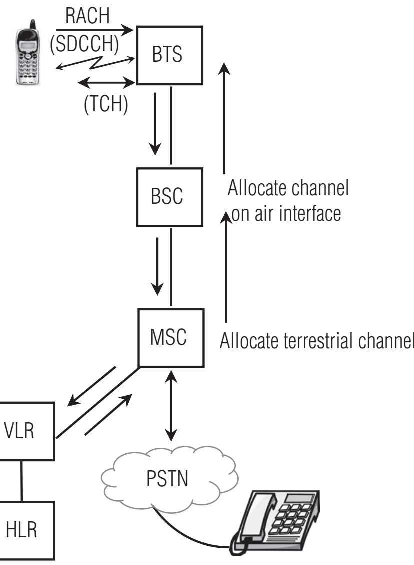


Figure 19.34 Call network

- MSC/VLR analyses the number and initiates a call setup with the PSTN.
- MSC/VLR asks the corresponding BSC to allocate a traffic channel (a radio channel and a time slot).
- The BSC allocates the traffic channel and passes the information to the MS.
- The called party answers the call and the conversation takes place.
- The MS keeps on taking measurements of the radio channels in the present cell and neighbouring cells and passes the information to the BSC. The BSC decides if handover is required, if so, a new traffic channel is allocated to the MS and the handover is performed. If handover is not required, the MS continues to transmit in the same frequency.

19.11.2 Call flow from the land to MS

The MSC receives its initial data message from the PSTN (via C7) and then establishes the location of the MS by referencing the HLR. It then knows which other MSC to contact to establish the call and that MSC then sets up the call via the BSS serving the MS's location.

When a PSTN subscriber calls a MS, the sequence of events is as follows:

- The gateway MSC receives the call and queries the HLR for the information needed to route the call to the serving MSC/VLR.
- The GMSC routes the call to the MSC/VLR.
- The MSC checks the VLR for the location area of the MS.
- The MSC contacts the MS via the BSC through a broadcast message, that is through a paging request.
- The MS responds to the page request.
- The BSC allocates a traffic channel and sends a message to the MS to tune to the channel. The MS generates a ringing signal and, after the subscriber answers, the speech connection is established.
- Handover, if required, takes place, as discussed in the earlier case.

The actual processes are, of course, considerably more complex than described above. In addition, at this moment we consider in detail just the MS-to-land and land-to-MS call sequences and the intra-MSC (inter-BSS) handover sequence. This will give you a good appreciation of the messaging that occurs in the GSM system, and how the PLMN interacts with the PSTN.

The MS codes the speech at 13 kbps for transmission over the radio channel in the given time slot. The BSC converts (or transcodes) the speech to 64 kbps and sends it over a land link or radio link to the MSC. The MSC then forwards the speech data to the PSTN. In the reverse direction, the speech is received at 64 kbps rate at the BSC and the BSC does the transcoding to 13 kbps for radio transmission.

In its original form, GSM supports 9.6 kbps data, which can be transmitted in one TDMA time slot. Over the last few years, many enhancements were done to the GSM standards (GSM Phase 2 and GSM Phase 2+) to provide higher data rates for data applications.

19.12 Summary

- GSM is a standard that ensures interoperability without stifling competition and innovation among suppliers to the benefit of the public in terms of both cost and service quality. For example, by using very large scale integration (VLSI) microprocessor technology, many functions of the MS can be built on one chipset, resulting in lighter, more compact, and more energy-efficient terminals.
- Telecommunications are evolving towards personal communication networks, whose objective can be stated as the availability of all communication services anytime, anywhere, to anyone, by a single identity number and a pocketable communication terminal.
- The GSM system and its sibling systems operating at 1.8 GHz (called DCS1800) and 1.9 GHz (called GSM1900 or PCS1900, and operating in North America) are a first approach at a true personal communication system.
- The SIM card is a novel approach that implements personal mobility in addition to terminal mobility.
- Together with international roaming, and support for a variety of services such as telephony, data transfer, fax, short message service, and supplementary services, GSM comes close to fulfilling the requirements for a personal communication system: close enough that it is being used as a basis for the next generation of mobile communication technology in Europe, the universal mobile telecommunication system (UMTS).
- Another point where GSM has shown its commitment to openness, standards, and interoperability is the compatibility with the integrated services digital network (ISDN) that is evolving in most industrialized countries and Europe in particular (the so-called Euro-ISDN).
- GSM is also the first system to make extensive use of the intelligent networking concept, in which services like 800 numbers are concentrated and handled from a few centralized service centres, instead of being distributed over every switch in the country. This is the concept behind the use of the various registers such as the HLR.
- In addition, the signalling between these functional entities uses signalling system number 7, an international standard already deployed in many countries and specified as the backbone signalling network for ISDN.
- GSM is a very complex standard, but that is probably the price that must be paid to achieve the level of integrated service and quality offered while subject to the rather severe restrictions imposed by the radio environment.

Review questions

1. What are the advantages of digital cellular systems over analogue?
2. What is GSM?
3. Explain the architecture of GSM.
4. What are the operations of GSM?
5. Discuss address identifiers of global system for mobile.
6. Describe various protocols of GSM.
7. What are the various types of logical channels? Explain in detail.
8. What are the channel modes of GSM?
9. What are the differences between GSM and CDMA mobile phone?
10. What is the main reason of call drop in GSM system?
11. Explain the concept of GSM superframe, multiframe, TDMA frame, and time slot in a GSM channel. Give suitable illustration for GSM frame hierarchy.
12. Explain in detail the GSM architecture. (Refer Section 19.3)
13. Explain about the traffic channels in GSM. (Refer Section 19.8)
14. What are the services offered by GSM channels? (Refer Section 19.2.7)
15. What are the different types of logic channels? How these differ from the physical channels? (Refer Section 19.8.2)
16. Explain how network switching subsystem (NSS) manages the communication between GSM users and other telecommunication users. (Refer Section 19.3.3)
17. What is base station subsystem (BSS)? Briefly explain its working. (Refer Section 19.3.2)
18. Explain the various logical GSM channels. (Refer Section 19.8.2.2)
19. Why HLR and VLR are required in network and switching subsystem? Differentiate them. (Refer Sections 19.3.3.2 and 19.3.3.3)
20. What are the major problems in AMPS system? How these can be overcome in GSM system? (Refer Sections 19.1 and 19.2.7)

Objective type questions and answers

1. ________ is a globally accepted standard for digital cellular communication. It is also the name of a standardization group founded in 1982 to create a common mobile telephone standard in Europe.
   (a) GSM (b) CDMA (c) both a & b (d) FDMA
2. Type of modulation that GSM uses
   (a) GMSK (b) BPSK (c) QPSK (d) b and c
3. The following technology does not require the SIM card
   (a) CDMA (b) GSM (c) both (d) FDMA
4. GSM is a ________ digital system that converts voice and access information to digital data, and communicates those data in bursts during brief time slots allocated to multiple subscribers sharing a radio channel.
   (a) CDMA (b) TDMA (c) FDMA (d) both c and d
5. 3G technology is a combination of
   (a) SDMA and TDMA (b) GSM and CDMA (WCDMA)
   (c) FDMA and GSM (d) None of the above

6. ________ technology offers general packet radio service
   (a) CDMA (b) GSM (c) both a and b (d) FDMA
7. The features and benefits expected in the GSM system are
   (a) superior speech quality (b) low terminal, operational, and service cost.
   (c) international roaming (d) all of the above
8. In GSM 900 with FDM: 124 frequencies (up/down) and TDM: 8 slots/frequency Max number of active users are
   (a) 124/8 (b) 8/124 (c) 124×8 (d) $124 \times 8/132$

Answers: 1. GSM 2. (a), 3. (a), 4. (b), 5. (c), 6. GPRS 7. (d), 8. (c)

Open book questions

1. Explain about GSM specifications?
2. Explain the channels in GSM.
3. Explain GSM services and features.
4. Explain the significance of SIM in mobile station?
5. What is the difference between interface and protocol?
6. Explain the different types of interfaces used to connect the units of base station subsystem in GSM?
7. Write short notes on modes in GSM channels.
8. What are the different types of physical channels? How they differ from logic channels.

Further reading

M. Appleby and F. Harrison, "Cellular Radio Systems," Chapter 47, in F. Mazda, ed., *Telecommunication Engineer's Handbook*. London: Butterworth Heinmann, 1995.

E. Dahlman, B. Gudmundson, M. Nilsson and J. Skold, "UMTS/IMT-2000 based on wideband CDMA," IEEE Communications Magazine, 36(9), 1998.

ETSI (1991c) Subscriber Identity Modules, Functional characteristics, European Telecommunication Standards Institute, Recommendation GSM 02.17.

GSM association (2002), http://www.gsmworld.com

Holtzman, "The GSM radio interface", *British Telecom Technological Journal*, Vol. 8, No. 1, pp. 31–43, January 1990.

A. Mehrotra, *Cellular Radio–Analogue and Digital Systems*. Boston: Artech House, 1994.

M. Mouly and M.-B. Pautet, *The GSM System for Mobile Communications*, Published by M. Mouly et Marie-B. Pautet, France,1992.

M. Mouly and M.-B.Pautet, "Current evolution of GSM systems" Personal Communications Magazine, Vol., No. 5, October 1995.

Global Mobile Satellite Systems (GMSS) and Personal Access Communication System (PACS) 20

20.1 Introduction

The existing terrestrial networks are used to provide a microwave link (MWL) between two locations. If the locations are in remote areas, it is not possible to communicate because the terrestrial coverage is not available to all over the world. So, satellites are used for communication. With the integration of satellite communication, the mobile phone can switch to satellites and provide worldwide connectivity to a customer. By using satellites operating in geostationary earth orbits (GEO), it is possible to provide communication between two remote locations on the earth. If the location does not have any terrestrial wire line or wireless coverage, a large costly terminal is required to make a phone call. This results in high call charges. Also, the round trip signal delay is very long since the GEO is about 36,000 km above the earth's surface. This results in degradation in service quality. To provide superior service quality, constellations of satellites operating in low earth orbits (LEO) or medium earth orbits (MEO) are needed. The basic purpose of satellites for mobile communication is to extend the area of coverage without replacing the existing mobile phone networks. With global mobile satellite systems, calls between any two locations can be made easier, affordable, and user-friendly.

The features of global mobile satellite systems include the following:

- It is possible to use a single mobile wireless phone anywhere in the world.
- Communication is possible where mobile and even basic telecommunication services are unavailable.

A number of mobile satellite services (MSS) blanket the globe with satellite telephone coverage from a constellation of over 1,000 satellites. These systems include Iridium, Globalstar, ICO, Teledesic, Odyssey, Thuyra, ACeS (ASIA Cellular Satellite), and Agrani. The primary applications for the first three systems are voice, fax, and messaging services for mobile communication subscribers. The Teledesic system's application is high-speed data and multimedia services using satellites.

Almost all these satellite services offer a combination of all-digital transparent voice, data, fax, and paging services to and from hand-held telephone devices. The systems will share an air interface standard called geostationary mobile satellite standard (GMSS) that is similar to global system for mobile communication (GSM). This means that satphone (satellite phone) customers will be able to use mobile phones that are compatible with satellite systems in any country where GMSS is offered.

This chapter introduces the various orbits of satellite system like LEO, MEO, HEO, and GEO and deals with global mobile satellite systems such as Iridium, Globalstar, ICO, and Teledesic. It also explains how these systems are used for communication.

20.2 Types of orbits for mobile and other satellite systems

The following are the four general system designs , which are differentiated by the type of orbit in which the satellites operate:

- Low Earth orbit (LEO)
- Medium Earth orbit (MEO)
- Highly elliptical orbit (HEO)
- Geostationary Earth orbit (GEO)

Each of these has various strengths and weaknesses in their ability to provide particular communications services. Figure 20.1 shows the orbital altitudes for satellite constellations.

20.2.1 Low earth orbit

The satellites in LEO have polar orbits. An LEO satellite system usually has a cellular type of access, similar to the cellular telephone system. Because LEO satellites are close to Earth, the round-trip time propagation delay is normally less than 20 ms, which is acceptable for audio communication.

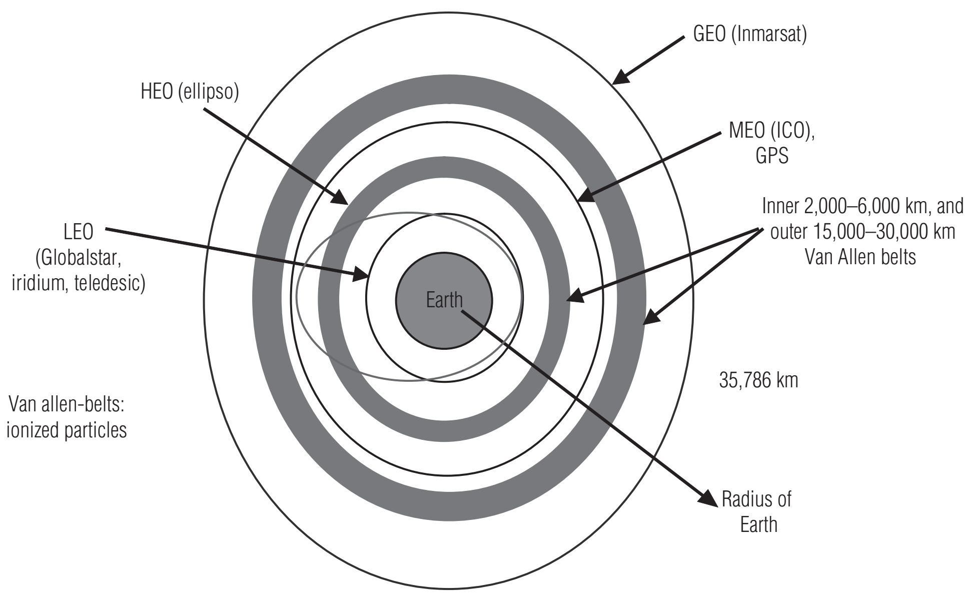


Figure 20.1 Orbital altitudes for satellite constellations

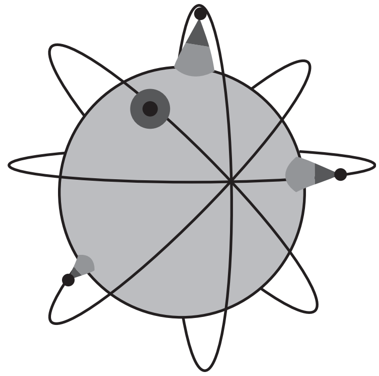


Figure 20.2 Footprints of LEO satellite

Each satellite in LEO system acts as a switch. Satellites that are close to each other are connected through intersatellite links (ISLs). A mobile system communicates with the satellite through a user mobile link (UML). A satellite can also communicate with an earth station (gateway) through a gateway link (GWL).

Footprint is the area where signals from the satellite can be received on earth. Figure 20.2 shows the footprint of LEO satellite.

Types of LEO systems

There are three types of LEO systems: little LEOs, big LEOs, and broadband LEOs:

- Little LEOs – Little LEO satellites are very small and use very little bandwidth for communications. These are used for low data rate messaging. They operate under 1 GHz. These systems often employ mechanisms to maximize capacity, such as frequency reuse schemes. Examples of little LEO systems include Orbcomm.
- Big LEOs – These operate from 1 to 3 GHz. Globalstar and Iridium systems are examples of big LEOs.
- Broadband LEOs – Broadband LEOs provide communication similar to fibre optic networks. Teledesic and Skybridge systems are examples of broadband LEO systems, optimized for packet-switched data rather than voice.

Important features of LEO satellite

- They have polar orbits.
- Orbital period is **90–120 min** (less than 2 h).
- Altitude is between **50 and 2,000 km**.
- Each LEO satellite will be visible from the earth for around 10 min.

Advantages of LEO systems

- Power required is very less (1 W).
- The delay for packets delivered via LEO is relatively low (~10 ms).
- Due to smaller footprint, they allow frequency reuse.
- Better global coverage due to higher elevation in polar regions.

Disadvantages of LEO systems

- More number of satellites are required for global coverage.
- Short lifetime of 5–8 years
- Due to smaller footprint, other factors are needed for routing of data packets from satellite to satellite. GEO system does not need this type of routing, as senders and receivers are mostly in the same footprint.

20.2.2 Medium earth orbit

The MEO is a compromise between the LEO and GEO. Compared to LEOs, MEOs require fewer satellites to provide global coverage. Compared to GEOs, MEOs can operate effectively with smaller mobile equipment and with less latency. One example of a MEO satellite system is the global positioning system (GPS). The footprints of MEO satellite are shown in the Figure 20.3. MEO constellations have 10–17 satellites distributed over two or three orbital planes. Most MEO systems offer phone services similar to the big LEOs. Before the MEO designation came into wide use, MEO systems were considered big LEOs. An example of MEO system includes ICO system.

Important features of MEO satellite

- Orbital period is **6–12 h.**
- Altitude is between **5,000 and 25,000 km**.
- Each MEO satellite will be visible from the earth for several hours.

Advantage of MEO system

- MEO systems require fewer satellites than LEOs. Therefore, it reduces the overall system complexity and cost. Depending on the inclination, a MEO can cover larger populations and hence it requires only a fewer handovers.

Disadvantages of MEO system

- Like LEOs, MEO satellites have much shorter life expectancy than GEOs. So, more frequent launches are required to maintain the system over time.
- The MEO satellites need higher transmitter power and larger antennas.

20.2.3 Highly elliptical orbit

HEO systems operate differently than LEOs, MEOs, and GEOs. As the name implies, the satellites orbit the earth in an elliptical path rather than circular paths as in LEOs and GEOs. The HEO path typically is not centred on the Earth, as LEOs, MEOs and GEOs are. This orbit causes the satellite to move around the earth faster when it is travelling close to the earth, and slower the farther away it gets. In addition, the satellite's beam covers most of the earth from farther away as shown in the Figure 20.4. Examples of HEO systems include Ellipso and the proposed Pentriad.

Important features of HEO satellite

- They have elliptical orbits.
- HEO systems do not offer continuous coverage over outlying geographic regions, especially near the South Pole.
- The orbits are designed to maximize the amount of time each satellite spends in view of populated areas.

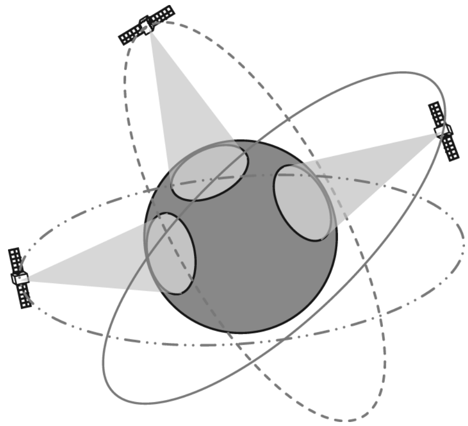


Figure 20.3 Footprints of MEO satellite

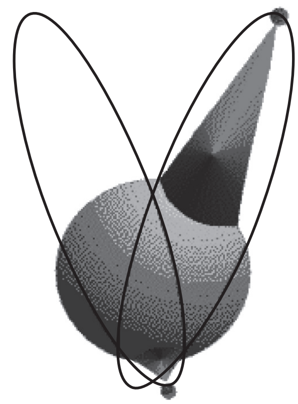


Figure 20.4 Illustration of HEO system

Advantage of HEO system

- Requires fewer satellites than LEOs and provides superior line-of-sight in comparison to most LEOs and GEOs.

Disadvantage of HEO system

- Coverage of a typical HEO system is not as complete as other orbital designs, although they provide good coverage over most population centres.

20.2.4 Geostationary earth orbit

Geostationary satellites have circular orbits that are orientated in the plane of the earth's equator. In a GEO, the satellite appears stationary, that is, in a fixed position, to an observer on the earth. More technically, a GEO is a circular prograde orbit in the equatorial plane with an orbital period equal to that of the earth. A satellite in a GEO will appear fixed above the surface of the earth, that is, at a fixed latitude and longitude. The footprint or service area of a geostationary satellite covers almost one-third of the earth's surface. So, global coverage can be achieved with a minimum of three satellites in orbit. By placing the satellite at an altitude where its orbital period exactly matches the rotation of the earth (approximately 35,800 km), the satellite thus appears to "hover" over one spot on the earth's equator and thus appears to stay stationary over the same point.

Important features of geostationary satellite

- They have circular orbits.
- Orbital period is 24 h (same as earth rotation period).
- Orbital height is 35,786 km (22,300 mi).
- GEOs support voice, data, and video services, most often providing fixed services to a particular region.

- GEO systems are less complicated to maintain because their fixed location in the sky requires relatively little tracking capability in the ground equipment.
- GEOs remain in orbit longer than the systems operating closer to earth.

Examples of GEO systems include INTELSAT, Inmarsat, and PanAmSat. The footprint of a geostationary satellite is shown in Figure 20.5. The footprint of a geostationary satellite is one-third of the earth. This means that if you are orbiting at 35,786 km (the altitude for geostationary satellites) from earth, then you can only see one-third of the earth's surface. LEOs (low orbit) have a smaller footprint because they are closer to the earth.

Advantages of GEO system

- Three GEO satellites are enough for a complete coverage of any spot on the earth.
- Senders and receivers can use fixed antenna positions. No adjustment is needed.
- Lifetime of GEO satellites is about 25 years.
- They have a large footprint.
- GEOs do not exhibit any Doppler shift because the relative movement is zero.
- GEO systems have significantly greater available bandwidth than the LEO and MEO systems. This permits them to provide two-way data, voice, and broadband services that may be unpractical for other types of systems.

Disadvantages of GEO system

- The transmit power needed is relatively high (10 W).
- These satellites cannot be used for small mobile phones.
- Due to large footprint, frequencies cannot be reused. GEO satellites need special antennas to focus on a smaller footprint.
- Transferring a GEO into orbit is very expensive.
- GEO systems require line-of-sight communication paths between terrestrial antennae and the satellites. But, due to fewer (fixed) satellites, the opportunities for line-of-sight communication are fewer than for systems in which the satellites "travel" across the sky. This is a significant disadvantage of GEO systems as compared to LEO and MEO systems, especially for mobile applications and in urban areas where tall buildings and other structures may block the line-of-sight communication for hand-held mobile terminals.

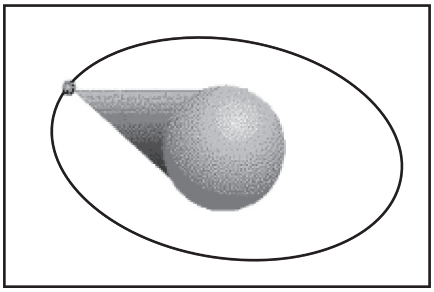


Figure 20.5 Footprint of geostationary satellite

Table 20.1 Comparison of the properties of LEO, MEO, HEO and GEO

Characteristic	LEO	MEO	HEO	GEO
Launch & Satellite cost	Maximum	Maximum	Medium	Medium
Satellite Life (years)	4–8	10–15	2–4	20–30
Orbital period	< 100 minutes	8–12 hours	½ Sidereal day	1 Sidereal day (23.9344696 hours)
Inclination	90°	45°	63.4°	Zero
Coverage	Global	Global	Near Global	Near Global
Satellite Visibility	Short	Medium	Medium	Continuous
Round-trip delay	Small	Small	Medium	High
Handover	Very Much	Medium	No	No
Network complexity	Complex	Medium	Simple	Simple
Propagation delay	Low	Medium	High	High

20.2.5 Comparison of the properties of LEO, MEO, HEO and GEO satellites

Table 20.1 shows the comparison of the properties of the satellites in LEO, MEO, HEO, and GEO. From the above table, it is observed that

- The orbital period is very less in LEO satellites, high in GEO satellites, and medium in MEO and HEO systems.
- Satellite visibility is short in lower orbits than in higher orbits.
- Network complexity in LEO systems is more because number of satellites required for global coverage in low orbits is more.
- The round trip delay of GEO satellites is more in comparison to lower orbits.

20.2.6 Integrating GEO, LEO, MEO, and terrestrial mobile systems

Until now, communication satellites have operated using a GEO lying about 36,000 km above the earth's surface. From this orbit, the satellite appears to be stationary (fixed) above a specific location from earth, thereby ensuring continuous, uninterrupted coverage to that location. The primary role of a geostationary communications satellite is to act as a wireless repeater station in space that operates in a broadcast mode and provides a MWL between two remote locations on earth. The key components of a communications satellite include various transponders, transceivers, and antennas that are tuned to the allocated frequency channels. Although the geostationary satellites have a large footprint, so that the entire surface of the earth can be covered by a few such satellites, their high altitude leads to very long round-trip signal delays and resultant degradation in service quality. To support a wide range of services and provide superior service quality comparable to that available from terrestrial wireless and wireline networks, constellations of satellites operating in LEO or MEO are considered more suitable. Table 20.2 provides some key characteristics of LEO, MEO, and GEO satellites.

Table 20.2 Characteristics of different satellite configurations

Characteristic	GEO	MEO	LEO
Altitude range	36,000 Km	10,000–20,000 Km	500–2,000 Km
Satellite visibility	24 hours	2–4 hours	10–20 min
Round trip delay	500 ms	40–80 ms	5–10 ms
Satellite lifetime	20–30 yrs	10–15 years	4–8 years
Satellite constellation	Low	Medium	High

A number of global mobile satellite systems are in various stages of planning and deployment, with the first global mobile satellite service initiated in 1998. Four such systems that are in advanced stages of planning and/or early implementation are Iridium, Globalstar, ICO, and Teledesic. While the primary target applications for the first three systems are voice, fax, and messaging services for mobile communication subscribers, Teledesic system's target application is high-speed data and multimedia services using satellites. Except for the ICO system which deploys a constellation of MEO satellites, these mobile satellite systems use various size constellations of LEO satellites.

To provide global coverage for mobile subscribers, LEO systems need to deploy a large number of satellites; they must either support ISLs (Iridium, Teledesic) or use a large number of ground stations (Globalstar). These factors, combined with the requirement for more frequent replacement of LEO satellites, may lead to overall higher costs for LEO systems than for MEO and GEO satellite systems. LEO systems also face additional technical challenges because of frequent switching of phone calls from one satellite to another (handoff) and potential susceptibility to shadowing (loss of signal due to shadows cast by buildings, etc.) associated with low orbits.

20.3 The Iridium system

The terrestrial cellular coverage is not available all over the world. With the integration of satellite communication, the mobile phone can switch to satellites and provide worldwide connectivity to a customer. If we use the satellites in GEO for communication, due to round trip signal delay, the quality of service (QOS) will be degraded. To provide superior service quality, constellations of satellites operating in LEO or MEO are needed.

The Iridium system is a satellite-based wireless personal communications network in LEO. It is proposed to extend the area of coverage than terrestrial cellular services. It is designed to permit any type of telephone transmission such as voice, paging, facsimile, or data to reach its destination anywhere on earth. It revolutionized communications for business professionals, travellers, residents of rural or undeveloped areas, and others who need the features and convenience of a wireless hand-held telephone with a single worldwide number. Unlike conventional telecommunications networks, the satellite-based system tracks the location of the telephone, providing global transmission even if the subscriber's location is unknown. In areas, where compatible cellular service is available, the dual-mode telephone provides the option of transmitting a call via the local cellular system.

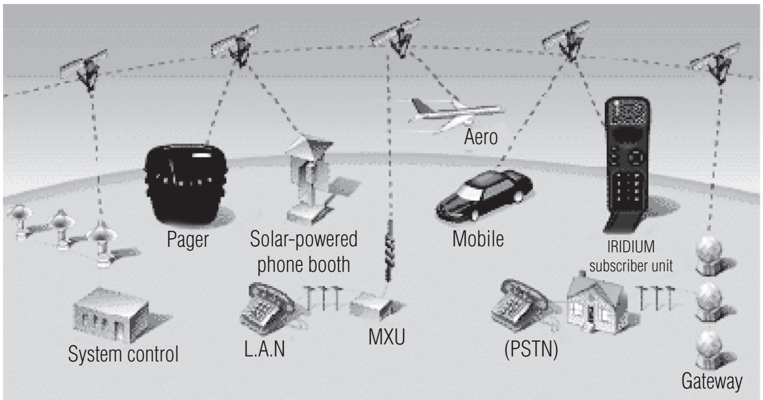


Figure 20.6 The Iridium system (courtesy Iridium)

The Iridium system provides the following:

- More capacity (large number of channels) and better QOS than the satellite system using GEO.
- Emergency service where the terrestrial cellular services are disabled in disaster situations such as earthquakes, fires, and floods.
- High-quality voice connections and interface with laptop computers, personal digital assistants, palmtop organizers, and other communications equipment. Figure 20.6 shows the Iridium system.

History

The concept for the Iridium system was proposed by Motorola engineers Ray Leopold, Ken Peterson, and Bary Bertiger in the year 1987. They envisioned a constellation of low orbiting satellites. After the original service started on 1 November 1998, Iridium LLC (founded in 1991 and having invested about $7 billion) announced the end of commercial service on 17 March 2000. The service had been relaunched on 28 March 2001.

Iridium system was funded by 19 investors, which include Iridium China (Hong Kong) Ltd., Iridium Africa Corporation, Iridium Middle East Corporation, Iridium Canada, Inc, and Motorola. Few more investors were involved in the operation and maintenance of 12 ground station gateways. The 12 gateway operators served as regional distributors of Iridium services in their designated commercial territories. The Iridium satellite is shown in the Figure 20.7. Initially 77 satellites are proposed for constellation. So the system was called Iridium, the element, which has 77 electrons in its orbit. Later it was decided that only 66 satellites are adequate to provide the required services.

20.3.1 Iridium system constellation

The Iridium system constellation is shown in Figure 20.8. The Iridium system constellation consists of 66 active satellites with spare satellites in-orbit to serve in case of failure. These are situated within six near-polar (inclination: 86.4°) orbital planes of 11 satellites each. The six planes have satellites revolving around the earth in the same direction, which means that there is a counter-rotating seam between planes 6 and 1. Planes 1 through 6 are each separated by 31.6°, whereas planes 6 and 1 are separated by 22.1° above the equator.

Satellites within each plane are separated by 32.7° and are moving at about 7.5 km/s, resulting in an orbital period of just over 100 min. Satellite orbits for the Iridium system are shown in the

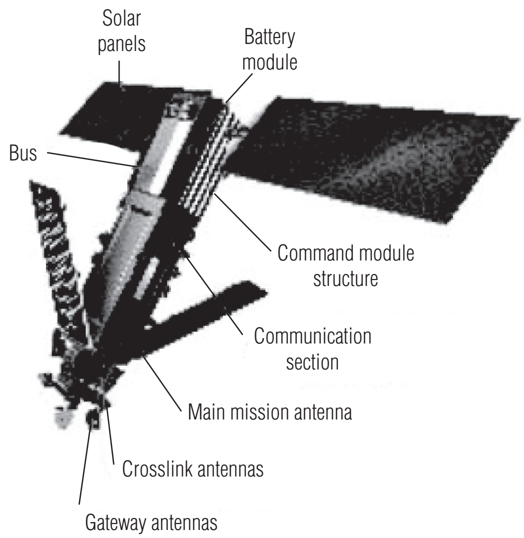


Figure 20.7 The Iridium satellite

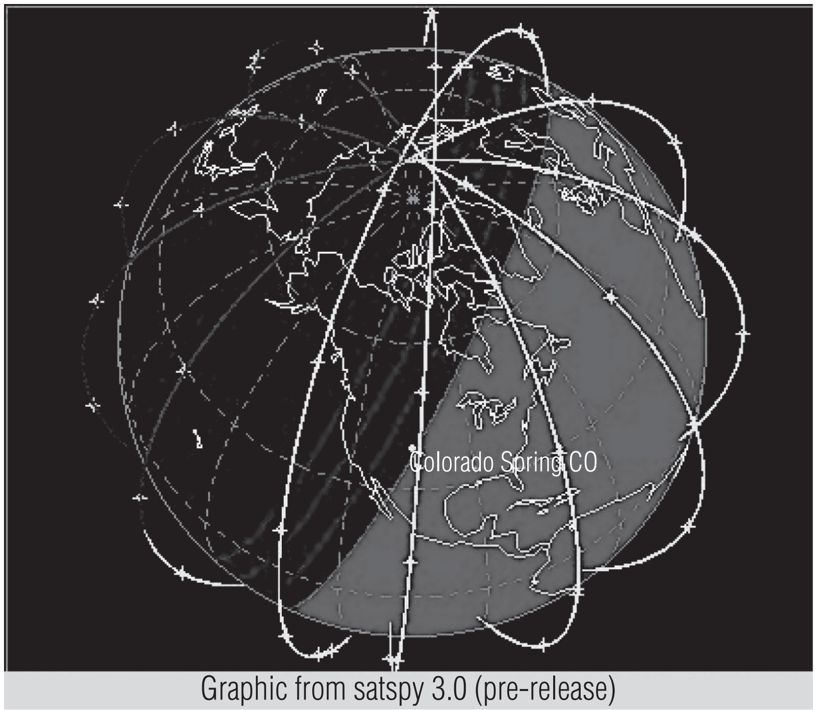


Figure 20.8 The Iridium system constellation

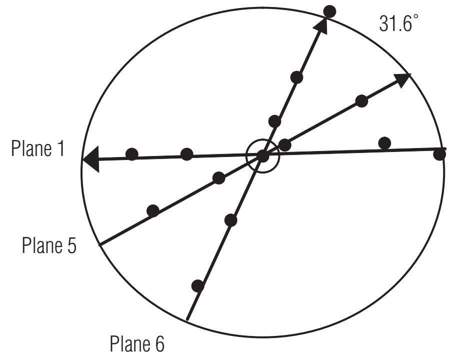


Figure 20.9 Satellite orbits for the Iridium system

Figure 20.9. The existing constellation of 66 satellites is expected to remain operational until at least 2014, with many satellites expected to remain in service until the 2020s. Iridium is planning a new generation of satellites with improved bandwidth to be operational by 2016. This system will be backward compatible with the current system.

Satellites orbit the earth at an altitude of 780 km. The footprint of a satellite is roughly 2,150 mi in diameter. Each satellite has 48 spot beams. Some of the beams are turned off as the satellite approaches the pole. This is to prevent self-interference when the satellite footprints begin to overlap with each other near the poles. The number of active spot beams at any moment is approximately 2,000. Each spot beam covers a cell on earth, which means that earth is divided into approximately 2,000 (overlapping) cells. The spot beams are provided by three-phased array panel antennas. In the Iridium system, communication between two users takes place through satellites. The iridium system provides direct worldwide communication using hand-held terminals similar to cellular telephony.

Spare satellites are usually held in a 667 km (410 mi) storage orbit. In case of a satellite failure, these will be boosted to the correct altitude and put into service. Recently one Iridium satellite failed in July 2008. Iridium system has 285,000 subscribers as of early August 2008. Calls to Iridium phones are notoriously expensive, ranging from US $3 to US $14 per min.

The Iridium system supports the following links as shown in Table 20.3.

20.3.2 Architecture of the iridium system

The architecture for the Iridium system is shown in the Figure 20.10. A call from ISU can be routed to another ISU located anywhere on the earth through satellite network or it can be connected to the public switched telephone network (PSTN) through an earth station for routing

Table 20.3 Supporting link of Iridium system

Direction	Frequency
Space Vehicle-Gateway	27.5–30.0 GHz (Ka band)
Gateway-Space Vehicle	18.8–20.2GHz (Ka band)
Space Vehicle-Space Vehicle	22.55–23.55GHz(Ka band)
Space Vehicle-ISU	1.616–1.625GHz(L band)
ISU-Space Vehicle	1.161–1.625GHz (L band)

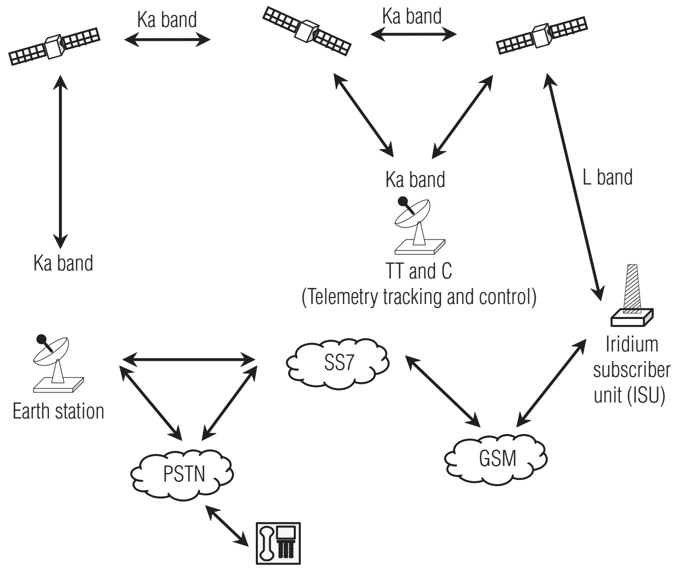


Figure 20.10 Network architecture for the Iridium system

and delivery through the PSTN. Connections between the Iridium satellites and the PSTN are provided via gateways, generally colocated with earth stations (ES). The ISLs can be used to route each call to the earth station closest to the origination or destination of the call. The dual mode mobile station of ISU can also provide public land mobile network.

A call flow between PSTN and ISU in the Iridium system is illustrated in Figure 20.11. First, the PSTN gateway sends Iridium subscriber unit's (ISUs) Mobile **station integrated services digital network number** (MSISDN) to the home gateway (GW). This is represented by path *a*. Then the home gateway sends a query to serving gateway for ISU location information. After receiving the query from the home gateway, the serving gateway returns the location information to the PSTN gateway. These two paths are indicated with *b* and *c*. Then the PSTN gateway routes a call to the serving gateway as shown with path *d*. The serving gateway alerts ISU (path *e*), then ISU answers (path *f*).

Figure 20.12 shows a call flow example between two ISU's in Iridium system. The ISU calling subscriber acquires space vehicle (satellite). Then it accesses the home gateway (calling) for authentication as shown in paths *a* and *b*. After authentication, the service transferred to serving gateway (calling). This is shown by path *c*. The serving gateway (calling) also receives request (path *d*). Then the serving gateway sends a query to serving gateway (called) for location information (path *e*).

General characteristics

Satellite weight – 700 kg (1,500 lb (pound))

- Spot beams – 48 per satellite
- Link margin (the difference between the receiver's sensitivity and the actual received power) – 16 dB (average)
- Life time – 5–8 years.

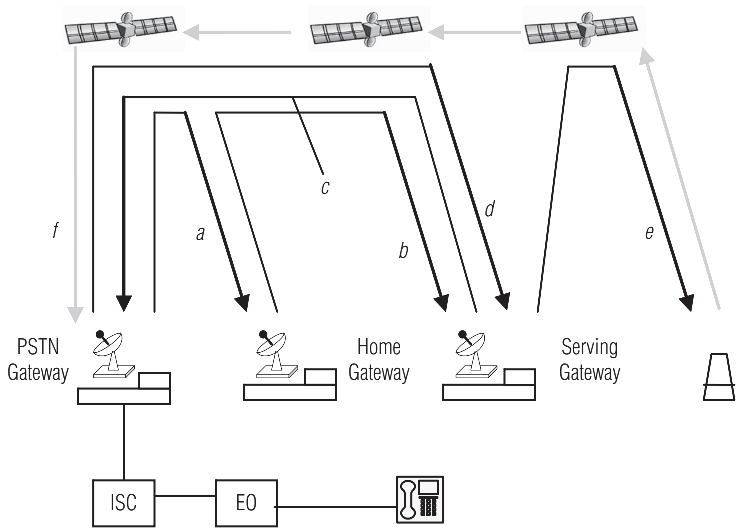


Figure 20.11 Call flow example between PSTN and ISU in Iridium

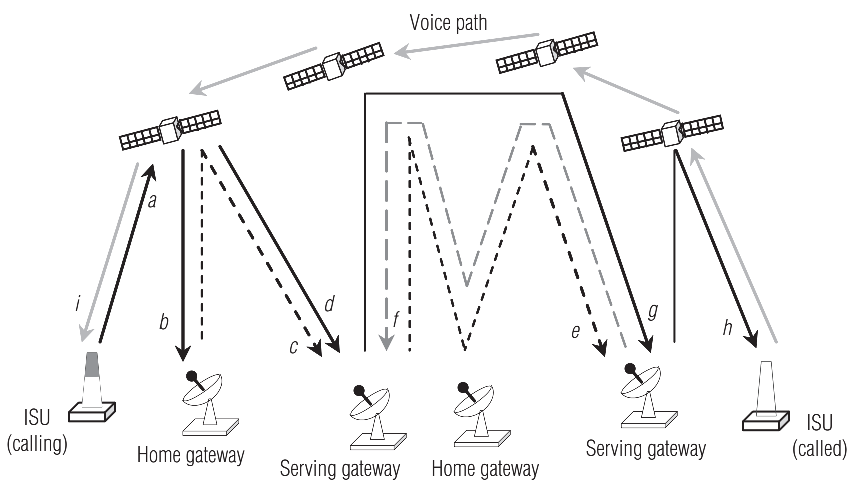


Figure 20.12 Call flow example between ISU and ISU in Iridium

Handoff

As a satellite travels over the horizon, calls are handed to adjacent spot beams. This occurs approximately every 50 s. A satellite only stays in view for 7 min at the equator. When the satellite disappears from view, an attempt is made to hand the call to another satellite. If no other satellite is in view then the connection is dropped. This may occur when the signal from either satellite is blocked by an obstacle. When successful, the inter-satellite handoff may be noticeable by a quarter-second gap.

Allocated Frequencies

The allocated frequencies for Iridium system are shown in Table 20.4.

Table 20.4 Frequency allocation different Iridium satellite system

Direction	Frequency
Iridium Phone-Satellite	1621.35–1626.5MHz(L band)
Satellite-Iridium Phone/Pager	1616–1626.5MHz (L band)
Satellite-Satellite	23.18–23.38GHz (Ka band)
Satellite-Gateway	19.4–19.6GHz (Ka band)
Gateway-Satellite	29.1–29.3GHz (Ka band)

20.4 Globalstar system

Globalstar is another LEO satellite system. Communication between two users in the Iridium system is through ISLs. Globalstar system does not use ISLs. Call routing and delivery is done through large number of interconnected earth stations or gateways. The use of gateway ground stations provides customers with localized regional phone numbers for their satellite handsets. But if there are no gateway stations to cover certain remote areas (such as areas of the South Pacific and the Polar Regions), service cannot be provided in these remote areas, even if the satellites may fly over them.

The Globalstar satellite is shown in Figure 20.13, and Figure 20.14 illustrates the Globalstar system.

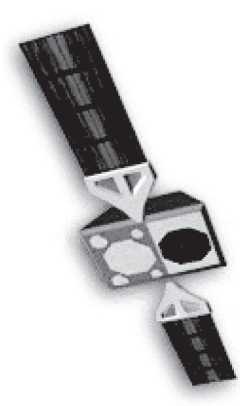


Figure 20.13 Globalstar satellite

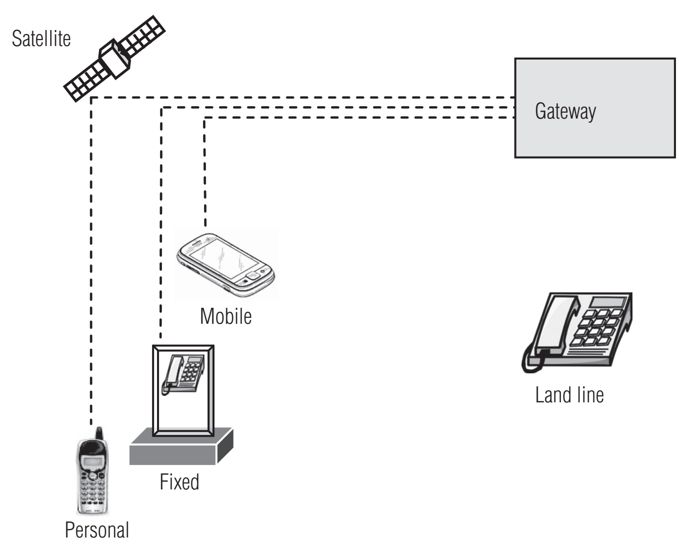


Figure 20.14 Illustration of globalstar system

The Globalstar system is designed to provide high-quality satellite-based services to a broad range of users, including

- Voice calling
- Short messaging service (SMS)
- Roaming
- Positioning
- Facsimile
- Data transmission

20.4.1 Globalstar system constellation

The system uses 48 satellites in six polar orbits. Each orbit has eight satellites with additional satellites in orbit as spares. Since it consists of fewer satellites, the Globalstar system is cheaper. The system operates at an altitude of 876 mi (1,414 km) in space. When all gateways are fully deployed, Globalstar constellation of 48 LEO satellites will pick up signals from over 80 per cent of the earth's surface, everywhere outside the extreme polar regions and some mid-ocean regions. Globalstar launched eight spare satellites in the year 2007.

- Globalstar orbits have an inclination of 52°. Therefore, Globalstar does not cover polar areas due to the lower orbital inclination.
- Globalstar orbits have an orbital height of approximately 1,400 km and latency is still relatively low (approximately 60 ms).
- Each Globalstar satellite consists of an antenna, a trapezoidal body, two solar arrays, and a magnetometer.

General characteristics

Total weight – 450 kg,
Number of spot beams – 16
Power – 1,100 W
Lifetime – 7.5 years

20.4.2 Architecture of the globalstar system

The architecture of a Globalstar system is shown in Figure 20.15. Calls from a Globalstar subscriber user (GSU) are routed to the nearest earth station /gateway via a satellite. From the earth station, they will be routed over the existing terrestrial network. There are more than 100 gateway stations distributed around the world. Each station is equipped with three to five antennas that track the trajectories of the satellites to provide interface between the terrestrial network and Globalstar satellites. A Globalstar gateway is designed to serve an area of 3,000 km in diameter.

The Globalstar uses two types of communication links:

- Communication between the terminals and the satellite using L band/S band
- Communication between the earth stations and the satellite using C band

The allocated frequencies for communication are shown in Table 20.5.

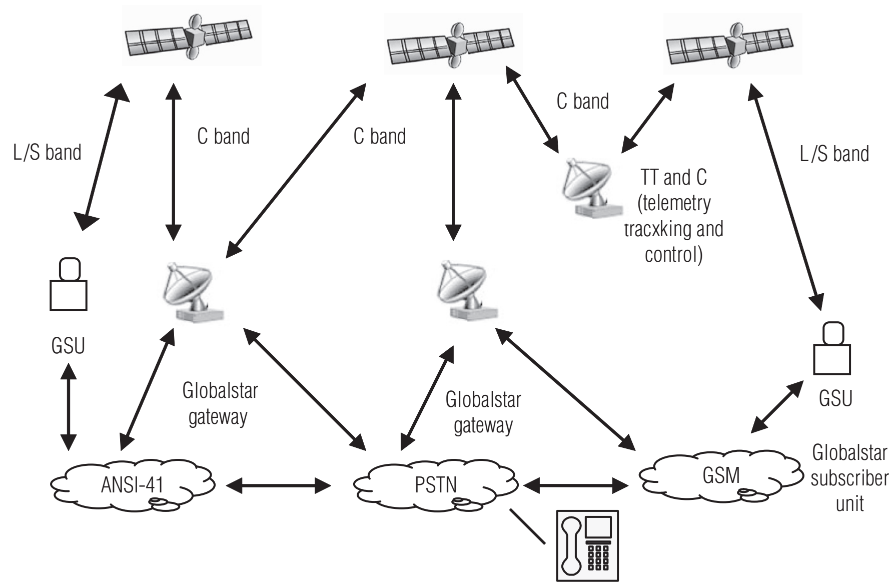


Figure 20.15 Network architecture for the Globalstar system

Allocated frequencies

Globalstar uses CDMA multiple access technology for the service links. It uses FDMA/FDD for the GWLs with QPSK modulation. CDMA provides increased capacity through its frequency reuse, voice activity detection, and spectrum sharing capabilities, and better performance through its support for multipath diversity.

Several satellites pick up a call, and this "path diversity" assures that the call does not get dropped even if a phone moves out of sight of one of the satellites. As soon as a second satellite picks up the signal and is able to contact the same terrestrial "gateway", it begins to simultaneously transmit. If buildings or terrains block your phone signal, this "soft-handoff" prevents call interruption. The second satellite now maintains transmission of the original signal to the terrestrial "gateway." Additional advantages of using LEO satellites within the Globalstar system include no perceptible voice delay and lighter/smaller all-in-one handsets. The satellites utilize "bent-pipe" architecture. On any given call, several satellites transmit a caller's signal via

Table 20.5 Frequency Allocation different Globalstar satellite system

Direction	Frequency
Globalstar Phone-Satellite	1610–1625.5 MHz (L band)
Satellite-Globalstar Phone	2483.5–2500 MHz (S band)
Satellite-Gateway	6875–7055 MHz (C band)
Gateway-Satellite	5091–5250 MHz (C band)

CDMA technology to a satellite dish at the appropriateGateway. Gateways process calls, and then distribute them to the existing fixed and cellular local networks. Terrestrial gateways are an important part of Globalstar's strategy to keep key technology and equipment easily accessible and to integrate our services as closely as possible with the existing local telephony networks. This makes the Globalstar system and its services simpler to manage, expand, and improve.

20.5 The ICO system

The ICO system is a TDMA MEO system. ICO is targeted primarily at users from the existing terrestrial cellular market, which travel to places where terrestrial cellular coverage is incomplete, patchy, or non-existent. The system is designed to offer digital voice, data facsimile, and short-targeted messaging services to its subscribers. The ICO system design is intended to integrate mobile satellite communications capability with the public land mobile networks like GSM.

20.5.1 ICO system constellation

The system uses 10 satellites and two spares in two inclined circular orbits with five satellites in each plane. The system operates at an altitude of 10,355 km above the earth surface. The inclination of orbital planes with respect to earth equator is 45°. At any time, two or more satellites are visible to the user and the satellite access node (SAN).

General characteristics

Total weight – less than 2,000 kg
Lifetime – 12 years

20.5.2 Architecture of ICO system

Figure 20.16 shows the architecture of ICO system. To provide link between ICO satellites and terrestrial networks, 12 SANs are located in various parts of the world. Each SAN consists of earth stations with multiple antennas to provide communication with satellites, public telephone, and

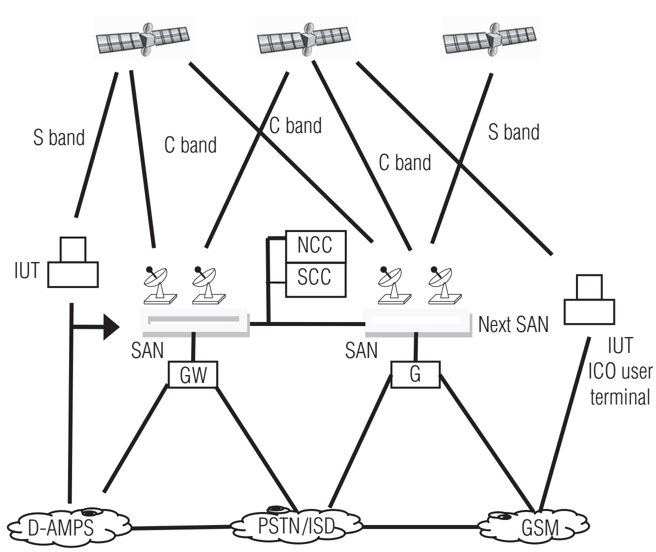


Figure 20.16 Architecture of ICO system

mobile networks. The interconnection to the public networks is through gateways. Besides the SANs, the ICO system deploys TT&C stations connected to a satellite control centre (SCC) for monitoring and controlling the satellites, as well as one or more network control centres (NCC) for overall management and control of the ICO system. The TT&C functions are associated with 6 of the 12 interconnected SANs.

A call between two ICO user terminals is routed via the inter-SAN network. For call origination/termination, a PSTN or cellular terminal uses the nearest SAN. Gateways provide the only terrestrial access between the PSTN and the SANs, and a gateway is connected to at least one SAN.

Allocated frequencies

Table 20.6 gives the frequencies of ICO satellite system.

Table 20.6 Frequency allocation of different ICO satellite system

Direction	Frequency
ICO user terminal-Satellite	2170–2200 MHz (S band)
Satellite-ICO user terminal	1980–2010 MHz (S band)
SAN-Satellite	5100–5250 MHz (C band)
Satellite-SAN	6935–7075 MHz (C band)

20.6 The teledesic system

Using a constellation of several hundred LEO satellites, Teledesic provides worldwide access to "fibre-like" telecommunications services such as broadband Internet access, digital voice, data, video-conferencing, and interactive multimedia. The target application for Iridium, Globalstar, and ICO systems is voice, with support of low bit data rate for facsimile and messaging for mobile subscribers. But Teledesic system provides worldwide connectivity to support multimedia, video, and high bit data rate services.

The Teledesic network can support a maximum capacity of one million full-duplex E1 connections. Simultaneously it can support millions of users. The capacity of this network can be improved by increasing the number of satellites. Assignment of channel bandwidth is dynamic and non-symmetrical. The range of uplink is from 16 Kbps to 2 Mbps whereas 28 Mbps on downlink. For gateway connections, Teledesic can also provide few high rate channels at 155 Mbps to 1.2 Gbps.

It requires low-power terminals and antennas with less cost, as it uses LEO satellites and high frequency.

20.6.1 Teledesic system constellation

The constellation is organized into 21 circular orbit planes that are staggered in altitude between 695 and 705 km. Each plane contains a minimum of 40 operational satellites plus up to four on-orbit spares spaced evenly around the orbit. As the satellite planes orbit north-to-south and south-to-north, the earth rotates underneath. The system design is costly. The Teledesic network is designed for dual-satellite visibility with at least one insight satellite at a minimum elevation of 40°.

20.6.2 Architecture of teledesic system

The architecture of the Teledesic system is shown in Figure 20.17. Communication between terrestrial networks (PSTN, PLMN, and private networks) and Teledesic satellites is routed via gateways and earth stations using Ka band.

Teledesic customers can also communicate directly with the satellite by using a specially designed Teledesic terminal using Ka band. The terminal can be mounted on a rooftop and connected inside a building with the customer's computer network.

Frequencies used for uplink:	28.6–29.1 GHz
Frequencies used for downlink:	18.8–19.3 GHz

Teledesic uses steerable antennas to minimize the number of handoffs due to the motion of the satellites and the earth's rotation. The entire earth's surface will be divided into a hierarchy of cells and supercells with 9 cells per supercell. A maximum of 64 supercells corresponding to a supercell per beam can be covered with a Teledesic satellite's footprint. The frequencies and time slots will be associated with each cell and managed by the serving satellite. So, during the duration of the call, more than one satellite may serve the mobile terminal. The channel resources assigned to the call will remain unchanged.

In a Teledesic system, a combination of time division multiple access (TDMA), frequency division multiple access (FDMA), space division multiple access (SDMA) , and asynchronous TDMA (ATDMA)) is used. Between cells, TDMA is used within a supercell. Here, a beam scans each cell in the supercell. Since all supercells receive transmissions from the satellite at the same time, SDMA is used between cells scanned simultaneously in adjacent supercells. Mobile terminals use FDMA for uplink and ATDMA for downlink within the time slots assigned to individual cells. The TDMA operation between cells implies that only one of the nine cells (in a supercell) can use all available frequencies at a time. This corresponds to a frequency reuse pattern of nine. In case of 20,000 supercells in a Teledesic system, the global frequency reuse factor is 2,222.

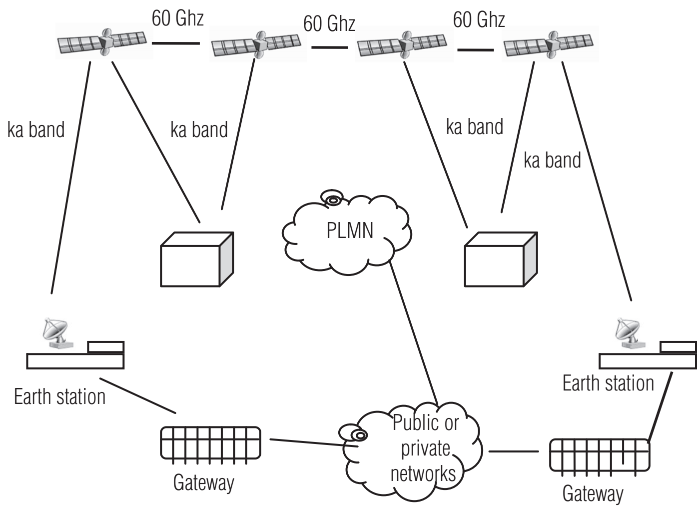


Figure 20.17 Architecture of the Teledesic system

20.6.3 Comparison of the characteristics of iridium, globalstar, ICO and teledesic systems

The Comparison of the characteristics of iridium, globalstar, ICO and teledesic systems are shown in Table 20.7.

Table 20.7 Comparison of the characteristics of Iridium, Globalstar, ICO and Teledesic systems

Characteristic	Iridium	Globalstar	ICO	Teledesic
Number of satellites	66 + 6 spares	48 + 8 spares	10 + 2 spares	288+ 12 spares
System type	LEO	LEO	MEO	LEO
Number of planes	6	8	2	12
Orbital altitude	783 km	1414 km	10,355 km	1350 km
ISL(Intersatellite links)	Yes	No	No	Yes
Inclination	86.4°	52°	45°	98.16°
Minimum elevation	8.2°	10°	10°	40°
Round trip delay	~10 ms	~10 ms	~200 ms	~10 ms
Satellite lifetime	5–8 years	7.5 years	12 years	10 years
Target services	Voice, fax, low speed data	Voice, fax, low speed data	Voice, fax, low speed data	Multimedia, high speed data
Satellite visibility	11.1 min	16.4 min	115.6 min	3.5 min

20.7 Personal communication services

Important features of personal communication services (PCS) include the following:

- A wide variety of wireless access and personal mobility services
- Can connect to PSTN and public data network (PDN)
- The main goal of PCS is to enable personal communication ability at any time, at any location, and in any terminal form

20.7.1 Categorization of PCS

- High-tier digital cellular system for vehicular and pedestrian services
  - Global system for mobile communication (GSM)
  - Digital communication system-1800 (DCS 1800)
  - Personal digital cellular (PDC)
- Low-tier telecommunication system for residential use, work place
  - Digital European cordless telephone (DECT)
  - Personal access communication system (PACS)
  - Personal handy phone system (PHS)

Table 20.8 compares the characteristics of high-tier cellular, low-tier PCS, and cordless systems.

Table 20.8 Comparison of the characteristics of high-tier cellular, low-tier PCS, and cordless systems

Systems	High-Tier Cellular	Low-Tier PCS	Cordless
Cell size	Large (0.5–35 Km)	Medium (50–500 m)	Small (50–100 m)
User Speed	High (≤257 Km/hr)	Medium (≤96 Km/h^{-1})	Low (≤48 Km/hr)
Coverage Area	Large(Macrocells)	Medium (micro- and pico-cells)	Small (Picocells)
Handset Complexity	High	Low	Low
Speech coding rate	Low (8–12 Kbps)	High (32 Kbps)	High (32 Kbps)
Delay	High (≤600 ms)	Low (≤10ms)	Low (≤20 ms)

20.8 Personal access communication system

Personal access communications system (PACS) is primarily based on the wireless access communication system (WACS) developed by Bellcore, which till recently was engaged in providing research support for the Regional Bell Operating Companies in the United States. Bellcore's WACS system not only addressed the radio component, but also proposed a network architecture known as the Network and Operations Plan (NOP), to permit the interconnection of WACS radio to a public or private network with intelligent network and mobility management functions. With the allocation of frequency spectrum in the 2 GHz band for PCS in the United States, multiple radio interfaces for PCS were standardized for operation in the paired licensed band, as well as for operation in the unpaired unlicensed band.

In essence, three versions of PACS have been standardized:

One for the licensed PCS frequency band (1,850–1,910 and 1,930–1,990 MHz FDD mode) and two for the unlicensed PCS band (1,910–1,930 MHz TDD mode). Whereas the term "PACS" is applied to the standard for the licensed band, systems for the unlicensed band are known as PACS-UB and PACS-WUPE (wireless user premises equipment). The former is based on the original WACS and the licensed band PACS, and the latter on the Japanese PHS (personal handyphone system).

The important features of PACS include the following:

- A fully integrated networked approach can be easily integrated with various cellular systems in a single handset
- Can support both private key and public key encryption for authentication and privacy
- Downlink pre-selection receiver antenna diversity and uplink full receiver diversity for better signal quality
- Downlink and uplink switched transmit antenna diversity for improving error-free signal transfers
- Automatic frequency assignment based on the quasi-static automatic frequency assignment (QSAFA) procedure support of sub-rate (16 and 8 kbs) and aggregated ($n \times 32$ kbs) channels
- Protocols to support messaging, circuit mode data, packet mode data, and interleaved speech/data services

20.8.1 Functional architecture for PACS

In one simplified architecture for PACS (Figure 20.18), multiple portable handsets are supported by individual radio ports (RPs) using TDMA with frequency division duplex (FDD) operation to distinguish between the uplink and downlink channels. Multiple RPs then subtend on individual radio port control units (RPCUs). The user traffic across the P interface between an RP and an RPCU is separated from the control signals for management of radio functions by using an embedded operations channel (EOC). The RP-to RPCU transmission links may take the form of Tl or El links, digital subscriber lines (DSL), or high speed digital subscriber lines (HDSL). Whereas the P interface for signalling between the RP and the RPCU is generally a proprietary or provider-specific interface, the C interface between the RPCU and the integrated services digital network/advanced intelligent network (ISDN/AIN) () switch is based on the ISDN Basic Rate Interface (BRI).

The radio ports in PACS have a simple design and function primarily as radio frequency (RF) modems. They are powered from the local switch (within 12,000 feet) by means of HDSL technology over copper wires. The RPs can be easily mounted on utility poles or building walls. The RPCU contains the necessary intelligence (electronics) to provide management and control functions for managing radio resources.

The access manager (AM), which may be integrated with a RPCU or may operate as a stand-alone unit, includes functions to support such tasks as remote database query, assisting in call setup and delivery, automatic link transfers (ALT) during handoff between two RPCUs, and management of multiple radio ports. The access manager functions may reside in such AIN elements as a Service Control Point (SCP), an Intelligent Peripheral (IP), or a switch adjunct. The mobility management functions, necessary databases, and protocols (like IS-41) need to be supported in the public or private network that serves the PACS radio port control units and the access managers.

20.8.2 PACS radio aspects

The basic frame for the PACS radio has duration of 2.5 ms (with 8 time slots/frame) at 400 frames per second. The uplink transmissions from the PACS subscriber unit (SU) utilize TDMA and the set receiver operates in the time division multiplex (TDM) mode for the downlink signals. PACS supports a system broadcast channel (SBC) that may be deployed as one of the following channels:

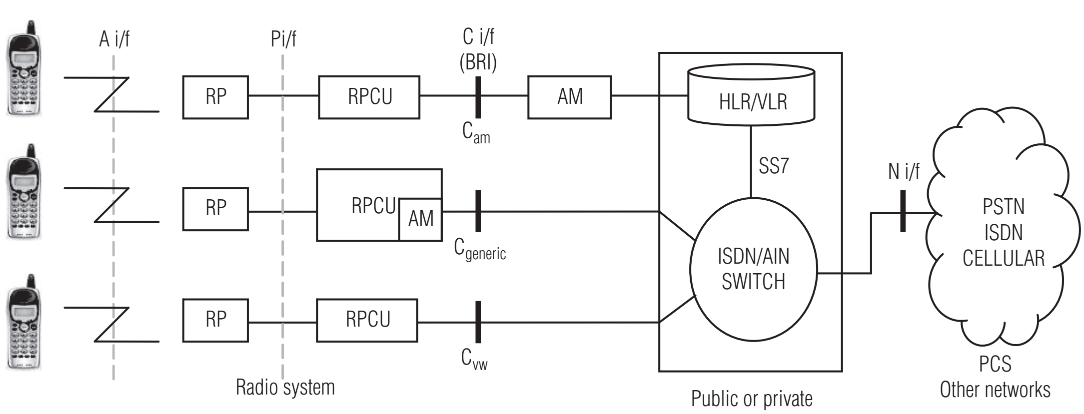


Figure 20.18 Functional architecture for PACS

- An alerting channel (AC) for alerting SUs to incoming calls
- A system information channel (SIC) to broadcast system information (including, e.g., subscriber/terminal identities, relevant timers, protocol parameters)
- A priority request channel (PRC) to be used by SUs for emergency call. The user information (voice or data) is carried in the 80-bit fast channel (FC) or 10-bit slow channel (SC).

The frame and burst structure for PACS radio is shown in Figure 20.19.

A time slot in the PACS frame consists of 120 bits, with 80 bits allocated for the payload (user and signalling traffic) and 40 bits for overhead in the downlink and uplink frames. In the downlink direction, the 40 overhead bits are deployed as follows:

- A synchronization channel (SYNC) (14 bits), which provides synchronization
- A SC (10 bits), which may be used for transporting additional synchronization patterns, indication of word errors, signalling information, or user data
- A 15-bit cyclic redundancy check (CRC)
- A power bit for optimizing power output at the subscriber unit

In the uplink direction, the initial 12 bits are used as a guard time between consecutive time slots, which are followed by 2 bits for priming differential decoder at the radio ports. The remaining bits provide the slow channel (10 bits), the fast channel (80 bits), and the CRC (15 bits), with remaining bit as reserve bit. The 80-bit fast channel available once every frame (2.5 ms) translates

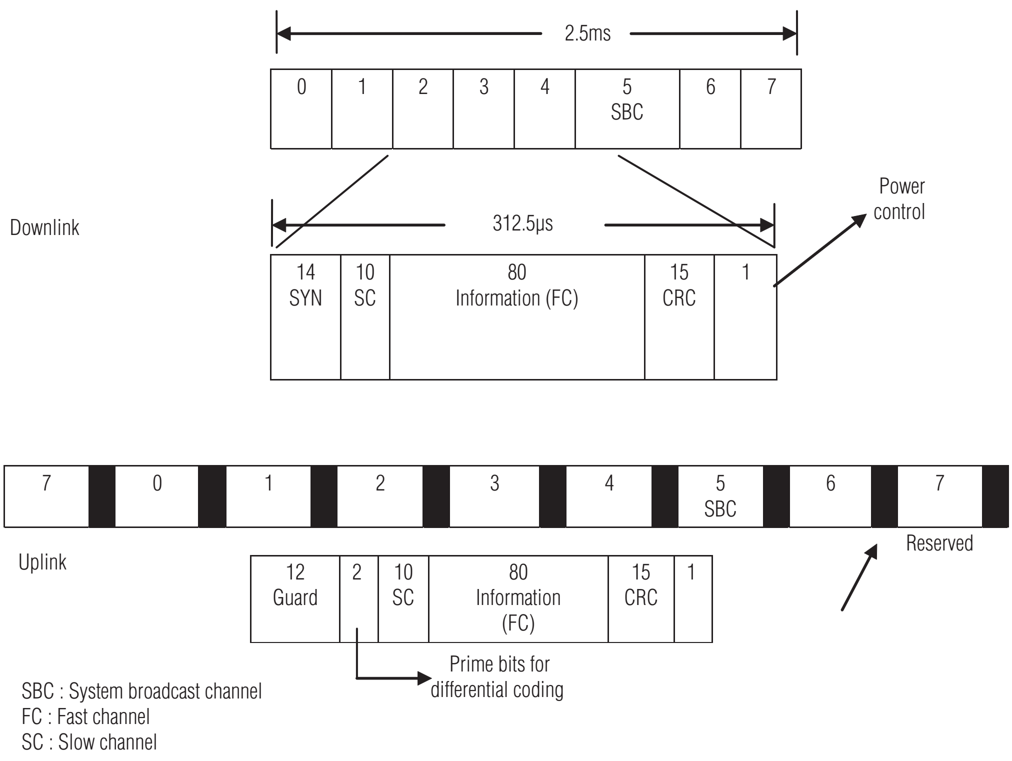


Figure 20.19 Frame structure for PACS

into a basic data rate of 32 kbs, which can drive a good quality speech coder. PACS can also provide sub-rate channels (16 or 8 kbs) by spacing the bursts every two or four frames, respectively.

Higher data rates can be achieved in PACS by aggregating two or more time slots in a frame. The relatively large bandwidth available for the alerting channel (AC) implies almost negligible blocking rates on the alerting channel for call delivery, even when the alerting/registration area (ARA) is very large (e.g., serving > 200,000 users). The use of large ARAs leads to considerably low registration traffic levels in PACS, as well as reduced network signalling between home location register (HLR) and visiting location register (VLR).

To enhance the signal quality at the subscriber units and radio ports, a PACS radio system deploys two types of diversity techniques. The pre-selection diversity at the subscriber unit takes advantage of the radio port's continuous transmission, which allows the subscriber unit to conduct antenna diversity measurements. Thus the subscriber unit can determine, immediately prior to an incoming burst, which antenna receiver will provide the best signal. Pre-selection diversity provides an economical, effective, and efficient diversity mechanism for pedestrian speed applications in PACS. As opposed to the pre-selection antenna diversity used at the SU, the radio port deploys full selection dual-receiver selection diversity, whereby the uplink signals received independently by the two diversity receivers are demodulated and the radio port selects the receiver with the better quality signal.

The pre-selection and full selection diversity methods described above provide improvements in the general signal quality received at the two ends of the radio interface. PACS also provides for a mechanism that enhances the probability of receiving an error-free burst. This is achieved by deploying switched transmitter antenna diversity both at the subscriber unit and at the radio port. In this method, each end (SU and RP) informs the other if the previous burst was received error free or in error. If the last burst was received in error, the transmitting end switches to the other diversity antenna for the next transmission.

20.8.3 Features of general systems in PACS

As mentioned earlier in this section, the PACS specification includes the radio interface as well as the necessary radio access and network functions to define a complete, implementable system. The three broad categories of functions that need to be supported are radio resource management functions, mobility management functions, and call and service control functions.

The PACS radio link maintenance feature is implemented to ensure adequate signal quality during a call. The procedures used for this purpose include ALT, time slot transfer (TST), and power control at theSU. Time slot transfer is a special case of ALT in which the call is handed off to another time slot within the same radio port.

The decisions to invoke an ALT and the choice of the receiving RP, as well as the decision for aTST, are made by the SU. The SU output power levels, on the other hand, are controlled by the RPCU, which utilizes the power control procedure to request an SU to adjust its output power level. Invocation and implementation of these procedures are based on the following RF measurements undertaken by the subscriber unit (when turned on) and the radio ports on every time slot.

- Radio signal strength indication (RSSI), which provides a measure of the cochannel interference power and noise.
- Quality indicator (QI), which provides an indication of signal-to-interference, and signal-to-noise ratios, including the effects of dispersion. QI is used in setting the power control bit for raising the SU output power level.

- Word error indicator (WEI), which provides an indication of occurrence of one or more bit errors (in a time slot) due to radio link degradation. WEI is used by the SU to indicate the uplink performance.

Automatic link transfer

In PACS, ALT or call handoff is initiated and controlled by the SU based on signal strength and quality measurements from the serving RP, as well as a number of candidate RPs suitable for possible handoff. PACS can support the following types of ALT or handoff procedure:

- Transfer from one time slot (channel) to another within the same RP, that is, TST
- Transfer from one RP to another RP within the same RPCU (intra-RPCU ALT)
- Transfer from one RP to another RP, the two RPs under different RPCUs but within the same switch (inter-RPCU ALT)
- Transfer from one RP to another RP, the two RPs under different RPCUs and different switches (inter-switch ALT)
- Transfer from one RP to another RP, the two RPs under different RPCUs and different AMs (inter-AM ALT)

The different types of ALT listed above are transparent to the SU that is, the SU uses the same procedure to invoke an ALT request. It is up to the network to recognize which type of ALT procedure is appropriate in a given instance. As an example, an inter-RPCU ALT is illustrated in Figure 20.20.

The steps involved in this type of handoff can be summarized as follows:

- The SU requests an ALT by sending a brief signal to the new RP-B by briefly interrupting the conversation on the traffic channel.
- On receipt of the request, the AM associated with RPCU-B transfers the session cipher key to the ciphering unit in the new RP-B.
- The RPCU-B requests the switch to create a new connection to the SU over the new RPCU-B and RP-B and bridging it to the old path (over RPCU-A and RP-A).
- The network tells the SU to transfer from the old channel to the new channel through both RP-A and RP-B.
- The SU confirms the transfer to the new channel and the network dismantles the old conversation path.

Location registration/deregistration

Location registration (and deregistration) is a key function under mobility management in any mobile network. It allows the network to maintain the current location of a subscriber unit as the subscriber moves around in the network so that incoming calls to the subscriber may be delivered. The registration function is invoked every time the SU is powered on or moves from one registration area (RA) into another, whereby the SU registers with the VLR. The VLR then informs the subscriber unit's HLR about its current address. In case of registration due to registration area change, the VLR on which the SU was previously registered may be informed about its new location to permit updating of records.

For incoming call delivery to the SU, its HLR is first consulted about its location, which in turn requests the serving VLR for a routing number/address. Whereas the registration procedure is similar to that used in cellular or PCS systems that utilize ANSI-41 networking protocols, the deregistration procedure in PACS is somewhat different. ANSI-41 utilizes an explicit deregistration

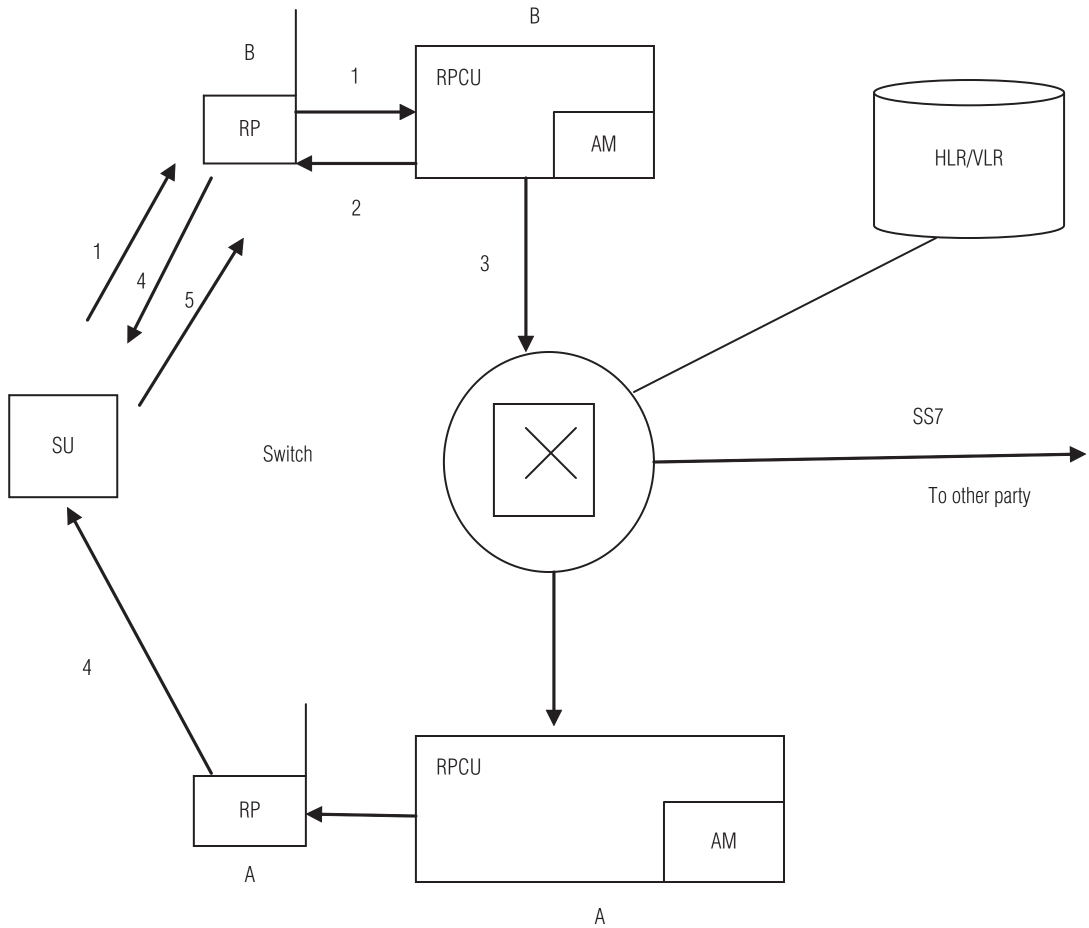


Figure 20.20 Illustrative example of inter-RPCU handoff in PACS

procedure. When an SU registers with a new VLR, the previous VLR is always informed about the move so that it may deregister the SU in its records. To reduce signalling load in the network, PACS suggests the use of deregistration by time-out or by polling. In the former case, the SU is deregistered automatically if the SU does not request re-registration within a specific time period since its last request. In the latter case (generally used in PACSUB), the network periodically sends an alerting message to the SU and if the SU does not respond to the polling signal within a time-out period, it is deregistered.

Authentication and ciphering

To prevent fraudulent use of the terminal and to maintain privacy of information over the radio traffic channel, PACS provides for explicit user authentication and ciphering over the radio link. The authentication and key agreement (AKA) protocol in PACS can support a security mechanism based on either a private key or a public key. The SU authentication in PACS requires that a unique identity be assigned to an SU (similar to mobile identity number in IS-136 systems or international mobile subscriber identity in GSM). The authentication procedure based on a private key is similar to that used in ANSI-41 systems or radio link ciphering. An algorithm that conforms to the U.S. Data Encryption Standard may be used on the PACS traffic channel. For a newly established channel, the required session key is generated by the AKA protocol. PACS supports a security menu concept, which provides the flexibility to use different options and combinations of AKA procedures and link ciphering algorithms. The security menu is implemented by means of a 4-octet field in the SIC: the first two octets indicate the available AKA procedures and the latter two indicate the link encipherment algorithms and modes of operation. As mentioned earlier, during an ALT or handoff of a call in progress, the session key is transferred to the new serving RPCU for the duration of the call or until the next handoff.

Call origination and delivery

The call and service control functions are required to support call originations from the SU, call delivery to the SU, and support of supplemental services or vertical features (e.g., three-way calling, call waiting, calling line identity). Whereas the call origination feature in PACS is similar to that in a mobile network utilizing ANSI-41 procedures, the call delivery feature in PACS follows a procedure that is more efficient in the use of network resources. Whereas in the ANSI-41-based call delivery procedures, the VLR returns the routing address to the HLR on request, in PACS, the VLR first pages the SU and a routing address is returned to the HLR only if the SU is not turned off. Thus, if the SU is turned off, the call will not be routed through the network and it may be given a different treatment (announcement, transfer to a voice mail box) based on the called party's profile.

20.9 Mobile and personal communications: past, present, and future

Radio communication can trace its origin to the discovery of electromagnetic waves by Hertz in 1888 and the subsequent demonstration of transatlantic radio telegraphy by Marconi in 1901. Mobile radio systems using simplex channels (push-to-talk) were introduced in the 1920s for police and emergency services.

The first public mobile radio system in the United States was introduced in 1946 and can perhaps be considered to be the beginning of the era for public mobile communication services. As illustrated in Figure 20.21, the evolution of public mobile and PCS may be divided into three broad periods. The development of the cellular concept in the 1970s was a defining event, which has played a significant part in the evolution of mobile communication systems and networks around the world. The delineation of a boundary between the mobile communications in the past and in the present is not very difficult, because implementation of analogue and digital cellular systems clearly represents a step change in the design and capabilities of mobile communication systems. However, a similar delineation of a boundary between the present and future mobile and personal communications systems is not so clear. Future mobile and personal communication systems will, to a large extent, represent evolution and enhancements of the present systems in many directions and on many fronts. These directions include the following:

- Increased capacity and coverage
- Global roaming and service delivery
- Interoperability between different radio environments
- Support of high bit rate data, the Internet, and multimedia services
- Wireless and wireline integration for mobile broadband services
- Global coverage using satellite constellations

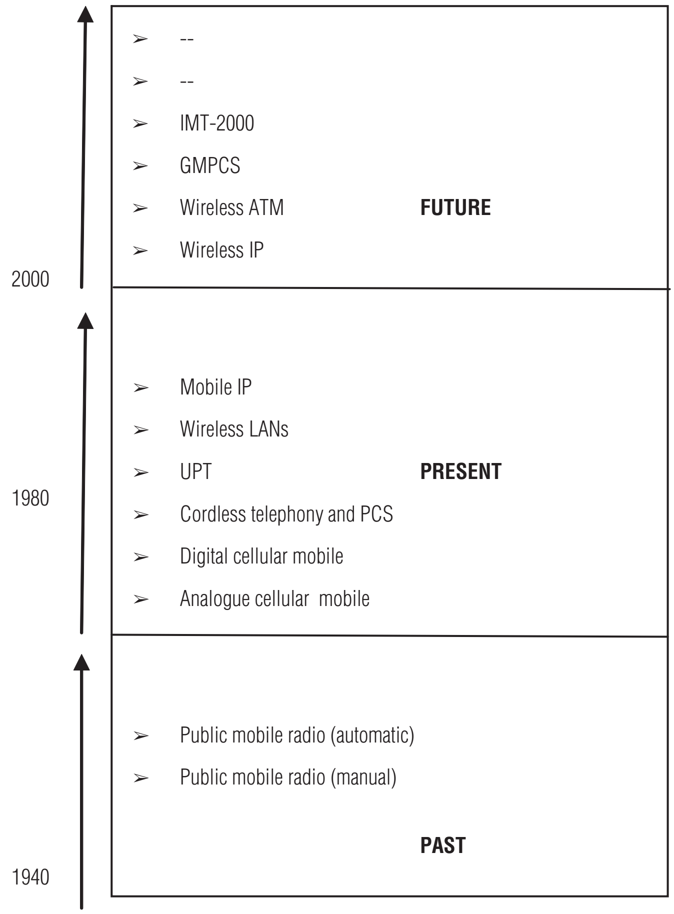


Figure 20.21 Evolution of public mobile and personal communication services

20.9.1 The past

From the introduction of public mobile radio in the United States in 1946 until the first analogue cellular system went into operation in Chicago in 1983, mobile radio systems were based on the trunking principle. In other words, the available frequency spectrum (in the 150 or 450 MHz band) was divided into a suitable number of frequency channels. A centralized, high power antenna was used to transmit signals to mobile receivers. Large mobile receivers were installed in automobiles (in the trunks) and the telephone sets also were rather large. A call originating from or terminating on a mobile terminal had to compete for one of the limited number of channels. The QOS in terms of call blocking probabilities was very high – in the order of 20–25 per cent. However, the users were willing to trade off the convenience of mobility against the poor QOS in terms of call blocking and received signal quality. These systems were also severely limited in terms of capacity and coverage.

To alleviate the high blocking problem in the early systems, efforts were made to allow call originations from the mobile telephones to wait for a free channel. In the so-called automated mobile telephone system (AMTS), the mobile telephone user would key in the called number and press the send button. The receiver system would then start scanning for an idle channel by cycling through all the channels in the system. In some systems, the number of scan cycles was restricted, so that if an idle channel was not found within the allowed number of scans, the call would be blocked. However, incoming calls to mobile terminals (mostly originating from fixed terminals in public switched telephone networks) had no mechanism for awaiting a free channel and were blocked on all-channels-busy condition. Though the improvement in the QOS in these systems was only marginal, they did provide some interesting performance modeling problems.

20.9.2 The present

Since the initial commercial introduction of advanced mobile phone system (AMPS) service in 1983, mobile communications has seen an explosive growth worldwide. Besides the frequency reuse capabilities provided by cellular operation, advances in technologies for wireless access, digital signal processing, integrated circuits, and increased battery life have contributed to exponential growth in mobile and PCS. Systems are evolving to address a range of applications and markets, which include digital cellular, cordless telephony, satellite mobile, and paging and specialized mobile radio systems. Data capabilities of these systems are also coming into focus with the increasing user requirements for mobile data communications, driven by the need for e-mail and Internet access. Whereas the analogue cellular mobile systems fall in the category of first-generation mobile systems, the digital cellular, low power wireless, and personal communication systems are now perceived as second-generation mobile/PCS systems. The first digital cellular system specification was released in 1990 by the European Telecommunications Standard Institute (ETSI) for the GSM system. The GSM, DCS 1800 (1,800 MHz version of GSM), and DECT (digital enhanced cordless telecommunications) systems developed by ETSI form the basis for mobile and PCS not only in Europe but in many other parts of the world including North America. The number of GSM subscribers worldwide exceeds 100 million and is growing rapidly.

In the United States, the implementation of digital cellular standards developed by the Telecommunications Industry Association (TIA) is progressing at a rapid rate. These standards are based on time (TDMA and CDMA) technologies. Unlike GSM, the systems are designed to operate with dual-mode terminals that can also support analogue AMPS service. The intent of the emerging PCS standards in the United States is to provide a combination of terminal mobility, personal mobility, and service portability to the end users utilizing a range of wireless technologies and network capabilities. The cellular mobile and PCS standardization activity in the United States reflects the highly competitive and open-market view of mobile and PCS and their evolution. Rather than a single agreed standard across the entire industry, multiple standards for radio systems and network implementations have emerged, and as expected, need for marketplace and end-user acceptance is driving the ultimate implementation decisions by the operators. A third digital cellular system called the personal digital cellular (PDC) was developed in Japan and is in full commercial operation in that country. To a large extent, the specifications for these second-generation cellular systems are being developed to meet the business and regulatory requirements in specific countries and/or regions, leading to incompatible systems that are unable to provide global mobility. Analogue cordless telephones have been in common use in residential applications, where the telephone cord is replaced by a wireless link to provide terminal mobility to the user within a limited radio coverage area. Low-power digital cordless telecommunication systems like CT2 (Cordless Telephony 2), DECT and Japan's PHS are intended to provide terminal mobility in residential, business, and public access applications where the users can originate and receive calls on their portable terminals as they change locations and move about at pedestrian speeds within the coverage area. It is also anticipated that the same terminal can be used in all three application environments at the residence, at the workplace, and at public locations (airports, train and bus stations, shopping centres, etc.).

While the initial focus of the current generation of mobile and personal communication systems has been circuit-switched voice and low bit rate data services, the demand for wide-area as well as local-area wireless data services is rapidly increasing. Reflecting market needs for better mobile data services are such standards as cellular digital packet data (CDPD) for support of packet data services on analogue cellular networks, high speed circuit-switched data (HSCSD) and general packet radio service (GPRS) for GSM, and IEEE 802.11 and high performance European radio LAN (HIPERLAN) for wireless LANs. The emerging industry view is that the main drivers for next generation wireless networks will be Internet and multimedia services. Evolution towards high bit rate packet mode capabilities is therefore a key requirement for present and future mobile and personal communication systems.

With respect to personal mobility services, such features as call forwarding, call waiting, automatic credit card calling, and personal number services represent ad hoc attempts by telecommunications network operators to provide a level of personal mobility to the users. Universal personal telecommunication (UPT), the emerging standard in the International telecommunication union's telecommunications standardization sector (ITU-T) for personal mobility, will utilize the intelligent network (IN) and integrated services digital network (ISDN) capabilities to provide network functions for personal mobility.

20.9.3 The future

With the rapidly increasing penetration of laptop computers, which are primarily used by mobile users to access Internet services like e-mail and World Wide Web (WWW) access, support of Internet services in a mobile environment is an emerging requirement. Mobile IP is an Internet protocol that attempts to solve the key problem of a developing mechanism that allows IP nodes to change physical location without having to change IP address. Asynchronous transfer mode (ATM) is now generally accepted as the platform for supporting end-to-end, broadband multimedia services with guaranteed QOS. Wireless ATM (WATM) aims to provide an integrated architecture for seamless support of end-to-end multimedia services in the wireline as well as the wireless access environment. Thus, WATM is expected to meet the needs of wireless users who are looking for a common networking solution that can meet their high speed data and multimedia service requirements with excellent reliability and service quality.

To complement the cellular and personal communication networks, whose radio coverage will be confined to populated areas of the world (less than 15% of the earth's surface), a number of global mobile satellite systems are in advanced stages of planning and implementation. These systems are generally referred as Global Mobile Personal Communications by Satellites (GMPCS). GMPCS systems like Iridium, Globalstar, and ICO use constellations of LEO or MEO satellites and operate as overlay networks for existing cellular and PCS networks. Using dual-mode terminals, they will extend the coverage of cellular and PCS networks to any and all locations on the earth's surface. On the other hand, a LEO satellite system like Teledesic aims to provide high capacity satellite links to enable delivery of high bit rate and multimedia services to every location on the earth. International Mobile Telecommunications-2000 (IMT-2000) is the standard being developed by the ITU to set the stage for the third generation of mobile communication systems. The IMT-2000 standard not only will consolidate different wireless environments (cellular mobile, cordless telephony, and satellite mobile services) under a single standard but will also ensure global mobility in terms of global seamless roaming and delivery of services. ETSI is also developing a third-generation mobile communication system called Universal Mobile Telecommunication System (UMTS), which will belong to the family of IMT-2000 systems.

20.10 Rake receiver

The basic idea of a rake receiver was first proposed by Price and Green and patented in 1956. The rake receiver uses a multipath diversity principle – it rakes the energy from the multipath propagated signal components and combines it. M-ray multi-path models can be used for this purpose as shown in Figure 20.22. Each of the M paths has an independent delay *t*, and an independent complex time variant gain G. Here $r(t)$ is the transmitted signal received at the receiver front-end and $r_p(t)$ is received signal after processing as shown in Figure 20.23, a rake receiver utilizes multiple correlators to separately detect M strongest multipath components, which experienced different delays during their travel over the channel. Each correlator detects a time-shifted version of the original transmission, and each finger correlates with a portion of the signal, which is delayed by at least one chip in time from the other fingers.

The outputs of each correlator are weighted to provide a better estimate of the transmitted signal than is provided by a single component. Outputs of the M correlators are denoted by $Z_1, Z_2, \ldots, Z_M$ and the outputs are weighted by $\alpha_1, \alpha_2, \ldots, \alpha_M$, respectively. Once these are correlated, weighted, and

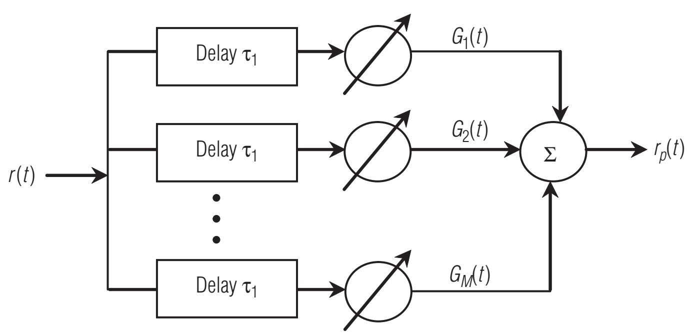


Figure 20.22 Rake receiver

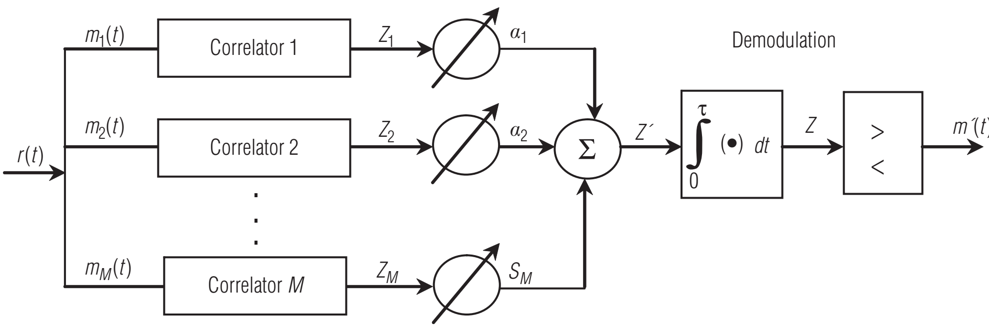


Figure 20.23 Rake with correlated and weighted sum decision

summed up, then the result is passed for demodulation and bit decisions. The weighting coefficients are based on the power or the signal-to-noise ratio from each correlator output. If the power or SNR from a correlator is small, then a small weighting factor will be assigned accordingly. Now if maximal-ratio combining is used, then the following equation can be written for the summed output.

The weighting coefficients, α_m, are normalized to the output signal power of the correlator

$$Z' = \sum_{m=1}^{M} \alpha_m Z_m \tag{20.1}$$

Choosing weighting coefficients based on the actual outputs of the correlator leads to better rake receiving performance with weighting coefficients, α_m given by

$$\alpha_m = \frac{Z_m^2}{\sum_{m-1}^{M} Z_2^2} \tag{20.2}$$

20.10.1 Applications of rake receiver

Apart from its use in a CDMA/WCDMA receiver, the rake receiver can also be used in a non-CDMA based receiver. Basically, a rake receiver "rakes" in the symbol energies from various multipath delays, and combines them using a maximal-ratio combining technique. An adaptive FIR equalizer will do the same job, basically equalizing the impulse response of the channel; this is equivalent to a matched filter that is matched to the different channel delays. There is nothing about a rake receiver that makes it applicable only for CDMA type signals. Things that need to be considered are as follows:

- The total length of the time window to collect the multipath, if the length of the time window is more, the possibility of providing many multipath will be more, but this leads to more complexity of the receiver and if the time window is shorter then we may not get enough multipath in that time window. On the basis of the the cell size and reflector objects in the cell, this may vary.
- Chip rate or symbol rate: if the chip rate is too low (e.g., wider time gap) and the cell size is small (smaller time to travel from transmitter to receiver over any path), then in this case, the multipath signals will be separated by a very small time gap and the chip/symbol from different multipath may overlap with each other at the receiver. Thus, resolving multipath will be hard.

20.10.2 Advanced rake receiver

Many technologies like high-speed downlink packet access technology (HSDPA), are used in WCDMA to support high data rate applications. The downlink data rates in WCDMA are improved using this HSDPA technology. But, in high-bit rate applications, self-interference limits performance of radio channel. To suppress such self-interference, Ericsson has developed advanced rake receivers.

G-rake receiver acts as an equalizer and suppresses self-interference. Traditional rake receiver architecture minimizes time to market and also the cost. G-rake reception can also be used in following applications:

- It can be used in voice service, where it can increase downlink capacity by 30 per cent.
- G-rake reception and two-antenna terminal platforms together can improve data and voice applications.
- It can also be used to improve performance when uplink data rates exceed 2 Mbps.

20.11 Mobility management

Mobility management is one of the major functions of a GSM or a UMTS network that allows mobile phones to work. The aim of mobility management is to track where the subscribers are, allowing calls, SMS, and other mobile phone services to be delivered to them.

20.11.1 Location update procedure

In a GSM network, a group of base stations is a given location area. It is the function of mobiles to detect the location area codes (LACs). When a mobile device moves from one location area to another, it finds that the LAC has changed. Then it rechecks by sending a location update request to the network. Mobile also sends its temporary subscriber identity (TMSI) and previous location to the network. This procedure is called as location update procedure.

Updated location information is required by the mobile in the following situations:

- To perform international mobile subscriber identity (IMSI) attach or detach location update procedure whenever a mobile is switched on or off.
- To perform periodic location update procedure
- When a mobile moves from one location area to another
- Due to signal fade, when a mobile chooses coverage from a different cell in another location area

Thus, the location updating procedure will allow a subscriber to have a reliable access to the network. It helps the subscriber to get or make a call anywhere within the whole coverage area.

20.11.2 Temporary mobile subscriber identity

Whenever a mobile is switched on, Visitor Location Register (VLR) randomly assigns a temporary identity to it and this identity is known as TMSI. It is 4 octet (8 bits) with hex digits and is transmitted between mobile and network. This TMSI is required for a process known as “Paging” which is one-to-one communication between the mobile and the base station.

Whenever the mobile enters a new geographical area, this TMSI is updated. This identity is also changed by the network at anytime to avoid anyone to listen to private conversations secretly. But, this random changing of TMSI makes it difficult to differentiate one mobile from another. To avoid this, another identity known as IMSI is sent to network, whenever the data in the mobile becomes invalid or when ever mobile is switched on. Proper care is taken in sending IMSI, so that mobile is not tracked and identified.

20.11.3 Roaming

Roaming is the ability of a mobile subscriber to use a visiting network for sending and receiving data, making and receiving voice calls, etc when present outside the geographical coverage area of home network. All the cellular networks support this fundamental mobility management procedure. Roaming is usually done either by using subscriber identity in visited network or by using a communicating terminal.

20.11.4 Location area

Location area is the name given to a group of base stations whose grouping is done to optimize signalling. Each location area is assigned a unique number known as LAC. Many base stations share a single base station controller, which receives measurements from mobile phones, controls handovers, and handles allocation of radio channels. Base transceiver station broadcasts LAC.

The location area should not be too small or too large. In large location area, if simultaneously many mobiles operate, it then leads to heavy paging traffic, because each and every paging request should be sent to all base stations in that area and the mobile has to listen to all messages for large amount of time. This results in wastage of power and bandwidth. If the location area is too small, the mobile should communicate with network whenever it gets in to new location area which wastes power.

20.11.5 Routing area

Certain mobile equipment have a new data transmission technology known as GPRS. GPRS is used for multimedia services and wireless services, whose data transmission is bursty in nature. It allows the users to connect to Internet service providers. When GPRS is involved, more paging messages are sent to each mobile and it is required to know the mobile location more precisely. To serve this purpose a subdivision of location area is defined as routing Area. Routing area update is done similar to location area update.

20.12 Network signalling

Telecommunications signalling is the transmission of data for the purpose of sharing information for network control and/or call control. The type of signalling used in a modern network is called signalling system number seven (SS7). SS7 is a critical component of modern telecommunications systems and its architecture is shown in Figure 20.24. SS7 is a communications protocol that provides signalling and control for various network services and capabilities. Being a layered protocol, SS7 provides various protocol levels for connection oriented and connectionless (database) signalling in fixed and mobile networks. SS7 comprises a series of interconnected network elements such as switches, databases, and routing nodes. Each of these elements is interconnected with links, each of which has a specific purpose. The routing nodes are the heart of the SS7 network and are called a signal transfer point (STP). STPs are connected to service switching points (SSP) that are switches equipped with SS7 control logic.

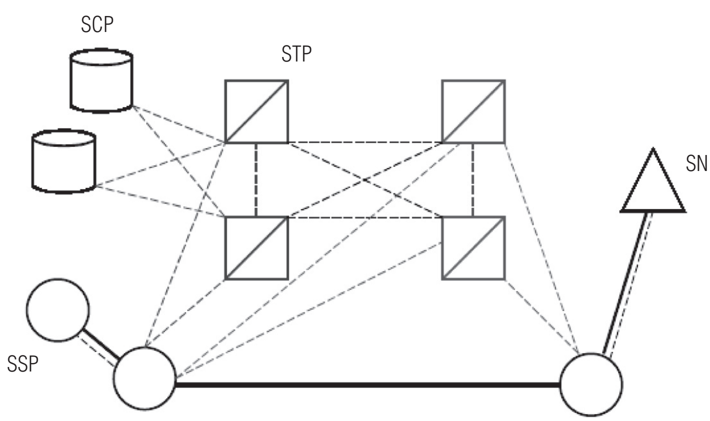


Figure 20.24 SS7 Architecture

20.12.1 Mobile networking

The above illustration depicts a SS7 network. The network elements are shown in Figure 20.25.

Each of these network elements has a specific role in the SS7 network.

- Service switching point (SSP): A telecommunications switch that contains the control logic (software) necessary to send/receive SS7 messages to other nodes in the network. If a mobile switching centre (MSC) has control logic, it is by definition considered an SSP.
- Signal transfer point (STP): This is the "heart" of the SS7 network. STP routes messages between other network elements.
- Service control point (SCP): This is the "brain" of the SS7 network. The SCP is nothing more than a database. However, utilization of a SCP offers profound enhancements for service delivery and network control.
- Service node (SN): Includes database functionality of the SCP along with additional capabilities such as voice interaction and control of voice resources. Generally speaking, SCPs work well with requirements that call for voluminous data transactions. SNs, on the other hand, are typically not designed for high volume data processing. Instead, SNs are best suited for special circumstance call processing involving voice resources and/or interaction.

SS7 involves two different types of signalling: connection oriented signalling and connectionless oriented signalling. Connection oriented signalling refers to the establishment of switch-to-switch facilities call inter-office trunks. These trunks carrier voice communications, The ISDN user part (ISUP) of the SS7 protocol is utilized to establish trunks between switches. In contrast, the transaction capability application part (TCAP) is utilized for connectionless signalling which typically entails switch-to-database or database-to-database communications. An example of connectionless signalling is TCAP signalling of HLR to VLR.

20.12.2 Mobile networking standards

There are two major types of inter-system signalling for mobile communications: **GSM mobile application part (MAP)** and **ANSI-41**. GSM MAP is the standard utilized for GSM, and ANSI-41 is the inter-system standard for all other mobile networks including CDMA, D-AMPS (IS-136), and AMPS.

Although the two standards have their differences, both have certain key things in common. They support three network elements that are required for mobile communications: MSCs, HLRs, and VLRs.

Mobile switching centre

A MSC is a telecommunications switch deployed in mobile communications networks to provide call control, processing, and access to the PST (fixed) network.

Home location register

The HLR is a database that is maintained by a user's home carrier or the mobile operator from whom the user has initiated service. The HLR stores information about the user, including the user profile (preferences), account status, features, and capabilities.

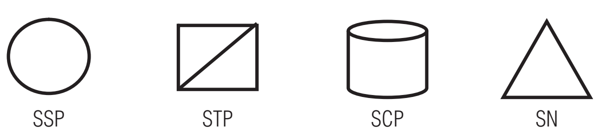


Figure 20.25 SS7 network elements

Visiting location register

The VLR is another database and is used by the serving carrier system to manage service requests from mobile users who are away from their home system as shown in Figure 20.26. The interaction between the MSC, HLR, and VLR is best understood by walking through the registration process.

In Figure 20.26, the mobile user's home system is depicted as System A and the mobile user is currently roaming in System B, the visited or serving system.

Step one: Detecting a potential user

The first step involves the serving system detecting your mobile phone. This occurs over the radio interface. Each mobile phone emits a unique identification that is detected by the RF equipment of the serving system. The mere fact that you have your mobile phone "powered on" triggers the serving system RF equipment to inform the equipment switching of your presence. The switching equipment in turn queries a database, which first determines whether you are in your "home" area or whether you are a "visitor." In this context, visitor means that you are not in your normal home location (the city/area where you signed up for service).

Step two: Exchanging the appropriate information about the user

The second step involves database interaction to determine appropriate handling of call requests. If you are in your home area, the HLR provides information necessary to handle requests for either call origination (making a call) or termination. If you are not in your home area, a VLR must request information from the HLR so that the visited (serving) system can process calls appropriately. Communication between the VLR in the serving system and the HLR of the home area is facilitated by mobile networking protocols and signalling based on SS7. In GSM networks, GSM MAP mobile networking protocol rides on top of SS7, allowing VLR to HLR (and HLR to VLR) communications.

This signalling and database communications typically occurs before any call is either placed or received, allowing the serving system to know exactly how to handle calls when a call is placed or received.

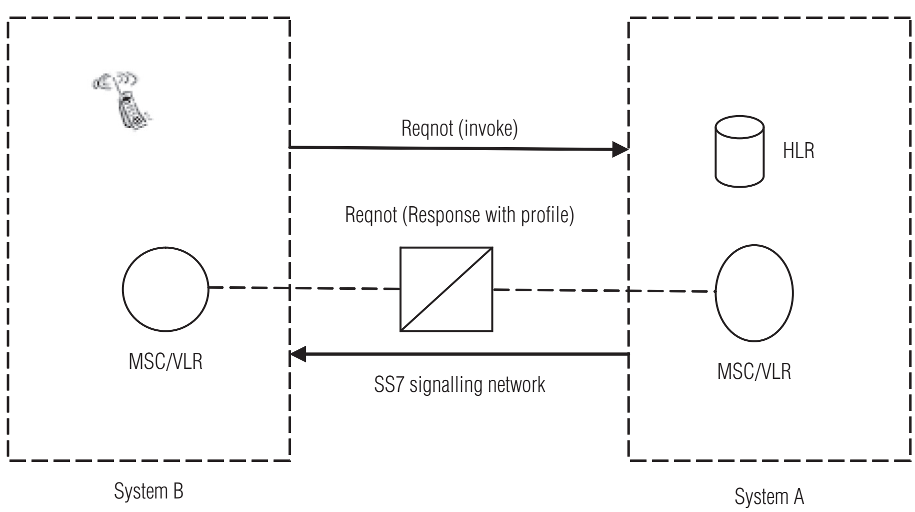


Figure 20.26 Mobile user home system

Step three: Handling calls

By the time the user either places or receives a call, this database interaction has occurred. The serving system now knows whether the mobile phone is associated with an account in good standing (bills all paid), user feature/service subscription, and the location of the user so that calls placed to the user may be delivered.

Placing a call

Upon detecting a request for a mobile phone user to originate a call, the serving system reviews the VLR record (established in step two above) to determine appropriate treatment. This allows the serving system to determine if the call is allowed and if any additional features/services should be made available. Once this investigation occurs (in a fraction of a second), the switch processes the call attempt as appropriate.

Receiving a call

When someone calls a mobile phone user, the home switch reviews the HLR. If the user is in the home area, the call is delivered immediately. If the mobile user is in another serving area (called “roaming”), the HLR record indicates which VLR is currently maintaining the mobile user’s records. The HLR uses SS7 and the appropriate mobile networking protocol to request delivery instructions from the VLR. The VLR provides these instructions to the HLR allowing the home switch to deliver the call to the serving switch and seamlessly terminate the call to the mobile phone as if the user were in the home area.

Mobile networking and applications

ANSI-41 and GSM MAP support various applications including basic services such as call waiting and conference calling. This support is provided in the form of profile updates between the serving system (VLR) and the home system (HLR). The HLR updates the VLR regarding subscriber services, allowing the serving system to provide appropriate services to the visitor.

The solved problems that follow use basic formulae pertaining to the field of satellite communications.

20.13 Summary

- A MWL between two remote locations with terrestrial network is not possible because it is not available all over the world.
- A communication satellite in GEO provides communication between two remote locations on the earth. But a large costly terminal is required to make a phone call from a location that does not have any terrestrial wireline or wireless coverage.
- To support a wide range of services and to provide superior service quality, constellations of satellites operating in LEO or MEO are needed. With the global mobile satellite systems, it is possible to use a single mobile wireless phone anywhere in the world. This chapter describes the satellite orbits such as LEO, MEO, HEO, and GEO.
- A number of satellite systems are being used for mobile services. These systems include Iridium, Globalstar, ICO, and Teledesic.
- First three systems support voice, fax, and messaging services for mobile communication subscribers and the Teledesic system supports high speed data and multimedia services.
- The satellite systems Iridium, Globalstar, and Teledesic use LEO constellations and the ICO system uses MEO constellation.

- In the Iridium system, ISLs are used to route each call to the earth station closest to the origination or destination of the call. Unlike Iridium system, Globalstar system does not use ISLs.
- It depends on large number of interconnected earth stations or gateways for efficient call routing and delivery over the terrestrial network.
- ICO's primary target customers are users from the existing terrestrial cellular systems who expect to travel to locations in which coverage is unavailable or inadequate.
- Service switching oint (SSP): A telecommunications switch that contains the control logic (software) necessary to send/receive SS7 messages to other nodes in the network. If a MSC has said control logic, it is by definition considered an SSP.
- Signal transfer point (STP): This is the "heart" of the SS7 network. STP routes messages between other network elements.
- Service control point (SCP): This is the "brain" of the SS7 network. The SCP is nothing more than a database. However, utilization of a SCP offers profound enhancements for service delivery and network control.
- Service node (SN): Includes database functionality of the SCP along with additional capabilities such as voice interaction and control of voice resources. Generally speaking, SCPs work well with requirements that call for voluminous data transactions. SNs, on the other hand, are typically not designed for high volume data processing. Instead, SNs are best suited for special circumstance call processing involving voice resources and/or interaction.

Example problem 20.1

Determine the orbital velocity of a satellite moving in a circular orbit at a height of 200 km above the surface of earth given that gravitation constant, $G = 6.67 \times 10^{-11}$ N-m^2/ kg, mass of earth, $M = 5.98 \times 10^{24}$ kg, radius of earth, $R_e = 6{,}372$ km.

Solution

Given that $R = 6{,}372$ km
$M = 5.98 \times 10^{24}$kg
$G = 6.67 \times 10^{-11}$ kg
$H = 200$ km

The orbital velocity (v) is given by

$$v = \sqrt{\mu/(R+H)}$$

where μ = GM

$\mu = 6.67 \times 10^{-11}\text{kg} \times 5.98 \times 10^{24}\text{kg} = 39.8 \times 10^{13}$ kg

By substituting the values in the above equation, we get

$$v = \sqrt{\frac{39.8\times10^{13}\,\text{kg}}{(6372+200)\times10^{3}}}$$

$$v = 7.89 \text{ km/s}$$

Example problem 20.2

A satellite in an elliptical orbit has an apogee of 30,000 km and a perigee of 1,000 km. Determine the semi-major axis of the elliptical orbit.

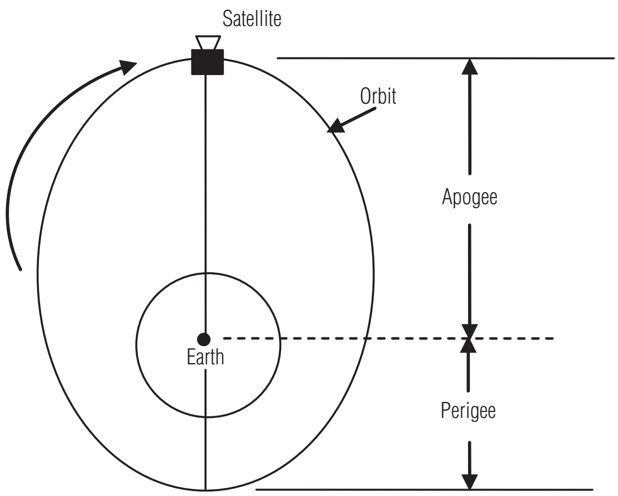
Solution

Given that Apogee = 30,000 km
Perigee = 1,000 km
The semi-major axis is calculated as

$$\text{Semi-major axis} = \frac{\text{apogee} + \text{perigee}}{2}$$

$$\text{Semi-major axis} = \frac{30,000 + 1,000}{2} = 15,500 \text{ km}$$

Example Figure 1 explains the problem.



Example Figure 1

Example problem 20.3

A satellite moving in an elliptical eccentric orbit has the semi-major axis of the orbit equal to 16,000 km (see Example Figure 2). If the difference between the apogee and the perigee is 30,000 km, determine the orbit eccentricity.

Solution

Given that Difference between apogee and perigee = 30,000 km
Semi-major axis of the ellipse = 16,000 km
Apogee = $a(1 + e)$
Perigee = $a(1-e)$

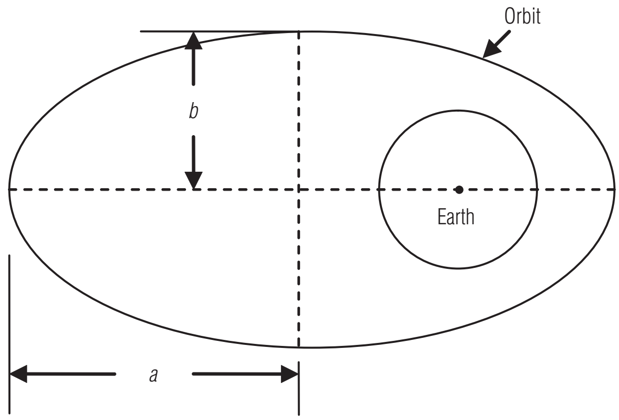
where a is the semi-major axis of the ellipse
e is the eccentricity

$$\text{Apogee} - \text{Perigee} = a(1 + e) - a(1 - e) = 2ae$$

or

$$\text{Eccentricity, } e = \frac{\text{Apogee} - \text{Perigee}}{2a}$$

$$e = \frac{30{,}000}{2 \times 16{,}000} = 0.93$$



Example Figure 2

Example problem 20.4

The farthest and the closest points in a satellite's elliptical eccentric orbit from earth's surface are 30,000 km and 200 km, respectively. Determine the apogee, the perigee, and the orbit eccentricity. Assume the radius of earth to be 6,370 km.

Solution

Apogee = 30,000 + 6,370 = 36,370 km
Perigee = 200 + 6,370 = 6,570 km

$$\text{Eccentricity, } e = \frac{\text{Apogee} - \text{Perigee}}{2a}$$

where

a is the semi-major axis of the elliptical orbit

$$\text{Also,} \quad a = \frac{\text{Apogee} + \text{Perigee}}{2}$$

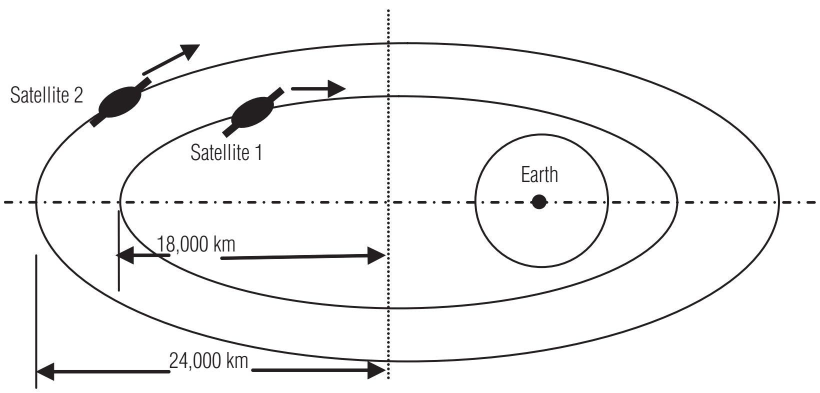
Or
$2a$ = Apogee + Perigee
Therefore,

$$\text{Orbit eccentricity} = \frac{\text{Apogee} - \text{Perigee}}{\text{Apogee} + \text{Perigee}}$$

$$= \frac{36370 - 6570}{36370 + 6570} = 0.693$$

Example problem 20.5

Satellite 1 in an elliptical orbit has the semi-major axis equal to 18,000 km and Satellite 2 in an elliptical orbit has a semi-major axis equal to 24,000 km, shown in the Example Figure 3. Determine the relationship between their orbital periods.



Example Figure 3

Solution

The orbital time period (T) is given by

$$T = 2\pi\sqrt{\frac{a^3}{\mu}}$$

where μ is the GM
  G is the earth's gravitational constant
  M is the mass of earth
  a is the semi-major axis of ellipse

If (a_1) and (a_2) are the values of the semi-major axis of the elliptical orbits of the satellites 1 and 2, (T_1) and (T_2) are the corresponding orbital periods, then

$$T_1 = 2\pi\sqrt{\frac{a_1^{\,3}}{\mu}}$$

$$T_2 = 2\pi\sqrt{\frac{a_2^3}{\mu}}$$

$$\frac{T_2}{T_1} = \left(\frac{a_2}{a_1}\right)^{3/2}$$

$$= \left(\frac{24000}{18000}\right)^{3/2} = 1.54$$

Thus, orbital period of Satellite 2 is 1.54 times the orbital period of Satellite 1.

Example problem 20.6

A satellite is moving in a near earth circular orbit at a distance of 640 km. Determine its orbital period. (Assume R = 6,360 km)

Solution

We know that $G = 6.67 \times 10^{-11}$ kg
$M = 5.98 \times 10^{24}$ kg
Given that R = 6,360 km
H = 640 km
The orbital velocity is given by

$$\text{Orbital velocity} = \sqrt{\frac{GM}{(R+H)}}$$

$$= \sqrt{\frac{6.67\times10^{-11}\times5.98\times10^{24}}{(6360+640)\times10^{3}}}$$

$$= \sqrt{\frac{39.8\times10^{13}}{7\times10^{6}}}$$

$$= 7.54 \text{ km/s}$$

$$\text{Orbital period} = \frac{2\pi\,(R+H)}{V}$$

$$= \frac{6.28\times7000}{7.54} = 5830\,\text{s} = 1\,\text{h}\ 37\,\text{min}$$

Example problem 20.7

Find the mass of the sun using the distance between the earth and the sun (1.496×10^{11} m) and the period of the earth's orbit (3.156×10^{7} sec). Apply Kepler's third law to find the mass of the sun.

Solution

Using equation

$$T^2 = \frac{4\pi^2}{GM}a^3$$

$$M_{sun} = \frac{4\pi^2}{GT^2}a^3$$

$$= \frac{4\pi^2(1.496\times10^{11})^3}{(6.67\times10^{-11})(3.156\times10^{7})}$$

$$= 1.99\times10^{30}\,\mathrm{kg}$$

Review questions

1. List the important features of the satellites in GEO.
2. What are the advantages and disadvantages of the satellites in LEO?
3. Compare the characteristics of LEO, MEO, HEO, and GEO satellites.
4. Write short notes on global mobile satellite systems.
5. Explain the constellation of the Iridium system.
6. What are the frequency bands used for uplink and downlink in the Globalstar system?
7. Explain the architecture of the ICO system.
8. Compare the Iridium, Globalstar, ICO, and Teledesic systems.
9. Explain the architecture of the Teledesic system.
10. Explain the call flow in the Iridium system from ISU to ISU.
11. List the frequencies used for uplink and downlink in the Teledesic system.
12. Explain the constellation of the Globalstar system.

Objective type questions and answers

1. Which of the following satellite orbit requires more number of satellites for global coverage?
   (a) LEO (b) MEO (c) HEO (d) GEO
2. How many satellites are required for global coverage in GEO orbits?
   (a) 3 (b) 6 (c) 66 (d) 10
3. Which of the following satellite systems requires less transmitter power?
   (a) LEO satellites (b) MEO satellites (c) HEO satellites (d) GEO satellites
4. Which of the following satellite systems have elliptical orbits?
   (a) LEO satellites (b) MEO satellites (c) HEO satellites (d) GEO satellites
5. What is the orbital period of geostationary satellite?
   (a) 2 h (b) 10 h (c) 24 h (d) 12 h
6. Which of the following satellite systems have smaller footprint?
   (a) LEO satellites (b) MEO satellites (c) HEO satellites (d) GEO satellites
7. Which of the following system provides high speed data and multimedia services?
   (a) Iridium (b) Globalstar (c) ICO (d) Teledesic

8. In which satellite system, the round trip delay is more?
   (a) Iridium (b) Globalstar (c) ICO (d) Teledesic
9. How many orbital planes are there in Teledesic system?
   (a) 2 (b) 6 (c) 12 (d) 8
10. Which of the following system belongs to the medium earth orbit?
   (a) Iridium (b) Globalstar (c) ICO (d) Teledesic
11. The orbital period of the following satellites is equal to earth's rotation period
   (a) LEO satellites (b) MEO satellites (c) HEO satellites (d) GEO satellites
12. Satellite visibility time or the station keeping time is more in the following system
   (a) Iridium (b) Globalstar (c) ICO (d) Teledesic
13. The orbital altitude of the following satellite system is more in
   (a) Iridium (b) Globalstar (c) ICO (d) Teledesic
14. Number of satellites in the Iridium system constellation are
   (a) 12 (b) 288 (c) 66 (d) 48
15. Number of satellites in the Globalstar system constellation are
   (a) 12 (b) 48 (c) 66 (d) 288
16. GEO satellites appears to be fixed with reference to a location on the earth's surface because
   (a) orbital period of the GEO satellites is twice the earth's rotation period
   (b) orbital period of the GEO satellites is equal to the earth's rotation period
   (c) orbital period of the GEO satellites is half of the earth's rotation period
   (d) orbital period of the GEO satellites is thrice the earth's rotation period
17. Satellite's orbital period is function of
   (a) satellite velocity and altitude
   (b) satellite inclination angle and altitude
   (c) satellite signal power and velocity
   (d) none of the above
18. In an elliptical orbit, the relation between semi-major axis, apogee, and perigee is
   (a) semi-major axis = (apogee + perigee)/2
   (b) semi-major axis = (apogee – perigee)/2
   (c) apogee + perigee
   (d) apogee – perigee
19. In the Iridium system, the frequency band used for uplink (Iridium phone to satellite) is
   (a) X-band (b) C-band (c) S-band (d) L-band
20. The frequency band used for downlink (Satellite to Globalstar Phone) in the Globarstar system is
   (a) X-band (b) C-band (c) S-band (d) L-band

Answers: 1. (a), 2. (a), 3. (a), 4. (c), 5. (c), 6. (a), 7. (d), 8. (c), 9. (c), 10. (c), 11. (d), 12. (c), 13. (c), 14. (c), 15. (b), 16. (b), 17. (a), 18. (a), 19. (d), 20. (c).

Open book questions

1. How are the HLR and VLR used?
2. Write about PACS frame structure. (Refer Figure 20.20)

Key equations

1. The weighting coefficients, α_m, are normalized to the output signal power of the correlator

$$Z' = \sum_{m=1}^{M} \alpha_m Z_m$$

2. The orbital velocity (v) is given by

$$v = \sqrt{\frac{\mu}{(R+H)}}$$

3. The semi-major axis is calculated as

$$\text{Semi-major axis} = \frac{\text{apogee} + \text{perigee}}{2}$$

4. Eccentricity is given by

$$e = \frac{\text{apogee} - \text{perigee}}{2a}$$

5. The orbital time period (T) is given by

$$T = 2\pi\sqrt{\frac{a^3}{\mu}}$$

Further reading

Brown, T. X., and Mohan, S. "Mobility Management for Personal Communications Systems," *IEEE Transactions on Vehicular Technology*, Vol. 46, No. 2, May 1997, 269–278.

GSM Specification Series 8.01–8.60, "BSS-MSC Interface, BSC-BTS Interface." GSM Specification Series 9.01–9.11, "Network Interworking, MAP."

Kaaranen, H., Ahtiainen, A., Laitinen, L. *UMTS Networks: Architecture, Mobility and Services*. John Wiley & Sons, June 2004.

Lam, D., Cox, D. C., and Widom, J. "Teletraffic Modeling for Personal Communications Services," *IEEE Communications Magazine*, February 1997, 79–87.

Lin, Y-B., and Chlamtac, I. *Wireless and Mobile Network Architecture*. John Wiley & Sons, 2001.

Logsdon T. *Mobile Communication Satellites-Theory and Applications*. New York: McGraw-Hill, 1995

Markoulidakis, G., et al. "Mobility Modeling in Third-Generation Mobile Telecommunications Systems," *IEEE Personal Communications*, Vol. 4, August 1997, 41–56.

McNair, J., Akyildiz, I. F., and Bender, M. An Intersystem Handoff Technique for IMT-2000 System. Tel Aviv, Israel: Infocom00. March 2000.

Rajpandya, *Mobile and Personal Communication Systems and Services*. 2000, PHI Publications.

Rao G. S., "*Global Navigation Satellite Systems with Essentials of Satellite Communications*", McGraw-Hill, New Delhi, India, May 2010.

Rappaport, T. S. *Wireless Communications Principles and Practice*, 2nd Edition, January 2002, PHI Publications.

Schiller, T. *Mobile Communications*, 2nd Edition, Pearson Education, September 2003.

Sharma, S. *Wireless & Cellular Communications*, S. K. Kataria & Sons publications, 2009.

www.ee.survey.ac.uk/personal/wood/constellations/

www.searchnetworking.com
www.cellular.co.za
www.iridium.com
www.satobs.org
http://en.wikipedia.org/wiki/Globalstar

Wireless Generations Technologies up to 3G 21

21.1 Introduction

For understanding the complex 3G mobile-communication systems today, it is also important to understand where they came from and how cellular systems have evolved from an expensive technology for a few selected individuals to today's global mobile-communication systems used by almost half of the world's population. Developing mobile technologies has also changed from being a national or regional concern to becoming a very complex task undertaken by global standards-developing organizations such as the *Third Generation Partnership Project* (3GPP) and involving thousands of people.

First generation (1G) mobile systems were introduced in early 1908s and designed to offer a single service, that is, speech only. 1G mobile systems usually offered handover and roaming capabilities but the cellular networks were unable to interoperate between countries. Another disadvantage of 1G mobile systems is that the base station and the mobile stations have to transmit at higher powers in order to communicate, thereby making mobile handsets infeasible.

With the emergence of digital communications, second generation (2G) mobile systems were introduced in the end of 1980s, supporting both data services (low bit-rate) and conventional voice services. In 2G mobile systems, the notion of *frequency reuse* was introduced to increase the system capacity. One well-known system is the Global System for Mobile Communications (GSM) introduced in Europe. New technologies have been developed based on the original GSM system, bringing about some more advanced systems known as 2.5 Generation (2.5G) systems. It is so called because 2.5G extended 2G with data service and packet switching capabilities, bringing the Internet into mobile personal communications. 2G was designed from the beginning as an evolving platform, from which emerged the High Speed Circuit-Switched Data (HSCSD), the General Packet Radio System (GPRS), and the Enhanced Data rates for GSM Evolution (EDGE).

With the rising demand of mobile communications, the third generation (3G) mobile systems were emerged by providing higher date rate (up to 384 Kbps) to facilitate real-time applications for example video, and new applications such as location-based services and wireless Internet access. 3G mobile system will be a significant step forward in the convergence of telecommunication and data communication industries, and the 3G mobile system refers to a family of new air interfaces multiple access networks. By exploiting Code Division Multiple Access (CDMA) techniques, 3G was developed in the late 1990s and is now being deployed globally.

3G systems are referred to as Universal Mobile Telecommunications System (UMTS) in Europe and International Mobile Telecommunications-2000 (IMT-2000) worldwide. Two main *standards **Wideband CDMA** and **CDMA2000** are selected as an interfaces multiple access techniques for the European-backed UMTS and IMTS-2000 systems respectively.*

2.5G networks currently offer true data speeds up to 28 Kbps. In comparison, the theoretical speed of 3G can be up to 2 Mbps, that is, approximately 200 times faster than previous 2G networks. This added speed and throughput will make it possible to run applications such as streaming video clips. It is anticipated that 4G speeds could be as high as 100 Mbps. At present, it is important for industries to develop a strong 3G offering that is palatable for the general public. Equally as important, industries must ensure that expectations are realistic and that services meet and exceed those expectations. The cellular evolution path from 1G to 3G and their technologies are discussed in this Chapter.

21.2 First generation (1G)

In mobile phone industry, as a general model of 10 years time is being taken for a generation change from one generation to the next generation. The 1G used analogue transmission techniques for traffic, which was almost entirely voice. It used FDMA technology to achieve radio communications. With the FDMA, the voice channels are carried by different radio frequencies. The technology and features of 1G are shown in Table 21.1.

21.2.1 1G standards

The most successful examples of 1G standards were *Nordic Mobile Telephone* (NMT), *Total Access Communications System* (TACS), and *Advanced Mobile Phone Service* (AMPS). For example, a total of 50 MHz in the band 824–849 MHz and 869–894 MHz is allocated for AMPS. This spectrum is divided into 832 frequency channels (416 downlinks and 416 uplinks).

Drawbacks in 1G: Analogue modulation is sensitive to interference from other users in the system, and the voice quality is quite vulnerable to various kinds of noise. International roaming was not possible in the 1G system due to different nations used different frequency bands and schemes. As a consequence of these, other means of capacity improvement such as efficient modulation schemes were sought for the second generation.

21.3 Second generation (2G)

The cellular industry introduced the second generation (2G) of mobile telephony 10 years later. The 2G services were launched in 1992 and since then they have been expanding and evolving continuously. The technology and features of 2G are shown in Table 21.2.

Table 21.1 Technology and features of 1G

Generation	Technology	Multiple access/ duplex	Bandwidth	Features
1G	AMPS, NMT, TACS	FDMA/FDD	9.6 Kbps	• Analog transmission • Voice service only • No data capability

Table 21.2 Technology and features of 2G

Generation	Technology	Multiple access/ duplex	Modulation	Bandwidth	Features
2G	GSM IS-95 based on CDMA	TDMA/FDD CDMA/FDD	GMSK QPSK	9.6–14.4 Kbps	Digital voice role transmission Global roaming Circuit switched data capability

The 2G mobile systems use digital radio transmission for traffic. The difference between 1G and 2G is 1G used analogue signalling while the latter used digital signalling. Thus, the boundary line between 1G and 2G systems is obviously the analogue/digital split.

21.3.1 Advantages of digital system use in 2G

- Digital data can be compressed and multiplexed much more effectively than analogue voice encodings.
- Digital cellular systems have many features, such as improved communication quality due to various digital signal processing technologies.
- Digital technology enables the use of signal processing techniques to increase robustness against interference.
- It also reduces the spectral bandwidth required for each user and hence provides higher capacity.
- 2G provides about 3–4 times the capacity of the 1G without adding new base stations. One frequency channel is simultaneously divided among several users (either by code or time division).
- Since digital systems are more immune to noise, a Signal to Interference Ratio (SIR) of 7 dB could be tolerated for a digital system, whereas 15 dB is required for the analogue systems under the same circumstances.
- Digital voice encoding allows for lower powered radio signals that require less battery power. This is possible because of CODEC introduction which encodes and decodes digital data stream or signal (Figure 21.1).
  - The digital voice encoding also allows digital error checking (increases sound quality and lowers the noise level).
  - Going to “all-digital” allowed for the introduction of digital data transfer (SMS–“short message service” and e-mail).

21.3.2 Disadvantages of digital system use in 2G

- Cellular towers had a limited coverage area.
- Abrupt dropped calls due to uneven decay curve.
- Spotty coverage, which is shown in Figure 21.2.

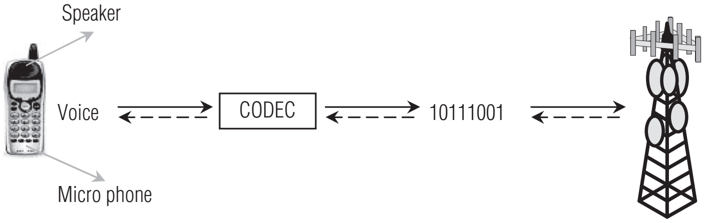


Figure 21.1 Digital voice encoding

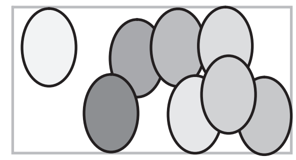


Figure 21.2 Spotty coverage

21.3.3 2G standards

Various 2G cellular systems were developed in the 1990s. There are four main standards for 2G systems:

- *Global System for Mobile* (GSM) communications and its derivatives.
- *Digital AMPS* (D-AMPS).
- *Code Division Multiple Access* (CDMA) IS-95 and
- *Personal Digital Cellular* (PDC).

The analogue cellular systems (1G) in each country used different frequency bands and schemes, which made interconnection impossible across national borders.

In 1982, the European Conference of Postal and Telecommunications Administrations (CEPT) established the Group Special Mobile, and development efforts were carried out under the leadership of the European Telecommunications Standards Institute (ETSI). GSM-based services were launched in 1992. The 2G standards discussed earlier can be divided into two groups:

- TDMA-based 2G standards.
- IS-95 standard-based CDMA technology.

21.4 TDMA-based 2G standards

Most of the 2G systems employ Time Division Multiple Access (TDMA), such as the Global System for Mobile (GSM) Communications, *Digital AMPS* (D-AMPS), the Digital Cellular System 1800 (DCS1800) in Europe, the Interim Standard (IS) 54 in the USA, and the Personal Digital Cellular (PDC) system in Japan.

With TDMA, the time axis is subdivided into different non-overlapping time slots where each user is exclusively assigned each time slots in which this user employs the total available bandwidth.

Thus, TDMA separates the users in the time domain. In practice, TDMA is combined with FDMA to reduce the hardware complexity of an otherwise extremely broadband system and to increase the flexibility of the system.

21.4.1 Global system for mobile (GSM) communications

GSM is the most successful and widely used 2G system. Originally designed as a pan-European standard, it was quickly adopted all over the world. GSM is widely used in Europe, Australia, and

Asia. In GSM, every frequency carrier is divided into fixed time slots that support up to eight voice channels. The speech coding rate is 13 Kbps in GSM. With TDMA, the radio hardware in the base station can be shared among multiple users. Only in America, the GSM has not reached a dominant position yet. In North America, Personal Communication System-1900 (PCS-1900; a GSM derivative, also called GSM-1900) has gained some ground, and in South America, Chile has a wide-coverage GSM system. The GSM standard operates at a different set of frequencies (uplink and downlink channels) worldwide and mainly at 850, 900, 1,800, and 1,900 MHz. The forward channel (downlink) and reverse channel (forward link) ranges of GSM 00 are 935–960 MHZ and 890–915 MHz, respectively.

21.4.2 Digital AMPS (D-AMPS)

With TDMA, the radio hardware in the base station can be shared among multiple users. In North America, however, the main design objective has been to make a smooth transition from the low-capacity analogue systems to high-capacity digital systems. This is possible since digital technology enables allocation of three TDMA channels on the same radio frequency as one FDMA channel in the AMPS system. Such a mixed system, known as ADC (D-AMPS or IS-54 standard), enhances the capacity of the system three times just by exchanging the analogue FDMA transceivers to digital TDMA transceivers. The speech coding rate is 7.95 Kbps in ADC. PDC and ADC systems have high modulation efficiency due to the use of QPSK modulation and low bit rate codec. Therefore, both systems have more system capacity than GSM.

21.4.3 Personal digital cellular (PDC)

PDC is the Japanese 2G standard. Originally it was known as *Japanese Digital Cellular* (JDC), but the name was changed to *Personal Digital Cellular* (PDC) to make the system more attractive outside Japan. However, this renaming did not bring about the desired result, and this standard is commercially used only in Japan. The specification is known as RCR STD-27, and the system operates in two frequency bands: 800 MHz and 1,500 MHz. It has both analogue and digital modes. Its physical layer parameters are quite similar to D-AMPS, but its protocol stack resembles GSM. PDC has been a very popular system in Japan. This success has also been one of the reasons that the Japanese have been so eager to develop 3G systems as soon as possible, as the PDC system capacity is quickly running out.

21.5 IS-95 (Code division multiple access (CDMA) or cdmaOne standard)

Parallel to the TDMA-based 2G standards, a new standard employing **CDMA** as multiple access technology known as IS-95, has been developed in North America which claims to have many advantages over TDMA technology, including improvement of capacity up to 10–12 times over the analogue systems. IS-95 standard introduced the CDMA technique in 1993.

A number of terms are used to refer to CDMA implementations. The original U.S. standard defined by QUALCOMM was known as IS-95, the IS referring to an Interim Standard of the Telecommunications Industry Association (TIA). IS-95 is often referred to as 2G or second generation cellular. The QUALCOMM brand name cdmaOne may also be used to refer to the 2G CDMA standard.

> The term ***cdmaOne*** is used to describe cellular networks based on IS-95A and IS-95B technology.

CDMA uses a different approach to air interface design. Instead of dividing a frequency carrier into short time slots as in TDMA, CDMA uses different codes to separate transmissions on the same frequency.

21.5.1 Modulation technique in CDMA

QPSK (Quadrate Phase Shift Key) is the form of digital modulation used in CDMA. The QPSK carrier will cycle through four output phases, with all transitions being allowed, including through the origin. Like π/4 Differential Quadrature Phase Shift Keying (DQPSK), QPSK does not use a constant signal envelope, so amplitude will change as the phase shifts from one state to another. Note that the symbol/phase correlation is not based on changes in phase (e.g., differential) but absolute phase position (i.e., each vector represents a specific symbol). Where π/4 DQPSK had four phase transitions, each representing 2 bits each, QPSK has four phase states, each again representing 2 bits each (Figure 21.3).

21.5.2 Multiple access in CDMA

The exploitation of the spread spectrum technique to enable multiple users a simultaneous access to the channel is called CDMA.

In CDMA, the frequency remains the same for all users in all cells in the network, and all users transmit at the same time (except in the case where an extra frequency is added, but the concept remains the same). Users are identified by particular code and can transmit and receive at any time (e.g., there is no time domain, other than synchronization and the use of logical channels).

With frequency reuse not being an issue at all, adding cells to the network becomes substantially easier. This makes filling in holes or adding extra cells to cover special events a simpler task. The separation of the user signals is performed in the code domain by applying user-specific matched filtering in the receiver, referred to as dispreading. The following are three user-specific spread spectrum principles that can be used in this standard:

- Direct sequence (DS)
- Frequency hopped (FH)
- Time hopped (TH) spreading

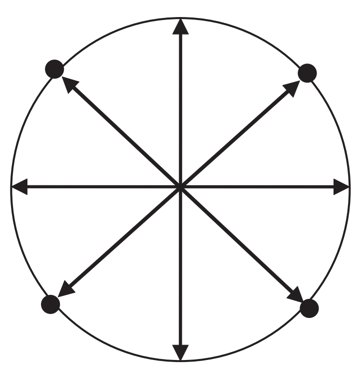


Figure 21.3 The block circles indicate the decision points in a QPSK signal. Each decision point represents 2 bits

With FH spreading and TH spreading, the user-specific spreading code is used to pseudo randomly hop the carrier frequency through a large bandwidth and to pseudo randomly hop the time slots through large time duration, respectively. With DS spreading, the user-specific spreading code introduces rapid phase transitions into the data stream, expanding the required bandwidth.

21.5.3 Cellular capacity in the CDMA standard

The primary design goal of the cellular technology is to increase the capacity. Shannon's equation is the reason why spread spectrum systems are becoming so popular:

$$\text{Capacity} = \text{Bandwidth} \times \log_2(1 + \text{Signal/Noise})$$

From the above equation, the *capacity* is related to the *S/N ratio* and the *bandwidth*. Because we are using the same frequency band throughout the entire network, it is easy to see how problems with power in one sector can easily interfere with overall capacity of the system. Because the same frequency is used throughout the entire system, other sectors and base stations will sound like noise if a phone is not set to listen to them. Also, remember that because of processing gain we can use a lower S/N ratio in spread-spectrum systems, while still maintaining the same quality of service.

21.5.4 Understanding the "code" in CDMA

As mentioned earlier, CDMA does not use frequency or time to separate users. Instead, it uses different digital sequences, or codes, and in some cases different timings on the same sequence. In CDMA, an often-used analogy is the "CDMA Cocktail Party." Imagine a lively party where everyone was speaking in a different language. Even with all of the people talking at the same time, you could still make out the person speaking your language, as long as there was only one person speaking it. Now, what are the catches to this concept? First, you would need to be located in the room where you would only clearly hear one person speaking your language, also, the other languages being spoken nearby would need to be substantially different than your own (e.g., if you speak English and someone nearby was speaking with a Scottish accent, and another near you was speaking with a New York accent, it would probably cause a problem as the languages [or codes] are not different enough from each other). Also, the overall noise in the room needs to be low enough, in relation to the level you receive from the person speaking your language (the S/N ratio). If the language was different enough, and the person spoke intentionally very slowly and clearly, you might be able to get away with a lower S/N ratio (processing gain).

Thus, to get separation we need to ensure the codes are different enough from each other. As discussed in Chapter 1, how well the two digital sequences compare to each other is called the correlation. Remember a correlation of one means the two sequences is completely alike. A correlation of zero means they are completely unlike exactly what we need for separation of users.

If two sequences have a correlation of zero, we call them orthogonal to each other. On the forward link, we separate users with 64 orthogonal codes, each being 64 bits long, called *Walsh functions* (Figure 21.4). Hence, Walsh code 0 is all zeros. Each user will be assigned one of these codes, and this code will separate everyone (as well as the overhead channels). The next logical question is: How do we use these orthogonal codes to separate?

- Repeat to the right and below.
- Inverse diagonally.
- Continue until 64 bits across and down.

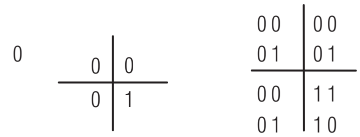


Figure 21.4 Walsh functions

In Figure 21.4, Walsh functions are uncorrelated to each other. Note that Walsh code 0 (will be the top sequence, from left to right) will end up being all zeros when the pattern is taken out 64 bits.

The principle behind spreading and de-spreading is that when a symbol is XORed with a known pattern, and the result is then XORed with a known pattern, the original data will be recovered. In cdmaOne, each symbol is XORed, or spread, with all 64 chips of the Walsh code. In Figure 21.5, the symbol of value one is spread with Walsh code 59, yielding a 64-chip representation of the symbol. In other words, for every one bit of data, we end up with 64 chips output. (Spread spectrum was discussed in Chapter 15.)

In Figure 21.5, we spread each symbol with an orthogonal code of four bits for illustrative purposes. Thus, we output four chips for every one symbol of user data. We do this by XORing the user data with the Walsh code. Remember, in the real implementation, the 4-bit Walsh function shown is actually 64 bits long. To recover the transmitted signal, we simply XOR with the correct Walsh code (the code used to originally spread) and then integrate the result.

If we were to use another Walsh code (which of course would be orthogonal to the actual code), you would see that it would not give us any usable data, and our system would immediately know that there is just noise on that channel. The second type of codes used extensively in CDMA is pseudo random noise (PN) codes, binary sequences that have the properties of randomness (i.e., equal numbers of zeros and ones, set number of runs). In CDMA, each information symbol will be spread with all 64 bits of the Walsh code sequence as shown in Figure 21.5.

We use three PN codes in CDMA: two *short codes* and one *long code*. Before we see how these codes are used, it is important to understand some of the traits of PN codes.

A very important feature of PN codes is that if we time shift the same versions of a PN code, we end up with two codes that are close to completely uncorrelated to each other (nearly

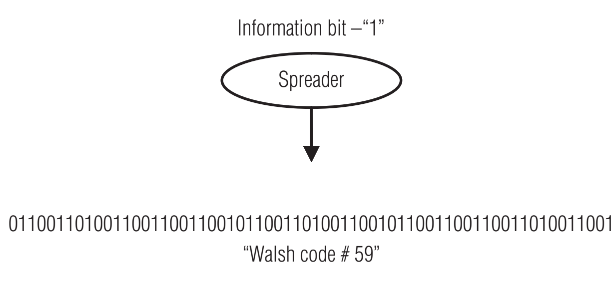


Figure 21.5 Representation of information bit "1" by the 64-chip sequence

1	0	0	1	1	User data
0110	0110	0110	0110	0110	4-bit Walsh function
1001	0110	0110	1001	1001	Result of spreading chips
0110	0110	0110	0110	0110	4-bit Walsh function
1111	0000	0000	1111	1111	Result of dispreading
1	0	0	1	1	Integrate-user data recovered

Figure 21.6 Process of spreading user data with a sequence (XORing), transmitting it, receiving it, and then dispreading (XORing) with the same sequence

orthogonal). Thus, we can use the same sequence and time shift the start of the sequence, ending up with nearly orthogonal sequences (Figure 21.6). The sequences that we use are all synchronized to specific triggers, which are derived from the GPS satellite system.

In order to offset a PN code to create nearly uncorrelated sequences, we use a masking system. The pattern will be the same, but the timing will be different. With different time offsets, each sequence will be nearly orthogonal to each other.

Remember the sequence remains the same, but it simply starts at a different point in the sequence. In the circuit shown, sequence of bits placed in the "mask registers" will offset the output sequence with a specific number of bits. The short PN sequence is 32,768 bits long, and we use increments of 64 bits to offset. This leaves us with 512 unique time offsets on the sequence. Also remember that because the short code is always transmitted at the CDMA data rate (1.2288 Mbps), the sequence always take 26.667 ms to completely cycle, regardless of the time offset.

As stated earlier, the start of the sequence is triggered by a unique trigger from GPS called the *even second clock*.

IS-95 is the only 2G CDMA standards so far to be operated commercially. It is used in the United States, South Korea, Hong Kong, Japan, Singapore, and many other East Asian countries. The forward channel (downlink) and reverse channel (forward link) range of IS-95 (CDMA) are 1,930–1,990 MHz and 1,850–1,910 MHz, respectively. CDMA-based multiple access systems are the most complex digital wireless systems. In CDMA systems, unlike separating users with different frequency channels as is the case with AMPS, or time slots as with TDMA systems, there is no limitation of frequency or time.

21.6 Two point five generation (2.5G)

"Generation 2.5" is a designation that includes all advanced upgrades for the 2G networks. The need for increased throughput data rates in data transfer (such as web browsing and e-mail) led to the evolution of 2.5G. These upgrades may in fact sometimes provide almost the same capabilities as the planned 3G systems. The boundary line between 2G and 2.5G is an unclear one. It is difficult to say when a 2G becomes a 2.5G system in a technical sense. In order for the evolution from 2G to 2.5G to be backwards compatible and cost bearable, 2.5G standard includes the following technologies:

- HSCSD
- GPRS
- EDGE
- *Interim Standard 95B* (IS-95B)

Table 21.3 Technology and features of 2.5G

2.5G		Technology	Modulation	Multiple access	Bandwidth	Features
	HSCSD	*High-Speed Circuit-Switched Data*	GMSK	TDMA	9.6–57.6 Kbps	Extension of GSM Higher data speeds
GSM	GPRS	*General Packet Radio Services*	GMSK	TDMA	9.6–115 Kbps	Extension of GSM Always on connectivity Packet–switched data
	EDGE	*Enhanced Data Rates for Global Evolution*	8-PSK	TDMA	64–384 Kbps	Extension of GSM Always on connectivity Faster than GPRS
IS-95B or CDMA 2000	CDMA	*High Data Rate (HDR), Code division multiple access*	QPSK	CDMA	64 Kbps–2.4 Mbps	Extension of cdma one or IS-95 Faster than 2.5 GSM standard

Generally, a *2.5G GSM system* includes at least one of the HSCSD, GPRS, and EDGE technologies. An IS-95 system is called 2.5G when it implements IS-95B, or CDMA2000 1xRTT upgrades. The technology and features associated with 2.5G are illustrated in Table 21.3.

Upgrade of 2G GSM to 2.5 GSM

The biggest problem with the 2G standard GSM is its low air interface data rates. The basic GSM could provide a maximum of 9.6 Kbps to 14.4 Kbps user data rate.

Anyone who has tried to web surf with these rates knows that it can be a rather desperate task. Various methods that are adopted for the upgradation of 2G GSM to 2.5 GSM are given in the following (Figure 21.7).

21.6.1 High-speed circuit-switched data (HSCSD)

HSCSD is the easiest way to speed things up. This means that instead of one time slot, a mobile station can use several time slots for a data connection. In current commercial implementations, the maximum is usually four time slots. One time slot can use either 9.6 Kbps or 14.4 Kbps speeds. The total rate is simply the number of time slots times the data rate of one slot. This is a relatively inexpensive way to upgrade the data capabilities, as it requires only software upgrades to the network (plus, of course, new HSCSD-capable phones), but it has drawbacks. The biggest

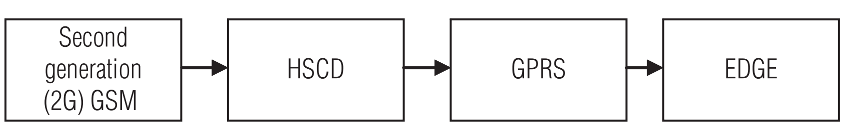


Figure 21.7 Upgrade of 2G GSM to 2.5 GSM

problem is the usage of scarce radio resources, because it is *circuit switched*, HSCSD allocates the used time slots constantly, even when nothing is being transmitted. In contrast, this same feature makes HSCSD a good choice for real-time applications, which allow for only short delays. The high-end users, which would be the most probable HSCSD users, typically employ these services in areas where mobile networks are already congested. Adding HSCSD capability to these networks certainly will not make the situation any better. An additional problem with HSCSD is that handset manufacturers do not seem very interested in implementing HSCSD. It can be seen that HSCSD will be only a temporary solution for mobile data transmission needs.

21.6.2 General packet radio services (GPRS)

The next solution is GPRS. With this technology, the data rates can be pushed up to 115 Kbps, or even higher if one can forget error correction. However, with adequate data protection, the widely quoted 115 Kbps is the theoretical maximum in optimal radio conditions with eight downlink time slots. A good approximation for throughput in "average" conditions is 10 Kbps per time slot. What is even more important than the increased throughput is that GPRS is packet switched, and thus it does not allocate the radio resources continuously but only when there is something to be sent.

The maximum theoretical data rate is achieved when eight time slots are used continuously. The first commercial launches for GPRS took place in 2001. GPRS is especially suitable for non-real-time applications, such as e-mail and web surfing. Also, bursty data is well handled with GPRS, as it can adjust the assigned resources according to current needs. It is not well suited for real-time applications, as the resource allocation in GPRS is contention based; thus, it cannot guarantee an absolute maximum delay.

21.6.3 Enhanced data rates for global evolution (EDGE)

EDGE was designed specifically as an upgrade to GPRS for integration into GSM network starting from the GSM community as a path to 3G. It uses the same basic GSM infrastructure with the difference being that it can use 8-PSK modulation in addition to the GMSK. It has nine different air interfaces called *Multiple Modulation Coding Schemes* (MCS), which is named from 1–9. Every MCS has a varying control protection and can use either GMSK for a low data rate (8.8–17.6 Kbps) or 8-PSK for a high data rate (22.4–59.2 Kbps) for each time slot. According to the level of error correction needed for the application every mobile user can adopt whatever MSC is suitable without the error protection and eight time slots taken when it theoretically connects with $8 \times 59.2 = 547.2$ Kbps. A minimum error control and network considerations limit the throughput at 384 Kbps. It requires new hardware (routers, gateways) and software updates at the base stations. Figure 21.8 presents the worldwide EDGE coverage.

The third 2.5G improvement to GSM is EDGE. Originally this acronym stood for Enhanced Data rates for GSM Evolution, but now it translates into Enhanced Data rates for Global Evolution as the EDGE idea can also be used in systems other than GSM. The idea behind EDGE is a new modulation scheme called *eight-Phase Shift Keying* (8PSK). It increases the data rates of standard GSM up to threefold. EDGE is an attractive upgrade for GSM networks, as it only requires a software upgrade to base stations if the RF amplifiers can handle the non-constant envelope modulation with EDGE's relatively high peak-to-average power ratio. It does not replace but rather coexists with the old *Gaussian Minimum Shift Keying* (GMSK) modulation, so mobile users can continue using their old phones if they do not immediately need the better service quality provided by the higher data rates of EDGE. It is also necessary to keep the old GMSK because 8PSK

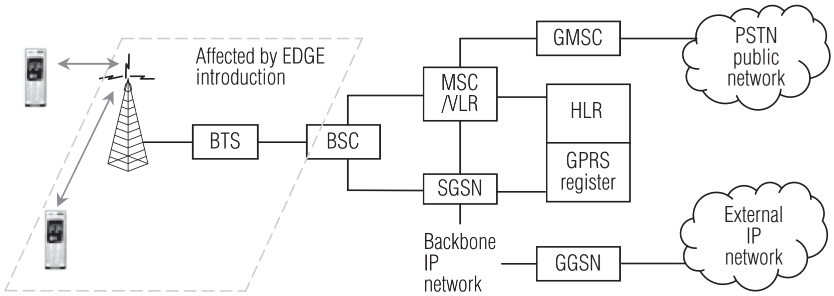


Figure 21.8 GSM/EDGE network structure

can only be used effectively over a short distance. For wide area coverage, GMSK is still needed. If EDGE is used with GPRS, then the combination is known as *enhanced GPRS* (EGPRS). The maximum data rate of EGPRS using eight time slots (and adequate error protection) is 384 Kbps. Note that the much-advertised 384 Kbps is thus only achieved by using all radio resources of a frequency carrier, and even then only when the mobile station is close to the base station. ECSD is the combination of EDGE and HSCSD, and it also provides data rates three times the standard HSCSD. A combination of these three methods provides a powerful system, and it can well match the competition by early 3G networks.

21.6.4 Interim standard 95B (IS-95B)

The IS-95 (CDMA) standard currently provides 14.4 Kbps data rates. It can be upgraded to IS-95B, which is able to transfer 64 Kbps with the use of multiple code channels. Interim Standard 95B (IS-95B) is the revision of IS-95 and IS-95A networks proposed in 1995.

> *The term **cdmaOne** is used to describe cellular networks based on IS-95A and IS-95B technology. cdmaOne provides packet and circuit switched data access through CDMA radio channels.*

The first IS-95B networks were originally deployed in September 1999 in Korea by dedicating eight different orthogonal user channels simultaneously, and theoretically it can achieve $8 \times 14.4 = 115.2$ Kbps. Due to this data speed the IS-95B is capable of reaching packet data service, therefore it is categorized as a 2.5G technology.

2G cdmaOne advantages

cdmaOne technology offers numerous benefits to the 2G cellular operators and their subscribers:

- Capacity increases from 8 to 10 times that of an AMPS analogue system and 4 to 5 times that of a GSM system.
- Improved call quality, with better and more consistent sound as compared to AMPS systems.
- Simplified system planning through the use of the same frequency in every sector of every cell.
- Enhanced privacy.
- Improved coverage characteristics, allowing for the possibility of fewer cell sites.
- Increased talk time for portables.
- Bandwidth on demand.

21.7 Third generation (3G) development

Need for 3G cellular systems: The need for high speed Internet access up to 2 Mbps, live video communications, fast web access, and the simultaneous data and voice transmission led to the development of 3G cellular networks. The new services have to be available in indoor and outdoor environments, where the integration of satellite links shall enable world-wide coverage. These requirements cannot be completely covered by the 2G systems that have a relatively low available bit rate in the range of 14.4 Kbps per user, primarily designed for speech transmission applications. Research activities concerning the standardization of the 3G mobile radio systems are in progress world-wide.

IMT-2000/UMTS

> *The International Telecommunications Union (ITU) created a project in 1998 for a common worldwide cellular standard under the name* ***International Mobile Telephone 2000 (IMT-2000).*** *IMT-2000 is also referred to as* ***Universal Mobile Telecommunications Systems (UMTS)*** *in Europe.*

The technology and features associated with the 3G are illustrated in Table 21.4.

21.7.1 Features of 3G mobile communications

1. With 3G, the information is split into separate but related packets before being transmitted and reassembled at the receiving end. Packet switched data formats are much more common than their circuit switched counterparts.
2. The World Wide Web (WWW) is becoming the primary communication interface. People access the Internet for entertainment, services, and information collection, the intranet for accessing enterprise information and connecting with colleagues, and the extranet for accessing customers and suppliers. These are all the derivatives of the WWW aimed at connecting different communities of interest. Information and other resources are being stored in remote Web servers, which serve the various needs of human beings through Web browsers at their ends.

Table 21.4 Technology and features of 3G

Generation		Technology	Multiple access/ duplex	Modulation	Bandwidth	Features
3G	IMT-2000	International mobile telecommunication 2000/	TDMA/FDD	GMSK	64–2,048	Always on Global roaming IP enabled
	UMTS	Universal mobile telecommunication	CDMA/FDD	QPSK		

3. Speeds of up to 2 Mbps are achievable with 3G. The data transmission rates will depend upon the environment; the call is being made in, however, only indoors and in stationary environments that these types of data rates will be available. For high mobility, the data rates of 144 Kbps are expected to be available.

21.8 3G Air interface technologies

Since the 2G cellular infrastructures already had different implementation variations, an important challenge for UMTS and IMT-2000 is the selection of an appropriate *multiple access scheme* to meet the demands of 3G mobile radio systems. The IMT-2000 standard accepts 5 possible radio interfaces based on 3 multiple access technologies (FDMA, TDMA, and CDMA) as illustrated in Figure 21.9. The five 3G air interface technologies are as follows:

- *Wideband CDMA* (WCDMA)
- *Code Division Multiple Access 2000* (cdma2000)
- *Time Division Synchronous Code Division Multiple Access* (TD-SCDMA)
- *Universal Wireless Communications-136* (*UWC-136*)
- *Digital Enhanced Cordless Telephone* (DECT)

Among the five air interfaces as shown in Figure 21.9, three air interfaces **WCDMA, CDMA2000,** and **TD-SCDMA** are the main interfaces and became popular.

- The first two interfaces ***WCDMA*** and ***CDMA2000*** are the variations of CDMA technology and are based on FDD.
- The third interface ***TD-SCDMA*** was proposed by the China and is based on TDD. It uses TDMA/TDD and CDMA techniques for the high-speed data transfer over GSM networks up to 384 Kbps. It has a 1.6 MHz radio channel and a 5 ms frame subdivided into seven time slots.
- The fourth interface ***UWC-136*** falls under the TDMA category, which is also known as EDGE.
- The fifth interface falls under the ***FD-TDMA*** category and is known as *DECT*. The DECT is mainly used in the indoor environments.

From the above explanation, it is clear that the overall infrastructure of 3G mostly includes WCDMA and CDMA techniques in terms of applicability and future potential.

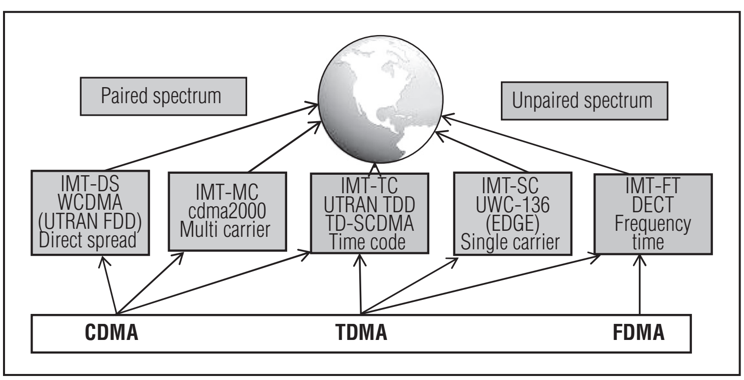


Figure 21.9 Five radio interfaces based on 3 technologies

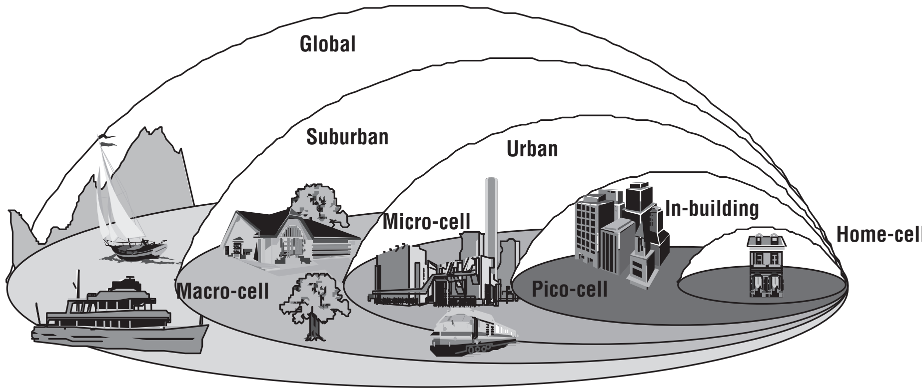


Figure 21.10 Different environments for UMTS

WCDMA-based UMTS will offer a common air interface covering home, office, car, train, aeroplane, or a pedestrian as shown in Figure 21.10. UMTS will integrate all the services offered by different mobile communication systems such as mobile telephone, cordless telephone, public air radio, satellite radio, etc in one service. It will allow users to roam during an existing connection between different types of communication networks. UMTS will offer broadband services, that is, it will be possible to transmit voice, text, data, and images over one connection.

21.9 3G spectrum

The channel spacings (bandwidths) used for making a call in 1G and 2G communication services are 30 kHz and 30–200 kHz, respectively. The amount of bandwidth needed for 3G services is 15–20 MHz, and you can see that there is as much as a 500 times increase in the amount of bandwidth required.

The 3G spectrum was first identified at World Administrative Radio Congress (WARC-92). Worldwide frequency spectrum identified for the 3G transmission is in the 1,885–2,025 MHz and 2,110–2,200 MHz bands. Of these 230 MHz of 3G spectrum, 2 × 30 MHz were intended for the satellite component of IMT–2000 and the rest for the terrestrial component. The main IMT-2000 standardization effort was to create a new air interface that would increase frequency usage efficiently. The 3G also identifies the paired and unpaired parts of the spectrum.

For example, in the IMT-2000 3G spectrum, the satellite service uses the bands 1,980–2,010 MHz for uplink and 2,170–2,200 MHz for downlink. This leaves the three bands 1,900–1,980 MHz, 2,010–2,025 MHz, and 2,110–2,170 MHz for terrestrial component.

Paired spectrum: Radio spectrum is often organized (and sold) as *paired spectrum* with a bit of spectrum in a lower frequency band and a bit of spectrum in an upper frequency band. Paired spectrum is often specified in a form like "2 × 15 MHz" meaning 15 MHz in a lower band and 15 MHz in an upper band. For the paired spectrum, the bands for uplink (mobile transmit) and downlink (base-station transmit) are identified for *Frequency Division Duplex* (FDD) operation.

The technique of two users talking to each other on two separate frequencies is called Frequency Division Duplex (FDD). WCDMA is an FDD technique.

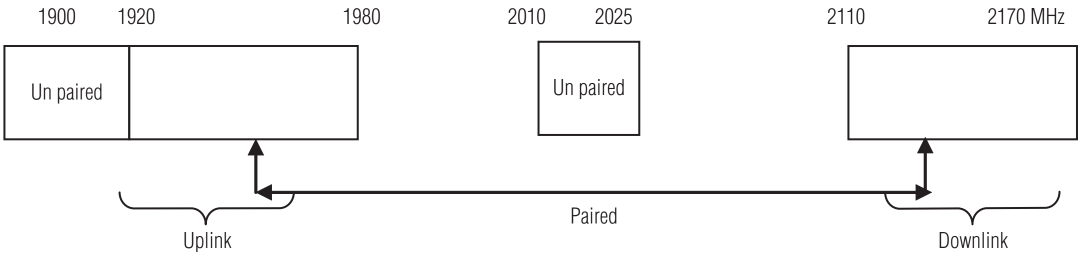


Figure 21.11 UMTS frequency band plan

Unpaired spectrum: The unpaired bands, for example, used for *Time Division Duplex* (TDD) operation. TD-CDMA is a TDD technique. In this duplex method, uplink and downlink transmissions are carried over by the same frequency band by using synchronized time intervals. Thus, time slots in a physical channel are divided into transmission and reception part. Note: The band that is most globally deployed for 3G is still 2 GHz.

The WCDMA air interface was selected for paired frequency bands (FDD operation) and TDCDMA (TDD operation) for unpaired spectrum.

Then, somewhat diverging arrangements are made between regions of the frequency bands assigned to 3G means that there is not even a single band that can be used for 3G roaming worldwide. The choice of frequency band for implementing in Europe and Asia is clear; for example, the UMTS. The 3G spectrum plan in Europe is shown in Figure 21.11.

We can observe from the earlier discussion that the UMTS FDD is designed to operate in paired bands, with uplink in the 1,920–1,980 MHz and downlink in the 2,110–2,170 MHz band. UMTS TDD was operated with the unpaired frequency bands 1,900–1,920 MHz and 2,010–2,025 MHz.

In case of the United States, these frequency bands are not available; they have already using these frequencies. Therefore, the three frequency bands suggested for implementation of UMTS in the United States are

- **806–890** MHz band
- **1,710–1,885** MHz band
- **2,500–2,690** MHz

21.10 Internet speeds of 2G, 2.5G, and 3G technologies

2.5G is the interim solution for the current 2G networks to have 3G functionality. 2.5G networks are being designed such that a smooth transition (software upgrade) to 3G can be realized. 2.5G networks currently offer true data speeds up to 28 Kbps. In comparison, the theoretical speed of 3G can be up to 2 Mbps, that is, approximately 200 times faster than previous 2G networks. This added speed and throughput will make it possible to run applications such as streaming video clips. It is anticipated that 4G speeds could be as high as 100 Mbps. Thus, 4G will represent another quantum leap in mobile Internet speeds and picture quality. Ericsson confirms that 4G

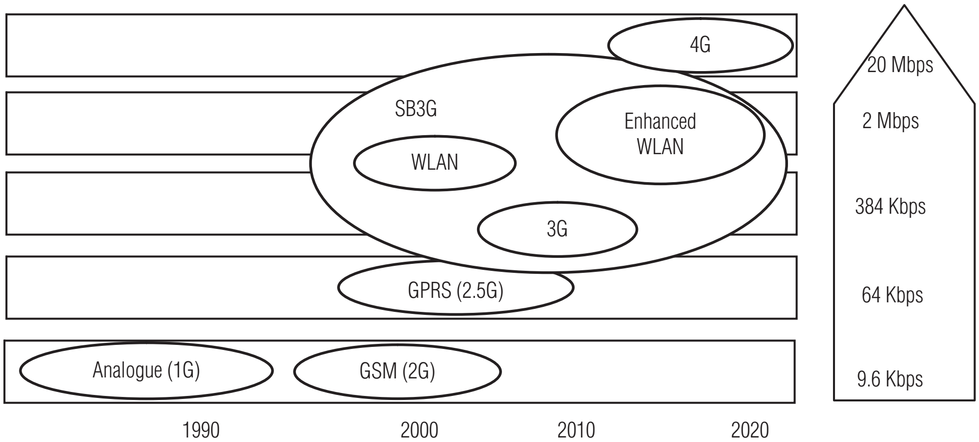


Figure 21.12 Mobile generation

could bring connection speeds of up to 50 times faster than 3G networks and could offer three-dimensional visual experiences for the first time. Figure 21.12 represents the typical progression of wireless communications.

21.11 Limitations of 3G

3G performances may not be sufficient to meet the needs of future high-performance applications like multimedia, full-motion video, and wireless teleconferencing. We need a network technology that extends 3G capacities by an order of magnitude.

1. There are multiple standards for 3G, making it difficult to roam and interoperate across networks. We need global mobility and service portability.
2. 3G systems are based on primarily a wide-area concept. We need hybrid networks that utilize both wireless LAN (hot spot) concept and cell or base-station wide area network design.
3. We need wider bandwidth.
4. Researchers have come up with spectrally more efficient modulation schemes that cannot be retrofitted into 3G infrastructures.
5. We need all digital packet networks that utilizes IP in its fullest form with converged voice and data capability.

21.12 Subscriber forecast for 3G in India

Subscriber forecast for GSM in India: By combining the historical trends with the current market conditions, the forecasting model predicts that the number of total GSM wireless subscribers in India will reach 876.6 million in 2013. Bharti-Airtel will continue to be the largest cellular operator in India. Its expected subscriber base will increase from 85.7 million in 2008 to 239.9 million in 2013. In addition, it is expected that Reliance Communications (RCom) will have 158.8 million mobile subscribers and Vodafone Essar will have 151.6 million mobile subscribers base by the end of 2013. We see some shift in market share among operators with BSNL losing market share to Bharti-Airtel, RCom, and Idea.

*Forecast for **3G** in India:* It is predicted that there will be 10 million (WCDMA & HSPA) 3G connections by the first half of 2011, and there will be 100 million 3G connections by the first quarter of 2014. By the end of 2014, there will be 150 million 3G connections. The state-owned BSNL and MTNL, which were the first carriers to launch, have managed to get a mere 1.5 million 3G connections. BSNL and MTNL are expected to control about 25 per cent of the market, the three largest non-state-owned carriers such as Bharti-Airtel, R-Com, and Vodafone comprise 43 per cent of the market. India is expected to have more than 600 million mobile connections by the end of the first half of 2010 and a billion connections by 2013.

The 3G frequency spectrum allocated for Aircel and Bharti-Airtel is 1,959–1,964 MHz, for Idea and Reliance it is 1,969–1,974 MHz, for STEL and Tata it is 1,974–1,979 MHz, and for Vodafone it is 1,964–1,969 MHz band.

21.13 Quality of services (QOS) in 3G

The term *quality of service (QOS)* designates simply a set of service requirements to be met by the network while transporting a traffic stream from source to destination. The QOS attributes are usually specified in terms of bit error rate (BER) and/or packet error rate (PER), transfer delay, and so on. In the QOS approach, a set of the explicit *QOS classes* is defined. A QOS class is a composition of set of admission control rules and a set of condition traffic rules. Main services featured in UMTS can be divided into four QOS classes primarily based on their ability to tolerate PER and transfer delay. *Conversational* and *streaming* classes preserve time relation between information entities of the stream. They are suitable to carry real-time traffic since they define an upper limit on transfer delay within their QOS profiles. The most well known use of conversational class is telephony speech, but it also covers applications like voice over Internet protocol (VoIP), videoconferencing etc. Real time streaming traffic has slightly flexible delay requirements and is convenient for real time or streaming video applications, for example. *Interactive* and *background* classes are mainly intended to represent conventional Internet applications (e.g., interactive web browsing, telnet, file transfer protocol FTP, and e-mail or file downloading) and also SMS and MMS services. Interactive applications have higher priority than background ones in terms of resource assignment to ensure responsiveness.

21.14 Summary

- Cellular evolution time line:
  - 1981—**1G** analogue cellular systems were introduced (e.g., standards: AMPS, NMT, TACS).
  - 1991—**2G** introduced with digital voice, low-speed circuit data (9.6 Kbps), SMS (e.g., standards: GSM, IS-54, PDC, cdmaOne) .
  - 1990–2000—**2.5G** introduced with packet data, improved voice, medium-speed circuit-switched and packet-switched data (~100 Kbps), and enhanced SMS (e.g., standards: GPRS, cdmaOne) .
  - 2002–2003—**3G** with improved voice quality, high-speed packet-switched data (384 Kbps–2 Mbps), improved spectral efficiency and capacity (e.g., standards: WCDMA, cdma2000 (1X, 3X), EDGE) .

- The difference between 1G and 2G: 1G used analogue signalling and 2G used digital signalling. Thus, boundary line between 1G and 2G systems is the analogue/digital split.
- There are four main standards for 2G systems: GSM, D-AMPS, CDMA or IS-95, and PDC. ***GSM*** is the most successful and widely used 2G system.
- The need for increased throughput data rates in data transfer (such as web browsing and e-mail) led to the evolution of 2.5G. 2.5G is a designation that includes all advanced upgrades for the 2G networks.
- 2.5G standard includes four technologies: HSCSD, GPRS, EDGE, and IS-95B.
- The need for high-speed Internet access up to 2 Mbps, live video communications, fast web access, and simultaneous data and voice transmission led to the development of 3G cellular networks.
- The 3G services have to be available in indoor and outdoor environments, where the integration of satellite links shall enable world-wide coverage.
- 3G systems are referred to as UMTS in Europe and IMT-2000 worldwide. Wideband CDMA and CDMA2000 are used as air interfaces in the European-backed UMTS and IMTS-2000 systems, respectively.
- Worldwide frequency spectrum identified for 3G is in the 1,885–2,025 MHz and 2,110–2,200 MHz bands. Of these 230 MHz of 3G spectrum, 2×30 MHz were intended for the satellite component and the rest for the terrestrial component.
- 2.5G networks currently offer true data speeds up to 28 Kbps. In comparison, the theoretical speed of 3G can be up to 2 Mbps, that is, approximately 200 times faster than previous 2G networks. This added speed and throughput will make it possible to run applications such as streaming video clips. It is anticipated that 4G speeds could be as high as 100 Mbps.

Review questions

1. What are the main standards of 2G? Write a short on them.
2. Explain the TDMA-based 2G standards in detail.
3. Explain EDGE.
4. List some 3G air interface technologies.
5. What are the limitations of 3G systems?
6. What are the advantages of digital systems used in 2G?
7. Write notes on digital cellular systems. (Refer Section 21.3.1)

Objective type questions and answers

1. What type of transmission techniques used in 1G
   (a) digital (b) analogue (c) both (a) and (b)
2. Bandwidth consideration in 1G is ________
   (a) 9.6 Kbps (b) 7.6 Kbps (c) 9.6 Mbps (d) 8.6 Mbps
3. QPSK means ________
4. Technologies used in 2G are ________
5. Technologies including in 2.5G are ________
6. The problem in 2.5G standards is ________
   (a) low data rate (b) less BW (c) low interference data rates

7. The idea behind EDGE is a new modulation scheme called ________
8. GMSK: ________
9. Modulations used in 3G are _________
10. WWWW: ____________

Answers: 1. (b), 2. (a), 3. quadrate phase shift keying, 4. GSM and IS-95 based on CDMA, 5. HSCSD, GPRS, EDGE, and IS-95B, 6. (c), 7. 5GMSK, 8. Gaussian minimum phase shift keying, 9. GMSK and QPSK, 10. Wireless World Wide Web.

Open book questions

1. What are the advantages of digital systems used in 2G?
2. Explain TDMA-based 2G standards in detail.
3. Explain HSCSD, GPRS, EDGE, and IS-95B.
4. Write a short note on 3G spectrum.

Further reading

Beach, M., Propagation and System Aspects, University of Bristol, Future Communication Systems Course, April 1994.

Classen, F., Meyr, H., and Sehier, P., "Maximum Likelihood Open Loop Carrier Synchronizer for Digital Radio," *IEEE proceedings,* pp. 493–497, May 1993, Geneva.

Djanpera, E., and Prasad, R. (Eds.), *WCDMA: Towards IP Mobility and Mobile Internet*, Artech House, 2001.

Firmanto, W. T., and Gulliver, T. A., "Code Combining of Reed–Muller Codes in an Indoor Wireless Environment," *Wireless Personal Communications*, 6, 1997, 359–371.

Garg, V. K., *IS-95 CDMA and CDMA2000*. Beijing: Publishing House of Electronics Industry, 2002.

Holma H., and Toskala, A. (Eds.), *WCDMA for UMTS*. John Wiley, 2001.

Muratore, F. (Ed.), *UMTS–Mobile Communications for the Future*. John Wiley, 2001.

Rappaport, T. S., *Wireless Communications Principles and Practice*. New York: IEEE Press, Prentice Hall, 1996, pp. 169–177.

Van Nee, R., and Prasad, R., *OFDM for Wireless Multimedia Communications*. Boston: Artech House, 2000, pp. 80–81.

WCDMA and CDMA2000 22

22.1 Introduction

3G mobile systems offer high bit rate services, high-quality videos, images, and fast web access. They differ significantly from the 2G technologies (global system for mobile communication [GSM] and CDMA1). The aim of 3G is to provide communication services from person to person at any place and at any time through any medium using a compact light-weight terminal with guaranteed quality of service (QoS) and security. The two standards of 3G technology that are most popular in the world are

- Wideband code division multiple access (WCDMA)
- Code division multiple access 2000 (CDMA2000).

Across the world, the following two major organizations are formed for the research and development of the above two interface technologies.

- 3G Project Partnership (3GPP) and
- 3G Project Partnership 2 (3GPP2)

The main aim of these two groups is to meet the standards defined by the International Mobile Telecommunication (IMT)-2000. IMT-2000 is defined by a single standard comprising a family of technologies intended to provide the users with the ability to communicate anywhere, at any time, and with anyone.

> IMT-2000 is a radio and network access specification defining several methods that meet the overall goals of the specification under the same brand name.

The responsibility of each group in the development process is as follows:

- **WCDMA** system specifications by 3GPP
- **CDMA2000** system specifications by 3GPP2

Europe and parts of Asian countries are using **WCDMA** for its 3G networks. United States is using the **CDMA2000** (the next generation of CDMA) for its 3G networks. It is suitable for the deployment of existing cellular and personal communications service (PCS) bands.

> *Universal mobile telecommunication system (UMTS) of Europe is using the WCDMA as a multiple access technology and Japan has been at the forefront in the research.* UMTS has

the capacity of 384 Kbps in the microcellular environment and 2 Mbps for indoor environment. It operates in the frequency band of around 2 GHz.

CDMA2000 of United States is a backward-compatible upgrade, whereas WCDMA is not backward compatible with CDMA. But both systems use many of the same concepts and features. This chapter presents the principles of ***WCDMA air interface***, referred also as UMTS terrestrial radio access (UTRA) and ***CDMA2000 air interface***, referred as IMT-2000 3G standard. In this chapter, the most popular worldwide cellular voice and data network technologies are also described.

22.2 CDMA

CDMA is the most preferred 3G technology for wireless communications because of the many advantages it offers, including *higher capacity, lower mobile transmit power, lower reuse factors, high-peak data rates, multipath diversity, and soft handoff*. As both WCDMA and CDMA2000 air interfaces technology mainly depend on the CDMA technique, a brief review of the CDMA principle that was already discussed in Chapter 17 is given below. The CDMA concept makes use of the following three important properties of spread spectrum technique:

- Transmission bandwidth is much larger than the information bandwidth.
- Bandwidth does not depend on the information signal.
- Processing gain = transmitted bandwidth/information bandwidth.

The CDMA technology can be classified as

- ***Direct sequence***: Data are scrambled by user-specific PRN code at the transmitter side.
- ***Frequency hopping***: Signal is spread by changing the frequency over the transmitted time of the signal.
- ***Time hopping***: Data are divided into frames and the frames are divided into time intervals. The data burst is hopped over the frames by utilizing code sequences.

CDMA allows every mobile device in a cell to transmit over the entire bandwidth at all times.

Each mobile phone has a unique and orthogonal code that is used to encode and recover the signal. The mobile phone digitizes the input voice data as it is received and encodes the data with the unique code for that phone. This is accomplished by taking each bit of the signal and then multiplying it by all bits in the unique code for the phone. Thus, one data bit is transformed into a sequence of bits of the same length as the code for the mobile phone. This makes it possible to combine with other signals on the same frequency range and still recover the original signal from an arbitrary mobile phone as long as the code for that phone is known. Once encoded, the data are modulated for transmission over the bandwidth allocated for that transmission. The data sending process in CDMA is shown in Figure 22.1(a).

The process for receiving a signal is shown in Figure 22.1(b). Once the signal is demodulated, a correlator and integrator pair recovers the signal based on the unique code from the cellular phone.

The correlator recovers the original encoded signal for the device, and the integrator transforms the recovered signal into the actual data stream.

The spreading and de-spreading, correlation and integration aspects of CDMA system are explained in the next section with an example for better understanding of the reader.

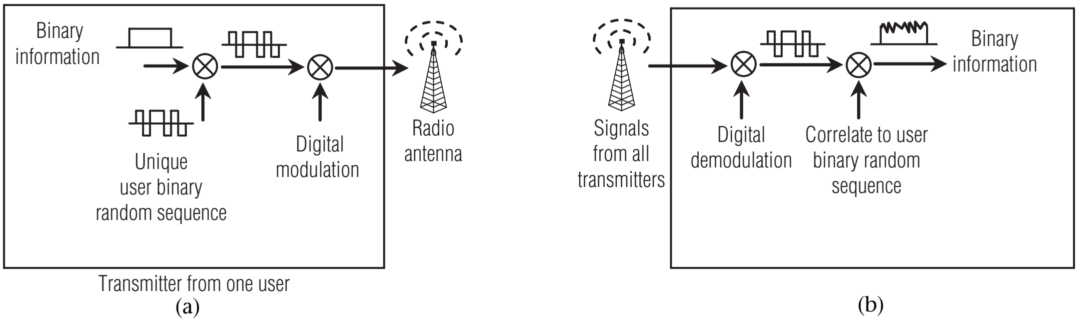


Figure 22.1 (a) Sending data using CDMA; (b) Receiving data using CDMA

22.2.1 Spreading and de-spreading, correlation and integration effects of CDMA

The basic principle of spreading and de-spreading for a DS-CDMA system is illustrated in Figure 22.2. Here, the user data are assumed to be a BPSK modulated bit sequence of rate R, the user data bits assuming the values of ±1. The spreading operation in this example is the multiplication of each user data bit with a sequence of 8 code bits, called chips. We assume the same method also for the BPSK spreading modulation. We can observe that the resulting spread data are at a rate of 8 × R and has the same random (pseudo-noise-like) appearance as the spreading code. In this case, we would say that we used a spreading factor of 8. This wideband signal would then be transmitted across a wireless channel to the receiving end.

During de-spreading, we multiply the spread user data/chip sequence, bit by bit, with the very same 8 code chips as we used during the spreading of these bits. As shown, the original user bit sequence has been recovered perfectly provided we also have (as shown in Figure 22.2) perfect synchronization between the spread user signal and the (de) spreading code. The increase of the signalling rate by a factor of 8 corresponds to a widening (by a factor of 8) of the occupied spectrum of the spread user data signal. Due to this virtue, CDMA systems are more generally called spread spectrum systems. De-spreading restores a bandwidth that is proportional to R for the signal.

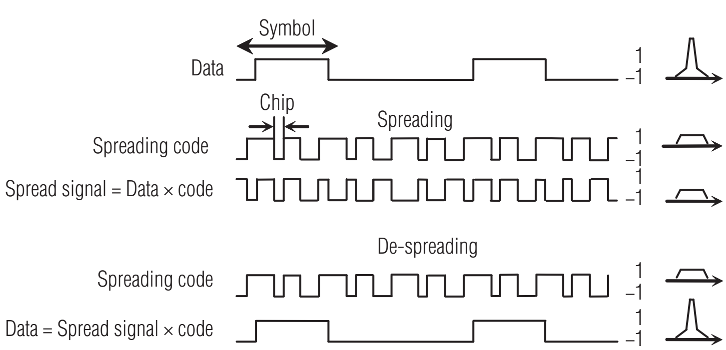


Figure 22.2 Spreading and de-spreading in DS-CDMA

The basic operation of the CDMA correlation receiver is shown in Figure 22.3(a). The upper half of the figure shows the reception of the desired signal. As in Figure 22.3(b), we see the de-spreading operation with a perfectly synchronized code. Then, the correlation receiver integrates (i.e., sums) the resulting products (data × code) for each user bit. The lower half of Figure 22.3(c) shows the effect of the de-spreading operation when applied to the CDMA signal of another user whose signal is assumed to have been spread with a different spreading code. The result of multiplying the interfering signal with the own code and integrating the resulting products leads to interfering signal values lingering around 0.

Processing gain in the CDMA system: From Figure 22.3(d), the amplitude of the own signal increases on an average by a factor of 8 relative to that of the user of the other interfering system, that is, the correlation detection has raised the desired user signal by the spreading factor, here 8, from the interference present in the CDMA system. This effect is termed as "processing gain" and is a fundamental aspect of all CDMA systems and in general of all spread spectrum systems.

High chip rate vs. processing gain: To utilize the entire available bandwidth (W Hz) the phase of the modulator is to be shifted pseudo randomly according to the pattern from the PRN generator at a rate of W times/sec.

High bit rate means less processing gain and higher transmit power or smaller coverage.

By keeping security in mind while designing the new system, the creators of 2G wireless were able to produce a usable system that is still in use today. Unfortunately, 2G technology is beginning to feel its age. Consumers now demand more features, which in turn require higher data rates than 2G can handle. A new system is needed that merges voice and data into the same digital stream conserving bandwidth to enable fast data access. By using advanced hardware and software at both ends of the transmission, 4G is the answer to this problem.

22.3 WCDMA evolution

As an upgrade to GSM, the European Telecommunication Standards Institute conceived UMTS in 1996. It was submitted to the IMT-2000 in 1998 as a worldwide standard under the name UTRA. A few years later, it merged with other CDMA proposals under the current name **WCDMA**. The technology evolution path of WCDMA is shown in Figure 22.4.

WCDMA is backward compatible with 2G systems GSM, IS-136, and PDC as well as with the 2.5G GPRS and EDGE.

It uses a variable direct sequence spread spectrum and requires a minimum spectrum allocation of 5 MHz. The chip rates can exceed 16 Mchip/sec/user. In addition, it is necessary to change the hardware at the base stations (BSs).

WCDMA supports demand bandwidth and two basic operation modes: ***Frequency Division Duplex (FDD)*** *and* ***Time Division Duplex (TDD)****. In FDD mode, you have 5 MHz for the uplink and downlink and TDD is time-shared (5 MHz) between uplink and downlink.*

In a WCDMA-based system, all users share the same frequency and time domain (as shown in Figure 22.5). Users are separated by the codes and the codes are orthogonal, that is,

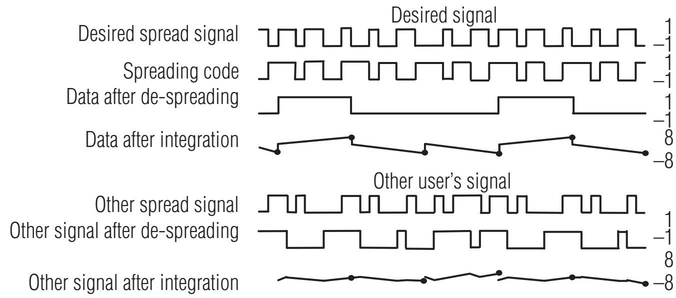


Figure 22.3(a) Principle of the CDMA correlation receiver

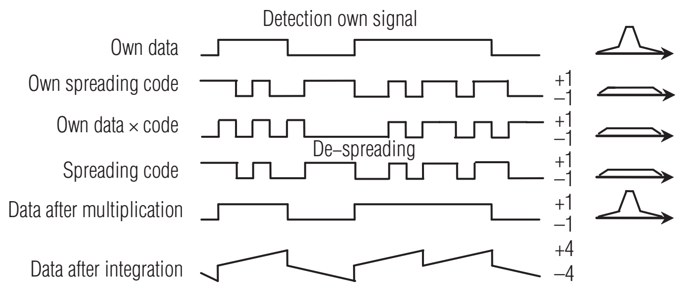


Figure 22.3(b) De-spreading operation with a perfectly synchronized code

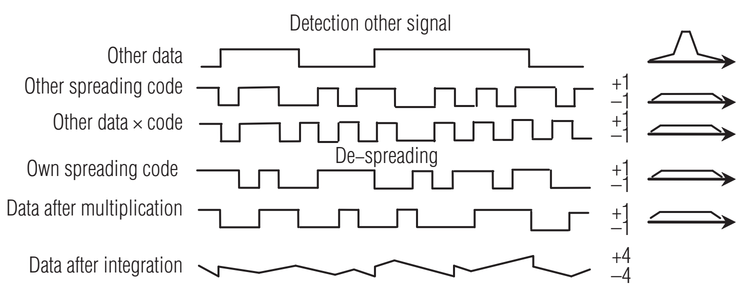


Figure 22.3(c) Effect of the de-spreading operation when applied to CDMA signal of another user

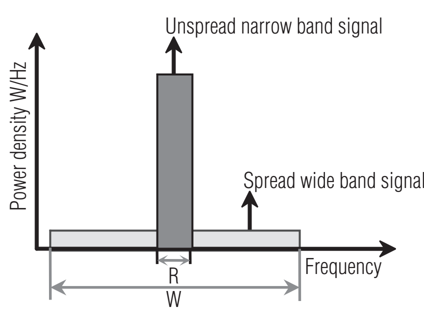


Figure 22.3(d) Processing gain and spreading

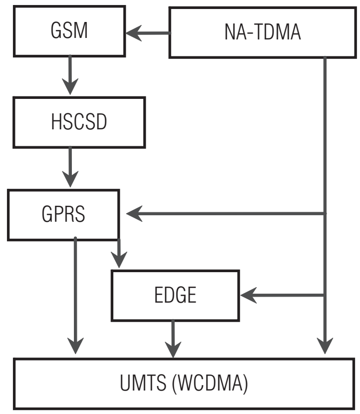


Figure 22.4 Technology evolution path of WCDMA

$$\int_a^b c_1(t)c_2(t)dt = 0$$

where, $c_1(t)$ and $c_2(t)$ are the spreading codes of user1 and user2, respectively.

There are two variants in WCDMA. The first one is known as *WCDMA FDD* and is a direct-sequence CDMA system with a nominal bandwidth of 5 MHz and it uses uplink and downlink in separate frequency bands. The second one is known as *WCDMA TDD* where the uplink and downlink are in the same frequency band.

In the FDD mode, separate 5 MHz carrier frequencies are used for the uplink and downlink respectively, whereas in TDD only one 5 MHz is time-shared between the uplink and downlink. Uplink is the connection from the mobile to the BS and downlink is that from the BS to the mobile.

22.3.1 Salient features of WCDMA

- WCDMA uses a new spectrum with a 5 MHz carrier and uses the DS-CDMA radio access (multiple access) technology. It provides 50 times higher data rate than in present GSM networks and 10 times higher data rate than in GPRS networks.
- WCDMA is a technology for wideband digital radio communications of Internet, multimedia, video, and other capacity demanding applications.
- WCDMA is the demanding 3G technology providing higher capacity for voice and data at higher data rates.
- The wider band makes it possible to divide and combine reception signals propagated through multipath-fading channels into more multipath components, which helps to improve the reception quality through RAKE time diversity.
- Its merits include the ability to accommodate a greater number of users who communicate at high speed (e.g., at 64 and 384 Kbps). It has also been verified in experiments that high-quality data transmission at 2 Mbps can be implemented using the 5 MHz bandwidth.
- The wide bandwidth of WCDMA gives an inherent performance gain over the previous cellular systems and it reduces the fading of the radio signal thus improving the performance.

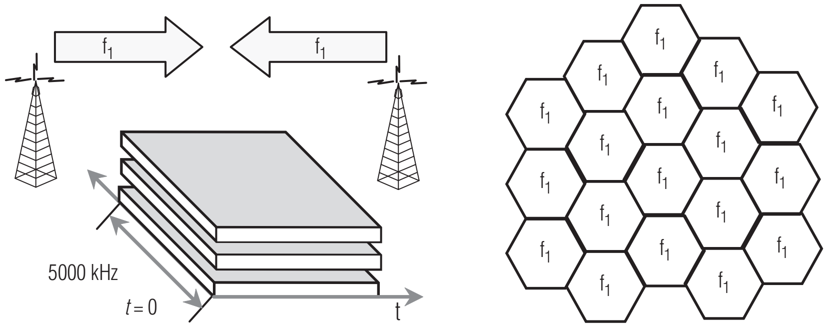


Figure 22.5 WCDMA system features

- WCDMA uses dual mode packet access scheme. Packet transfer can take place on both the common and dedicated channels. Owing to this phenomenon, packet access can be optimized for fast access response as well as for maximum throughput.
- The advance form of WCDMA is high-speed downlink packet access (HSDPA). HSDPA is a technology that leads to the cost-effective delivery of the most advance data services and significantly improves the network capacity.

22.3.2 Difference between WCDMA and 2G

The future generation of mobile communications is 3G, which means an enormous change with respect to the 2G systems. The principal objectives are global roaming, high-speed data transmission, commutation of circuits and packages, supports IP technology and multimedia applications. Differences between WCDMA and 2G are given in Table 22.1.

Table 22.1 Differences between WCDMA and 2G

Characteristic	WCDMA	Second generation
Carrier spacing	5 MHz	200 kHz
Frequency reuse factor	1	1–18
Power control frequency	1600 Hz	2 Hz or lower
Quality control	Radio resource management algorithms	Frequency hopping
Packet data	Load-based packet scheduling	Time slot based scheduling with GPRS
Downlink transmit diversity	Supported for improving downlink capacity	Not supported by the standard but can be applied

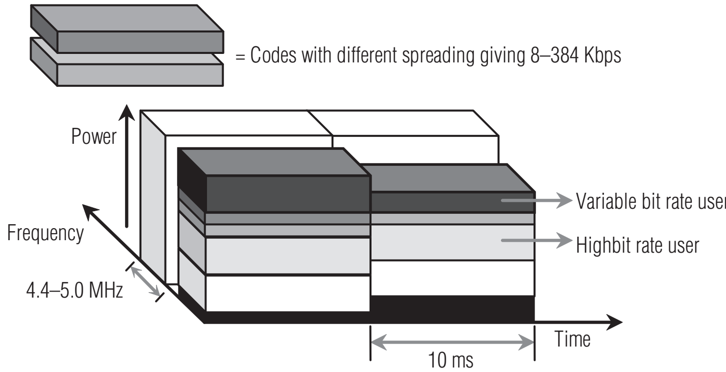


Figure 22.6 Allocation of bandwidth in WCDMA in the time–frequency–code space

22.4 WCDMA system design

The important parameters used in the WCDMA system design are described in this section. WCDMA uses the DS-CDMA system, that is, user information bits are spread over a wide bandwidth by multiplying the user data with quasi-random bits (called chips) derived from CDMA spreading codes. To support very high bit rates (up to 2 Mbps), the use of a variable spreading factor and multicode connections is supported. This arrangement is shown in Figure 22.6.

The chip rate of 3.84 Mcps (Million chips per second) leads to a carrier bandwidth of approximately 5 MHz. DS-CDMA systems with a bandwidth of about 1 MHz, such as IS-95, are commonly referred to as narrowband CDMA systems. The inherently wide carrier bandwidth of WCDMA supports high user data rates and also has certain performance benefits such as increased multipath diversity. Subject to the operating licence, the network operator can deploy multiple 5 MHz carriers to increase capacity, possibly in the form of hierarchical cell layers. This feature is shown in Figure 22.6. The actual carrier spacing can be selected on a 200 kHz grid approximately between 4.4 and 5 MHz depending on interference between the carriers.

WCDMA supports highly variable user data rates, in other words, the concept of obtaining bandwidth on demand (BoD) is well supported. The user data rate is kept constant during each 10 ms frame. However, the data capacity among the users can change from frame to frame. Figure 22.6 also shows an example of this feature. This fast radio capacity allocation can be typically controlled by the network to achieve optimum throughput for the packet data services. The parameters of WCDMA are given in Table 22.2.

22.4.1 Processing gain in WCDMA

Processing gain gives the CDMA systems the robustness against self-interference that is necessary to reuse the available 5 MHz carrier frequencies over geographically close distances. Let us take an example with real WCDMA parameters. Speech service with a bit rate of 12.2 Kbps has a processing gain of 25 dB = $10 \times \log_{10}(3.84e6/12.2e3)$. After de-spreading, the signal power needs to be typically a few decibels above the interference and noise power. The required power density over the interference power density after de-spreading is designated as $E_b = N_0$ in this book, where E_b is the energy or power density per user bit and N_0 is the interference or noise power density. For speech service, $E_b = N_0$ is typically in the order of 5.0 dB and the required wideband signal-to-interference ratio (SIR) is therefore 5.0 dB minus the processing gain = –20.0 dB. In other words,

Table 22.2 Parameters of WCDMA

Multiple access method	**WCDMA**
Duplexing method	Frequency division duplex/time division duplex
Channel bandwidth	5 MHz
Base station synchronisation	Asynchronous operation
Spreading modulation	Balanced QPSK (downlink) Dual-channel QPSK(uplink) Complex spreading circuit
Data modulation	QPSK (downlink) BPSK (uplink)
Chip rate	3.84 Mcps
Frame length	10 ms
Service multiplexing	Multiple services with different quality of service requirements multiplexed on one connection
Multirate concept	Variable spreading factor and multicode
Detection	Coherent using pilot symbols or common pilot
Multiuser detection, smart antennas	Supported by the standard, optional in the implementation

the signal power can be 20 dB under the interference or thermal noise power and the WCDMA receiver can still detect the signal. The wideband SIR is also called the carrier-to-interference ratio (C/I). Owing to spreading and de-spreading, C/I can be lower in WCDMA than, for example, in GSM. A good quality speech connection in GSM requires C/I = 9–12 dB.

Since the wideband signal can be below the thermal noise level, its detection is difficult without knowledge of the spreading sequence. For this reason, spread spectrum systems originated in military applications where the wideband nature of the signal allowed it to be hidden below the omnipresent thermal noise. Note that within any given channel bandwidth (chip rate), we will have a higher processing gain for lower user data bit rates than for high data bit rates. In particular, for user data bit rates of 2 Mbps the processing gain is less than 2 (=3.84 Mcps / 2 Mbps = 1.92, which corresponds to 2.8 dB) and some of the robustness of the WCDMA waveform against interference is clearly compromised.

- The processing gain together with the wideband nature suggests a frequency reuse of 1 between different cells of a wireless system (i.e., a frequency is reused in every cell per sector). This feature can be used to obtain high spectral efficiency.
- Having many users to share the same wideband carrier for their communications provides interferer diversity, that is, the multiple access interference from many system users is averaged out and this again will boost capacity compared to systems where one has to plan for the worst-case interference.
- However, both the above benefits require the use of tight power control and soft handover to avoid one user's signal blocking the others' communications. Power control and soft handover are explained later in this chapter.
- With a wideband signal, the different propagation paths of a wireless radio signal can be resolved at higher accuracy than with signals at a lower bandwidth. This results in higher diversity content against fading and thus improved performance.

22.4.2 Power control

Near–Far Effect: In a cellular CDMA system, a strong signal received at a BS receiver from a nearer mobile phone will mask the weak signal received from another mobile station (MS). In general, the nearer mobile phone is the mobile phone that is closer to the receiver. This interference is known as the ***near–far effect*** **(**Figure 22.7**)**.

To combat the near–far effect, a power control scheme is applied. With this scheme, the power transmitted from the mobile phone is adjusted according to its received power in the BS. In a perfect power control situation, all the signals from the MS or mobile phone reach the BS at equal power. If there are M simultaneous users in a given BS receiver and each of them produces signals with equal power, P, at a BS, then the SIR (noise) can be written as

$$\text{SNR} = \frac{\text{User power}}{\text{Interfering power}} = \frac{P}{(M-1)P} \tag{22.1}$$

For a system with a data bit rate R and bandwidth W, we may rewrite the above equation in terms of bit energy (E_b) to noise density (N_0) ratio to obtain

$$\frac{E_\mathrm{b}}{N_0} = \frac{P/R}{(M-1)P/W} = \frac{W/R}{M-1} \tag{22.2}$$

If the required E_b/N_o is given, then the maximum number of users can be easily calculated. From Equation (22.2) the total number of simultaneous active users in a system for a given E_b/N_o is

$$M \le \frac{W/R}{E_b/N_0} + 1 \tag{22.3}$$

In Equation (22.3), W/R represents a parameter known as the spread spectrum processing gain. For the Qualcomm CDMA system (IS-95), the processing gain is 128. E_b/N_o represents the operating energy per bit to noise ratio requirement and is related to the bit error rate (BER) performance of the system. In our analysis, we use the limiting value of BER = 0.01 as in previous sections.

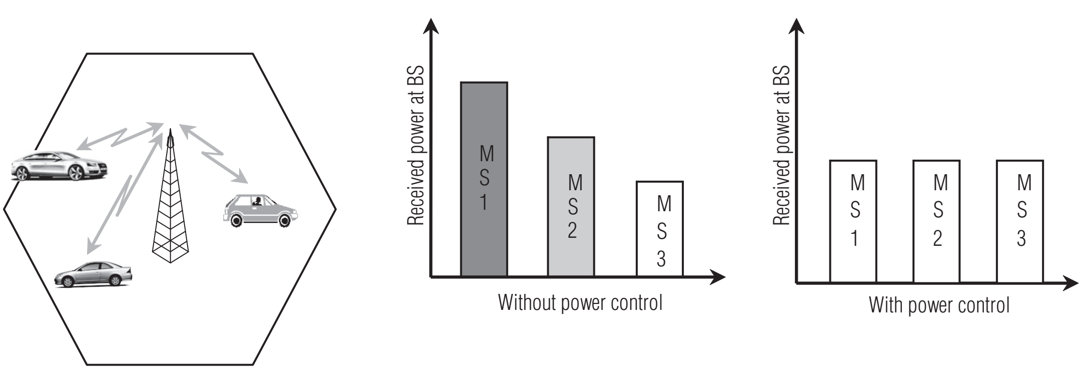


Figure 22.7 Near-far effects

Power control in WCDMA

Tight and fast power control is perhaps the most important aspect in WCDMA, in particular on the uplink. Without it, a single overpowered mobile could block a whole cell. Figure 22.7 depicts the problem and the solution in the form of closed loop transmission power control. Mobile stations MS1 and MS2 operate within the same frequency, separable at the BS only by their respective spreading codes. It may happen that MS1 at the cell edge suffers a path loss, say 70 dB above that of MS2, which is near the BS. If there were no mechanism for MS1 and MS2 to be power controlled to the same level at the BS, MS2 could easily overshoot MS1 and thus block a large part of the cell, giving rise to the so-called near–far problem of CDMA. The optimum strategy in the sense of maximizing capacity is to equalize the received power per bit of all MSs at all times. While one can conceive the open loop power control mechanisms that attempt to make a rough estimate of path loss by means of a downlink beacon signal, such a method would be far too inaccurate. The prime reason for this is that the fast fading is essentially uncorrelated between uplink and downlink owing to the large frequency separation of the uplink and downlink bands of the WCDMA FDD mode. Open loop power control is however used in WCDMA, but only to provide a coarse initial power setting of the MS at the beginning of a connection.

The solution to power control in WCDMA is fast closed loop power control, also shown in Figure 22.8. In closed loop power control in the uplink, the BS performs frequent estimates of the received SIR and compares it to a target SIR. If the measured SIR is higher than the target SIR, the BS will command the MS to lower the power; if it is too low, it will command the MS to increase its power. The cycle 'measure-command-react' is executed at a rate of 1,500 times per second (1.5 kHz) for each MS and thus operates faster than any significant change of path loss could possibly happen and indeed even faster than the speed of fast Rayleigh fading for low to moderate mobile speeds. Thus, closed loop power control will prevent any power imbalance among all the uplink signals received at the BS.

The same closed loop power control technique is also used on the downlink, though here the motivation is different, on the downlink there is no near–far problem due to the one-to-many scenario. All the signals within one cell originate from the one BS to all mobiles. It is, however, desirable to provide a marginal amount of additional power to MSs at the cell edge, as they suffer from increased other-cell interference. Also, on the downlink a method of enhancing weak signals caused by Rayleigh fading with additional power is needed at low speeds when other error-correcting methods based on interleaving and error correcting codes do not yet work effectively.

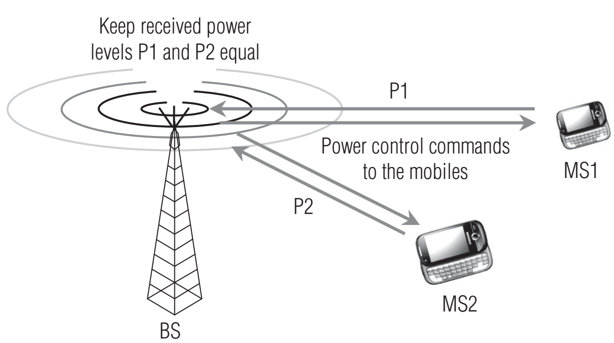


Figure 22.8 Closed loop power control in CDMA

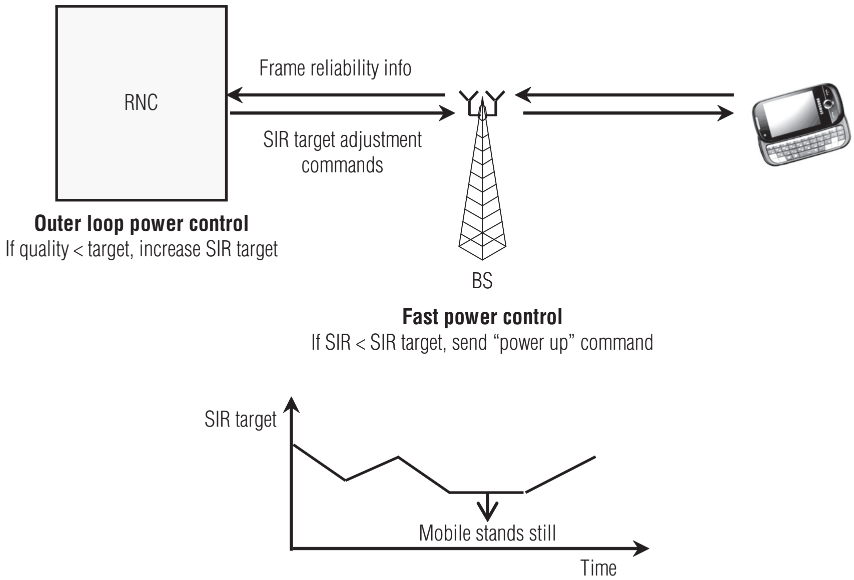


Figure 22.9 Outer loop power control

Figure 22.9 shows how uplink closed loop power control works on a fading channel at low speed. Closed loop power control commands the MS to use a transmit power proportional to the inverse of the received power (or SIR). Provided the MS has enough headroom to ramp the power up, only very little residual fading is left and the channel becomes an essentially non-fading channel as seen from the BS receiver. While this fading removal is highly desirable from the receiver point of view, it comes at the expense of increased average transmit power at the transmitting end. This means that an MS in a deep fade, that is, using a large transmission power will cause increased interference to other cells. Figure 22.8 illustrates this point.

Before leaving the area of closed loop power control, we mention one more related control loop connected with it: outer loop power control. Outer loop power control adjusts the target SIR set point in the BS according to the needs of the individual radio link and aims at a constant quality usually defined as a certain target BER or block error rate (BLER). Why should there be a need for changing the target SIR set point? The required SIR (there exists a proportional $E_b = N_0$ requirement), say, BLER = 1% depends on the mobile speed and the multipath profile. Now, if one were to set the target SIR set point for the worst case, that is, high mobile speeds, one would waste much capacity for those connections at low speeds. Thus, the best strategy is to let the target SIR set point float around the minimum value that just fulfils the required target quality. The target SIR set point will change over time as shown in the graph in Figure 22.9 as the speed and propagation environment changes.

Outer loop control is typically implemented by having the BS tag each uplink user data frame with a frame reliability indicator, such as a CRC check result obtained during decoding of that particular user data frame. Should the frame quality indicator indicate to the radio network controller (RNC) that the transmission quality is decreasing, the RNC in turn will command the BS to increase the target SIR set point by a certain amount. The reason for having outer loop control reside in the RNC is that this function should be performed after a possible soft handover combining. Details of soft handover are presented in the next section.

22.4.3 Softer and soft handovers

During softer handover, an MS is in the overlapping cell coverage area of two adjacent sectors of a BS. The communications between MS and BS take place concurrently via two air interface channels, one for each sector separately. This requires the use of two separate codes in the downlink direction so that the MS can distinguish the signals. The two signals are received in the MS by means of Rake processing, very similar to multipath reception, except that the fingers need to generate the respective code for each sector for the appropriate de-spreading operation. Figure 22.10 shows the softer handover scenario.

In the uplink direction, a similar process takes place at the BS. The code channel of the MS is received in each sector, then routed to the same baseband Rake receiver, and the maximal ratio is combined there in the usual way. During softer handover only one power control loop per connection is active. Softer handover typically occurs in about 5–15% of connections.

Figure 22.11 shows soft handover. During soft handover, an MS is in the overlapping cell coverage area of two sectors belonging to different BSs. As in softer handover, the communications between MS and BS take place concurrently via two air interface channels from each BS separately. As in softer handover, both channels (signals) are received at the MS by maximal ratio

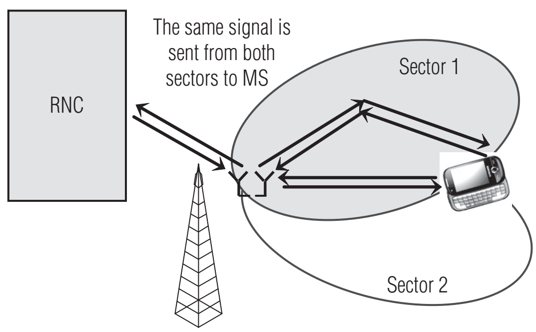


Figure 22.10 Softer handover

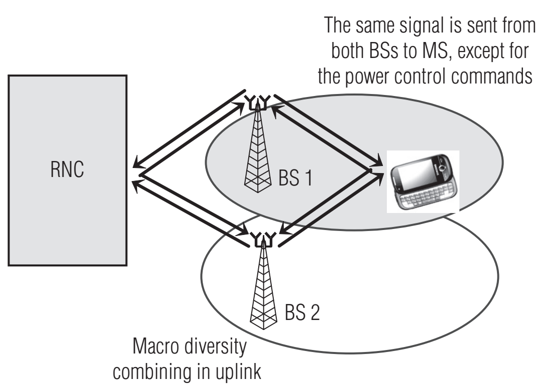


Figure 22.11 Soft handover

combining Rake processing. Seen from the MS, there are very few differences between softer and soft handover.

However, in the uplink direction soft handover differs significantly from softer handover, that is, the code channel of the MS is received from both BSs, but the received data are then routed to the RNC for combining. This is typically done so that the same frame reliability indicator as provided for outer loop power control is used to select the better frame between the two possible candidates within the RNC. This selection takes place after each interleaving period, that is, every 10–80 ms. Note that during soft handover two power control loops per connection are active, one for each BS. Soft handover occurs in about 20–40% of connections. In addition to soft/softer handover, WCDMA provides other handover types as follows:

- Inter-frequency hard handovers that can be used, for example, to handover a mobile from one WCDMA frequency carrier to another. One application for this is high-capacity BSs with several carriers.
- Inter-system hard handovers that take place between the WCDMA FDD system and another system, such as WCDMA TDD or GSM.

22.5 Capacity of a WCDMA system

Consider a single cell CDMA system where a number of mobiles are simultaneously transmitting at the same frequency. Here, each mobile is assigned a unique PN code sequence.

Let us assume that

P is the carrier power,

E_b is the energy per bit,

B_c is the spread spectrum signal bandwidth,

f_{data} is the information bit rate,

I is the power due to interference, and

N_o is the noise power per bit.

Then, $E_b = P/f_{data}$

$$\frac{E_b}{N_o} = \frac{P}{N_o f_{data}} \tag{22.4}$$

We have

$$N_o = \frac{I}{B_o}$$

$$\frac{E_b}{N_o} = \frac{P}{I}\frac{B_c}{f_{data}} = \frac{P}{I} \times G_P \tag{22.5}$$

Here G_p is the RF bandwidth divided by the information bit rate. In the CDMA system being discussed here, the signal is quadrature phase shift key (QPSK)-modulated, where the RF bandwidth is approximately equal to the chip rate. In other words, if f_{chip} is the chip rate, then the RF bandwidth $B_c = f_{chip}$, and in that case $G_p = f_{chip}/f_{data}$ is called the *process gain*. For a given BER, E_b/N_0 is fixed.

Consequently, the greater the process gain, the larger the allowable interference (i.e., I/P) for that BER will be. If there are N transmitters, all transmitting at the same power and using the same chip rate, then

$$\mathrm{I} = (N-1)\mathrm{P}$$

So, using Equation (22.4),

$$\frac{I}{P} = \frac{(N-1)P}{P} = N-1 = \frac{G_P}{E_b/N_o}$$

$$N = 1 + \frac{G_P}{E_b/N_o} \approx \frac{G_P}{E_b/N_o} \tag{22.6}$$

for large values of N.

Notice that for a fixed BER (i.e., a fixed value of E_b/N_0), the greater the process gain, the larger the capacity, N, of the system will be. Similarly, with a fixed process gain, the capacity increases if the value of E_b/N_0 required to provide a satisfactory operation decreases.

The capacity given by the previous equation is achieved only under ideal conditions. In actual practice, it may be significantly less for a number of reasons. For example, the capacity will decrease if the power control is not perfect. Similarly, in a multicell system, where each cell operates at the same frequency, transmissions in other cells may cause the interference to be increased by 60–85%.

Because the system is interference limited, the capacity of the system can be increased by reducing the interference. There are a number of ways of doing this. First, the interference due to other users can be reduced by replacing an omnidirectional antenna with a directional one. For example, a three-sector antenna would increase the capacity by a factor of about 2–3.

Second, human conversation is characterized by talk bursts followed by silence periods. If the transmitter is turned off during these silence periods, the interference to other transmitters will decrease and consequently the overall system capacity will increase. Thus, actual capacity may be given by

$$N = 1 + \frac{G_P}{E_b/N_o}\frac{\alpha}{(1+\beta)v} \tag{22.7}$$

where,

α is the correction factor due to imperfect power control,

β is the effect of co-channel interference from other cells in a multicell system, and

v is the voice activity factor.

Table 22.3 gives some typical values of these parameters.

Table 22.3 Typical values of parameters that affect the system capacity

Parameter	**Average values**
α	0.5–1.0
v	0.4–0.6
β	0.5–0.9. A typical value for a three-sector cell is 0.85. For an omni-directional antenna, it is 0.6.

Example problem 22.1

Find the capacity of CDMA system with the following parameters.

α = correction factor due to imperfect power control = 1, β = effect of co-channel interference from other cells in a multicell system = 0.85 for a three-sector cell, data rate = 9.6 Kbps for an 8 Kbps vocoder and chip rate = 1.2288 Mc/sec. The required E_b/N_0 = 7 dB.

Solution

Given:

$E_b/N_0 = 10^{0.7} = 5.01$

$G_p = 1.2288 \times 106/9{,}600 = 128.$

$$\text{Channel capacity} = N = 1 + \frac{G_P}{E_b/N_o}\frac{\alpha}{(1+\beta)v} = 1 + (128/5.01)(1/1.85)(1/0.4) = 35.$$

Notice that the capacity can be increased by simply reducing E_b/N_0, but that would result in increased BERs for all users. On the other hand, it is possible to minimize E_b/N_0 without necessarily running the risk of increasing the BER. One way to do this is to select an appropriate modulation technique. For example, if the desired BER is 10.5, the required E_b/N_0 is 12.6 dB with *Binary Frequency Shift Keying* (BFSK), whereas it is only 9.6 dB for *Binary Phase Shift Keying* (BPSK) or QPSK using coherent detection. Because the BER increases as the SIR is minimized, it is necessary to use an error-correcting code. The coding that is normally used in CDMA and WCDMA systems is convolutional coding where it is possible to achieve a coding gain of 4–6 dB with hard decision sequential and soft decision Viterbi decoding. Thus, the capacity of a CDMA system can be increased by using channel coding. It is interesting to know the minimum SNR that one can possibly use. The maximum attainable data rate R_{max} on a channel with infinite bandwidth in the presence of Gaussian noise is given by Shannon's channel capacity theorem.

$$\text{Highest data rate attainable } (R_{max}) = \frac{P}{N_o \ln 2} \tag{22.8}$$

Comparing Equations (22.5) and (22.6), we see that the minimum SNR is ln 2 = 0.693, that is, 10 log(0.693) or −1.6 dB. The maximum data rate is determined not only by this SNR but also by the transmitter power.

22.6 Air interface in UMTS

WCDMA with an FDD modulation technique and a carrier pair of 5 MHz bandwidth is used in UMTS. The uplink channels use a frequency range from 1,920 to 1,980 MHz and downlink from 2,110–2,170 MHz, respectively. Each user's data are spread with the use of a channelization code to the chip rate. Therefore, the data streams of each user are separated. Multiple users' data are then scrambled with the use of a channelization code at the chip rate. Every user has a unique scrambling code and many users may have common channel codes. The chip rate is 3.68 Mcps. The air interface allocates resources every 10 ms. Figure 22.12 presents the UMTS radio block.

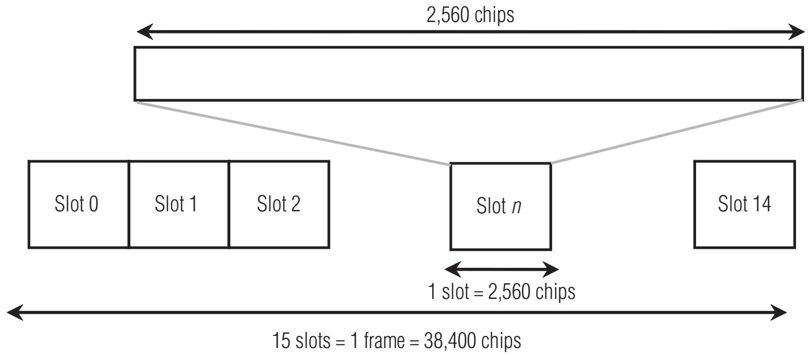


Figure 22.12 UMTS air interface radio block

22.7 UMTS network architecture

UMTS is evolved from GPRS to replace the radio access network (RAN). The UMTS terrestrial radio access network (UTRAN) consists in B Node the 3G term for BTS, and RNCs connected by ATM network. The 3G mobile network evolved from the 2G systems such as GSM and GPRS.

Some of the UMTS elements in the networks are: (i) The *UMTS subscriber identity module* similar to the GSM SIM card, (ii) the *Node B*, analogous to the GSM BTS, (iii) the RNC, analogous to the GSM BSC, (iv) the *call state control* function, (v) the *multimedia resource* function, (vi) the *media gateway* (MGW), (vii) the *transport media gateway,* (viii) the *roaming signalling gateway*, and (ix) the *media gateway control function*. The above network elements communicate in the following predefined interfaces: (i) the *Iub interface*, between RNC and Node B, (ii) the *Iur interface*, between RNCs, (iii) the *Gr interface*, between HLR and GGSN, and (iv) the *Gi interface*, between GGSN and MGW or other packet-based networks. Figure 22.13 presents the UMTS release five network topology.

22.8 WCDMA architecture

WCDMA is based on a hierarchical architecture with the different nodes and interfaces as illustrated in Figure 22.14. A terminal also referred to as *user equipment* in 3GPP terminology communicates with one or several Node B. In the WCDMA architecture, the term *Node B* refers to a logical node responsible for physical-layer processing such as error-correcting coding, modulation and spreading, as well as conversion from baseband to the radio frequency signal transmitted from the antenna. A Node B is handling transmission and reception in one or several cells. Three-sector sites are common, where each Node B is handling transmissions in three cells, although other arrangements of the cells belonging to one Node B can be thought of, for example, as a large number of indoor cell or several cells along a highway belonging to the same Node B. Thus, a BS is a possible implementation of a Node B.

The RNC controls multiple Node B's. The number of Node B's connected to one RNC varies depending on the implementation and deployment, but up to a few hundred Node B's per RNC is not uncommon. The RNC is in charge of call setup, QoS handling, and management of the radio resources in the cells for which it is responsible. The ARQ protocol, handling retransmissions

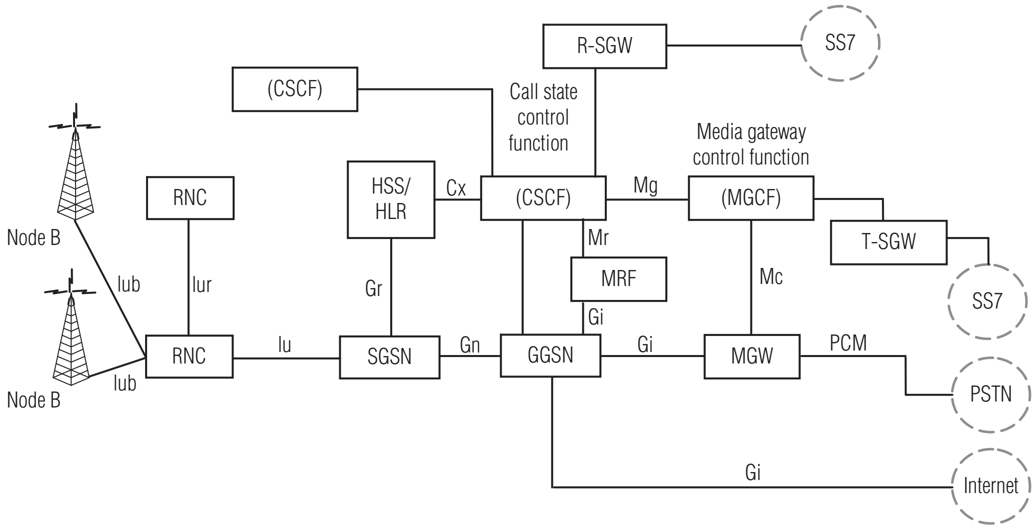


Figure 22.13 UMTS network architecture

of erroneous data, is also located in the RNC. Thus, in Release 99 most of the "intelligence" in the radio access network resides in the RNC, while the Node B's mainly acts as modems. Finally, the RNCs are connected to the Internet and the wired telephony network through the core network.

22.9 CDMA2000

22.9.1 CDMA deployments

CDMA is the fastest growing wireless technology and it will continue to grow at a faster pace than any other technology. It is the platform on which 2G and 3G advanced services are built. The CDMA air interface is used in both 2G and 3G networks. 2G CDMA standards are branded.

CDMA is the foundation for 3G services: the two dominant IMT-2000 standards, CDMA2000 and WCDMA, are based on CDMA. Wireless communication standards based on CDMA technique are presented in Table 22.4.

Table 22.4 Worldwide wireless communication standard based on CDMA technique

Standard	Type	Year	Multiple access	Frequency band (MHz)	Modulation	Channel BW (kHz)
IS-95/ CDMA1	Cellular/ PCS	1993	CDMA	824–894 1800–2000	QPSK/ BPSK	1250
cdma2000	Cellular	2000	CDMA	450, 800, 1700,1900, 2100	QPSK	2000

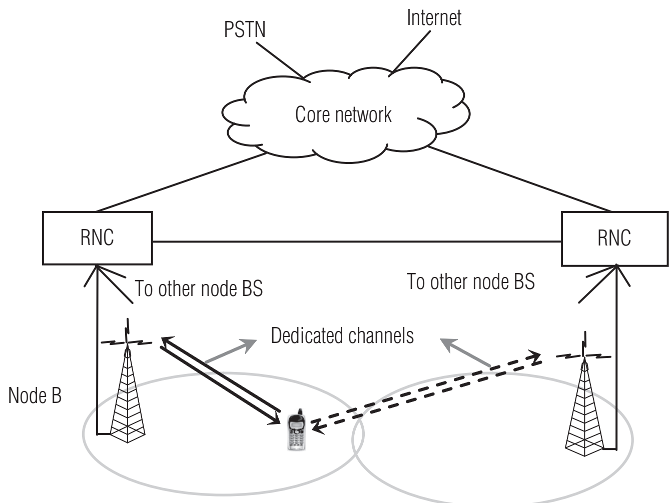


Figure 22.14 WCDMA radio access network architecture

22.9.2 CDMA 2000 evolution

CDMA2000 radio access technique evolved as a cellular standard under the name IS-95 and later became a part of the IMT-2000 family of technologies. This radio access technique is backward compatible with IS-95/IS-95A/IS-95B and is based on the original 1.25 MHz channel bandwidth per user. The CDMA2000 standard is going through an evolution similar to that of WCDMA/HSPA as shown in Figure 22.15.

From the original 1G IS-95 standard to the 3G CDMA2000 standard evolution, there are six primary steps involved.

- ***IS-95***: The standard 2G format is heavily deployed.
- ***IS-95B***: This has been deployed in Asia but not widely deployed in the United States. Incorporates several enhancements as well as packet-switched data.
- ***3G-1x***, also known as radio transmission technology with one carrier (1xRTT), or the first phase: This is often called 1x. It is the first 3G phase of CDMA, expected to be widely deployed initially because of capacity gains and for its high-speed data capabilities. Many are calling this as a 2.5G standard.
- ***1xEV-DO***, also known as Qualcomm *high data rate* (HDR): This is a high-speed data-only solution that can be integrated relatively easily into existing CDMA networks. Again, many consider this as a 2.5G interim solution. There is a thought that 1xEV would be deployed in parallel with 1x, thus using 1x for voice services and 1xEV for data services.

Figure 22.15 CDMA evolution path to 3G CDMA2000

- ***1xtreme or 1xEV-DV***: An enhanced version of 1xEV-DO, it also allows for simultaneous voice services.
- ***3G-3x (3xRTT):*** The complete 3G version of CDMA. Currently expectations are that it will only see limited deployment because of high deployment costs and demand and competitive issues.

The key to the six versions is the integration aspect. Wireless operators should be able to deploy these versions relatively easily with relatively low investment. Upgrading is neither a free process nor is it completely plug and play technically. Compared to the completely new deployment of a brand-new technology, such as many GSM operators face with deploying WCDMA, the cost and complexity should be lower.

IMT-2000 is a global standard designed to harmonize the 3G standards. One of the primary goals of this standard is to standardize the type of backhaul network on which the radio formats can operate. You can recall from earlier chapters that IS-95 is designed to operate on the IS-41 network, while GSM networks work on the MAP networks. As CDMA moves forward in evolution, it will be possible to operate it on either GSM-MAP or IS-41 networks (Figure 22.16).

22.9.3 Overview of 1x-RTT and 3x-RTT

Almost every CDMA operator in the world today is planning to deploy at least the first advanced phase of CDMA2000, known as 1x. The 1x implementation should be relatively simple. Because the bandwidth of the signal remains identical (1.2288 MHz) and the system is completely backward compatible, carriers can deploy the technology – in most instances as a circuit card and software upgrade to existing BSs – and realize the benefits of the deployment as they swap out the older handsets for 1x-compatible handsets. As mentioned, the benefits of deploying 1x infrastructure can only be realized if the end users are also given 1x capable handsets; otherwise, the network will simply perform as an IS-95 network. Given the ease of deployment, the reasons for deploying 1x-RTT are quite compelling. Under IS-95A, operators could deploy voice services utilizing up to 64 Walsh codes. In addition, the capability is available for circuit-switched data

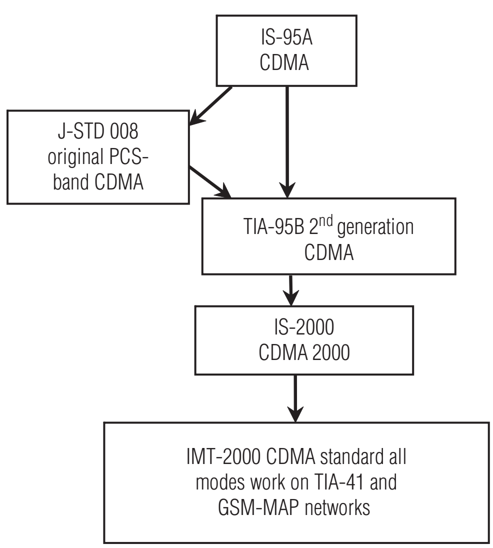


Figure 22.16 CDMA evolution to IMT-2000 the global standard

albeit at very low speeds. IS-95B offered some improvements. While it still only allowed 64 Walsh codes, it allowed for packet-switched data by combining Walsh codes for higher data rates. In this way, data rates of 64 Kbps or 96 Kbps could be deployed and offered. The obvious problem with this is the limit of Walsh codes – if several Walsh codes were used for data, this would have an adverse effect on capacity of that carrier.

The 1x technology takes a large leap from IS-95A and IS-95B. The number of Walsh codes available has been increased to 128; there has also been an increase in the length of the Walsh codes. This is coupled with several performance improvements, including stronger error coding, faster and more improved forward power control and transmit diversity. The end result is the ability to possibly double voice capacity as well as offer high-speed packet-switched data services (up to 153 Kbps). There are several other improvements as well, including several new physical channels that will improve performance (e.g., the quick paging channel, which can increase the handset's idle time substantially which would mean longer battery lives).

The 3x technology would take another leap from 1x. The concept is basically to use three 1.2288 MHz channels together (hence, it uses what is called *spreading rate 3*) (Figure 22.17). In 3x, the Walsh codes can be up to 256 chips long, allowing for much more voice capacity in addition to very high-speed data rates (perhaps up to 2 Mbps). In addition, 3x will have some degree of world standardization as far as the core network it works off. It will be able to operate on ANSI-41, which is the primary core network many American time division multiple access (TDMA) and CDMA networks operate on today, or GSM-MAP, the core network more standard throughout the world for GSM and its future upgrade paths.

In Figure 22.17, the 3x technology uses three 1x carriers hence the spreading rate is 3. The 3x business case, however, may result in this technology not being widely deployed. While it may offer higher data rates, the cost of deployment may be quite high and 1x performance may be enough for some time. For that reason, most of this text will focus on 1x technology although the 3x is not very different.

There are several evolution phases in CDMA2000 networks as given below:

- CDMA2000 1RTT
- CDMA2000 1xEV
- CDMA2000 3xRTT

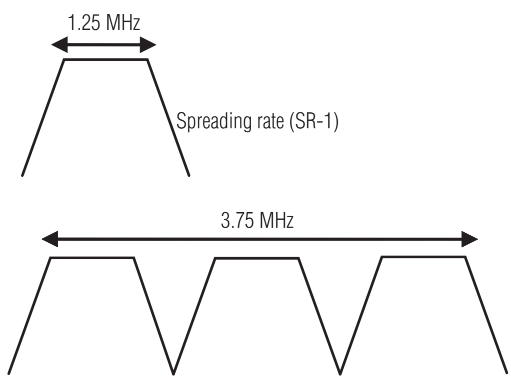


Figure 22.17 Spreading rate 3 (SR-3)

- ***CDMA2000 1xRTT*** is one of several types of radio access techniques included in the CDMA2000 initiative. The CDMA2000 standard that originally supported single carrier (1x) and multicarrier (3x) mode was adopted by ITU-R under the name IMT-2000CDMA *multicarrier* (MC). The RTT stands for round trip time. The *CDMA2000 1xRTT* improved the 2G and 2.5G standards by introducing a rapidly adaptable signalling rate and chipping rate for each user. It supports instantaneous data theoretical throughput of 307 Kbps and typically 144 Kbps per user.
- ***CDMA2000 1xEV*** *The CDMA2000 1xEV* was developed by the Qualcomm Company to be compatible with the WCDMA. In August 2001, the ITU included it in the IMT 2000 project. There are two types of systems:
    - ***CDMA2000 1xEV-DO*** *(data only)* that dedicates radio channels strictly for data and supports instantaneous data throughput of 2,400 Kbps on a specific CDMA channel.
    - ***CDMA2000 1xEV-DV*** (data and voice) provides integrated voice and data simultaneous high-speed packet data multimedia services at speeds of up to 3.09 Mbps. *CDMA2000 1xEV-DO and 1xEV-DV are both backward compatible with CDMA2000 1x and CDMA1.*
- ***CDMA2000 3xRTT*** The *CDMA2000 3xRTT* is using three 1.25 MHz radio channels and provides data throughput of 2,000 Kbps. If the three radio channels are not adjacent and used individually, then need exists for added hardware at the BSs. If the three channels are adjacent to manipulate an instant 3.75 MHz channel, additional hardware at the BSs is needed. Figure 22.18 represents a graphical representation of CDMA2000 and WCDMA. When CDMA2000 1xRTT is used, each user communicates within a single 1.25 MHz carrier and data are spread with 1.2288 Mcps. When CDMA2000 3xRTT is used, each user can communicate through three consecutive carriers and data are spread with 3.6864 Mcps. In contrast, the WCDMA user resides within a single 5 MHz carrier and data are spread with 4.096 Mcps.
- ***CDMA2000 3xEV-DO*** (data only) that dedicates radio channels strictly for data and supports instantaneous data throughput of 2,400 Kbps on a specific CDMA channel.
- ***CDMA2000 3xEV-DV*** (data and voice) that dedicates radio channels for data and voice and supports throughput of 144 Kbps on a specific CDMA channel.

Qualcomm has its own proprietary high-speed standard called HDR, to be used in IS-95 networks. It will provide a 2.4 Mbps data rate. A standard for HDR has been formulated in IS-856. The *1x evolved data optimized* (1xEV-DO) term is used when referring to the non-proprietary form of this advanced CDMA radio interface. The 1xEV-DO adds a TDMA component beneath the code components to support highly asymmetric, high-speed data applications. A more detailed

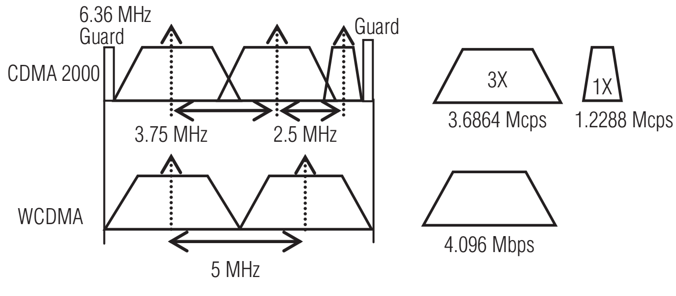


Figure 22.18 Comparison of cdma2000 and WCDMA carrier

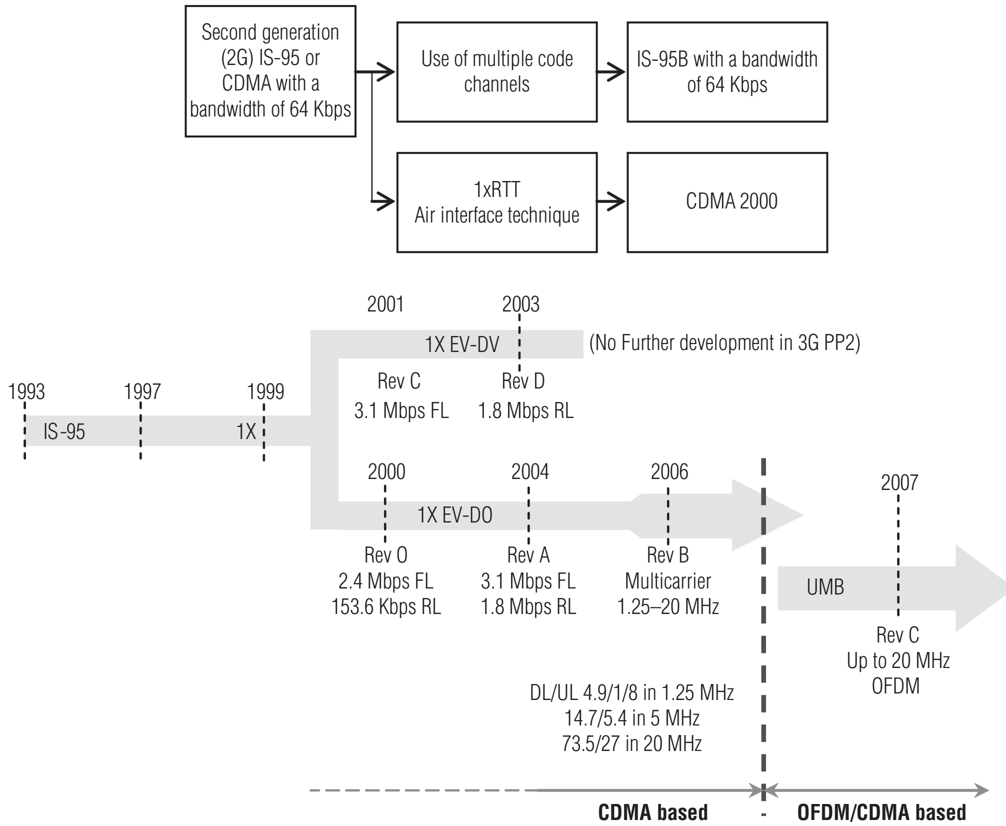


Figure 22.19 The evolution from IS-95 to CDMA2000 1x and 1x EV-DO

discussion on how the IS-95 system is evolved into a full CDMA2000 system with all the intermediate phases is given below.

The evolution steps of CDMA2000 are shown in Figure 22.19. After the CDMA2000 1x standard was formed as an input to ITU for IMT-2000, two parallel evolution tracks were initiated for better support of data services. The first one was EV-DO (*evolution–data only*) 1 which has continued to be the main track as further described below. It is also called *high-rate packet data*. A parallel track EV-DV (*evolution for integrated data and voice*) was developed to give parallel support of data and circuit-switched services on the same carrier. It is at the moment not developed further within 3GPP2.

22.10 Key features of CDMA2000

CDMA2000 builds on the inherent advantages of CDMA technologies and introduces other enhancements, such as orthogonal frequency division multiplexing (OFDM and OFDMA), advanced control and signalling mechanisms, improved interference management techniques, end-to-end QoS, and new antenna techniques such as multiple inputs multiple outputs (MIMO) and space division multiple access (SDMA) to increase data throughput rates and QoS, while significantly improving network capacity and reducing delivery cost.

- ***Leading performance***: CDMA2000 performance in terms of data speed, voice capacity, and latencies continue to outperform in commercial deployments and other comparable technologies.
- ***Efficient use of spectrum***: CDMA2000 technologies offer the highest voice capacity and data throughput using the least amount of spectrum, lowering the cost of delivery for operators and delivering superior customer experience for the end users.
- ***Support for advanced mobile services:*** CDMA2000 1xEV-DO enables the delivery of a broad range of advanced services such as high-performance VoIP, push-to-talk, video telephony, multimedia messaging, multicasting, and multiplaying online gaming with richly rendered 3D graphics.
- ***Devices selection***: CDMA2000 offers a broad selection of devices and has a significant cost advantage compared to other 3G technologies to meet the diverse market needs around the world.
- ***Seamless evolution path***: CDMA2000 has a solid and long-term evolution path, which is built on the principle of backward and forward compatibility in-band migration, and support of hybrid network configurations.
- ***Flexibility:*** CDMA2000 systems have been designed for urban as well as remote rural areas for fixed wireless, wireless local loop, limited mobility and full mobility applications in multiple spectrum bands including 450, 800, 1,700, 1,900, and 2,100 MHz.

22.10.1 CDMA2000 advantages

- Superior voice clarity
- High-speed broadband data connectivity
- Low end-to-end latency
- Increased voice and data throughput capacity
- Differentiated value-added services such as VoIP, PTT, multicasting, position location, and so on
- Flexible network architecture with connectivity to ANSI-41, GSM-MAP, and IP-based networks and flexible Backhaul connectivity
- Application, user and flow-based QoS
- Flexible spectrum allocations with excellent propagation characteristics
- Robust link budget for extended coverage and increased data throughputs at the cell edge
- Multimode, multiband, global roaming
- Improved security and privacy
- Lower total cost of ownership

22.11 Comparison of WCDMA and CDMA2000

CDMA2000 and UMTS were developed separately and are two separate ITU approved 3G standards. CDMA2000 1xRTT, CDMA2000 1xEV-DO,and future CDMA2000 3x were developed to be backward compatible with CDMA1. Both 1x types have the same bandwidth, chip-rate and it can be used in any existing CDMA1 frequency band and network. 3G WCDMA and CDMA2000 Standards are presented in Table 22.5

UMTS was developed mainly for countries with GSM networks because these countries have agreed to free new ranges of UMTS networks. UMTS uses new technology, new frequency band,

Table 22.5 3G WCDMA and CDMA2000 standards

UMTS-WCDMA	CDMA2000
No backward compatibility	Backward compatibility with CDMA2000
Cell site not synchronised	Cell site synchronised through GPS timing
Each cell site with different scrambling code for spreading	Adjacent cell sites use different time offset of same scrambling code for spreading
Complex code 38,400 chips, frame of 10 msec.	Pseudo Random (PN) sequence of length $2^{15}-1$ chips; period of 26.67 msec; different site offset of 64 chips
OVSF codes	Walsh codes

and new RAN. UMTS phones are dual mode and are compatible with GSM systems. Table 22.6 shows the comparative differences between the different technologies and it shows the clear difference between them. The CDMA2000 is widely used for the fixed Internet and data services but the WCDMA is the new technology designed for both the data and voices and it performs better operations.

The comparison between WCDMA and CDMA2000 technologies is given in Table 22.6

Table 22.6 Comparison of WCDMA and CDMA2000 technologies

Parameter	W-CDMA	CDMA2000
Carrier spacing/Channel Bandwidth	5 MHz	3.75 MHz (1.25 × N MHz. Initially, N may be 1, 2, or 3, but later could be 6, 9, or 12).
Spectrum Allocation	FDD mode 1920–1980 MHz uplink 2110–2170 MHz downlink TDD mode 1900–1920 MHz 2010–2025 MHz	1850–1910 MHz uplink 1930–1990 MHz downlink
Chip rate	3.84 Mcps	1.2288 Mcps
Data modulation	BPSK	FW – QPSK; RV – BPSK
Spreading	Complex (OQPSK)	Complex (OQPSK)
Power control frequency	1500 Hz	800 Hz
Variable data rate implement.	Variable SF; multi code	Repeat., puncturing, multi code
Frame duration	10 ms	20 ms (also 5, 30, 40)
Coding	Turbo and convolutional	Turbo and convolutional
Base stations synchronized?	Asynchronous	Synchronous
Base station acquisition/detect	Three step; slot, frame, code	Time shifted PN correlation
Forwards link pilot	TDM dedicated pilot	CDM common pilot
Antenna beam forming	TDM dedicated pilot	Auxiliary pilot

22.12 Summary

- In Europe and Asia, the current phase in cellular mobile communication systems focuses on the operation and optimization of 3G systems known as the UMTS.
- UMTS or as it is often termed, WCDMA is being widely deployed. It offers many advantages over GSM, GPRS, and EDGE in terms of much higher data rates and greater flexibility. WCDMA was selected as air interface multiple access technique for UMTS.
- UMTS WCDMA FDD is a direct-sequence CDMA system with a nominal bandwidth of 5 MHz. The second system, UMTS WCDMA TDD, also uses CDMA with a bandwidth of 5 MHz, but now the frequency band is time-shared in both directions. One half of the time is used for transmission in the forward direction and the other half of the time in the reverse direction.
- CDMA2000 is a multicarrier, direct-sequence CDMA FDD system. LikeCDMA1, its first phase is expected to use a single carrier with a bandwidth of 1.25 MHz.
- There are two approaches being used to develop 4G access techniques: 3xRTT (currently 1xRTT for 2.5 and 3G) and WCDMA. These disparate access techniques currently do not interoperate. This issue may be solved with software defined radios.
- There are several evolution phases in CDMA2000 networks as given below:
  - CDMA2000 1RTT
  - CDMA2000 1xEV
  - CDMA2000 3xRTT
- CDMA is suited for data transfer with bursty behaviour and where delays can be accepted. It is therefore used in wireless LAN applications; the cell size here is 500 ft because of the high frequency (2.4 GHz) and low power. The suitability for data transfer is the reason for why WCDMA seems to be “winning technology” for the data portion of 3G mobile cellular networks.
- GSM standard is a specification of an entire network infrastructure, the CDMA interface relates only to the air interface – the radio part of the technology.
- CDMA2000 builds on the inherent advantages of CDMA technologies and introduces other enhancements, such as OFDM and OFDMA, advanced control and signalling mechanisms, improved interference management techniques, end-to-end QoS, and new antenna techniques such as MIMO, and SDMA to increase data throughput rates and QoS, while significantly improving network capacity and reducing delivery cost.
- CDMA2000 is the next generation of CDMA and is a backward-compatible upgrade and is discussed in further chapters.
- WCDMA, while not backward compatible with CDMA, uses many of the same concepts and features.
- TDD-WCDMA is a radio interface technology that combines CDMA and TDMA, as well as TDD.

Review questions

1. Describe the WCDMA architecture.
2. What is near–far effect?
3. Describe the WCDMA/FDD and WCDMA/TDD systems.
4. What are the salient features of WCDMA?
5. Compare the 3G Cellular air interface technologies WCDMA and CDMA2000.
6. Discuss about the frequency spectrum allocation in respect of WCDMA and CDMA2000.

7. What are the various CDMA2000 advantages?
8. Describe the key features of CDMA2000.
9. Describe the CDMA2000 3xRTT and CDMA2000 1xEV-DO air interface technologies.
10. Describe the various evolution phases in CDMA2000 networks.
11. Describe the softer and soft handovers in respect of WCDMA.
12. Write about the processing gain in WCDMA.
13. What are the basic units of a cellular system?
14. Mention the functions of BS and MS.
15. Give the names of two standards of 3G technology.
16. State the different types of handoffs.
17. Write some features of CDMA.

Objective type questions and answers

1. Power control frequency of WCDMA is
   a) 900 Hz (b) 1,500 Hz (c) 700 Hz (d) 1,400 Hz
2. Chip rate of CDMA2000 is
   (a) 4.096 MHz (b)5.6864 MHz (c) 3.0864 MHz (d) 3 MHz
3. Codes that are used in UMTS WCDMA are
   (a) OVSF codes (b) Walsh codes (c) Both a and b
4. 1x EV-DO is also known as
   (a) Bandwidth on Demand (BoD) (b) Qualcomm High Data Rate (QHDR
   (c) QPSK (d) Radio network controller (RNC)
5. CDMA2000 3x-RTT is using ________ radio channels
   (a) Two 1.25 MHz (b) Three 2.25 MHz
   (c) Two 2.25 MHz (d) Three 1.25 MHz
6. CDMA2000 year of introduction, channel Bandwidth (kHz) are
   (a) 2000, 1,250 (b) 2000, 2,000 (c) 1998, 1,250 (d) 2002, 2,000
7. A good quality speech connection in GSM requires C/I = ________dB.
   (a) 9–11 dB (b) 8–12 dB (c) 11–14 dB (d) 9–12 dB
8. DS-CDMA uses
   (a) FDD (b) TDD (c) Both a and b (d) Duplexing method
9. Quality control of WCDMA is ________
10. Maximum number of users (M) in a channel that can be used simultaneously is

    (a) $M \le \frac{W/R}{E_b/N_o} + 1$ (b) $M \le \frac{W/R}{E_b/N_o} - 1$

    (c) $M \le \frac{W/R}{E_b/N_o} + 2$ (d) $M \le \frac{W/R}{E_b/N_o} - 2$

11. The standard used in 3G system is
    (a) WCDMA (b) CDMA (c) DS (d) UMTS

Answers: 1. (b), 2. (c), 3. (a), 4. (b), 5. (d), 6. (b), 7. (d), 8. (c), 9. Radio Resource Management Algorithm, 10. (d), 11. (a).

Exercise problems

1. Determine the maximum number of simultaneous users in a system given by $\frac{E_b}{N_o} = 10\ db$ and bandwidth = 20 KHz at a given bit rate of 100 bits/sec.

   [Hint: $M \le \frac{W/R}{E_b/N_o} + 1$]

2. Determine the channel capacity of CDMA system with the following specifications: $\alpha = 1$ and $\beta = 0.75$ with a data rate of 10 Kbps and for a 10 Kbps vocoder and $f_c = 1.5$. Mcps and $\frac{E_b}{N_o} = 10\ db$.

3. Determine the ratio of $\frac{E_b}{N_o}$ if carrier power = 5 W and power due to interference = 2.4 W and $f_{chip} = 50$ Mcps and $f_{data} = 25$ Mbps.

Open book questions

1. What are the features of WCDMA?
2. What are the important parameters that are used in the WCDMA system design?
3. What is the purpose of control in WCDMA?
4. Describe how the 3G CDMA2000 standard is evolved from the original 1G IS-95 standard.

Further reading

"CDMA Evolution from IS-95, IS-2000, to 1xTREME," Technology Transfer Training Class, Motorola, Inc., Colarado Springs, July 2000.

Cooper, G. and McGillem, C., Modern Communications and Spread Spectrum, New York, McGraw-Hill, 1998.

Dixon, R., Spread Spectrum Systems with Commercial Applications, New York, John Wiley & Sons, 1994.

Harri, H. and Antti, T., WCDMA for UMTS: Radio Access for Third Generation Mobile Communications, New York: John Wiley & Sons, 2000.

Jakes, W., Microwave Mobile Communications, New York, John Wiley & Sons, 1974.

Manjit S. and Manoneet S., "3G Wireless with Respect to IMT-2000 and Beyond," Telecom 99.

Peterson, R. L., "Third Generation Personal Communications: Physical Layer Status," Presentation at Clemson University, Scotland, February 2001.

Saunders, S., Antennas and Propagation for Wireless Communication Systems, UK, JohnWiley & Sons, 1999.

Viterbi, A., Principles of Spread Spectrum Communication, Addison-Wesley, Reading, Massachusetts, 1997.

Ziemer, R. and Peterson, R., Introduction to Digital Communications, Upper Saddle River, New Jersey, Prentice Hall, Chapter 10, 2001.

Next Generation Cellular Technology 4G 23

23.1 Introduction

Over the years, there has been an exponential increase in the capabilities of networks evolving from wired networks to wireless. However, the current demand for mobile communication facilities and the dramatic increase in its growth rate reveal that even the third generation systems cannot be expected to fulfil all demands. Even though 3G data rates were already real in theory, the initial systems like UMTS-based WCDMA did not meet the requirements in their practical deployments. Hence, new standards are needed to improve or to meet or to even exceed the demands. The combination of High Speed Downlink Packet Access (HSDPA) and the High Speed Uplink Packet Access (HSUPA) led to the development of the technology referred to as High Speed Packet Access (HSPA) or, more informally, 3.5G.

But why is this next generation technology needed when 3G or 3.5G networks seem to be sufficient to cater for subscriber demands for high data rates and quality of service? The answer is that the present 3G or 3.5G capability is considered to be substantially lower than the predicted future requirements and applications. Also, future systems should be much cheaper for consumers. These concepts can be summarized as follows:

- Next generation networks will provide subscribers with a higher bandwidth and a mobile data rate of 100 Mbps and more.
- It is expected that 3G networks will not be able to meet the needs of services like video-conferencing, full motion video, etc. in terms of quality of service (QOS).
- There will be greater mobility and lower costs.
- It will be possible to integrate WLAN and WAN.

Another goal for the next generation of mobile communication system is to seamlessly integrate wide variety of communication services such as high speed data, video and multimedia traffic, as well as voice signals. The technology needed to meet these challenges is popularly known as fourth generation (4G) mobile systems. 4G is also known as beyond 3G (B3G) and is used to describe the next step in wireless communications.

A 4G system will be able to provide a comprehensive Internet Protocol (IP) solution where voice, data, and streamed multimedia can be given to users on an "anytime, anywhere" basis, and at higher data rates than previous generations. 4G is also expected to converge different wireless systems (WLL, WLAN, PAN, WiMax, UMTS UTRAN, etc.). The 4G version is illustrated in Figure 23.1.

> *4G is not one defined technology or standard, but rather a collection of technologies and protocols aimed at creating fully packet-switched networks optimized for data.*

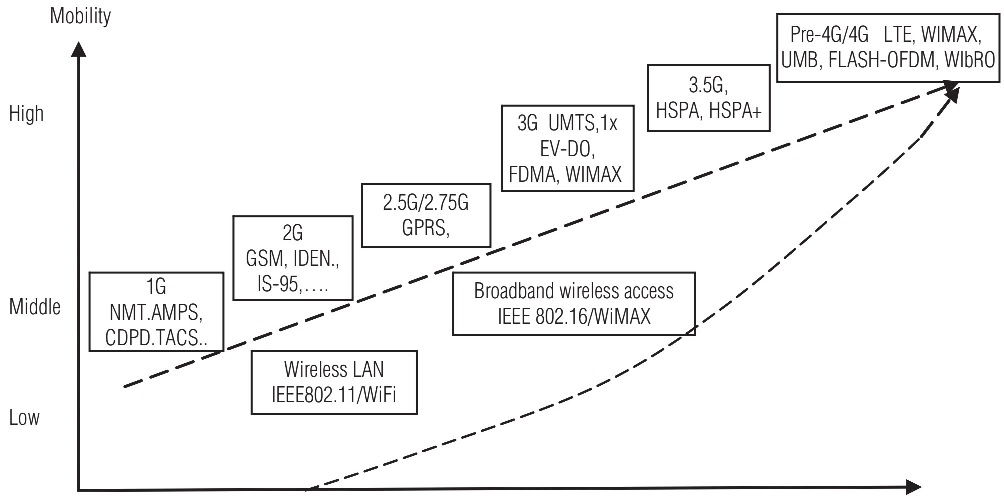


Figure 23.1 Evolution of technology to 4G wireless

For replacing 3G wireless systems, several technologies have been considered such as Ultra Wideband (UWB), Orthogonal Frequency Division Multiplexing (OFDM), Multi-Input Multi-Output (MIMO) antenna systems, and millimetre wireless. The existing Wi-Fi, WiMAX, and Long-Term Evolution (LTE) data networks, which are known as B3G systems, use OFDM technology. These technologies are the reason behind the high performance of 4G systems as is explained in the following:

- UWB is of particular interest for short range, high data rate, and/or low power.
- It makes use of long-term channel prediction schedule amidst users in frequency as well as time.
- Smart antennas are united with power control and adaptive modulation.
- Frequency band: 2–8 GHz. This range enables a mobile to be accessed from any part of the world, that is, worldwide roaming.

The main radio access design parameters of this new system include OFDM waveforms in order to avoid the inter-symbol interference that typically limits the performance of high-speed systems, and MIMO techniques to boost the data rates. At the network layer, an all-IP flat architecture supporting QOS has been defined.

The chapter provides an in-depth view on the technologies being considered for 4G evolution. First, the evolution from 3G to 4G is described in terms of performance requirements and main characteristics. The advantages of 4G over 3G and hardware and software aspects of 4G technology are explained in detail.

23.2 4G evolution

The existence of several diverse 3G standards (e.g., WCDMA and CDMA2000) limits seamless ***global roaming*** between different cellular networks for a mobile user with a single handset. In addition, there is a fundamental difference between wireless cellular networks (1G, 2G, or 3G) and wireless data networks such as WLANs and PANs.

> *The difference between wireless cellular networks (1G, 2G, or 3G) and* ***wireless data*** *networks (WLANS, PANs) is that wireless cellular systems are circuit-switched while wireless data networks are packet-switched.*

As mentioned in the previous chapter, all over the world, the majority of mobile communication service providers are operating the networks using two different families of standards:

- **3GPP** based (i.e., GSM, EDGE, UMTS, HSDPA, HSPA+)
- **3GPP2** based (i.e., CDMA IS-9, CDMA IS-2000, 1x Data Only Revision 0 (1xDO Rev0), 1x Data Only Revision A (1xDO RevA))

Both families of standards were set to evolve to separate 4G technologies:

- 3GPP family to LTE
- 3GPP-2 family to Ultra Mobile Broadband (UMB)

In addition, a brand-new wireless communication standard, IEEE802-16e1, was proposed by IEEE in 2005. Service providers have been participating in the development of LTE and have committed to adopting the LTE technology. With no adoption, work on UMB has been halted and mobile WiMAX has seen very little adoption by mobile communication service providers. Therefore, there is an ongoing convergence of mobile communication standards after many years of diverse standards deployed in various geographical locations (most of the North American networks are based on the 3GPP2 standards and Europe has deployment of networks that are based on the 3GPP standards). Both Americans and Europeans have continuous commitment to the adoption of a single standard (i.e., LTE) by both 3GPP-based and non-3GPP-based service providers.

Convergence issues for these differences between the wireless cellular systems and the wireless data networks will be addressed in the design of 4G cellular networks. It is envisioned that 4G networks, possibly running on a common IP-based backbone, will provide users with seamless wireless access to voice, data, and video services, irrespective of which wireless network they belong to. The fundamental reason for the transition to the all-IP is to have a common platform for all the technologies that have been developed so far, and to harmonize with user expectations of the many services to be provided.

The data rates of 100–155 Mbps (or speeds) are expected in the 4G cellular and to handle these data rate, OFDM, and among the other proposed technologies, UWB and MIMO antenna systems are emerging as the access method of choice. OFDM is spectrally more efficient than existing access methods such as TDMA, FDMA, CDMA, WCDMA, or CDMA2000. We focus on UWB radio due to its unique advantages: high data rate, low power, and resilience to multipath fading effects. Table 23.1 illustrates a timeline for 1G to 4G cellular systems.

OFDM is a successor to frequency hopping and direct sequence CDMA. It has the capability to cancel multipath distortion in a spectrally efficient manner without requiring multiple local oscillators. It uses the principles of IFFT and FFT. Also OFDM uses frequency orthogonality as compared to code orthogonality used in CDMA.

23.2.1 Description of 4G technology evolutionary path

Figure 23.2 shows the technology evolutionary path for 4G technology. The development from first generation analogue systems (1985) to second generation (2G) digital GSM (1992) was the heart of the digital revolution. But much more than this, it was a huge success for standardization emanating from Europe and gradually spreading globally.

Table 23.1 A timeline for 1G to 4G cellular systems

Technology	1G	2G	3G	4G
Design began	1970	1980	1990	2000
Implemented	1983	1991	2003	?
Standards	AMPS, TACS NMT	GSM,CDMA,PDC 2.5G standards are: GPRS, EDGE, 1xRTT, IS95B	WCDMA, CDMA2000	One standard
Service	Analog voice	Digital voice Short messages	Higher capacity Broadband	Higher capacity, Complete IP Oriented, multimedia
Data bandwidth	1 Kbps	14.4 Kbps	2 Mbps	1,000 Mbps
Multiplexing	FDMA	TDMA,CDMA,	CDMA	OFDM

However, world-wide roaming still presents some problems with standards IS-95 (a CDMA rather than a TDMA digital system) and IS-136 (a TDMA variant) still entrenched in some countries. Extensions to GSM (2G) via General Packet Radio Service (GPRS) and EDGE (E-GPRS), as well as WAP and i-mode (so called 2.5G), will allow the transmission of higher data rates as well as speech prior to the introduction of 3G.

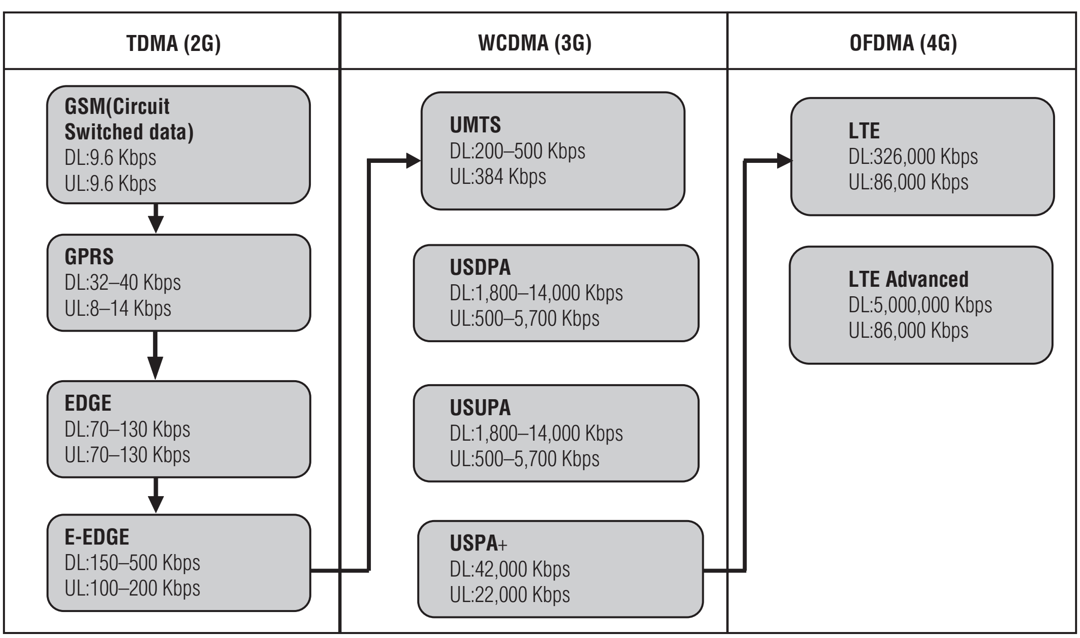


Figure 23.2 Evolutionary path for 4G technology

GSM was designed for digital speech services or for low bit rate data that could fit into a speech channel (e.g., 9.6 Kbps). It is a *circuit* rather than a packet-oriented network and hence is inefficient for data communications. To address the rapid popularity increase of Internet services, GPRS is being added to GSM to allow packet IP communications at up to about 100 Kbps.

3G systems were standardized in 1999. These include International Mobile Telecommunications 2000 (IMT-2000), which was standardized within ITU-R and includes the Universal Mobile Telecommunications System (UMTS) European standard from European Telecommunications Standards Institute (ETSI), the US-derived CDMA2000 and the Japanese NTT DoCoMo Wideband Code Division Multiple Access (W-CDMA) system. Such systems extend services to (multi-rate) high-quality multimedia and to convergent networks of fixed, cellular, and satellite components.

The radio air interface standards are based upon W-CDMA (UTRA FDD and UTRA TDD in UMTS, multicarrier CDMA2000, and single carrier UWC-136 on derived US standards). The core network has not been standardized, but a group of three evolved GSM (MAP), evolved ANSI-41 (from the American National Standards Institute), and IP-based are all candidates. 3G is also about a diversity of terminal types, including many non-voice terminals such as those embedded in all sorts of consumer products. Bluetooth (another standard not within the 3G orbit, but likely to be associated with it) is a short-range system that addresses such applications. Thus, services from a few bits per second up to 2 Mbps can be envisioned.

Long-term Evolution (LTE): LTE is an emerging technology for higher data rates. It is also referred to as 3.9G or super 3G technology. LTE is developed as an improvement to Universal Mobile Telecommunication System by 3rd Generation Partnership Project (3GPP). LTE uses Orthogonal Frequency Division Multiple Access (OFDMA). The download rate in LTE is 150 Mbps and it utilizes the available spectrum in a very sophisticated way. In LTE, the IP packet delay is less than 5 ms which provides the experience of wired broadband Internet access in wireless environment. The mobile TV broadcast is facilitated by LTE over LTE network.

23.3 Objectives of the projected 4G

4G will be a fully IP-based integrated system. This will be achieved after wired and wireless technologies converge and will be capable of providing between 100 Mbps and 1 Gbps speeds, both indoors and outdoors, with quality and high security. 4G will offer all types of services at an affordable cost. The following are the objectives of the 4G wireless communication standard:

- A spectrally efficient system (in bps/Hz and bps/Hz/site).
- High network capacity: more simultaneous users per cell.
- 4G systems support streaming video, voice calls, Internet, and many more broadband services.
- A nominal data rate of 100 Mbps while the client physically moves at high speeds relative to the station, and 1 Gbps while client and station are in relatively fixed positions.
- Smooth handoff across heterogeneous networks.
- Seamless connectivity and global roaming across multiple networks.
- High QOS for next generation multimedia support (real-time audio, high-speed data, HDTV video content, mobile TV, etc).
- Interoperability with existing wireless standards.
- IP-based mobile technology.

23.4 Advantages of 4G network technology over 3G

With the 3G network deployment yet to pick up speed, 4G technology is already in view. If the predictions of the mobile industry experts prove to be true, 4G network deployment may start anytime in this coming decade. Trials are already being conducted by some mobile operators and vendors. Table 23.2 gives a comparison of few key features of 3G and 4G technologies.

The need for the 4G technology is that the present 3G capability is considered to be substantially less than predicted future requirements and applications. Also, future systems should be much cheaper for consumers.

From Table 23.2, we can summarize the 4G concepts as following:

- 4G networks will provide subscribers with a higher bandwidth and a mobile data rate of 100 Mbps and more.
- It is expected that 3G networks will not be able to meet the needs of services like video-conferencing, full motion video, etc. in terms of QOS.
- There will be greater mobility and lower costs.
- It will be possible to integrate WLAN and WAN.

Table 23.2 Comparison of 3G and 4G network technologies

Feature	3G networks	4G networks
Driving force	Predominantly voice driven, data is secondary concern	Converged data and multimedia services over IP
Network architecture	Wide area cell based	Integration of wireless LAN and wide area networks
Frequency band	1.6–2.5 GHz	2–8 GHz
Data rate	385 Kbps–2 Mbps	20–100 Mbps and 1Gbps for stationary
Bandwidth	5 MHz	100+ MHz
Switching technique	Circuit and/Packet switched	Completely digital with packet voice
Radio access technology IP	*CDMA family* (WCDMA/ CDMA2000) Multiple versions	*OFDMA family* (MC-CDMA or OFDM) IPv6.0
Major characteristic	Predominantly voice-data as add-on	Converged data and VoIP
Component design	Optimized antenna; multi-band adapters	Smart antennas; SW multi-band; wideband radios
Forward error correction	Convolution code 1/2, 1/3; turbo	Concatenated coding
Mobile top speed	200 kmph	200 kmph

The technology may see some peculiar features, such as cell phones operating in very high speed vehicles (e.g., trains running at more than 210 km/h). Present subscriber requirements include downloading videos and music etc., but the future seems to be moving towards applications like online games that demand immense capacity, greater QOS, and very low costs. 4G system must be capable of providing highly efficient and cost-effective solutions for wireless network users.

23.5 Applications of 4G

4G offers three-dimensional visual experiences. Thus, 4G will represent another quantum leap in mobile Internet speeds and picture quality.

4G will have better support of roaming and handoffs across heterogeneous networks. Users, even in today's wireless market, demand service transparency and roaming. 4G may support interoperability between disparate network technologies by using techniques such as LAS-CDMA signalling. Other solutions such as software-defined radios could also support roaming across disparate network technologies in 4G systems.

One of the most notable advanced applications for 4G systems is location-based services. 4G location applications would be based on visualized, virtual navigation schemes that would support a remote database containing graphical representations of streets, buildings, and other physical characteristics of a large metropolitan area. This database could be accessed by a subscriber in a moving vehicle equipped with the appropriate wireless device, which would provide the platform on which would appear a virtual representation of the environment ahead. For example, one would be able to see the internal layout of a building during an emergency rescue. This type of application is sometimes referred to as "Telegeoprocessing," which is a combination of Geographical Information Systems (GIS) and Global Positioning Systems (GPS) working in concert over a high-capacity wireless mobile system.

Telegeoprocessing over 4G networks will make it possible for the public safety community to have wireless operational functionality and specialized applications for everyday operations, as well as for crisis management. The emergence of next generation wireless technologies will enhance the effectiveness of the existing methods used by public safety. 3G technologies and beyond could possibly bring the following new features to public safety:

Virtual navigation: As described, a remote database contains the graphical representation of streets, buildings, and physical characteristics of a large metropolis. Blocks of this database are transmitted in rapid sequence to a vehicle, where a rendering program permits the occupants to visualize the environment ahead. They may also "virtually" see the internal layout of buildings to plan an emergency rescue, or to plan to engage hostile elements hidden in the building.

Tele-medicine: A paramedic assisting a victim of a traffic accident in a remote location could access medical records (e.g., x-rays) and establish a video conference so that a remotely based surgeon could provide "on-scene" assistance. In such a circumstance, the paramedic could relay the victim's vital information (recorded locally) back to the hospital in real time, for review by the surgeon.

Crisis-management applications: These arise, for example, as a result of natural disasters where the entire communications infrastructure is in disarray. In such circumstances, restoring communications quickly is essential. With wideband wireless mobile communications, both limited and complete communications capabilities, including Internet and video services, could be set up in a matter of hours. In comparison, it may take days or even weeks to re-establish communications capabilities when a wire line network is rendered inoperable.

23.6 4G technologies

CDMA technology allows every mobile phone in a cell to transmit over the entire bandwidth at all times. Each mobile device has a unique and orthogonal code that is used to encode and recover the signal. The mobile phone digitizes the voice data as it is received, and encodes the data with the unique code for that phone. This is accomplished by taking each bit of the signal and multiplying it by all bits in the unique code for the phone.

CDMA has been patented in the United States by Qualcomm, making it more expensive to implement due to royalty fees. This has been a factor for cellular phone providers when choosing the system to implement. Consumers now demand more features, which in turn require higher data rates than 3G can handle. A new system is needed that merges voice and data into the same digital stream, conserving bandwidth to enable fast data access. By using advanced hardware and software at both ends of the transmission, 4G is the answer to this problem.

4G wireless systems mainly differ from 3G in the following aspects:

- They entirely consist of packet-switched networks.
- All network elements are digital.
- Wide bandwidth is necessary to provide multimedia services like streaming video.
- Network security is more important.

Thus far, several potential wireless technologies have been considered for 4G wireless systems:

- UWB
- OFDM
- MIMO antenna systems

The following subsections describe all the above three technologies briefly. However, UWB is particularly attractive due to its high data rate, low power, and robustness to harsh multipath environments.

23.6.1 Ultra-wideband networks

UWB is a spread spectrum technique that will play a vital role in wireless technologies. The advantages of UWB include low power consumption, short range wireless connectivity, availability of high bit rates, and location capabilities.

Implementation of 4G networks mainly uses UWB transmission technology, which is typically detected as noise. This type of noise could not cause interference with present radio frequency devices. While decoded by other device that will recognize UWB and reassemble it as a signal, it can use the frequency spectrum, that is, it can use the frequencies which are presently used by other radio frequency devices.

UWB signals are generated using sub-nanosecond pulses thus spreading energy over a very large frequency band. Due to the very large bandwidth, no spectrum can be allocated to UWB exclusively. Thus, UWB band overlaps with many other narrowband systems. Therefore, to guarantee existing systems from UWB emissions, UWB operates in the 3.1–10.6 GHz frequency band. Figure 23.3(a) shows the bandwidth comparison of different types of wireless systems with the UWB and Figure 23.3(b) illustrates the UWB spectrum usage.

The pulse can be called “shaped noise” because it is not flat, but curves across the spectrum. On the other hand, actual noise would look the same across a range of frequencies – it has no shape. For this reason, regular noise that may have the same frequency as the pulse itself does

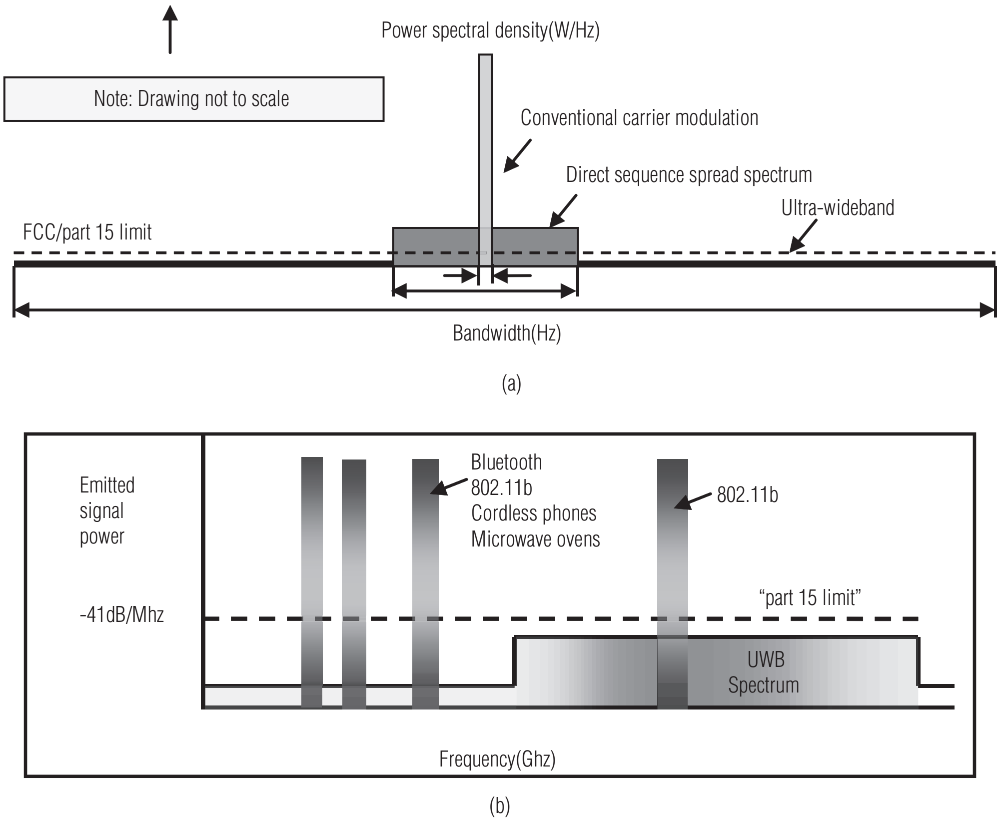


Figure 23.3 (a) UWB bandwidth comparison of different types of wireless systems; (b) UWB spectrum usage

not cancel out the pulse. Interference would have to spread across the spectrum uniformly to obscure the pulse.

Principle of UWB

To understand how UWB works, we start with a generalized UWB signal without defining any modulation and channelization schemes. This signal can be extended for inclusion of data modulation and channelization. A pulse train of a generalized UWB signal can be represented as a sum of pulses which takes the form

$$s(t) = A\sum_{i=-\infty}^{+\infty} \sum_{j=0}^{N_S-1} v(t - jT_f),\ (i-1)T_b < t \le iT_b. \qquad (23.1)$$

where

A is the amplitude of the pulse

N_s is the number of pulses required to transmit a single information bit.

T_b is the bit duration, where $T_b = N_s T_f$.

T_f is the frame time, also known as average pulse repetition period.

The reciprocal of T_f, called the pulse repetition frequency, has a closed effect on system design. $v(t)$ is the basic UWB pulse of duration T_v. The following sections illustrate each of the components in Equation (23.1) and their relationships in more detail.

Ultra-short pulses

The starting point of UWB is to generate the short pulses with which the system communicates. Let us start with an initial waveform of a Gaussian pulse which has the familiar form given below [5, 6]:

$$p(t) = e^{\frac{-t^2}{\tau^2}} \tag{23.2}$$

Taking the first derivative of this equation yields a Gaussian monocycle, which has the form

$$v(t) = \alpha t e^{\frac{-t^2}{\tau^2}} \tag{23.3}$$

where α is the parameter related to the amplitude of the pulse. A typical waveform for a $t = 0.5$ ns width pulse is shown in Figure 23.4.

It can be seen from the waveform in Figure 23.4 that there is a single zero crossing point. Taking additional derivatives yields waveforms with additional zero crossing points, one additional zero crossing for each additional derivative. According to Fourier transform theory, as we take additional derivatives, the relative bandwidth decreases, while the centre frequency increases (for a fixed value of τ). The equivalent of taking derivatives is filtering. This gives us a choice of waveforms to be used as short pulses for UWB, which will depend on the system performance and application requirements. The standard bodies such as FCC and NTIA may also influence the choice of the waveform. For example, if the requirement were to eliminate the signal energy for low frequencies to protect GPS and GSM systems, which operate in relatively low frequency bands, at least one derivative of the Gaussian waveform should be considered, as by taking derivatives, the spectrum tends to move to the higher frequency bands.

A Gaussian monocycle sequence, or "pulse train" can then be generated for data modulation purpose. The pulse train is acting somewhat like a "carrier" which can be used for the purpose of

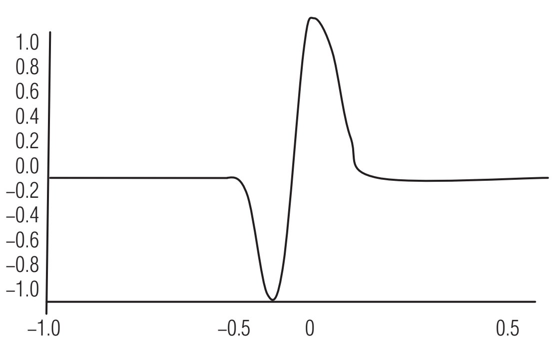


Figure 23.4 Gaussian monocycle pulse

modulation and transmission. In the frequency domain, a pulse train with regular time intervals will produce energy spikes, which might interfere with conventional radio systems. These energy spikes should be minimized to reduce the interference level.

Data modulation

The regular monocycle pulse train contains no information and produces energy spikes. In order to transmit information, the monocycle pulse train needs to be modulated by data. Information transmission can be achieved using a number of ways, including amplitude, time, and phase modulation of the UWB pulses. The modulation needs to reduce energy spikes, thereby minimizing the PSD as required by the regulation. The choice of the modulation schemes also affects the bit error performance.

Three of the popular modulation schemes proposed for UWB transmission are pulse position modulation (PPM), pulse amplitude modulation (PAM), and phase shift keying (PSK). For binary data modulation, PSK is also known as bi-phase modulation. The PPM scheme is discussed in the following.

Pulse position modulation

PPM is based on the encoding information by modifying the time shift between the pulses. Figure 23.5 illustrates a PPM scheme, where the pulse-frame length T_f (pulse repetition period), the symbol length T_b, and the chip length T_c are also indicated. The introduction of the chip length T_c is for the purpose of multi-user communication. As illustrated in Figure 23.5, PPM changes the time of transmission of every monocycle in a data symbol by a time shift, δ. For a binary data sequence, each bit in the data stream is sampled by N_s monocycles.

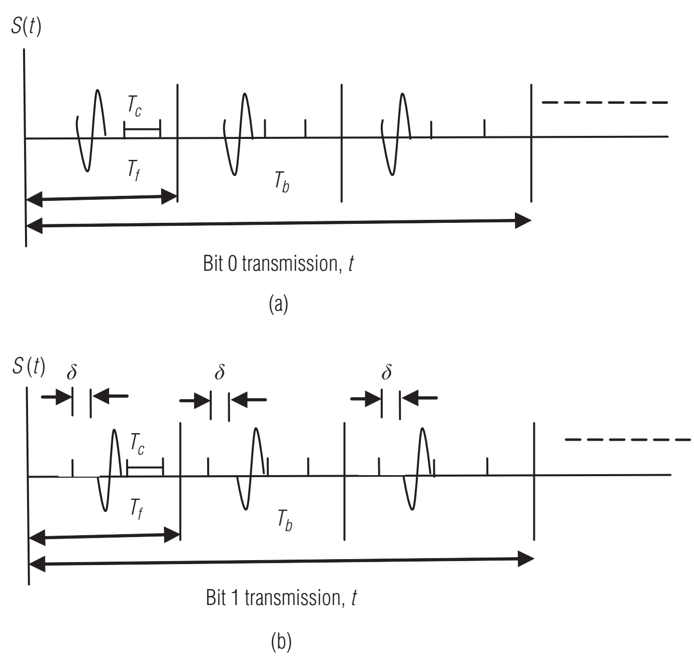


Figure 23.5 Pulse position modulation

In the system shown in Figure 23.5, transmitting three monocycles represents each data symbol. When the data symbol is 0, the transmission of data symbol starts a nominal position of T_c. Because of the time shift δ, PPM distributes the signal energy more uniformly across the spectrum. For the PPM modulation example shown in Figure 23.5, there are three frames per data symbol ($N_s = 3$, the number chosen for the purpose of illustration only). The other two modulation schemes PAM and PSK will use the same assumption. In real applications, there are a few hundred frames per data symbol. This increases the robustness of signal reception, because detection of only a few frames per symbol for a receiver requires very high receiver sensitivity.

The frame length is only slightly longer than the pulse width, which is limited by the graphic illustration. In real systems, the ratio of the frame length to the duration of a monocycle is much larger, resulting in a low-duty-cycle pulse. Usually the chip duration is larger than the pulse width because the pulse width is very short to ensure wide frequency occupancy.

Main features of UWB

- *UWB provides greater bandwidth–as much as 60 Mbps, which is six times faster than today's wireless networks.*
- *It also uses significantly less power, since it transmits pulses instead of a continuous signal.*
- *UWB uses all frequencies from high to low, thereby passing through objects like the sea or layers of rock. Nevertheless, because of the weakness of the UWB signal, special antennas are needed to tune and aim the signal.*

23.6.2 Orthogonal frequency-division multiplexing

It is thought that OFDM will be able to fulfil the three most important requirements of 4G mobile networks: higher coverage and capacity, with desired QOS at minimum cost. OFDM is a bandwidth-efficient signalling scheme for wideband digital communications. Currently, OFDM is used in wireless data networks such as Wi-Max (IEEE802.16) and Wi-Fi (IEEE802.11a/g).

> *OFDM is a frequency-division multiplexing (FDM) technique that is used to transmit large amounts of data on a radio signal. Basically, a large radio signal is sub-divided into smaller signals and then transmitted to the receiver using different frequencies.*
>
> *The main difference between FDM and OFDM is that in OFDM, the spectrums of the individual carriers mutually overlap.*

The biggest advantage of the OFDM technique is the mutual orthogonality of its carriers, which provides high spectral efficiency. This is possible because there is no guard band and carriers can be packed very close together. Most of the alternative techniques like FDMA require guard bands (Figure 23.6). In OFDM, even without a guard band, there is no interference because the carriers are orthogonal. The spectrum for OFDM lies between 200 MHz and about 3.5 GHz, with a spectral efficiency of about 1 bps/Hz.

Coverage in CDMA systems is limited by the phenomenon of *cell breathing* (described elsewhere in this book), as an increasing number of users decreases the area covered owing to an

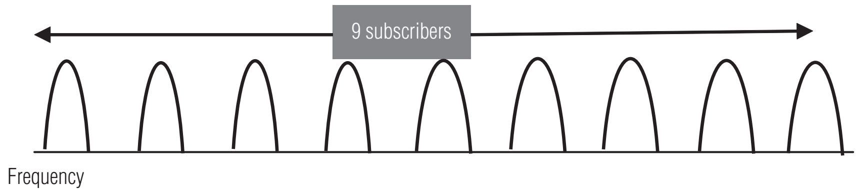


Figure 23.6(a) FDMA with nine sub-carriers using filters

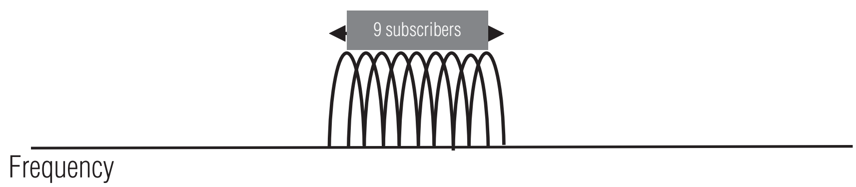


Figure 23.6(b) OFDM with nine sub-carriers

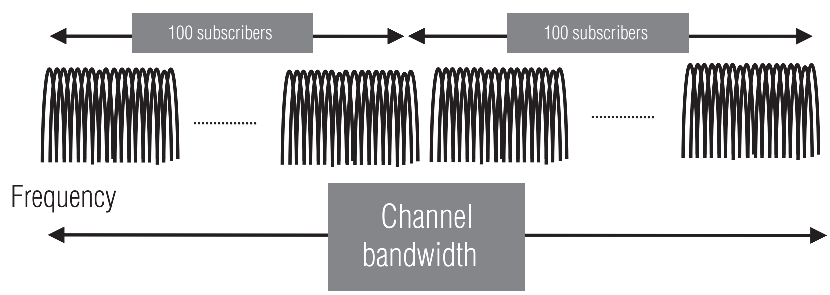


Figure 23.6(c) FDMA and OFDM

increase in interference. In an OFDM system, the *cell overlay technique* is used (similar to that in GSM), thereby reducing co-channel interference.

Network planning for an OFDM system is quite similar to that for GSM/GPRS. This is because frequency re-use is reintroduced (unlike in WCDMA, where the frequency re-use factor was 1, theoretically). For this reason, the power control feature in OFDM networks is not as essential as in WDCMA networks. In WCDMA radio networks, power control and spread spectrum are required for reducing interference. In OFDM radio networks, accurate estimation of frequency offset is required.

Increasing the number of transmitting and receiving antennas can increase capacity. MIMO antenna systems can be used. Network planning for OFDM networks is simpler than for CDMA networks. OFDM reduces the amount of crosstalk in signal transmissions. Thus, in a nutshell,

Table 23.3 Comparison of CDMA and OFDM

Attribute	CDMA	OFDM
Transmission bandwidth	Full system bandwidth	Variable up to full system bandwidth
Frequency selective scheduling	Not possible	A key advantage of OFDM although it requires accurate real time feedback of channel conditions from receiver to transmitter
Symbol period	Very short-inverse of the system bandwidth	Very long-defined by sub-carrier spacing and independent of system bandwidth
Equalization	Difficult above 5 MHz	Easy for OFDM since signal representation is in frequency domain
Resistance to multipath	Difficult above 5 MHz	Completely free of multipath distortion upto the CP length
Suitability for MIMO	Requires significant computing power due to signal being defined in the time domain	Ideal for MIMO due to signal representation in the frequency domain and possibility of narrowband allocation to follow real time variations in the channel
Sensitivity to frequency domain distortion and interference	Averaged across channel by spreading process	Vulnerable narrow band distortion and Interference
Separation of users	Scrambling and orthogonal spreading codes	Frequency and time although scrambling and spreading can be added as well

we can see that OFDM clearly has an edge over CDMA, making it the preferred air-interface technology for future mobile networks.
The advantages of OFDM over the 3G UMTS CDMA technology are illustrated in Table 23.3.

23.6.3 Smart antennas

Multiple “smart antennas” can be employed to help find, tune, and turn up signal information. Since the antennas can both “listen” and “talk,” a smart antenna can send signals back in the same direction that they came from. This means that the antenna system can hear many times louder and respond more loudly and directly as well. There are two types of smart antennas:

- *Switched beam antennas*
- *Adaptive array antennas*

Switched beam antennas (radiation pattern is shown in Figure 23.7) have fixed beams of transmission, and can switch from one pre-defined beam to another when the user with

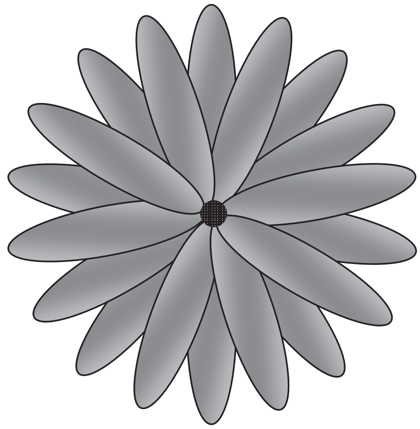


Figure 23.7 Switched beam antenna

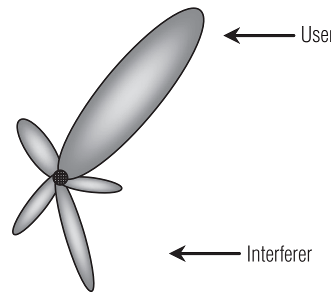


Figure 23.8 Adaptive array antenna

the phone moves throughout the sector. *Adaptive array antennas* (Figure 23.8) represent the most advanced smart antenna approach to date using a variety of new signal processing algorithms to locate and track the user, minimize interference, and maximize intended signal reception.

Smart antennas can thereby

- Optimize available power
- Increase base station range and coverage
- Reuse available spectrum
- Increase bandwidth
- Lengthen battery life of wireless devices

23.7 Smart antenna techniques

MIMO systems are an example of smart antenna technique. These systems use multiple antennas at both the transmitter and receiver to increase the capacity of the wireless channel (Figure 23.9). With MIMO systems, it may be possible to provide in excess of 1 Mbps for 2.5G wireless TDMA EDGE and as high as 20 Mbps for 4G systems.

With MIMO, different signals are transmitted out of each antenna simultaneously in the same bandwidth and then separated at the receiver. With four antennas at the transmitter and receiver, this has the potential to provide four times the data rate of a single antenna system without an increase in the transmitted power or bandwidth.

MIMO techniques can support multiple independent channels in the same bandwidth, provided that there is a direct line-of-sight between the transmitter and receiver. For example, if the number of transmitting antennas are M, and the number of receiving antennas are N, where $N \geq M$, we consider four cases:

- Single-input, single-output (SISO)
- Single-input, multiple-output (SIMO)
- Multiple-input, single-output (MISO)
- Multiple-input, multiple-output (MIMO)

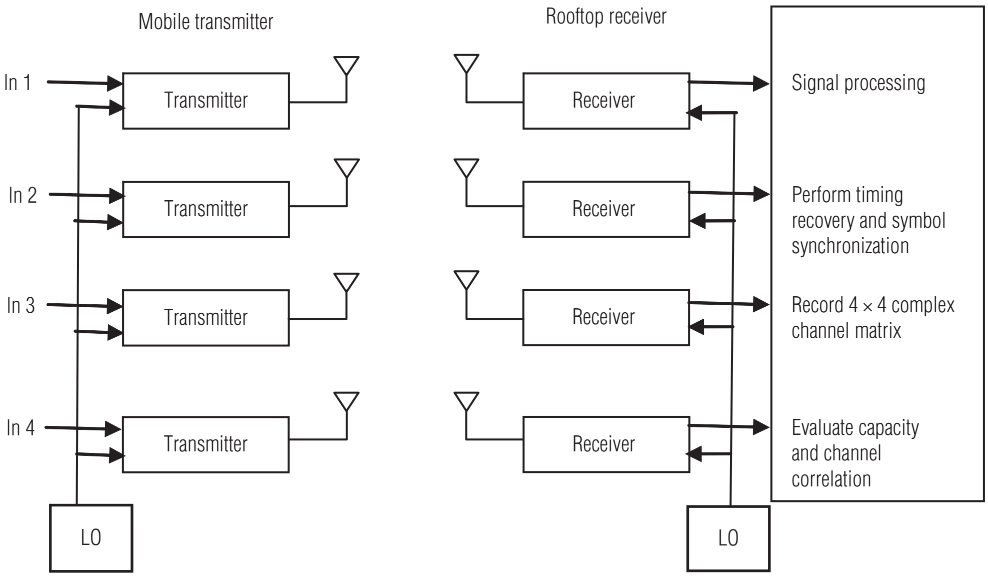


Figure 23.9 MIMO system

Single-input, single-output: If the channel bandwidth is B, the transmitter power is P_t, the signal at the receiver has an average signal-to-noise ratio of SNR_0, then the Shannon limit on channel capacity C is

$$C = B \log_2 (1 + \mathrm{SNR}_0) \text{ bps} \tag{23.4}$$

Single-input, multiple-output: There are N antennas at the receiver. If the signals received on the antennas have on average the same amplitude, then they can be added coherently to produce an N^2 increase in signal power. There are N sets of noise sources that are added coherently and result in an N-fold increase in noise power. Hence, the overall increase in SNR will be

$$\mathrm{SNR} = \frac{N^2 \times (\text{signal power})}{N \times \text{noise}} = N \times \mathrm{SNR}_0 \tag{23.5}$$

The capacity for this channel is approximately equal to

$$C = B \log_2 (1 + N \times \mathrm{SNR}_0) \text{ bps} \tag{23.6}$$

Multiple-input, single-output: We have M transmitting antennas. The total power is divided into M transmitter branches. If the signals add coherently at the receiving antenna, we get an M-fold increase in SNR as compared to SISO. Because there is only one receiving antenna, the noise level is same as SISO. The overall increase in SNR is approximately

$$\mathrm{SNR} = \frac{M^2 [(\text{signal power})/M]}{\text{noise}} = M \times \mathrm{SNR}_0 \tag{23.7}$$

Multiple-input, multiple-output: MIMO systems can be viewed as a combination of MISO and SIMO channels. In this case, it is possible to achieve approximately an *MN*-fold increase in the average SNR_0 giving a channel capacity equal to

$$C_{single} = B \log_2 (1 + (M \times N \times SNR_0) \text{ bps} \tag{23.8}$$

Assuming $N \geq M$, we can send different signals using the same bandwidth and still be able to decode correctly at the receiver. Thus, we are creating a channel for each one of the transmitters. The capacity of each one of these channels is roughly equal to

$$C_{single} = B \log_2 (1 + (N/M) \, SNR_0) \text{ bps} \tag{23.9}$$

Since we have M of these channels (M transmitting antennas), the total capacity of the system is

$$C_{single} = MB \log_2 (1 + (N/M) \, SNR_0) \text{ bps} \tag{23.10}$$

We get a linear increase in capacity with respect to the transmitting antennas. As an example, we assume SNR_0 is equal to 10 dB, $M = 4$, $N = 5$, and bandwidth *B,* (MHz) and list the system capacity for each channel type in Table 23.4.

Although UWB and smart antenna technology may play a large role in a 4G system, advanced software will be needed to process data on both the sending and receiving side. This software should be flexible as the future wireless world will likely be a heterogeneous mix of technologies.

23.8 4G software

4G will become a unification of different wireless networks, including WLAN technologies (e.g., IEEE 802.11), public cellular networks (2.5G, 3G), and even personal area networks (PANs). Under this umbrella, 4G needs to support a wide range of mobile devices that can roam across different types of networks. These devices would have to support different networks, meaning that one device should have the capability of working on different networks. One solution to this "multi-network functional device" is a software-defined radio.

Table 23.4 Comparison of channel capacity for different channel types

Channel Type	Capacity (Mbps)	Normalized capacity with respect to SISO
SISO	3.45 B	1.0
SIMO	5.66 B	1.64
MISO	5.35 B	1.55
MIMO (with same input)	7.64 B	2.21
MIMO (with different inputs)	15 B	4.35

23.8.1 Software-defined radio

A software-defined radio is one that can be configured to any radio or frequency standard through the use of software. For example, if one was a subscriber of Sprint and moved into an area where Sprint did not have service, but Cingular did, the phone would automatically switch from operating on a CDMA frequency to a TDMA frequency. In addition, if a new standard were to be created, the phone would be able to support that new standard with a simple software update. With current phones, this is impossible. A software-defined radio in the context of 4G would be able to work on different broadband networks and would be able to transfer to another network seamlessly while travelling outside the user's home network.

A software-defined radio's best advantage is its great flexibility to be programmed for emerging wireless standards.

By using new software, user can update the programme effectively without any modifications in the hardware and infrastructure. As with software-defined radio, users can download the interface automatically while entering into the new territory and the data should be formatted to some standard, no issues with roaming. Data formation will be chosen by packet layer, which splits into small "packets."

23.8.2 Packet layer

The packet layer separates the information from where it is being transmitted. Packets will follow rules for formatting the data. Until they follow these rules, they can be of any size, any kind of information, and carry this information from one network to another network. Transfer of files, pictures, videos, and other information over Internet is done in the form of packets. Presently, a little fault tolerance is built in the cellular systems. If there is a loss in voice data or interference from other devices occur while transferring between the locations, some effect can be done there. Even though there is a negligible loss, it can cause major issue with sensitive gadgets. All these problems provide to a low QoS.

23.8.3 Implementation of packets

Nowadays, **IP version 4 (IPv4)** is used to locate the devices. Address format of IPv4 is **xxx.xxx. xxx.xxx**, where each set of three digits ranges from 0 to 255(165.128.16.110) that allows around 4.2 billion unique addresses (2^{12}). A number of appliances like cars, refrigerators, phones, etc. are being connected to the Internet by people every day, which will require a larger address space.

IPv6

IP version 6 (IPv6) will be used in the next generation for addressing systems to locate the devices. Due to the lack of IP address space in IPv4, the IPv6 solution was defined. IPv6 has a much larger address space. It uses an address in the following format.

x:x:x:x:x:x:x:x

where x = hexadecimal value that makes up one-eighth of the address. This address format allows 3.40×10^{38} unique addresses which is 8.05×10^{28} times larger to the IPv4 address space. For every phone, there are enough addresses for unique address. In future, the user can use the phone through VoIP service over Internet. The transition from IPv4 to IPv6 can be achieved with the following methods:

- By using hosts with the *Dual Ipv4/IPv6 Stack* approach in which a device can communicate with either an IPv6 or IPv4 in the network layer74.

- By using *Tunnelling* in which IPv6 intermediate routers when needed to forward traffic in IPv4 routers, encapsulate the IPv6 datagram inside an IPv4 datagram74 above.
- By using *IPv4 to IPv6 Protocol Translators* with the use of Network Address Translation-Protocol Translation75 (NAT-PT).

Figure 23.10 presents a schema for the transition phases from IPv4 to IPv6. Initially, all nodes support IPv4. In the first phase, the first IPv6 Intranets appear and some mobile devices support the dual stack IPv4/IPv6. In the second phase, Ipv6 is widely deployed in the Internet but tunnelling is still used for the remaining IPv4 Intranets. The NAT-PT routers are responsible for traffic between the IPv4 and IPv6 networks. In the third and final phase, IPv6 is the dominant protocol and mobile devices no longer need to support a dual stack.

Voice over IP

Current standard used for voice communication over data networks is Voice over IP (VoIP). As several standards came into existence for VoIP, International Multimedia Telecommunications Consortium Standard H.323 is the primary one. PBX-based systems are already being replaced by VoIP in many offices and companies that offer inexpensive long distance phone calls over Internet, like Go2Call and Net2Phone. It supports packet-based communication. As there is an interconnection between the data Internet and telephone network, VoIP customers can communicate with each other as well as with the users of old telephone system.

VoIP allows slow transition from direct and connection-based communication to VoIP communication. Old style telephone users are allowed to connect to their central office (CO) by replacing their backbones. Then CO will connect to an IPv6 Internet backbone, and then connect to the destination CO. At the user end, there is no difference but the communication will occur primarily through a packet-based system. Impertinent users detect all the data, including voice, should they be encrypted while in transit.

23.9 Limitations of 4G

4G is still passing through research and, therefore, there are some problems that need to be fixed in order for the users to benefit from it fully. Still there are some limitations for 4G communications.

- Operating area is one of the major limitations. Although 2G networks are frequently used, still there are many areas not served. This drawback passes over into future generation.
- The extensive publicity to 3G networks nowadays gives unrealistic expectations in communications available at anytime, anywhere. The public should understand that high speed data communications will not be equivalent to the wired Internet.
- When 3G and 4G network services are deployed, there may be a great deal of disappointment and perceptions could become negative. If this were to happen, neither 3G nor 4G may realize its full potential.
- Cost is another limitation. The equipment required to implement a next generation network is still very expensive. Carriers and providers have to plan carefully to make sure that expenses are kept realistic.
- Pay-per-use model of services are currently being applicable in the Asian networks which are difficult to implement in United States, where the public is used to a service for free model (e.g., the Internet).

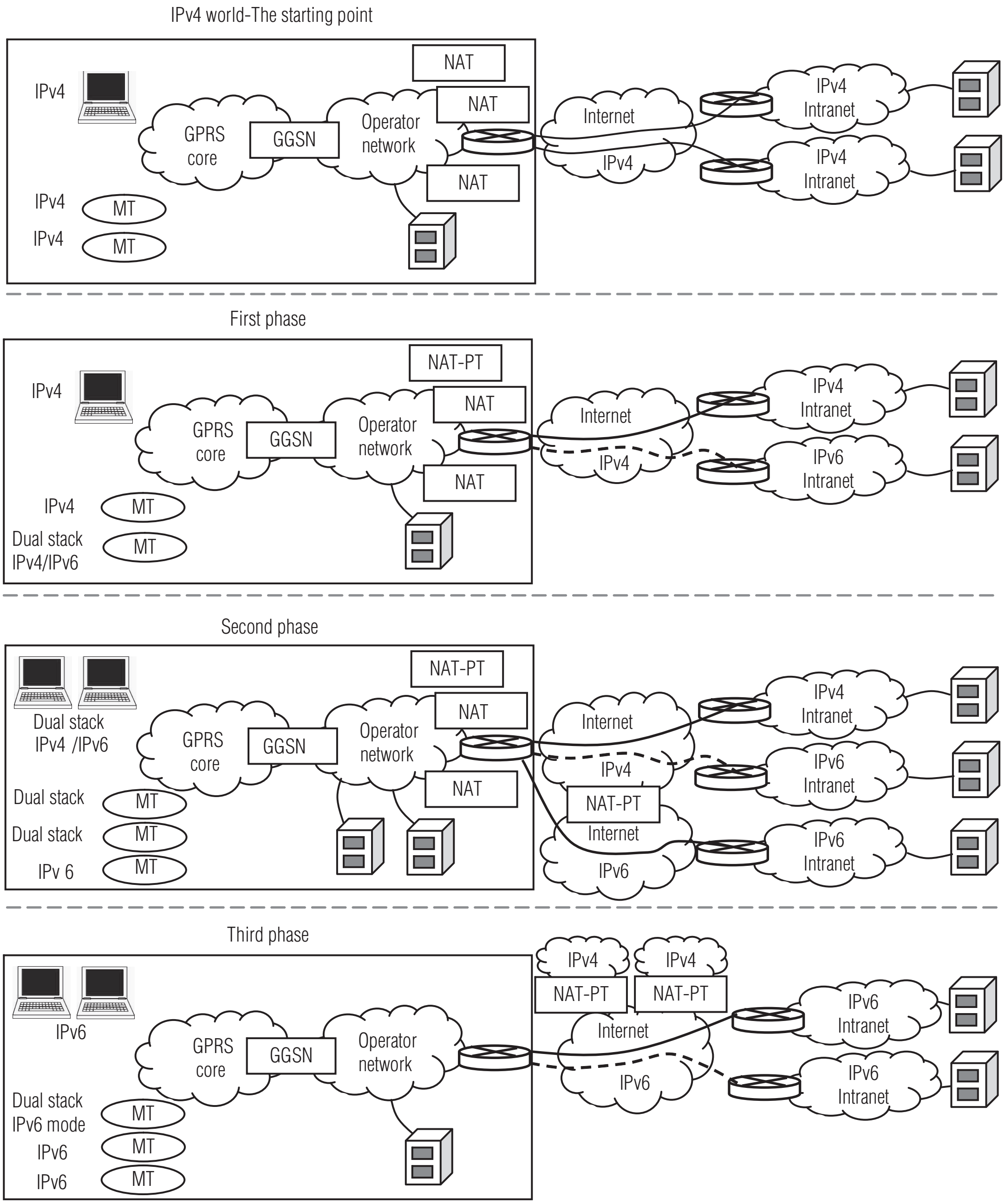


Figure 23.10 IPv4 to IPv6 Proposed Transition Phases

23.10 New technologies in cellular data networks

This section looks at a few new generation technologies in cellular data networks. Technologies introduced in this section include High-speed OFDM Packet Access, UMB, Worldwide Interoperability for Microwave Access, and Mobile Broadband Wireless Access/IEEE802.20 Personal Communication System and Virtual Private Networking.

23.10.1 High-speed OFDM packet access

High-speed OFDM Packet Access (HSOPA) is a proposed part of 3GPP's LTE upgrade path for UMTS systems also called Super 3G, but HSOPA is an entirely new air-interface system and unrelated and incompatible with W-CDMA. HSOPA has a flexible bandwidth usage of 1.25 MHz–20 MHz and has an increased spectral efficiency of 2–4 times compared to 3GPP release 6. The peak transfer rates can approach 100 Mbps for downlink and 50 Mbps for uplink. The roundtrip latency times from terminal to radio access network is around 20 ms, better than W-CDMA and almost the same as combined HSDPA/HSUPA system. New core technologies adapted in HSOPA are OFDM and MIMO. These two give HSOPA the ability to enlarge the users' number by 10 times compared to W-CDMA. HSOPA also decreases the processing power consumption on each handset. Theoretical maximum data transmission rate is 40 Mbps. The best experimental performance is around 37 Mbps in the downlink over a 5-MHz channel currently.

23.10.2 Ultra mobile broadband

UMB is a project within 3GPP2 to improve the CDMA2000 standard for next generation applications and requirements, based upon Internet (TCP/IP) networking technologies with peak rates of up to 280 Mbps. UMB has OFDMA-based air interface within a scalable bandwidth between 1.25–20 MHz. It supports mixed cell sizes including macro-cells, micro-cells, and pico-cells. Advanced antenna techniques like MIMO, SDMA (Spatial-Division Multiple Access), and beam forming are used to get significantly higher data rates and to reduce the latencies. It has a higher RL (reverse link) sector capacity with quasi-orthogonal reverse link. Adaptive interference management and dynamic fractional frequency reuse are implemented to increase the data rates.

Distributed RL power control is based on other cells' interference. Real-time services are enabled by fast seamless L1/L2 handover. Independent RL & FL (forward link) handover provide better airlink and handover performance. Power optimization is applied with quick paging and semi-connected state. Flexible airlink resource management and RL CDMA control channels are used. New scalable IP architecture supports inter-technology handover. New handover mechanisms support real-time services throughout the network. Fast acquisition and efficient multi-carrier operation are applied by using beacons. Multi-carrier configuration supports incremental deployment, mix of low complexity, and wideband devices.

23.10.3 Worldwide interoperability for microwave access/IEEE802.16

WIMAX is a broadband wireless solution that enables convergence of mobile and fixed broadband networks through a common wide area broadband radio access technology and flexible network architecture. It is based on the IEEE802.16 standard. In the MAC layer, it uses a scheduling algorithm for which the subscriber station needs to compete once. After that, it is allocated an access slot by the base station. The time slot can either be enlarged or remain assigned to the specific subscriber station. This scheduling algorithm can also be more bandwidth efficient. In recent updated 802.16e, scalable orthogonal frequency-division multiple access (SOFDMA) and MIMO brings potential benefits in terms of coverage, self installation, power consumption, frequency re-use, and bandwidth efficiency. The inclusion of MIMO antenna techniques along with flexible sub-channelization schemes, Advanced Coding and Modulation, all enable WIMAX to support peak DL data rates up to 63 Mbps per sector and peak UL data rates up to 28 Mbps per sector in a 10 MHz channel. QOS in 802.16 defines Service Flows which can map to DiffServ code points or MPLS flow labels that enable end-to-end IP. A flexible mechanism is also provided for optimal scheduling of space, frequency, and time resources over the air interface on a frame-by-frame basis. WIMAX is designed to work in different channels from 1.25 to 20 MHz to comply with varied worldwide requirements as efforts proceed to achieve spectrum harmonization in the longer term. This benefit helps WIMAX make affordable Internet access within various specific geographic needs from rural settings to the metro and suburban areas. The features provided for Mobile WIMAX security aspects are best in class with EAP-based authentication, AES-CCM-based authenticated encryption, and CMAC and HMAC based control message protection schemes. User credentials exist including SIM/USIM cards, smart cards, digital certificates, and username/password schemes. WIMAX supports optimized handover schemes with latencies less than 50 ms to ensure real-time applications without service degradation. Flexible key management schemes guarantee that security is maintained during handover.

23.10.4 Mobile broadband wireless access/IEEE802.20

IEEE802.20 or MBWA aims to prepare a formal specification for a packet-based air interface designed for IP-based services which is of low-cost and is always connected. The air interface will operate in bands below 3.5 GHz and with a peak data rate of over 1 Mbps. IEEE802.20 applies OFDM and MIMO in PHY to well use the resource in time, frequency, and space so as to increase the system spectrum efficiency. MBWA is based on pure IP infrastructure to deal with high demanding data service. The capability is better than 3G and equal 3.5G (HSPDA, EV-DO). Thus, the costs in application and deployment are decreased. High mobility with 250 km/h may lead 802.20 and IEEE into the mobile communication field. High frequency efficiency exceeding 1 bps/Hz/cell makes a clear selling point and is better than 0.5 bps/Hz/cell in CDMA 2000 1x. IEEE 802.20 fully supports both real and non-real time services. There is no such distinction between circuit domain and network domain in air interface. 802.20 can maintain and keep continuous connection, make the multiplexing in the same frequency. It supports seamless handover between zones and sectors and also the Media Independent Handover between other wireless technical solutions (e.g., 802.16, 802.11). 802.20 supports same end-to-end QOS as the core network level and Ipv4 and Ipv6 protocols which have QOS guarantee. 802.20 is able to auto-select the optimal data transfer rate based on the changes of channel environment in order

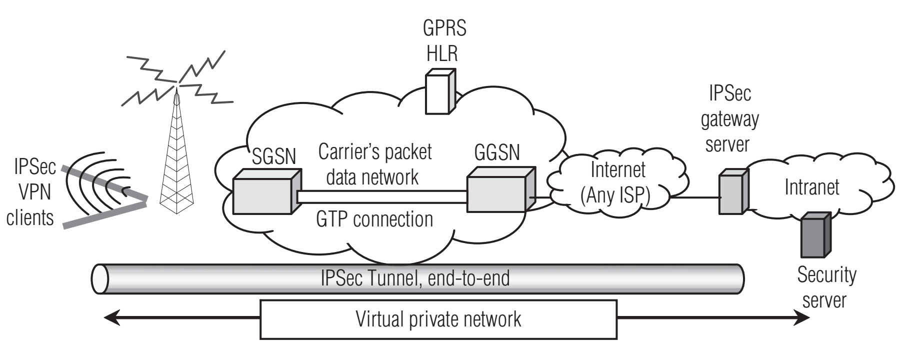


Figure 23.11 VPN between mobile client and gateway server

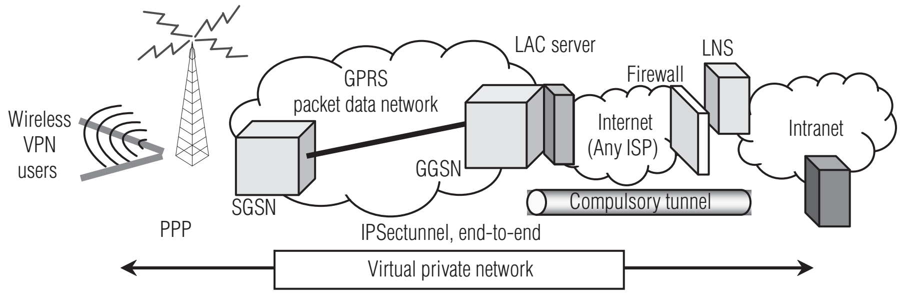


Figure 23.12 VPN between GGSN and corporate firewall

to fast reallocate the resource for UL and DL. Authentication mechanism is offered for the end-to-end network. 802.20 can co-exist with those existing cellular mobile communication systems.

23.10.5 Personal communication system

Another research initiative was launched by the European Community to develop an advanced communication network for Europe which is intended to incorporate the same service on fixed as well as on mobile radio networks. The idea is to establish a PCS that allows mobility of both users and services.

> *The main feature of PCS is the concept of personal mobility. Whether a subscriber is in the house, in the car, or in the office, they should be able to use the same terminal, at any time with any of the allowed access methods using their personal identification number.*

In addition, the advanced service features and variety of data transmission types will have to be supported by such systems. It is anticipated that PCS need an enormous capacity, which must be met with new technology. The mixture of applications implies that new access methods must be negotiated in order to host different data types, such as speech and video. Also the cell sizes have to become smaller to allow higher capacity in city environments. Since the users of cellular phones are getting used to pocket sized telephones, one cannot expect that the phones of PCS to be any larger. Therefore, the batteries have to be efficient enough to allow the use of high power transmissions in rural areas.

There are some situations in which providing radio coverage with terrestrial-based cellular networks is neither economically viable (such as remote, sparsely-populated areas) nor physically practical (such as over large bodies of water). In these cases, satellite-based cellular systems can be the best solution. By the use of many LEO satellites, a complete coverage of the world is possible with low power telephones. Motorola's Iridium project is an example of such a system. It seems that PCS will consist of a mixture of technologies, and the mobile terminal must be able to switch between systems so that a system that fits the user's occasion best is used. In this case, another kind of handover, to be referred as *intersystem handover,* will be essential.

23.10.6 Virtual private networking

One of the most useful applications for enterprise users using 2.5G/3G data networks is a VPN link between their laptop (which is connected to the mobile phone) and to any host within their corporate Intranet. VPN is creating secure communication by encrypting data traffic in OSI layer 2 (data link layer) with the use of *Point to Point Tunnelling Protocol* (PPTP69) or the Layer 2 Tunnelling Protocol (L2TP70). For security in OSI layer 3 (network layer), the IPSec71 is used. The latter can be implemented between a laptop and a corporate Intranet gateway server or between a cellular provider's GGSN and a corporate firewall (Figure 23.12).

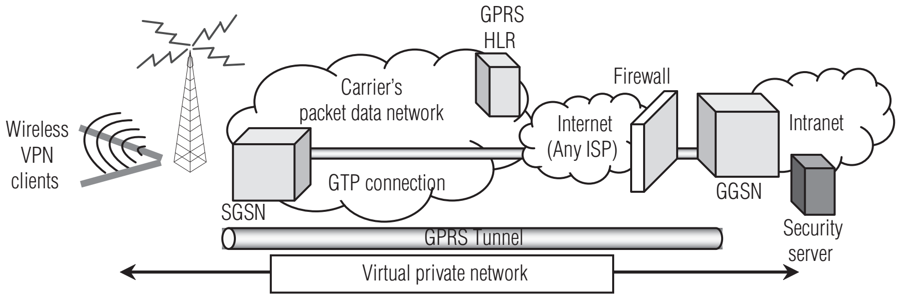


Figure 23.13 VPN between SGSN and GGSN behind the corporate firewall

A proposed alternative is an implementation between a cellular provider's SGSN and a GGSN server located behind the corporate firewall for GTP tunnelling over the Internet (Figure 23.13).

23.11 Summary

- 4G is not one defined technology or standard, but rather a collection of technologies and protocols aimed at creating fully packet-switched networks optimized for data. 4G networks are projected to provide speeds of 100 Mbps while moving and 1 Gbps while stationary.
- 3G networks provide the ability to transfer voice data and non-voice data (music downloads, e-mails and instant messaging) over the same network simultaneously. 3G networks deliver broadband capacity and support greater numbers of voice and data customers at lower incremental costs than 2G.
- 3G versus 4G: 4G provides less complexity and faster transmission. Unlike the 3G networks which are a combination of circuit switched and packet switched networks, 4G will be based on packet switching only. This will allow low-latency data transmission.
- 4G will provide a comprehensive IP solution where voice, data, and multimedia can be given to user on an anytime, anywhere basis.
- Technologies used in 4G are OFDM, UWB, Smart antennas, and Pv6
- OFDM transmits large amounts of digital data over a radio wave.
- OFDM works by splitting the radio signal into multiple smaller sub-signals that are then transmitted simultaneously at different frequencies to the receiver.
- By inserting a cyclic prefix between adjacent OFDM signal, inter-signal interference is virtually eliminated if the maximum channel delay spread is less than the time interval of cyclic prefix.
- In OFDM, the sub-carrier pulse used for transmission is rectangular.
- Here modulation can be performed by an IDFT, which can be generated very efficiently as an IFFT. So, the receiver only needs an FFT to reverse this process.
- UWB: An advanced technology that can be used in 4G technology. It is typically detected as noise.
- It can use any part of the frequency spectrum, which means that it can use frequencies that are currently in use by other radio frequency devices.
- It uses a frequency of 3.1–10.6 Hz.
- It uses less power, since it transmits pulse instead of continuous signal.
- Special antennas are needed to tune and aim the signal

- Smart antenna can send signal back in the same direction that they come from. There are two types of smart antennas: switched beam and adaptive array.
- Switched beam antenna has fixed beams of transmission, and switch from predefined beam to another when the user with the phone moves throughout the sector.
- Adaptive array antenna: It represents the most advanced smart antenna approach to data using a variety of new signal. It represents the most advanced smart antenna approach to date using a variety of new signal processing algorithms to locate and track the user, minimize interference, and maximize intended signal reception.
- A software-defined radio is one that can be configured to any radio or frequency standard through the use of software. The phone should automatically switch from operating on a CDMA frequency to a TDMA frequency whenever it is required. Roaming can be an issue with different standards, but with a software-defined radio, users can just download the interface upon entering new territory, or the software could just download automatically.

Review questions

1. Describe the salient features of 4G systems.
2. Compare the 3G and 4G systems.
3. What is the software-defined radio system?
4. Describe the IPv6 addressing scheme in 4G system.
5. Describe the role of smart antennas in the 4G systems.
6. What is a MIMO system? Explain.
7. What is OFDM and describe how it is useful in an air interface system in the 4G?
8. Compare the performance of UWB with respect to wideband and narrowband systems.
9. What is the basic difference between cellular mobile networks and wireless data networks such as WLANS?

Objective type questions and answers

1. What is the bandwidth of 4G technology?
   (a) 100–150 Mbps (b) 2 Mbps (c) 14.4 Kbps (d) 1 Kbps
2. What is the multiplexing used in 4G technology?
   (a) CDMA (b) WCDMA (c) OFDM (d) FDMA
3. 3G wireless data network uses?
   (a) circuit switched (b) packet switched (c) both (d) none
4. What is the switching technology used in 4G technology?
   (a) circuit switched (b) packet switched
   (c) digital with packet voice (d) both (a) & (b)
5. Which type of antenna is used in OFDM technique?
   (a) Yagi uda (b) smart antenna (c) log periodic (d) micro strip antenna
6. What is the main application of 4G technology?
   (a) virtual navigation (b) geo mapping
   (c) high speed data transmission (d) all the above

Answers: 1. (a), 2. (c), 3. (c), 4. (c), 5. (b), 6. (d).

Open book questions

1. Define the fourth generation (4G) cellular system.
2. What are the various advantages of UWB?
3. What are the salient features of 4G wireless systems?
4. Compare the features of various wireless networks starting from 1G to 4G technologies.

Key equations

1. A pulse train of a generalized UWB signal can be represented as a sum of pulses which takes the form

$$s(t) = A\sum_{i=-\infty}^{+\infty}\sum_{j=0}^{N_S-1} v(t - jT_f), \quad (i-1)T_b < t \le iT_b$$

2. The Shannon limit on channel capacity C is

$$C = B\log_2(1 + \mathrm{SNR}_0)\ \text{bps}$$

Further reading

Wang, S., and Cui, Y. "Intelligent Mobility Support for IPv6," 33rd IEEE Conference on Local Computer Networks, LCN 2008, Montreal, Quebec, 2008, pp. 403–410.

Misra, A., and Das, S. "IDMP-based Fast Handoffs and Paging in IP-based 4G Mobile Networks," *Communication Magazine*, Vol. 40, No. 3, 2002, 138–145.

Bell Labs, "The Bell Labs Security Framework: Making the Case for End-to-End Wi-Fi Security," September, 2006.

S-MIP: A Seamless Handoff Architecture for Mobile IP, Robert Hsieh, Zhe Guang Zhou, Aruna Seneviratne School of Electrical Engineering and Telecommunications, The University of New South Wales Sydney, Australia

Muraleedharan, R., and Osadciw, L. A. "Increasing QOS and Security in 4G Networks Using Cognitive Intelligence," IEEE Globecom Workshops, 2007.doi:10.1109/GLOCOMW.2007.4437837.

Lilliberg, J., and Prasad, R., "Research Challenges for 3G and Paving the Way for Emerging New Generations," *Wireless Personal Communications: An International Journal*, Vol. 17, No. 2–3, June 2001, 355–362.

Varshney, U., and Jain, R. "Issues in Emerging 4G Wireless Networks," *IEEE Computer*, Vol. 34, No. 6, June 2001, 94–96.

Wideband Orthogonal Frequency division Multiplexing (W-OFDM), Wi-LAN Inc., version 1.0, 2000.

Chang, R. W., and Gibby, R. A., "A Theoretical Study of Performance of an Orthogonal Multiplexing Data Transmission Scheme," *IEEE Transactions on Communications*, Vol. 16, No. 4, August 1968, 529–540.

Cruz-Perez, F. A., and Ortigoza-Guerrero, L. "Resource Sharing for Data Services (RSDS)," *Proceedings of the IEEE Vehicular Technology Conference*, Vol. 2, October 2001, 995–999.

Whale, P. C. "Challenges and Trends in 3G Devices," *Proceedings of the International Conference on 3G Mobile Communication Technologies*, Vol. 494, June 2003, 173–177.

Wang, J. *Broadband Wireless Communications: 3G, 4G and Wireless LAN*. Norwell: Kluwer Academic Publishers, 2001.

Wireless Local Area Networks 24

24.1 Introduction

Wireless technology is an alternative to wired technology, which is commonly used for connecting devices in wireless mode. Wireless local area networks (WLANs) transfer data through the air using radio waves instead of cables. The key advantages of the wireless LAN are that it eliminates the laying of cables and wiring cost, and that it accommodates mobile PC workstations. WLANs use the industrial scientific and medical (ISM) frequency bands 900 MHz, 2.4 GHz, and 5 GHz, for wireless LAN products and they need not obtain license to operate in this band. Wireless networks are standardized by IEEE. Under IEEE 802 committee, the wireless LAN and metropolitan area network (MAN) standards are developed. The first wireless network standard was created by IEEE in 1997 and had been named as 802.11. It uses 2.4 GHz frequency and the supported maximum network bandwidth is 2 Mbps. Later, IEEE 802.11b was created in July 1999 and the supported network bandwidth is 11 Mbps. The 802.11b uses radio frequency (2.4 GHz) as same as the original 802.11 standard.

> *Wireless fidelity (Wi-Fi) is a generic term that refers to the IEEE 802.11 communications standard for WLANs. Wi-Fi is used to connect computers to communicate each other, to the wired network, and to the Internet.*

An updated version of the original 802.11 standard was created and called 802.11a. Due to its higher cost, 802.11a is usually found on business networks whereas 802.11b better serves the home market. 802.11g was created to combine the best of both 802.11a and 802.11b and supports network bandwidth up to 54 Mbps. 802.11n is the newest IEEE standard in the 802.11 family. It improves on 802.11g in the amount of bandwidth by using multiple wireless signals and antennas instead of one. This technology is called multiple-input multiple-output (MIMO). It will also operate on the 2.4 GHz band.

The IEEE 802.11 standard is relating to physical (PHY) and medium access control (MAC) layer. There is only one standard for MAC layer, but different standards have been proposed for the PHY layer. The standard defines several different modulation methods: infrared, direct-sequence spread spectrum (DSSS), frequency-hopping spread spectrum (FHSS), orthogonal frequency-division multiplexing (OFDM), and also defines three different PHY layer technologies: IEEE 802.11a, IEEE 802.11b, and IEEE 802.11g.

Wired Ethernets offer data rates of 100 Mbps, and the performance gap between wired and wireless LANs is likely to increase over time without additional spectrum allocation. *Despite the* *big data rate differences, wireless LANs are becoming the preferred Internet access method due to their convenience and freedom from wires.*

For better understanding of how 802.11 and its variants (802.11a, IEEE 802.11b (Wi-Fi), IEEE 802.11g, and IEEE 802.11n) work, it is important to understand the basic building blocks and technologies used in the standards. The 802.11 architecture, wireless topologies, high-performance radio LANs (HIPERLANs), wireless personal area network (WPAN) standards, and wireless local loop (WLL) technology are further discussed in this chapter.

24.2 Advantages and disadvantages of wireless local area network

The following are some specific advantages of wireless LANs over wired LANs:

- *Mobility:* Wireless LANs support mobility. This improves the real-time access to information even when the user is moving from one place to another within the range of an access point (AP).
- *Different topologies:* Different wireless networks are configured in two different modes: ad-hoc mode and infrastructure mode.
  "*Ad-hoc*" mode provides peer-to-peer communication between wireless devices. "*Infrastructure*" mode provides communication between wireless device and a central node, which in turn can communicate with wired nodes on that LAN. In wired LAN, Ethernet cables must be run from each computer to another computer or to the central device. It can be time-consuming and difficult to run cables under the floor or through walls, especially when computers are placed in different rooms.
- *Flexible architecture:* It is easier to add or remove workstations.
- *Cost effective:* Although the initial investment required for wireless local area network (WLAN) hardware can be similar to the cost of wired LAN hardware, installation expenses can be significantly lower.

The following are the disadvantages of WLANs:

- *Less security*: Wireless LANs are less secure than wired LANs, because wireless communication signals travel through the air and can easily be intercepted by others using the same frequency band and by multipath fading. *Automatic repeat request (ARQ) and forward error-correction (FEC) techniques are used to increase reliability.*
- *Low data rates*: The data transfer rate decreases with increase in the number of devices.
- *Need for energy efficient*: In mobile applications, battery power is a scarce resource. Therefore, the devices must be designed to be energy efficient.
- *Limited coverage*: Devices will only operate at a limited distance from an AP, with the distance determined by the standard used and buildings and other obstacles between the AP and the user.

24.3 WLAN topologies

Wireless network topology is the configuration in which a mobile terminal (MT) communicates with another. WLANs can be built with either of the following topologies:

- Peer-to-peer (ad-hoc) topology
- Infrastructure topology

24.3.1 Ad-hoc network topology

Ad-hoc or peer-to-peer network topology applies to reconfigurable networks that can operate without need for a fixed infrastructure. This is the easiest WLAN mode to configure and requires the least hardware.

An ad-hoc mode WLAN is comprised of two or more computers communicating directly with each other using wireless network cards.

The ad-hoc network topology is shown in Figure 24.1. These networks are primarily used by the military and also in a few commercial applications for voice and data transmission.

Multi-hop ad-hoc networks: In some ad-hoc networking applications, where users may be distributed over a wide area, a given user terminal may be able to reach only a portion of the other users in the network due to transmitter signal power limitations. In this situation, user terminals will have to cooperate in carrying messages across the network between widely separated stations. Networks designed to function this way are called multi-hop ad-hoc networks. In an ad-hoc multi-hop network, each terminal should be aware of the neighbouring terminals in its coverage range.

24.3.2 Infrastructure network topology

In the infrastructure topology, there is a fixed (wired) infrastructure that supports communication between MTs and between MTs and fixed terminals.

The infrastructure networks are often designed for large coverage areas and multiple base station or AP operations. In this WLAN mode, a hardware or software AP is configured as part of the WLAN design. This AP then provides connectivity for all of the systems on the WLAN. The

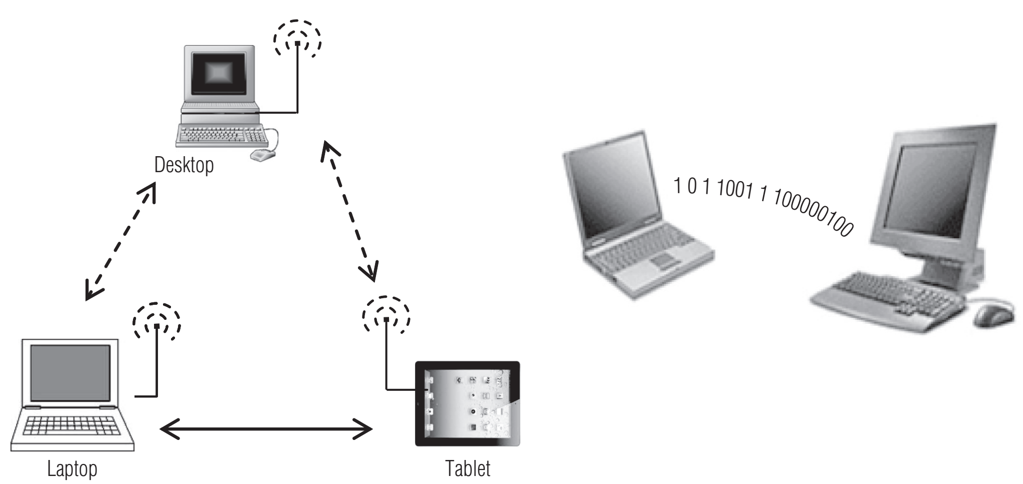


Figure 24.1 Ad-hoc network topology

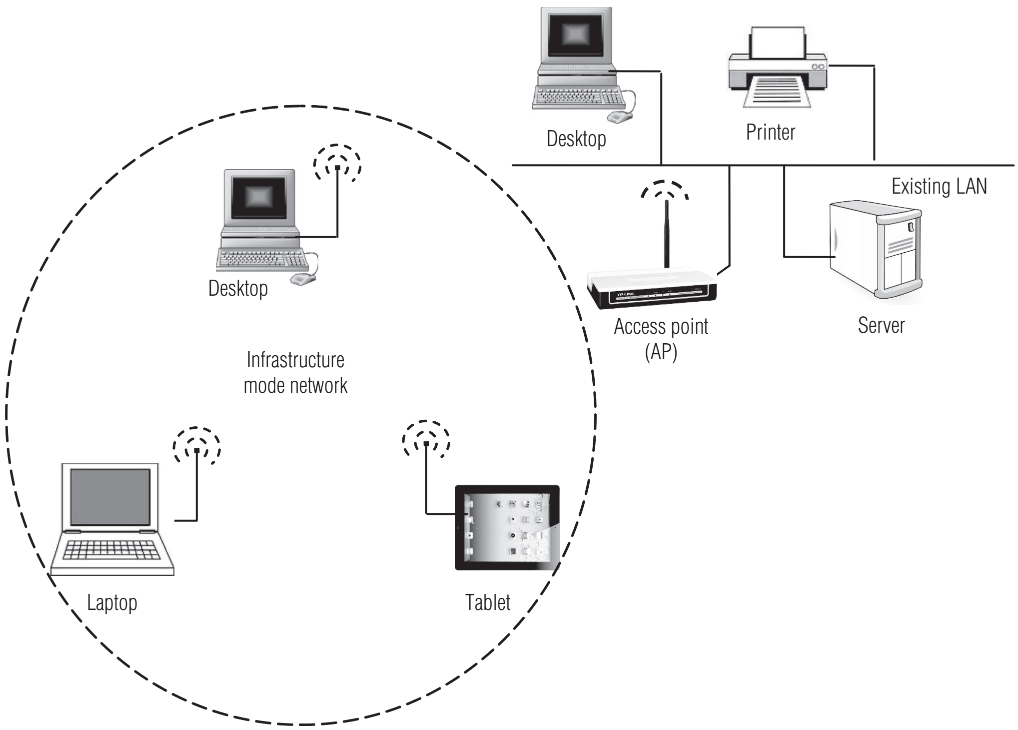


Figure 24.2 Infrastructure network topology

wireless network card on each computer is configured to use a specific AP to connect to a specific WLAN and all traffic to other computers on the WLAN is directed through the AP.

> *Access point : The AP is a wireless LAN transceiver or "base station" that can connect one or many wireless devices simultaneously to the Internet.*

The AP coordinates transmission and reception from multiple wireless devices within a specific range; the range and number of devices depend on the wireless standard being used and vendor's product. In infrastructure mode, there may be multiple APs to cover a large area or only a single AP for a small area such as a single home or small building. Figure 24.2 shows the basic operation of an infrastructure network with a single AP.

All standardized cellular mobile telephone and wireless data systems use an infrastructure network topology to serve MTs operating within the coverage area of any base station.

> *Infrastructure topology uses APs to bridge traffic onto a wired (Ethernet or token ring) or a wireless backbone.*

24.3.3 Comparison of ad-hoc and infrastructure network topologies

Ad-hoc mode WLANs are very easy to configure and do not require a great deal of effort to set up. Whereas infrastructure mode WLANs are slightly more difficult to set up than ad-hoc-mode WLANs.

In ad-hoc or peer-to-peer single-hop networks, expansion is always limited to the coverage of the radio transmitter and receiver. In multi-hop ad-hoc networks, as the number of terminals increases, the potential coverage of the network is increased. However, the traffic handling capacity of the network remains the same. To connect an ad-hoc network to the backbone of a wired network, one needs to use proxy server with a wireless connection as a member of the ad-hoc network. In practice, all terminals supporting ad-hoc networking operate in a dual mode that also supports the infrastructure operation. Wireless infrastructure networks are inherently scalable. To scale up with a wireless infrastructure network, the number of base stations or APs is increased to expand the coverage area or to increase the capacity while using the same available spectrum. Therefore, for wide area coverage and for applications with variable traffic loads, infrastructure networks are always used.

Flexibility

Operation of infrastructure networks requires deployment of a network infrastructure which is very often time-consuming and expensive. Ad-hoc networks are inherently flexible and can be set up instantly. Therefore, ad-hoc networks are always used for temporary applications where flexibility is of prime importance.

Controllability

To coordinate proper operation of a radio network, one need to centrally control certain features such as time synchronization, transmitted power of the mobile stations operating in a certain area, and so on. In an infrastructure network, all these features are naturally implemented in the base station or AP. In an ad-hoc network, implementation of these features requires more complicated structures demanding changes in all terminals.

Routing complexity

In multi-hop peer-to-peer networks, each terminal should be able to route messages to other terminals. This compatibility requires each terminal to monitor the existence of other terminals and be able to connect to those available in the immediate neighbourhood. For this, there is a need for a routing algorithm that directs information to the next appropriate terminal. Implementation of these features adds to the complexity of the terminal and the network operation. In infrastructure and peer-to-peer single-hop ad-hoc networks, this problem does not exist.

Reliability

Infrastructure networks are "single failure point" networks. If the base station or AP fails, the entire communications network is destroyed. This problem does not exist in ad-hoc peer-to-peer configurations.

24.4 Introduction to wireless local area network standard IEEE 802.11

The IEEE 802.11 is an international standard that describes WLAN characteristics. The Wi-Fi corresponds to the certification name issued by the Wi-Fi Alliance group. The IEEE 802.11 standard ensures the compatibility between the hardware devices.

Wi-Fi provides high-speed connections to laptops, desktop computers, and personal digital assistants (PDAs) located within a radius of 20–50 m for indoor and 100 m for outdoor communication applications. Wi-Fi providers started providing the Internet APs in public locations such as train stations, airports, and hotels with wireless networks. These Internet access areas are referred to as “Wi-Fi hotspots”.

The 802.11 is a specific standard that defines the MAC and PHY layers of a WLAN. The original 802.11 standard is a MAC standard plus a low data rate PHY which supports only 1 and 2 Mbps data rates. This first version of the standard operates at the 2.4 GHz ISM (industrial, scientific, and medical) band and allows the vendors to choose between a direct-sequence spread spectrum (DSSS) and a frequency-hopping spread spectrum (FHSS) implementations. As mentioned above, 802.11b is a PHY extension to the original 802.11 standard. It also operates at the 2.40 GHz band and allows for higher data rates of 5.5 and 11 Mbps. It uses a technique known as complementary code keying (CCK).

The 802.11a is another PHY extension to the 802.11 standard. It operates at the 5 GHz unlicensed band and allows for data rates of 6–54 Mbps. It uses a technique known as orthogonal frequency-division multiplexing (OFDM).

The 802.11g was the next extension to the 802.11 standard. It operates at the 2.4 GHz ISM band and allows for data rates ranging from 1 to 54 Mbps. The 1 and 2 Mbps rates are operated in the DSSS mode whereas the 5.5 and 11 Mbps rates are operated in CCK mode. In addition, rates at 6–54 Mbps are operated in OFDM mode. The 802.11g standard borrows the OFDM technique and data rates from the 802.11a standard but operates at the 2.4 GHz ISM band. It can therefore operate at very high data rates while being backward compatible with the 802.11b standard.

In addition to these standards, which have already been approved, the 802.11 committee has “working groups” to evolve and enhance the standard.

All three versions share the same MAC layer that uses carrier sense multiple access with collision avoidance (CSMA/CA) for contention data, a request-to-send/clear-to-send (RTS/CTS) mechanism to accommodate the hidden terminal problem, and an optional mechanism called point coordination function (PCF) to support time-bounded applications. The 802.11 standard supports both infrastructure WLANs connection through an AP (AP) and ad-hoc operation allowing peer-to-peer communication between terminals.

Compared with wired LANs, WLANs operate in a difficult medium for communication, and they need to support mobility and security. The wireless medium has serious bandwidth limitations and frequency regulations. It suffers from time and location dependent multi-path fading. It is subject to other interference from other WLANs, as well as other radio and non-radio devices operating in the vicinity of a WLAN. Wireless standards need to have provisions to support mobility that is not shared in the other LAN standards. The IEEE 802.11 body had to examine connection management, link reliability management, and power management – none of which were concerns for the 802 standards. In addition, WLANs have no PHY boundaries, and they overlap with each other, and therefore a standard is needed to define provisions for security of the links.

24.4.1 IEEE 802.11 architecture

The architecture comprises of the *station*, *AP*, *wireless medium*, *basic service set* (BSS), *distribution system* (DS), and *extended service set* (ESS). The architecture also includes station services and distribution services.

- *Station:* The component mobile, portable, or stationary that connects to the wireless medium in a network is referred to as station. All stations are equipped with wireless network interface controllers (WNICs) and consist of MAC and PHY.
- ***BSS:*** The BSS is a set of all stations that can communicate with each other. There are three types of BSS:

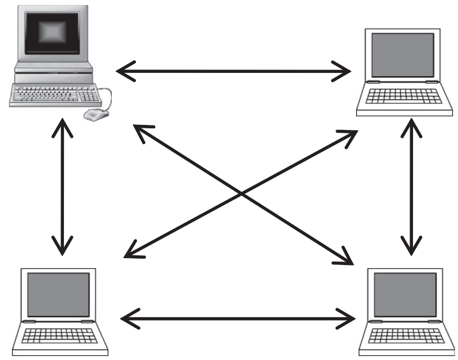


Figure 24.3 Independent basic service set (IBSS)

- Independent BSS (IBSS)
- Infrastructure BSS
- Extended service set (ESS)

Independent BSS**:** An IBSS is an ad-hoc network that contains no APs, which means they cannot connect to any other BSS. As its name implies, ad-hoc networks are temporary in nature, which are typically created and maintained as needed without prior administrative arrangement. Ad-hoc networks can be formed anywhere spontaneously and can be disbanded after a limited period of time. All the mobile stations in an IBSS can communicate directly with others. This does not mean that every mobile station communicates with every other mobile station, but they are all part of the same IBSS. The mobile stations must be in direct communication range to communicate with each other. The IBSS is shown in Figure 24.3.

Infrastructure BSS**:** A base station with an AP is called an infrastructure BSS or simply referred to as BSS (Figure 24.4). In an infrastructure BSS, all mobile stations communicate with the AP. The AP provides connection to both the wired LAN, if any, and the local relay functions for the BSS. Each mobile station must communicate with other mobile stations through the AP. Every BSS has an identification (ID) called the BSSID, which is the MAC address of the AP servicing the BSS. The infrastructure mode BSS is shown in Figure 24.5.

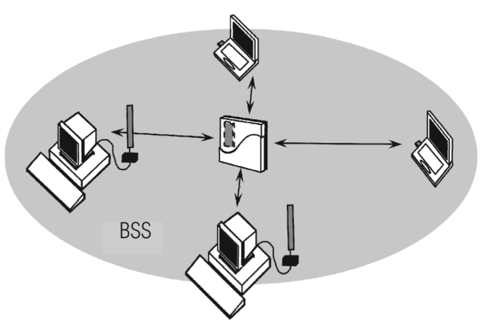


Figure 24.4 Infrastructure BSS

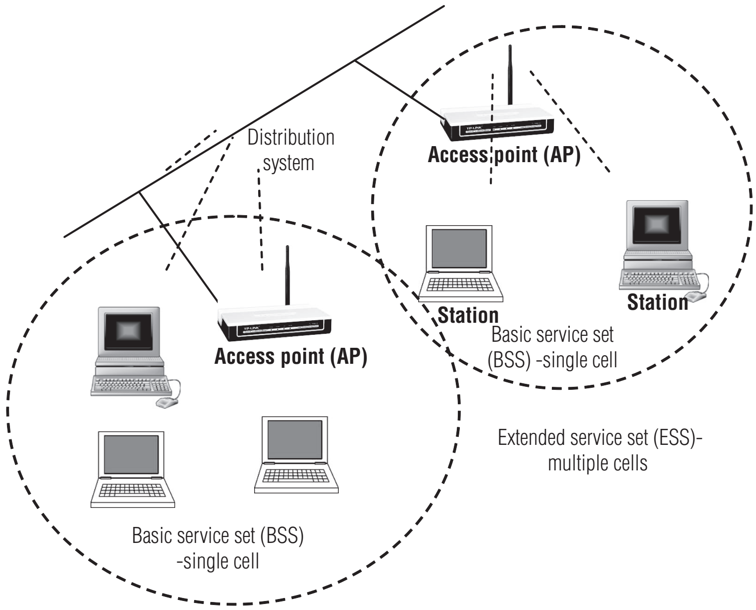


Figure 24.5 Infrastructure mode BSS

***Extended service set (ESS)*:** ESS is referred to as *a set of one or more BSS interconnected by a distribution system (DS*). DS is a system that connects a set of BSSs. In ESS, traffic always flows via AP as shown in Figure 24.6. This concept of DS increases network coverage. Each ESS has an ID called the SSID which is a 32 byte (maximum) character string. ESS configurations consist of multiple BSS cells that can be linked by either wired or wireless backbones. IEEE 802.11 supports ESS configurations in which multiple cells use the same channel and also use different channels to boost aggregate throughput. Communications received by an AP from the DS are transmitted to the BSS to be received by the destination mobile station. An ESS can also provide gateway access for wireless users into a wired network.

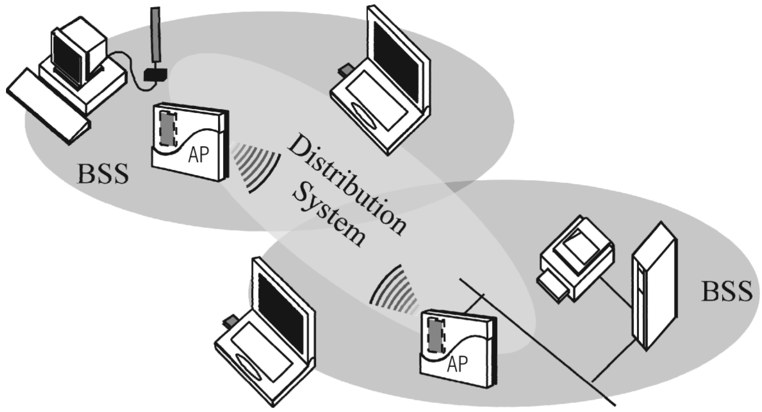


Figure 24.6 Extended service set (ESS) with wireless distribution system

The entire network in the ESS looks like an IBSS to the logical link control (LLC) layer. This means that stations within the ESS can communicate or even move between BSSs transparently to the LLC.

IEEE 802.11 architecture and services

The IEEE 802.11 architecture specifies nine services. These services are divided into two groups: station services and distribution services. There are four station services and five distribution services.

1. *Station Services*
   Services that are common to all stations are referred to as station services. The four station services are authentication, de-authentication, privacy, and data delivery. The authentication and de-authentication services allow only the authorized users to use the network. The authentication service provides the identity of one station to another. So a station without the identity is not allowed to use WLAN services. The de-authentication service is used to eliminate a previously authorized user from accessing the services of the network. The privacy service of IEEE 802.11 protects the data only as it traverses the wireless medium. It is not designed to provide complete protection of data between applications running over a mixed network. Data delivery services ensure that data are transported reliably over the wireless medium. This service provides reliable delivery of data frames from the MAC in one station to the MAC in one or more other stations, with minimal duplication and reordering of frames.
2. *Distribution Services*
   These services are also known as the DS services (DSS). There are five different services, and these services are provided across a DS. The five distribution services are association, disassociation, distribution, integration with wired network like LANs, and re-association. These services allow the users to move freely within an ESS and allow an IEEE 802.11 WLAN to connect with the wired LAN infrastructure. The distribution services determine how to forward frames within the IEEE 802.11 WLAN and also how to deliver frames from one IEEE 802.11 WLAN to network destinations outside of the WLAN.

   The wireless station uses the *association and disassociation* services to gain access and remove access to WLAN services. The *association* service is used to make a logical connection between a mobile station and an AP. Each station must become associated with an AP before it is allowed to send data through the AP onto the DS. The connection is necessary in order for the DS to know where and how to deliver data to the mobile station. The mobile station invokes the association service once and only once, typically when the station enters the WLAN.

   The *disassociation* service is used either to force a mobile station to eliminate an association with an AP or for a mobile station to inform an AP that it no longer requires the services of the WLAN. When a station becomes disassociated, it must begin a new association by invoking the association service. An AP may force a station or stations to disassociate because of resource restraints, the AP is shutting down or being removed from the network for a variety of reasons. When a mobile station is aware that it will no longer require the services of an AP, it may invoke the disassociation service to notify the AP that the logical connection to the services of the AP from this mobile station is no longer required.

   The *re-association* service enables a station to change its current association with an AP. The re-association service is similar to the association service, with the exception that it includes information about the AP with which a mobile station has been previously associated. A mobile station will use the re-association service repeatedly as it moves throughout the ESS, loses contact with the AP with which it is associated, and needs to become associated with a new AP. By using the re-association service, a mobile station provides information to the AP to which it will be associated and information pertaining to the AP from which it will be disassociated. This allows the newly associated AP to contact the previously associated AP to obtain frames that may be waiting there for delivery to the mobile station as well as other information that may be relevant to the new association. The mobile station always initiates re-association.

Distribution is the primary service used by an 802.11 station. A station uses the distribution service every time it sends MAC frames across the DS. The distribution service provides the distribution with only enough information to determine the proper destination BSS for the MAC frame. The three association services (i.e. association, re-association, and disassociation) provide the necessary information for the distribution service to operate. Distribution within the DS does not necessarily involve any additional features outside of the association services, though a station must be associated with an AP for the distribution service to forward frames properly.

The *integration* service connects the 802.11 WLAN to other LANs, including one or more wired LANs or IEEE 802.11 WLANs. The integration service translates 802.11 frames-to-frames that may traverse another network, as well as translates frames from other networks to frames that may be delivered by an 802.11 WLAN.

IEEE 802.11 family and its standards

The open system interconnection (OSI) reference model is a seven-layer model for the functions that occur in a communication process. Each layer performs a number of related functions in the communications process. The same model applies to both source and the destination of the information and the layers are paired.

The layers at the same level in the protocol stack are referred to as peer processes. The PHY connection between two layers of the protocol stacks occurs at the PHY layer. At each of the remaining layers, there is a virtual connection between peers. The protocol at *n*th layer communicates with a peer at the same level. The dashed line in the figure indicates an example in which layers of network layer on one side communicates with the network layer on the other side (Figure 24.7).

The 802.11 is a specific standard that defines the MAC and PHY layers of a bottom two levels of the ISO model (Figure 24.8). These two layers are explained briefly in the following paragraphs.

PHY layer: This layer provides a PHY mechanism for transmitting binary bits between any pair of nodes. The module for performing this function is called a *modem*.

The key resource of WLAN is the radio spectrum. At the PHY layer the emphasis is on modulation, source coding, channel coding, and detection techniques to maximize the use of the radio spectrum.

***Data link layer*:** This layer performs error detection and correction in order to provide a reliable error free link to higher layers. The role of the data link layer is more complicated when multiple nodes share the same media for transmission/reception. The component of the data link layer that controls multiple access communications is the MAC sublayer. The MAC allows frames to be sent over the shared media without undue interference from other nodes. At the link layer, the emphasis is on how spectrum is shared, in either time, frequency, area, or angular direction.

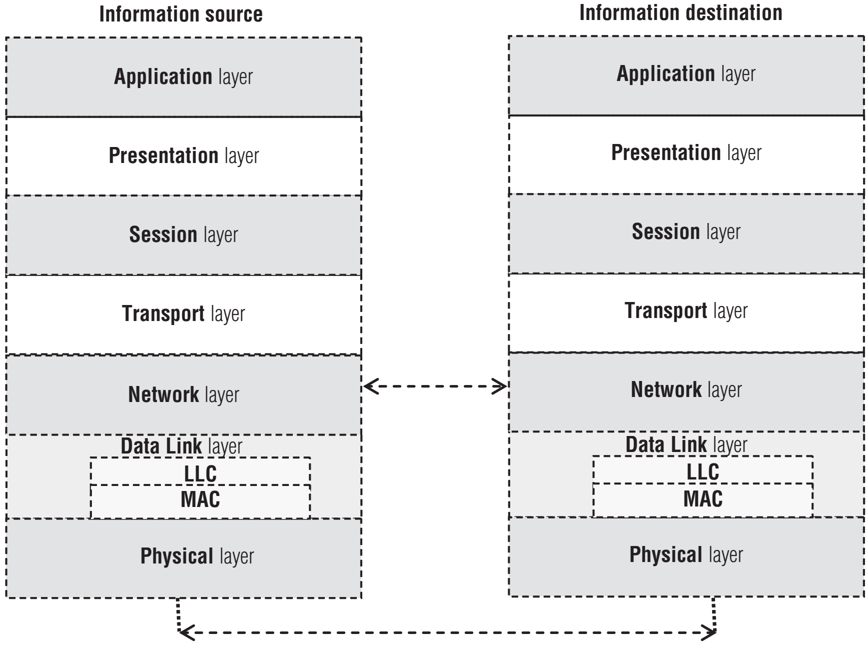


Figure 24.7 OSI reference model

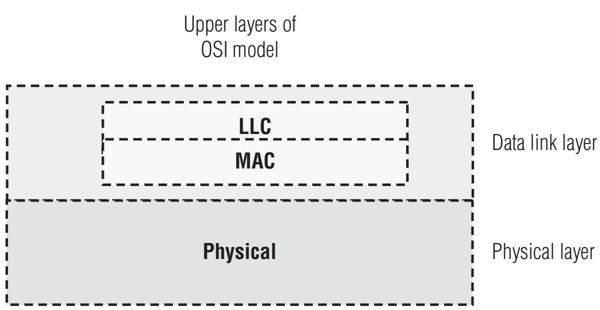


Figure 24.8 802.11 standard focuses on the bottom two levels of the ISO model: PHY and MAC

The original 802.11 standard is a MAC standard plus a low data rate PHY which supports only 1 and 2 Mbps data rates. This first version of the standard operates at the 2.4 GHz ISM band and allows the vendors to choose between a DSSS and a FHSS implementations. The 802.11b is a PHY extension to the original 802.11 standard. It also operates at the 2.40 GHz band and allows for higher data rates of 5.5 and 11 Mbps. It uses a technique known as complementary code keying (CCK).

The 802.11a is another PHY extension to the 802.11 standard. It operates at the 5 GHz unlicensed national infrastructure for information (UNII) band and allows for data rates of 6–54 Mbps. It uses a technique known as orthogonal frequency-division multiplexing (OFDM).

The 802.11g was the next extension to the 802.11b standard. It operates at the 2.4 GHz ISM band and allows for data rates ranging from 1 to 54 Mbps. The 1 and 2 Mbps rates are operated in the DSSS mode, whereas the 5.5 and 11 Mbps rates are operated in CCK mode. In addition, rates at 6–54 Mbps are operated in OFDM mode. The 802.11g standard borrows the OFDM technique and data rates from the 802.11a standard but operates at the 2.4 GHz ISM band. It can therefore operate at very high data rates while being backward compatible with the 802.11b standard. In addition to these standards, which have already been approved, the 802.11 committee has "working groups" to evolve and enhance the standard.

802.11 a and g standards

The frequency standards used in IEEE 802.11b, Bluetooth, and IEEE 802.11a are shown in the Table 24.1.

The three commonly known versions of the 802.11 PHY are 802.11a, 802.11b, and 802.11g. As described earlier, the 802.11a and 802.11g standards offer much higher speed than 802.11b. However, the advent of 802.11a and 802.11g will not necessarily result in the demise of 802.11b in the immediate future. There are applications that would require the lowest power consumption and/or the lowest system cost, and in such cases a stand-alone 802.11b solution may still be the best solution in the immediate future. On the other hand, most system vendors have migrated to 802.11g solutions, which are backward compatible with 802.11b and allow the higher data rates.

As an alternative to 802.11b and 802.11g, if the operator requires a higher data rate, higher user density, and network capacity, he or she would have to choose 802.11a because of the availability of a much wider spectrum at the 5 GHz band and the higher data rates offered by 802.11a.

For longer ranges and higher data rate applications, the operator would probably choose 802.11g. The 802.11g offers the added benefit of being backward compatible with 802.11b, which has the largest existing base. In terms of data rate, the 802.11b and 802.11g have an advantage, with rates up to 54 Mbps. In terms of range of operation, the 802.11b and 802.11g have the advantage because they operate at the lower frequency of 2.4 GHz. Since typically propagation losses are lower at lower frequencies, 802.11b and 802.11g systems would be able to operate over longer distances as compared to their 802.11a counterpart for a given transmit power and receiver sensitivity.

The free-space loss for cases in which the receiver-to-transmitter distance is much larger than the wavelength is given by the relation

$$L = \left(\frac{4\Pi d}{\lambda}\right)^2 = \left(\frac{4\Pi df}{c}\right)^2 \tag{24.1}$$

Table 24.1 Comparison of physical characteristics of various data network standards operating in ISM bands

	IEEE 802.11b	Bluetooth	IEEE 802.11a
Frequency band	2.4 GHz	2.4–2.4385 GHz	5.2 GHz
Channel bandwidth	54 MHz	80 MHz	20 MHz
Data rates	up to 11 Mbps	<1Mbps	up to 54 Mbps
Access strategy	CSMA/CA	FH/TDD	FDMA/CSMA
Network topology	Point to multipoint	Point-to-point Connection and connectionless	Point to multipoint

where

L is the propagation loss
d is the distance between the transmitter and the receiver
λ is the wavelength of the radio-frequency (RF) signal
f is the frequency of the signal
and c is the speed of light

Antenna gains, absorption losses, reflective losses, and several other factors are not considered in the Equation (24.1).

For example, at a distance of 10 m in free space and with the assumptions listed above, a 802.11g system operating at 2.4 GHz would experience 60 dB of propagation attenuation, whereas an 802.11a system operating at 5.8 GHz would experience 68 dB of propagation losses.

From a spectrum availability point of view, the 802.11a has several hundreds of megahertz of bandwidth available to it (although the exact frequencies would depend on the country of operation). In most countries, on the other hand, there is not more than 100 MHz available for users in the 802.11b or 802.11g bands.

From a power consumption point of view, 802.11b would win against the other standards. This is because it utilizes the simplest modulation technique among the three and therefore does not require a high-performance radio front-end or a sophisticated signal processing baseband. In particular, an 802.11b modulated signal has a small peak to average ratio, and therefore one can use higher efficiency (but lower linearity) power amplifiers on the transmit side.

From a system cost point of view, currently 802.11b offers the lowest system cost. However, the difference in the cost between 802.11g systems and 802.11b systems has been reducing quickly, and today most users are willing to pay slightly higher cost for 802.11g system for the significant gains in throughput.

Other IEEE 802.11 standards

802.11e: Tasked to improve quality of service (QOS). The inclusion of a QOS protocol is essential for tasks that require low latency such as VOIP.

802.11i: Tasked to improve encryption. A reliable and hard-to-break encryption technique is essential for the wide adoption of WLAN by the enterprise customer.

802.11f: Would allow for an inter access protocol for easy communication between APs.

802.11h: Allows for dynamic frequency selection (DFS), and transmit power control. By utilizing DFS, interference between various users would be reduced, and therefore the effective capacity of the cell and therefore the network would increase. Further, by utilizing transmit power control, the minimum required transmit power would be utilized in communication between the APs and the mobile units. This would also reduce co-channel interference and therefore increase the network capacity.

802.11n: Increases the maximum raw data rate from 54 Mbps to 600 Mbps using multiple-input multiple-output (MIMO). MIMO is a technology which uses multiple antennas to coherently resolve more information than possible using a single antenna. One way it provides this is through spatial division multiplexing (SDM). SDM spatially multiplexes multiple independent data streams, transferred simultaneously within one spectral channel of bandwidth. MIMO SDM can significantly increase data throughput as the number of resolved spatial data streams is increased. Each spatial stream requires a discrete antenna at both the transmitter and the receiver. In addition, MIMO technology requires a separate radio-frequency (RF) chain and analogue-to-digital converter for each MIMO antenna which translates to higher implementation costs compared to non-MIMO systems.

The IEEE 802.11 standard defines two bottom layers in the open systems interconnect (OSI) model, namely the MAC layer and the PHY layer. The MAC layer controls and regulates the access to the shared wireless medium using specified channel access mechanisms and the PHY layer manages the transmission of data between the AP and the client.

Example problem 24.1

Consider an IEEE 802.11a WLAN system in which OFDM baseband modulation scheme is used. The OFDM system has 52 subcarriers out of which four subcarriers are used as pilot subcarriers and the remaining as data subcarriers. OFDM symbol duration including guard interval for ISI mitigation is 4 μs. If the system uses three-fourth of the FEC code rate and 64-QAM carrier modulation scheme then show that the achievable transmission data rate is 54 Mbps.

Solution

Given data:
Total number of subcarriers in OFDM system = 52
Number of subcarriers used as pilot subcarriers = 4
FEC code rate = ¾
Type of carrier modulation used = 64 QAM
OFDM data symbol duration = 4 μs
Number of data subcarriers = 52 – 4 = 48 subcarriers
As 64-level or 26-QAM technique corresponds to 6 bits per symbol, then
Number of data bits transmitted per OFDM symbol = 6 × ¾ × 48
Number of data bits transmitted per OFDM symbol = 216 bits
Transmission data rate = 216 bits / 4 μs
Hence, transmission data rate = 54 Mbps

24.4.2 IEEE 802.11 physical layer

The PHY layer is the interface between the MAC and wireless media that provides transmission and reception of data frames over a shared wireless medium. The IEEE 802.11 standard defines four different transmission techniques with four different PHY implementations as follows:

- Direct-sequence spread spectrum (DSSS)
- Orthogonal frequency-division multiplexing (OFDM)
- Frequency-hopping spread spectrum (FHSS)
- Infrared (IR)
- Narrow band microwave LANs

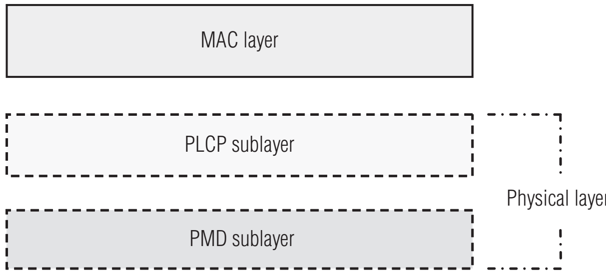


Figure 24.9 OSI model for IEEE 802.11 WLAN

Note that diffuse infrared and FHSS have received little attention in the market. For this reason, we will focus on the more popular interfaces based on DSSS and OFDM, used under the IEEE 802.11a/b/g specifications. The PHY layer is divided into two sub-layers:

- Physical layer convergence procedure (PLCP)
- Physical medium dependent (PMD)

The PLCP sublayer minimizes the dependence of the MAC layer on the PMD sublayer by mapping MAC protocol data unit (MPDU) into a frame format suitable for transmission by the PMD (Figure 24.9). Under the direction of the PLCP, the PMD provides actual transmission and reception of PHY entities between two STAs through the wireless medium. To provide this service, the PMD interfaces directly with the air medium and provides modulation and demodulation of the frame transmissions. The PLCP and PMD communicate using service primitives to govern the transmission and reception functions.

Functions of PHY layer

- Provides a frame exchange between the MAC and PHY under the control of the PLCP sublayer.
- PHY uses signal carrier and spread spectrum modulation to transmit data frames over the media under the control of the PMD sublayer.
- PHY layer provides a carrier sense indication back to the MAC to verify activity on the media.

Spread spectrum LAN

Spread spectrum is a type of modulation that spreads the data transmission across the available frequency band, in excess of the minimum bandwidth required to send the information. Spreading the data cross the frequency spectrum makes the signal resistant to noise, interference, and eavesdropping. Spread spectrum modulation schemes are commonly used with personal communication devices such as digital cellular phones as well as with WLANs.

The spread-spectrum signals are distributed over a wide range of frequencies and then collected onto their original frequency at the receiver. Just as they are unlikely to be intercepted by a military opponent, so are they unlikely to interfere with other signals intended for business and consumer users even though the signals are transmitted on the same frequencies.

Most wireless LAN systems use spread-spectrum technology, a wideband RF technique developed by the military for use in reliable, secure, mission-critical communications systems. Spread-spectrum is designed to trade-off bandwidth efficiency for reliability, integrity, and security. In other words, more bandwidth is consumed than in the case of narrowband transmission, but the trade-off produces a signal that is, in effect, louder and thus easier to detect, provided that the receiver knows the parameters of the spread-spectrum signal being broadcast. If a receiver is not tuned to the right frequency, a spread-spectrum signal looks like background noise. There are two types of spread spectrum radio: frequency hopping and direct sequence.

Spread-spectrum radio communications is also used in the military, because it resists jamming and is hard for an enemy to intercept, and is now on the verge of potentially explosive commercial development.

Applications for commercial spread spectrum range from “wireless” LAN’s (computer-to-computer local area networks), to integrated barcode scanner/palmtop computer/radio modem devices for warehousing, to digital dispatch, to digital cellular telephone communications, to “information society” city/area/state or country wide networks for passing faxes, computer data, e-mail, or multimedia data.

Direct-sequence spread spectrum (DSSS) PHY

In the DSSS PHY, data transmission over the media is controlled by the PMD sublayer as directed by the PLCP sublayer. The DSSS PMD takes the binary bits of information from the PLCP protocol data unit (PPDU) and transforms them into RF signals for the wireless media by using carrier modulation and DSSS techniques.

- *PMD transmitter and receiver:* Figures 24.10 and 24.11 shows the PMD transmitter and receiver. All information bits transmitted by the DSSS PMD are scrambled using a self-synchronizing 7-bit polynomial. An 11-bit Barker code (1, –1, 1, 1, –1, 1, 1, 1, –1, –1, –1) is used for spreading. In the transmitter, the 11-bit Barker code is applied to a modulo-2 adder together with each of the information bits in the PPDU. The output of the modulo-2 adder results in a signal

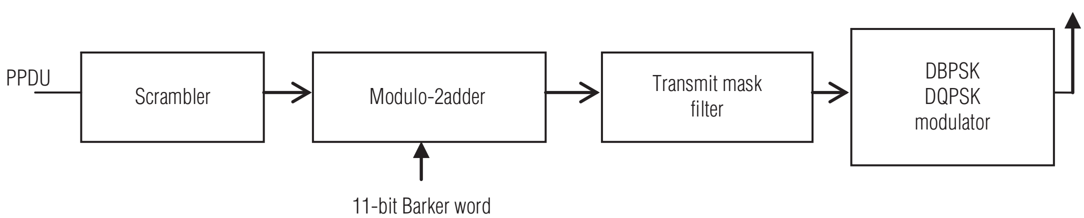


Figure 24.10 DSSS PMD transmitter

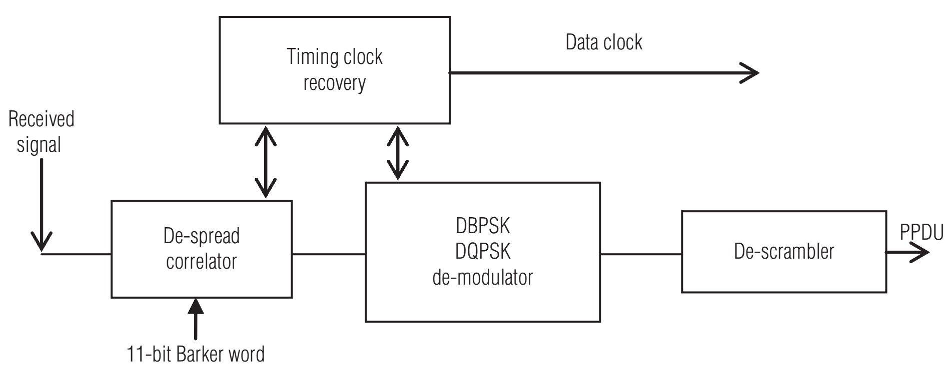


Figure 24.11 DSSS PMD receiver

with a data rate that is 10 times higher than the information rate. The result in the frequency domain is a signal that is spread over a wide bandwidth at a reduced RF power level. Every station in the IEEE 802.11 network uses the same 11-bit sequence. At the receiver, the DSSS signal is convolved with the same 11-bit Barker code and correlated.

- *DSSS PLCP Sublayer:* The PPDU is unique to the DSSS PHY layer. The PPDU frame consists of a PLCP preamble, PLCP header, and MPDU as shown in Figure 24.12. The receiver uses the PLCP preamble to acquire the incoming signal and synchronize the demodulator. The PLCP header contains information about MPDU from the sending DSSS PHY. The PLCP preamble and the PLCP header are always transmitted at 1 Mbps using differential binary phase-shift keying (DBPSK), and the MPDU can be sent at 1 Mbps DBPSK or 2 Mbps differential quadrature phase-shift keying (DQPSK), depending upon the content in the signal field of the PLCP header.
- *Sync*: This field is 128 bits (symbols) in length and contains a string of 1's which are scrambled prior to transmission. The receiver uses this field to acquire the incoming signal and synchronize the receiver's carrier tracking and timing prior to receiving the start of frame delimiter (SFD).
- *Start of frame delimiter (SFD):* This field is always 1111001110100000 (F3A0 hex) and defines the beginning of a frame.
- *Signal*: This field identifies the data rate of the 802.11 frame, with its binary value equal to the data rate divided by 100 kbps. For example, the field contains the value of 00001010 (0A hex) for 1 Mbps DBPSK, 00010100 (14 hex) for 2 Mbps DQPSK, and so on. The PLCP fields, however, are always sent at the lowest rate, which is 1 Mbps. This ensures that the receiver initially uses the correct demodulation mechanism, which changes with different data rates.
- *Service*: This field is always set to 00000000 (00 hex), and the 802.11 standard reserves it for future use.
- *Length*: This field represents the number of microseconds that it takes to transmit the contents of the PPDU, and the MAC layer uses this information to determine the end of the frame.
- *CRC*: In order to detect possible errors in the PHY layer header, the standard defines this field for containing 16-bit cyclic redundancy check (CRC) result. The receiver performs the calculation on the incoming signal, service and length fields, and compares the results against the transmitted value. If an error is detected, the receiver's MAC makes the decision if incoming PPDU should be terminated. CRC is also known as frame check sequence (FCS).

At the end portion of the PPDU, the MPDU is embedded. This field contains a 32-bit CRC, which protects the information in the PLCP service data unit (PSDU).

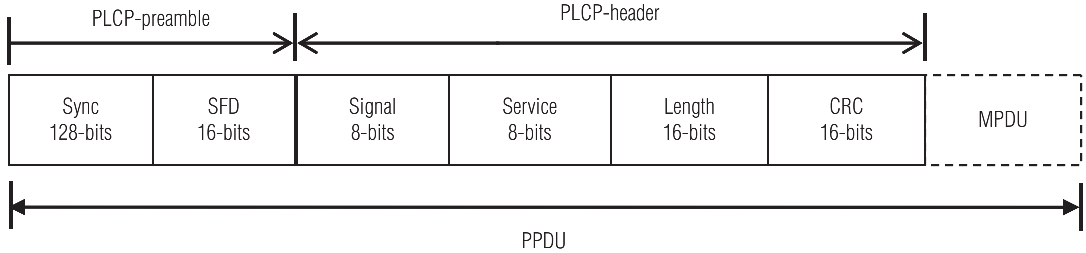


Figure 24.12 DSSS PHY PLCP preamble, header and MPDU

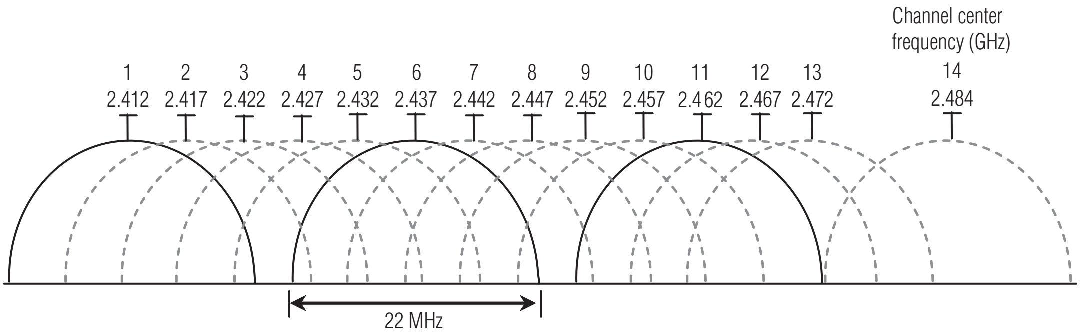


Figure 24.13 Graphical representation of Wi-Fi channels in 2.4 GHz band

- *DSSS operating channels and transmit power requirements: Each DSSS PHY channel occupies 22 MHz of bandwidth*, and the spectral shape of the channel represents a filtered sinX/X function. The DS channel transmit mask in IEEE 802.11 specifies that spectral products be filtered to –30 dBm from the centre frequency and all other products be filtered to –50 dBm. This allows for three non-interfering channels spaced 25 MHz apart in the 2.4 GHz frequency band (see Figure 24.12). In addition to frequency and bandwidth allocations, transmit power is a key parameter that is regulated worldwide. The maximum allowable radiated emissions for the DSSS PHY varies from region-to-region. Today in market wireless products have selected 100 mW as the nominal RF transmit power level.

The 802.11 divide each of the above-described bands into channels, analogously to how radio and TV broadcast bands are subdivided but with greater channel width and overlap. The 2.4000–2.4835 GHz band is divided into 13 channels each of width 22 MHz but spaced only 5 MHz apart, with channel 1 centred on 2.412 GHz and 13 on 2.472 GHz to which Japan adds a 14th channel 12 MHz above channel 13. This is shown in Figure 24.13.

Frequency-hopping spread spectrum (FHSS) PHY

Frequency-hopping utilizes a set of narrow channels and “hops” through all of them in a pre-determined sequence. For example, the 2.4 GHz frequency band is divided into 70 channels of 1 MHz each. Every 20 to 400 ms the system “hops” to a new channel following a pre-determined cyclic pattern.

The 802.11 FHSS PHY uses the 2.4 GHz RF band, operating with at 1 or 2 Mbps data rate. In FHSS PHY, data transmission over the media is controlled by the FHSS PMD sublayer as directed by the FHSS PLCP sublayer. The FHSS PMD takes the binary bits of information from the whitened PSDU and transforms them into RF signals for the wireless media by using carrier modulation and FHSS techniques.

- *FHSS PLCP sublayer: The PLCP preamble*, PLCP header, and PSDU make up the PPDU as shown in the Figure 24.14.

The PLCP preamble and PLCP header are unique to the FHSS PHY. The PLCP preamble is used to acquire the incoming signal and synchronize the receiver’s demodulator. The PLCP preamble and PLCP header are transmitted at 1 Mbps rate.

- *Synchronization:* This field contains a string of alternating 0s and 1s patterns and is used by the receiver to synchronize the receiver’s packet timing and correct for frequency offsets.

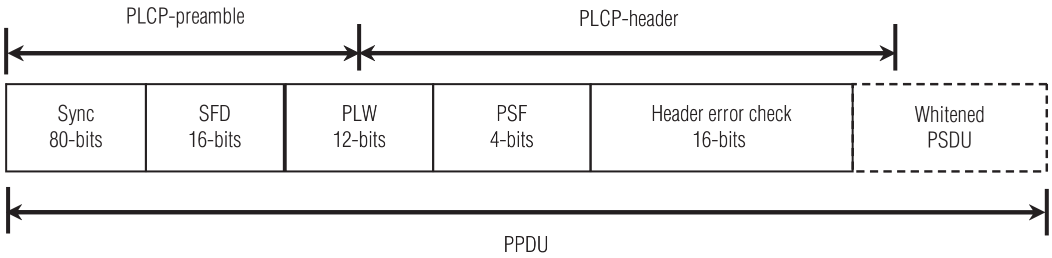


Figure 24.14 FHSS PHY PLCP preamble, header and PSDU

- *Start frame delimiter (SFD):* This field contains information marking the start of a PSDU. A common SFD is specified for all IEEE 802.11 FHSS radios using the following bit pattern: 0000110010111101.
- *PLCP_PDU length word (PLW):* This field specifies the length of the PSDU in octets and is used by the MAC to detect the end of a PPDU frame.
- *PLCP signalling field (PSF):* The PSF identifies the data rate of the whitened PSDU ranging from 1 Mbps to 4.5 Mbps in increments of 0.5 Mbps. The PLCP preamble and header are transmitted at the basic rate, 1 Mbps. The optional data rate for the whitened PSDU is 2 Mbps.
- *Header error check (HEC):* This field contains the results of a calculated FCS from the sending station. The calculation is performed prior to data whitening. The CCIT CRC-16 error detection algorithm is used to protect the PSF and PLW fields. The MAC makes the determination of the correct reception of PPDU frame by looking FCS which is embedded at the end of the PSDU portion of the PPDU.
- *FHSS modulation*: The *1997 version* of IEEE 802.11 uses two-level Gaussian frequency-shift key (GFSK) in the FHSS PMD to transmit the PSDU at the basic rate of 1 Mbps. The PLCP preamble and PLCP header are always transmitted at 1 Mbps. However, four-level GFSK is an optional modulation method defined in the standard that enables the whitened PSDU to be transmitted at a higher rate. The value contained in the PSF field of the PLCP header is used to determine the data rate of the PSDU.

GFSK is a modulation technique used by the FHSS PMD, which deviates (shifts) the frequency either side of the carrier hop frequency depending on if the binary symbol from the PSDU is either 1 or 0. A bandwidth bit period (BT) = 0.5 is used. The changes in the frequency represents symbols containing PSDU information. For two-level GFSK, a binary 1 represents the upper deviation frequency from the hopped carrier, and a binary 0 represents the lower deviation frequency. The deviation frequency (f_d) shall be greater than 110 KHz for IEEE 802.11 FHSS radios. The carrier frequency deviation is given by the following:

Binary 1 = $F_c + f_d$ – Carrier-hopped frequency plus the upper deviated frequency.
Binary 0 = $F_c - f_d$ – Carrier-hopped frequency minus the lower deviated frequency

Four-level GFSK is similar to two-level GFSK and used to achieve a data rate of 2 Mbps in the same occupied frequency bandwidth. The modulator combines two binary bits from the whitened PSDU and encodes them into symbol pairs (10, 11, 01, 00). The symbol pairs generate four frequency deviations from the hopped carrier frequency, two upper and two lower. The symbol pairs are transmitted at 1 Mbps, and for each bit sent, the resulting data rate is 2 Mbps.

- *FHSS channel hopping: A set of hop sequences* are defined in IEEE 802.11 for use in the 2.4 GHz frequency band. The channels are evenly spaced across the band over a span 83.5 MHz. Hop channels differs from country to country. Channel hopping is controlled by the FHSS PMD. The FHSS PMD transmits the whitened PSDU by hopping from channel-to-channel in a pseudorandom fashion using one of the hopping sequences.

Infrared (IR) PHY

The infrared PHY utilizes infrared light to transmit binary data either at 1 Mbps (basic access rate) or 2 Mbps (enhanced access rate) using a specific modulation technique for each. For 1 Mbps, the infrared PHY uses a 16-pulse-position modulation (PPM). The 1 Mbps version employs pulse-position modulation with 16-PPM and the 2 Mbps version uses 4-PPM.

Infrared LANs typically use the wavelength band between 780 and 950 nm, which is somewhere between the visible spectrum of light and microwaves. This is due to the ready availability of inexpensive, reliable system components. The infrared signals from a transceiver-equipped mobile or desktop computer go to a similarly equipped LAN access node, which translates the infrared signals into electrical signals suitable for transmission over the network in standard LAN formats. A line-of-sight connection is needed between transmitters and receivers because infrared will not penetrate walls or windows.

Infrared standards

Infrared products for computer connectivity conform to the standards developed by the Infrared Data Association (IrDA), an industry consortium. The IrDA Serial Infrared Data Link Standard (IrDA-SIR) was developed with the following advantages in mind:

Low-cost implementation: No special or proprietary hardware is required. The standard was developed to make use of components costing only a few dollars per device. With integrated chips that include IrDA functionality, the use of common opto-electronic components adds less than a dollar to the cost of components.

Low-power requirements: IrDA-SIR is designed to be power efficient so that it will not be a drain on the batteries of portable devices like notebook computers, PDAs, mobile phones, and other handheld devices. The use of directed IR, rather than diffuse IR, results in very low power consumption when transmitting.

Directed, point-to-point connectivity: The use of a directed IR beam avoids unintentional "spilling" of the transmitted data to nearby devices. However, the angular spread of the IR beam does not require the user to align the handheld device perfectly at the target device to achieve an IR link.

High-noise immunity: IrDA-SIR is specified to achieve bit error rates of better than 1 in 10 at ranges of up to 1 m, while still providing a high level of noise immunity within a typical office environment illuminated with fluorescent light, as well as in environments with full sunlight.

Optimized for data transfers: IrDA-SIR is a half-duplex system with the maximum universal asynchronous receiver/transmitter (UART)-based data rate of 115.2 kbps. Because the design can be driven by a standard UART, its data rate can be easily programmed from software to a lower data rate to match with slower devices. Of note is that version 2.0 of the IrDA-SIR specification also defines non-UART environments.

The IrDA-SIR PHY hardware is very simple. It consists of an encoder/decoder (which performs the IR transmit encoder and IR receiver decoder) and the IR transducer (which consists of the output driver and IR emitter for transmitting and the receiver/detector). The encoder/decoder interfaces to the UART, which most computers already have.

Narrow band microwave LANs

Narrow band microwave LANs use microwave RF band for signal transmission, with relatively narrow bandwidth. This band is wise enough to accommodate the signal. Most of the microwave LAN products use the licensed microwave band but some of the LAN products use the ISM band.

- *Licensed Narrow Band Microwave LANs*
  Microwave narrow band radio frequencies usable for voice, data, and video transmission are licensed and coordinated within specific geographic areas to avoid potential interference between the systems.
  A narrow band scheme makes use of the cell configuration where adjacent cells use non-overlapping frequency bands within the overall 18 GHz band. The advantage of the licensed narrow band microwave LAN is it provides inference free communication.
- *Unlicensed Narrow Band Microwave LANs: The unlicensed narrow band microwave LANs* use unlicensed ISM spectrum. These LAN products work at 10 Mbps in the 5.8 GHz and makes use of the peer-to-peer configuration. These LAN products automatically elect one node as the dynamic master based on parameter such as location, interference, and signal strength. The identity of the master can automatically change as the conditions change. The LAN also includes a dynamic relay function, which allows each station to act as a repeater to move the data between the stations that are out of reach of each other.

Example problem 24.2

IEEE 802.11 WLAN system operates at 2 Mbps. Determine the data transfer time of a 20 kb file.

Solution

Given data: Transmission data rate = 2 Mbps or 2,000 kbps
Size of a file to be transferred = 20 kb
Using 1 byte = 8 bits, size of the file = 20 × 8 = 160 kB
Therefore, data-transfer time = 160 kB / 2,000 kbps
Hence, data-transfer time = 80 ms

Example problem 24.3

The IEEE 802.11 WLAN system operates at 2 Mbps data transmission rate. Compute the size of the file transferred in 16 s.

Solution

Given data: Transmission data rate = 2 Mbps or 2,000 kbps
Data transfer time = 16 seconds
Therefore, size of a file transferred = 2 Mbps × 16 s
Hence, size of a file transferred = 32 Mb or 4 MB (megabytes)

24.5 IEEE 802.11 medium access control

The MAC is a sublayer of the data link layer specified in the seven-layer OSI model (Layer 2). It provides addressing and channel access control mechanisms that make it possible for several terminals or network nodes to communicate within a multi-point network, typically a local area network (LAN) or metropolitan area network (MAN). Each node in an 802.11 network is identified by its MAC address (exactly the same as Ethernet, a 6 byte [48 bit] value). Receiving nodes recognize their MAC address.

The functions of MAC layer are as follows:

- It provide a reliable data delivery service to the users of the MAC over wireless media through a frame exchange protocol
- It controls the access to the shared wireless medium through two different access mechanisms
  (a) The basic access mechanism called the distributed coordination function (DCF)
  (b) The centrally controlled access mechanism called the point coordination function (PCF)
- It protects the data delivered by nodes connecting to the network. The 802.11 provides a privacy service called wireless equivalent privacy (WEP), which encrypts the data sent over the wireless medium.

24.5.1 MAC frame exchange protocol

The IEEE 802.11 implements a frame exchange protocol to allow the source of a frame to determine when the frame has been successfully received at the destination. The minimal MAC frame exchange protocol consists of two frames, a frame sent from the source to the destination and an acknowledgement from the destination that the frame was received correctly. The frame and its acknowledgement are an atomic unit of the MAC protocol and they cannot be interrupted by the transmission from any other station.

According to the rules of the basic access mechanism, the source will attempt to transmit the frame again, if it does not receive the acknowledgment. This happens when the destination does not send the acknowledgement due to errors in the original frame or due to the corruption of acknowledgement. This retransmission of frames by the source effectively increases the bandwidth consumption.

24.5.2 Hidden node problem

If any node is not able to communicate directly with every other node in WLAN, then that node is termed as hidden node. In wireless networking, the hidden node problem or hidden terminal problem occurs when a node is visible from a wireless AP, but not from other nodes communicating with that AP. This leads to difficulties in media access control.

To explain the hidden node problem in wireless networks, let us consider the three nodes A, B, and C shown in Figure 24.15. Node A can communicate only with node B and node B can communicate with nodes A and C and node C can communicate only with node B.

Assume that node A is sending data to node B. In the middle of this transmission, node C also has data to send to node B. However, node C is out of the range of A and transmissions from A cannot reach C. Therefore, C thinks that the medium is free and sends its data to node B, which results in a collision at B because this node is receiving data from both A and C. In this case, we say that the nodes A and C are hidden from each other with respect to B. Hidden nodes can reduce the capacity of the network because of the possibility of collision. The solution to the hidden node problem is the use of the handshake frames request-to-send (RTS) and clear-to-send (CTS).

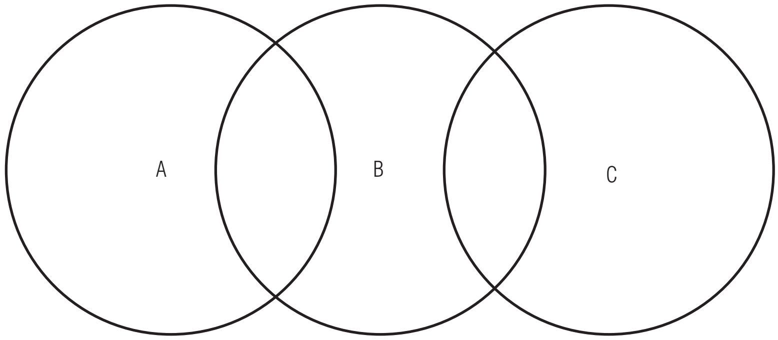


Figure 24.15 Hidden node problem

The source sends a RTS to the destination. The destination returns CTS to the source. Each of these frames contains information that allows other stations receiving them to be notified of the upcoming frame transmission and to delay any transmissions of their own. The RTS and CTS frames serve to announce the impending transmission from the source to destination to all stations in the neighbourhood of both the source and the destination. When the source receives the CTS from the destination the real frame is send to the destination. If that frame is correctly received at the destination, the destination will return an acknowledgment, completing the frame exchange protocol. Depending on the configuration of a station and its determination of local conditions, a station may choose when to use the RTS and CTS frames.

The four frames in this exchange are also an atomic unit of the MAC protocol. They cannot be interrupted by the transmissions of other stations. If this frame exchange fails at any point, the state of the exchange and the information carried in each of the frames allows the stations that have received these frames to recover and regain control of the medium in a minimal amount of time. A station in the neighbourhood of the source station receiving the RTS frame will delay any transmissions of its own until it receives the frame announced by the RTS. If the announced frame is not detected, the station may use the medium. Similarly, a station in the neighbourhood of the destination receiving the CTS frame will delay any transmissions of its own until it receives the acknowledgement frame. If the acknowledgement frame is not detected, the station may use the medium. In the source station, a failure of the frame exchange protocol causes the frame to be retransmitted.

24.5.3 Retry counters

There are two retry counters associated with every frame the MAC attempts to transmit: a short retry counter and a long retry counter. There is also a lifetime timer associated with every frame the MAC attempts to transmit. Between these counters and the timer, the MAC may determine that it is no longer worthwhile to continue attempting to transmit a particular frame. When the MAC makes that determination, it may cancel the frame's transmission and discard the frame. If a frame is cancelled, the MAC indicates this to the MAC user, through the MAC service interface.

The retry counters limit the number of times a single frame may be retransmitted. There are two counters so that the network designer may choose to allow more or fewer retries to shorter frames, as compared to longer frames.

RTS/CTS is not a complete solution and may decrease throughput even further.

The other methods that can be employed to solve hidden node problem are as follows:

- Increase transmitting power from the nodes
- Use omni-directional antennas
- Remove obstacles
- Move the node
- Use protocol enhancement software
- Use antenna diversity

24.5.4 Frame formats

The IEEE 802.11 MAC accepts MSDUs from higher layers in the protocol stack for the purpose of reliable sending of those MSDUs to the equivalent layer of the protocol stack in another station. To accomplish this task, MAC adds information to the MSDU in the form of headers and trailers to create a MPDU. The MPDU is then passed to the PHY layer to be sent over the wireless medium to the other stations. In addition, MAC may fragment MSDUs into several frames, increasing the probability of each individual frame being delivered successfully. The header and trailer information, combined with the information received as the MSDU, is referred to as the MAC frame. This frame contains, among other things, addressing information, IEEE 802.11-specific protocol information, information for setting the network allocation vector (NAV), and a FCS for verifying the integrity of the frame.

The IEEE 802.11 MAC frame format is shown in Figure 24.16. The frame begins with a MAC header. The start of the header is the frame control field. A field that contains the duration information for the NAV or a short identifier follows it. Three addressing fields follow that field. The next field contains frame sequence information. The final field of the MAC header is the fourth address field. Following MAC header is the frame body. The frame body contains the MSDU from the higher layer protocols. The final field in the MAC frame is the FCS.

- *Frame Control*: The frame control field is a 16-bit field that comprises the information the MAC requires to interpret all the subsequent fields of the MAC header. The frame control field is as shown in Figure 24.17. The subfields of the frame control field are *protocol version, frame type and subtype, To DS, From DS, more fragments, retry, power management, more data, wired equivalent privacy (WEP), and order.*

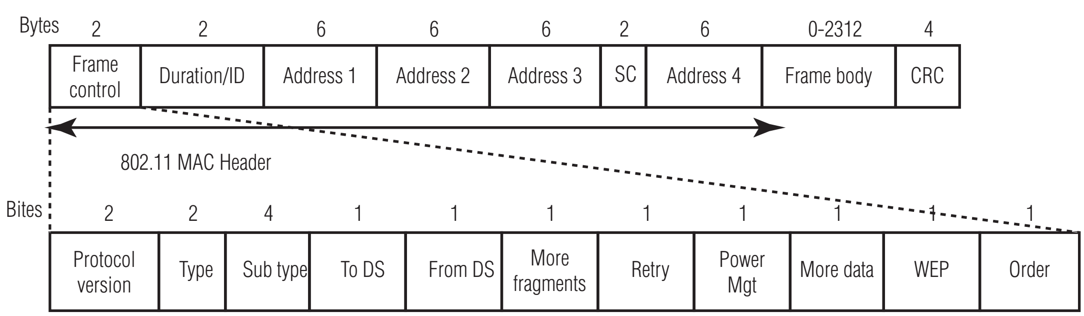


Figure 24.16 IEEE 802.11 MAC FRAME format

Bites	2	2	4	1	1	1	1	1	1	1	1
	Protocol version	Type	Sub type	To DS	From DS	More fragments	Retry	Power Mgt	More data	WEP	Order

Figure 24.17 Frame control field

The subfields of the frame control field are described below.

(a) *Protocol version*: This subfield is 2 bits in length and is used to identify the version of the IEEE 802.11 MAC protocol used to construct the frame. This field is set to zero in the current version of the standard. If the protocol version indicates that the frame was constructed by a version of the IEEE 802.11 MAC protocol that the station does not understand, the station must discard the frame and not generate any response on the medium or any indication to higher layer protocols that the frame was received.

(b) *Frame type and subtype*: Identifies the function of the frame and which other MAC header fields are present in the frame. There are three frame types: control, data, and management. The fourth frame type is reserved. Within each frame type there are several subtypes.

(c) *To DS and From DS subfields*: To DS subfield is 1 bit length. It is used only in data type frames to indicate that the frame is destined for the DS. It will be set in every data frame sent from a mobile station to the AP. This bit is zero in all other types of frames.

- The From DS subfield is a single bit in length. It is also used only in data type frames to indicate that the frame is being sent from the DS. This bit will be set in every data frame sent from the AP to a mobile station and is zero in all other types of frames.
- When both subfields are zero, the frame is a direct communication between two mobile stations. When the To DS subfield is one and the From DS subfield is zero, the frame is a transmission from a mobile station to an AP. When the To DS subfield is zero and the From DS subfield is one, the frame is a transmission from the AP to a mobile station. When both the subfields are one, it is used for a special case where an IEEE 802.11 WLAN is being used as the DS, that is the frame is being sent from one AP to another, over the wireless medium.

(d) *More fragments subfield*: This subfield is a single bit in length. This subfield is used to indicate that this frame is not the last fragment of a data or management frame that has been fragmented. This subfield is zero in the last fragment of a data or management frame that has been fragmented, in all control frames, and in any data or management frame that is not fragmented.

(e) *Retry subfield*: This subfield is a single bit in length. It is used to indicate whether a data or management frame is being transmitted for the first time or if it a retransmission. When this subfield is zero, the frame is being sent for the first time. When this subfield is one, the frame is a retransmission.

(f) *Power management subfield*: This subfield is a single bit in length. A mobile station uses the power management subfield to announce its power management state. A zero in this subfield indicates that the station is in the active mode and will be available for future communication. A one in this subfield indicates that the station will be entering the power management mode and will not be available for future communication. The station may not change its power management state until it has completed a successful frame exchange.

(g) *More data subfield*: This subfield is a single bit in length. The AP uses this subfield to indicate to a mobile station that there is at least one frame buffered at the AP for the mobile station. When this subfield is one, there is at least one frame buffered at the AP for the mobile station. When this subfield is zero, there are no frames buffered at the AP for the mobile station. In multicast frames, the AP may also set this subfield to one to indicate that there are more multicast frames buffered at the AP.

(h) *WEP subfield*: The WEP subfield is a single bit in length. When set to one, it indicates that the frame body of the MAC frame has been encrypted using the WEP algorithm. This subfield may be set to one only in data frames and management frames of subtype authentication. It is zero in all other frame types and subtypes.

(i) *Order subfield*: This subfield is a single bit in length. When set to one, this subfield indicates that the content of the data frame was provided to the MAC with a request for strictly ordered service. This subfield provides information to the AP and DS to allow this service to be delivered.

- *Duration/ID field*: The duration/ID field is 16 bits in length. It alternately contains duration information for updating the NAV or a short ID, called the association ID, used by a mobile station to retrieve frames that are buffered for it at the AP. When bit 15 of the field is zero, the value in bits 14–0 represent the remaining duration of a frame exchange. This value is used to update the NAV, preventing a station receiving this field from beginning a transmission that might cause corruption of the ongoing transmission.
- Address fields: The MAC frame format contains four address fields. The address format (48 bit) normally used to identify the source and destination MAC addresses contained in a frame.
  - *Source address*: The source address (SA) is the address of the MAC that originated the frame. This address is always an individual address. This address does not always match the address in the TA field because of the indirection that is performed by the DS of an WLAN. It is the SA field that should be used to identify the source of a frame when indicating a frame has been received to higher layer protocols.
  - *Destination address*: The destination address (DA) is the address of the final destination to which the frame is sent. This address may be either an individual or group address. This address does not always match the address in the RA field because of the indirection that is performed by the DS.

An IEEE 48-bit address comprises three fields: *a single-bit individual/group field, a single-bit universal/local field, and a 46-bit address field*. The *individual/group field* defines whether the address is that of a single MAC or a group of MACs. When the individual/group field is set to one, the remainder of the address is that of a group. If, in addition, all of the remaining bits in the address are set to one, the group is the broadcast group and includes all stations. When the individual/group bit is zero, the remainder of the address identifies a single MAC. The *universal/local field* defines whether the address is administered globally or locally by the IEEE. When the universal/local field is zero, the address is a globally administered address and should be unique. When the universal/local field is set to one, the address is locally administered and may not be unique.

In addition to the SA and DA, IEEE Standard 802.11–1997 defines three additional address types: *the transmitter address (TA), the receiver address (RA), and the BSS identifier (BSSID).*

- *BSS identifier*: The BSSID is a unique identifier for a particular BSS of an IEEE 802.11 WLAN. Its format is identical to that of an IEEE 48-bit address. In an infrastructure BSS, the BSSID is the MAC address of the AP. Using the MAC address of the AP for the BSSID ensures that the BSSID will be unique and also simplifies the address processing in the

AP. In an IBSS, the BSSID is a locally administered, individual address that is generated randomly by the station that starts the IBSS. The generation of this address from a random number provides some assurance that the address will be unique. However, there is a finite probability that the address generated is not unique. In both infrastructure and IBSSs, the BSSID must be an individual address.

- *Transmitter address*: The TA is the address of the MAC that transmitted the frame onto the wireless medium. This address is always an individual address. The TA is used by stations receiving a frame to identify the station to which any responses in the MAC frame exchange protocol will be sent.
- *Receiver address*: The RA is the address of the MAC to which the frame is sent over the wireless medium. This address may be either an individual or group address.

- *Sequence control field*: The sequence control field is a 16-bit field comprising of two subfields: *sequence number subfield* and *fragment number subfield*. The subfields are a 4-bit fragment number and a 12-bit sequence number. This field is used to allow a receiving station to eliminate duplicate received frames.
    - *Sequence number subfield*: The sequence number subfield contains a 12-bit number assigned sequentially by the sending station to each MSDU. This sequence number is incremented after each assignment and wraps back to zero when incremented from 4,095. The sequence number for a particular MSDU is transmitted in every data frame associated with the MSDU. It is constant over all transmissions and retransmissions of the MSDU. If the MSDU is fragmented, the sequence number of the MSDU is sent with each frame containing a fragment of the MSDU.
    - *Fragment number subfield*: The fragment number subfield contains a 4-bit number assigned to each fragment of an MSDU. The first, or only, fragment of an MSDU is assigned a fragment number of zero. Each successive fragment is assigned a sequentially incremented fragment number. The fragment number is constant in all transmissions or retransmissions of a particular fragment.
- *Frame body field*: The frame body field contains the information specific to the particular data or management frames. This field is variable in length. It may be as long as 2,304 bytes, without WEP encryption, or 2,312 bytes, when the frame body is encrypted using WEP. The value of 2,304 bytes as the maximum length of this field was chosen to allow an application to send 2,048 byte pieces of information, which can then be encapsulated by as many as 256 bytes of upper layer protocol headers and trailers.
- Cyclic redundancy check: The CRC (also known as FCS) frame is 32 bits in length. It contains the result of applying the CCITT CRC-32 polynomial to the MAC header and frame body. The CRC-32 polynomial is represented by the following equation:

$$G(x) = x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} + x^{10} + x^{8} + x^{7} + x^{5} + x^{4} + x^{2} + x + 1 \tag{24.2}$$

This is the same polynomial used in other IEEE 802 LAN standards.

24.5.5 Overview of MAC

The IEEE 802.11 MAC supplies the functionality required to provide a reliable delivery mechanism for user data over noisy, unreliable wireless media. In order to allow multiple users to access a common channel, the IEEE 802.11 standard has defined two different access mechanisms: the basic access mechanism called the distributed coordination function (DCF) and a centrally controlled access mechanism called the PCF. The basic access mechanism is CSMA/CA with binary exponential backoff.

In this mechanism, when a station listens to the medium before beginning its own transmission and detects an existing transmission in progress, the listening station enters a deferral period determined by the binary exponential backoff algorithm. It will also increment the appropriate retry counter associated with the frame. The binary exponential backoff mechanism chooses a random number which represents the amount of time that must elapse while there are not any transmissions, that is the medium is idle before the listening station may attempt to begin its transmission again. The random number resulting from this algorithm is uniformly distributed in a range, called the contention window, the size of which doubles with every attempt to transmit that is deferred until a maximum size is reached for the range. Once a transmission is successfully transmitted, the range is reduced to its minimum value for the next transmission. Both the minimum and maximum values for the contention window range are fixed for a particular PHY.

In order to avoid collisions, the IEEE 802.11 MAC implements a NAV which is a virtual carrier sensing mechanism that indicates to a station the amount of time that remains before the medium will become available. By examining the NAV, a station may avoid transmitting, even when the medium does not appear to be carrying a transmission by the PHY carrier sense. By combining the virtual carrier sensing mechanism with the PHY carrier sensing mechanism, the MAC implements the collision avoidance portion of the CSMA/CA access mechanism.

24.5.6 IEEE 802.11 MAC layer DCF operation

When the MAC receives a request to transmit a frame, a check is made by the PHY and virtual carrier sensing mechanisms. If both mechanisms indicate that the medium is not in use for an interval of Distributed Inter-Frame Space (DIFS), the MAC may begin transmission to the frame. If either the PHY or virtual carrier sense mechanisms indicate that the medium is in use during the DIFS interval, the MAC will select a backoff interval using the binary exponential backoff mechanism and increment the appropriate retry counter. The MAC will decrement the backoff value each time the medium is detected to be idle by both the PHY and virtual carrier sense mechanisms for an interval of one slot time. Once the backoff interval has expired, the MAC begins the transmission. If the transmission is not successful, that is the acknowledgement is not received, a collision is considered to have occurred. In this case, the contention window is doubled, a new backoff interval is selected, and the backoff countdown is begun, again. This process will continue until the transmission is successful or it is cancelled. Short inter-frame spacing is used to separate transmission belonging to a single dialogue. Each frame in IEEE 802.11 is composed of additional delay created by inter-frame spacing and backoff period. The IEEE 802.11b MAC layer CSMA/CA operation is shown in Figure 24.18.

24.5.7 Point coordination function

The original 802.11 MAC defines another coordination function called the PCF. This is available only in “infrastructure” mode, where stations are connected to the network through an AP. A point coordinator (PC) located at AP controls the PCF.

The PCF uses a virtual carrier-sense mechanism aided by an access priority mechanism. The PCF shall distribute information within Beacon management frames to gain control of the medium by setting the NAV in STAs (Stations). In addition, all frame transmissions under the PCF may use an inter-frame space (IFS) that is smaller than the IFS for frames transmitted via the DCF. The use of a smaller IFS implies that point-coordinated traffic shall have priority access to the medium over

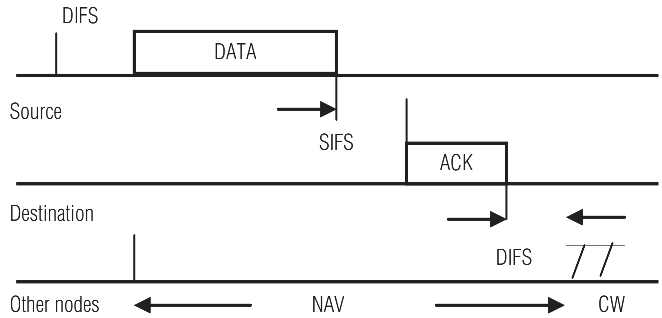


Figure 24.18 The IEEE 802.11b MAC layer CSMA/CA operation

STAs in overlapping BSSs operating under the DCF access method. The access priority provided by a PCF may be utilized to create a contention-free (CF) access method. The PC controls the frame transmissions of the STAs so as to eliminate contention for a limited period of time.

24.6 Comparison of IEEE 802.11 a, b, g, and n standards

Table 24.2 gives the comparison of IEEE 802.11a, b, g, and n standards in terms of frequency, bandwidth, data rate, modulation, and range.

Table 24.2 Comparison of 802.11 protocols

802.11 protocol	Freq. (GHz)	Bandwidth (MHz)	Data rate per stream (Mbps)	Allowable MIMO streams	Modulation	Approx. indoor range (m)	Approx. outdoor range (m)
802.11	2.4	20	1,2	1	DSSS, FHSS	20	100
a	5	20	6, 9, 12, 18, 24, 36, 48, 54	1	OFDM	35	120
b	2.4	20	5.5, 11	1	DSSS	38	140
g	2.4	20	6, 9, 12, 18, 24, 36, 48, 54	1	OFDM, DSSS	38	140
n	2.4	20	7.2, 14.4, 21.7, 28.9, 43.3, 57.8, 65, 72.2	4	OFDM	70	250

24.7 Wireless PANs

A wireless personal area network (WPAN) is a network for interconnecting devices centred around an individual person's workspace – in which the connections are wireless. These might include a mobile phone, a laptop computer, pagers, PDAs, and a personal stereo. There are many potential advantages of these devices being able to communicate to each other, especially without wires.

Typically, a WPAN uses Bluetooth technology, which was used as the basis for a new standard, IEEE802.15. Bluetooth technology has been adopted as the IEEE 802.15.1 WPAN standards which are commercially available in numerous devices ranging from cell phones, PDAs, laptops to wireless mouses, and cameras.

For example, today it is possible to link the laptop to a mobile phone through a cable connector and to wirelessly link into a remote data network to retrieve e-mails or perform other actions. It permits communication within about 10 m, which enables the use of low power, low cost, and extremely small-sized devices.

The concept of personal area networks (PANs) is that if each of these devices had a short-range communications tool built into them, they could exchange information without wires and without any intervention from the user. For example, the laptop, which was ostensibly in sleep mode, stored in a briefcase, could periodically talk to the cell phone clipped to the user's belt and ask it to check for e-mails. The cellular phone could retrieve these and send them to the laptop over the short-range link. The laptop could then store them so that when the user turned the computer on, all the e-mails would be available on the computer. As the user performs actions such as sending e-mails, the computer would talk with the mobile phone and request transmission of these to the data network.

The key concept in WPAN technology is known as *plugging in*. When any two WPAN-equipped devices come into close proximity (within several metres of each other) or within a few kilometres of a central server, they can communicate as if connected by a cable. Another important feature is the ability of each device to lock out other devices selectively, preventing needless interference or unauthorized access to information. The proposed operating frequencies for WPAN are around 2.4 GHz in digital modes. The objective is to facilitate seamless operation among home or business devices and systems. Every device in a WPAN will be able to plug in to any other device in the same WPAN, provided they are within PHY range of one another. In addition, WPANs will be interconnected worldwide.

See Chapters 25 and 27 for detailed description of *IEEE 802.15.4 Low-Rate WPAN and IEEE 802.15.1 WPAN/Bluetooth technology*, respectively.

> *The IEEE 802.15 standards is a family of protocols to address the needs of WPAN at different data rates in 2.4 GHz ISM band, same as defined in IEEE 802.11 WLAN standards.*

So far, three IEEE 802.15 protocols (IEEE 802.15.1, IEEE 802.15.3, and IEEE 802.15.4) have been developed based on data rates, technology, frequency band, channel access scheme, modulation scheme, and technology used. They are referred to as Bluetooth, high-rate WPAN, and low-rate WPAN. Table 24.3 summarizes the major parameters of these three WPANs.

24.7.1 WPAN applications

The IEEE 802.15/WPAN technology applications are given below:

- WPANs are used to replace cables between a computer and its peripheral devices.
- WPANs can be used for transmitting images, digitized music, and other data.
- WPAN, popularly known as Bluetooth technology, is an extremely short-range wireless network, formed around the personal operating space of a wireless terminal with built-in Bluetooth device.

Table 24.3 Major parameters of these three WPANs

Parameter	IEEE 802.15.1/ Bluetooth technology	IEEE 802.15./ High rate WPAN technology	IEEE 802.15.4/Low rate WPAN technology
Channel access scheme	TDD with master-slave polling	CSMA/CA with guaranteed time slot in a superframe structure	CSMA/CA with guaranteed time slot in a superframe structure
RF spectrum	2.4 GHz ISM band	2.402–2.480 GHz ISM band	2.4 GHz ISM band; 868 MHz band; 915 MHz band
Maximum data rate	Up to 1 Mbps	11–55 Mbps	250 Kbps at 2.4 GHz; 20 Kbps at 868 MHz; 40 Kbps at 915 MHz
Power consumption	1–60 mA	< 80 mA	20–50 μA
Modulation scheme	FHSS @ 1600 hops/s	Trellis-coded or uncoded QPSK or 16/32/64-QAM	DSSS with BPSK or MSK

- Bluetooth is the only WPAN technology which is commercially available and is an essential component in a series of devices ranging from laptops to wireless mouses to cameras and cell phones.
- With the help of Bluetooth technology, ad-hoc wireless piconets can be formed, which are local area networks with a very limited coverage (about 10 m) and without the need for an infrastructure, offering asynchronous data, and synchronous voice services at data rate of 1 Mbps.
- Bluetooth also provides a universal bridge to existing data networks and a mechanism to form small private mobile ad-hoc networks (MANETs).
- WPANs help in the interworking of wireless technologies to create heterogeneous wireless networks. For instance, WPANs and WLANs will enable an extension of devices without direct cellular access to 3G cellular systems.

24.8 Hiper LAN

The need for mobile broadband communications has increased rapidly in recent years placing new demands for the WLANs. The requirements of mobile broadband communication include support for *QOS, security, handover, and increased throughput*. To meet these requirements,

> *the European Telecommunications Standards Institute (ETSI) has come up with **HI**gh **Per**formance **R**adio **LAN** (HIPERLAN), which is an alternative for the IEEE 802.11 WLAN standards.*

The HIPERLAN standards provide features and capabilities similar to those of the IEEE802.11 WLAN standards. In HYPERLAN, there are a number of base stations, and devices can communicate either with the base station or directly with each other. The base stations, or APs, can automatically configure their frequency so that there is no need for manual frequency assignment.

The HIPERLAN standard family has four different versions: HIPERLAN/1, HIPERLAN/2, HIPERACCESS, and HIPERLINK.

HIPERLAN/1 provides communications up to 20 Mbps in the 5 GHz range of the RF spectrum. HIPERLAN/2 operates up to 54 Mbps in the same RF band. HIPERLAN/2 is compatible with 3G WLAN systems for sending and receiving data, images, voice communications, and intends to accommodate ATM as well as IP-type access with QOS support. HIPERLAN/2 has the potential, and is intended, for worldwide implementation in conjunction with similar systems in the 5 GHz RF band.

24.8.1 Wireless asynchronous transfer mode

Asynchronous transfer mode (ATM) is one of the leading technologies in fixed high-capacity networks. In most situations, ATM is implemented in optical fibre links, cables, or fixed microwave point-to-point links. The concept of wireless asynchronous transfer mode relates to the extension of ATM services to other scenarios through the use of wireless transmission and features mobility. It includes the wireless mobile ATM, which is the basis for providing services in the order of tens of megabits per second to mobile users, satellite ATM (where the large delays are significant), and WLANs.

24.8.2 HIPERLAN/1

HIPERLAN/1 is mainly designed to work without the need of any infrastructure. Two nodes may exchange data directly, without any interaction from a wired (or radio-based) infrastructure. Thus, the simplest HIPERLAN/1 consists of two nodes. Further, if two HIPERLAN/1 nodes are not in radio contact with each other, they may use a third node which must forward messages between the two communicating nodes. Figure 24.19 shows the overall system architecture of an ad-hoc HIPERLAN/1.

A multi-hub topology is considered to allow overlay of two HIPERLANs to extend the communication beyond the radio range of a single node. There are two overlapping HIPERLANs,

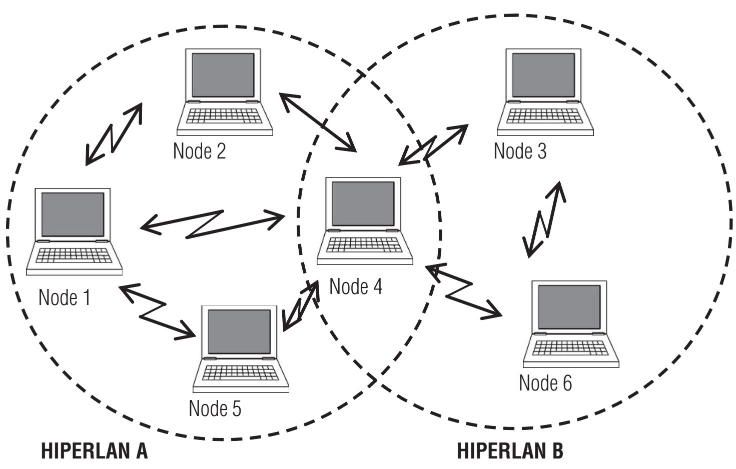


Figure 24.19 HIPERLAN/1 ad-hoc system architecture

A and B, and the node 4 acts as a bridge between the two. Each node is designated either as a forwarder (F) node or a non-forwarder (NF) node. In Figure 24.19, nodes 1, 4, and 6 are forwarder nodes and these have forwarding connections.

A forwarder node retransmits the received packet to other nodes in its neighbourhood, if the packet is not meant for it. Nodes 2, 3, and 5 are non-forwarder nodes, which simply accept the packet that is meant for it. Each non-forwarder node should select at least one of its neighbour nodes as a forwarder node. Inter-HIPERLAN forwarding needs mutual agreement and cooperation and should exchange regular update messages to support proper routing and maintenance.

24.8.3 HIPERLAN/2

HIPERLAN/2 has a very high-transmission rate up to 54 Mbps. This is achieved by making use of a modularization method called orthogonal frequency digital multiplexing (OFDM). HIPERLAN/2 connections are time-division multiplexed and connection-oriented, either bidirectional point-to-point or unidirectional point-to-multipoint connections. There is also a dedicated broadcast channel through which the traffic from an AP reaches all terminals.

The HIPERLAN/2 APs have a built-in support for automatic transmission frequency allocation within the APs coverage area. This is performed by the DFS function. This network supports authentication and encryption. Both the AP and the MT can authenticate each other to ensure authorized access to the network or to a valid network operator. In HIPERLAN, each communicating host or a node is given a HIPERLAN ID (HID) and a node ID (NID). The combination of these two IDs uniquely identifies any station, and restricts the way it can connect to other HIPERLAN nodes. All nodes with the same HID can communicate with each other using a dynamic routing mechanism denoted intra-HIPERLAN forwarding.

The support for handover enables mobility of MTs. The handover scheme is MT initiated, that is the MT uses the AP with the best signal as measured for instance by signal-to-noise ratio, and as the user moves around, all established connections move to the AP with the best radio transmission performance, while the MT stays associated to the HIPERLAN/2 network.

The HIPERLAN/2 architecture is easily adapted and integrated with a variety of fixed networks. All applications running over a fixed infrastructure can also run over a HIPERLAN/2 network.

The power save mechanism in HIPERLAN/2 is based on MT-initiated negotiation of sleep periods. The MT requests the AP for a low power state and a specific sleep period. At the expiration of the sleep period, the MT searches for a wake up indication from the AP, and in the absence of that sleeps the next period, and so forth. The MT receives any pending data as the sleep period expires. Different sleep periods are supported depending on the requirements.

24.8.4 Comparison of various WLAN technologies

Table 24.4 provides a comparison between HIPERLAN2 and the IEEE 802.11 variants.

24.9 Wireless local loop

Wireless local loop (WLL) is a system that connects subscribers to the local telephone station wirelessly. The other names of WLL are *radio in the loop or fixed-radio access*. The detailed description of wireless local loop is described in Section 27.8.

Table 24.4 A Comparison between different WLAN technologies

Characteristic	802.11	802.11b	802.11a	HiperLAN2
Spectrum	2.4 GHz	2.4 GHz	5 GHz	5 GHz
Max PHY data rate	2 Mbps	11 Mbps	54 Mbps	54 Mbps
Max user data rate	1.2 Mbps	5 Mbps	32 Mbps	32 Mbps
MAC	CSMA/CA	CSMA/CA	To be determined	TDD/TDMA
Connection	Connectionless	Connectionless	Connectionless	Connection oriented
Frequency selection	Frequency hopped (FH) or direct sequence spread spectrum (DSSS)	DSSS	OFDM	OFDM with dynamic frequency selection

24.10 Summary

- A wireless LAN (WLAN) is a local area network (LAN) without wires. The key advantages of the wireless LAN are that it eliminates the laying of cables and wiring cost, and that it accommodates mobile PC workstations. The 802.11 is a specific standard that defines the MAC and PHY layers of a WLAN.
- The basic multiple access control (MAC) used in IEEE 802 is carrier sense multiple access with collision avoidance (CSMA/CA).
- Wi-Fi (wireless fidelity) is a generic term that refers to the IEEE 802.11 communications standard for wireless LANs. Wi-Fi network connects computers to each other, to the Internet, and to the wired network.
- The IEEE 802.11 standard defines two bottom layers in the OSI model, namely the MAC layer and the PHY layer. The MAC layer controls and regulates the access to the shared wireless medium using specified channel access mechanisms and the PHY layer manages the transmission of data between the AP and the client.
- Wireless LANs support mobility. Wi-Fi networks can be configured in two different ways: "ad-hoc" mode allows wireless devices to communicate in peer-to-peer mode with each other. "Infrastructure" mode allows wireless devices to communicate with a central node that in turn can communicate with wired nodes on that LAN. Ad-hoc mode WLANs are very easy to configure and do not require a great deal of effort to set up.
- Access point (AP) – The AP is a wireless LAN transceiver or "base station" that can connect one or many wireless devices simultaneously to the Internet. The AP also serves as the WLAN hub that functions as a bridge and relay point between wireless terminals and the existing wired LAN.
- Wireless LANs are less secure than wired LANs and the data transfer rate decreases with increase in number of devices.
- BSS is a group of wireless terminals controlled by a single coordination function provided by an AP.
- ESS is a group of one or more interconnected BSSs and integrated IEEE 802 wired LANs that appear as a single BSS to the LLC layer at any wireless terminal associated with one of these BSSs.

- Infrastructure networks are "single-failure-point" networks. If the base station or AP fails, the entire communications network is destroyed.
- This *first version* of the standard (IEEE 802.11) operates at the 2.4 GHz ISM band which supports only 1 and 2 Mbps data rates using DBPSK and DQPSK. 802.11b is a PHY extension to the original 802.11 standard. It also operates at the 2.40 GHz band and allows for higher data rates of 5.5 and 11 Mbps using CCK.
- The 802.11a is another PHY extension to the 802.11 standard and uses OFDM technique. It operates at 5 GHz unlicensed band and allows for data rates of 6–54 Mbps.
- The 802.11g was the next extension to the 802.11b standard. It operates at the 2.4 GHz ISM band and allows for data rates ranging from 1 to 20 Mbps. It also uses OFDM technique.
- HIPERLAN2 offers the potential of much higher data rate transmission than 802.11. PHY transmission can take place at up to 54 Mbps, resulting in a user data rate after error correction and other overheads of 25 Mbps. The air interface used for this is OFDM, which segments the incoming data stream into a number of subsidiary streams and transmits them on subchannels. This avoids some of the problems associated with intersymbol interference when wideband transmissions are utilized.

Review questions

1. Compare the salient features, advantages, and disadvantages of WLANs and wired LAN technologies.
2. Explain various WLAN topologies with neat diagrams.
3. Compare ad-hoc and infrastructure mode WLAN topologies.
4. Write short notes on IEEE 802.11 WLANs.
5. Explain the architecture of IEEE 802.11 WLAN.
6. What is BSS and ESS with respect to WLANs?
7. List various IEEE 802.11 standards and the modulation techniques used in each standard.
8. Write short notes on IEEE 802.11b, IEEE 802.11a, and IEEE 802.11g standards.
9. Give the functions of PLCP and PMD sublayers of PHY layer.
10. Write short notes on DSSS PHY.
11. Give the frame format of PPDU in DSSS PHY.
12. Write short notes on FHSS PHY.
13. Give the frame format of PPDU in FHSS PHY.
14. Give the functions of MAC layer.
15. Explain the hidden node problem in WLANs.
16. Write short notes on retry counters.
17. Give the IEEE 802.11 MAC frame format and describe each subfield.
18. Explain the IEEE 802.11 MAC DCF and PCF operation.
19. Explain the IEEE 802.11b MAC layer CSMA/CA operation.
20. Compare IEEE 802.11802.11a, b, g, and n standards.
21. The IEEE 802.11 WLAN system operates at 2 Mbps. Determine the data transfer time of a 40 kB file (Ans: 0.16 s)
22. The IEEE 802.11 WLAN system operates at 2 Mbps data transmission rate. Compute the size of the file transferred in 32 s (Ans: 64 MB or 8 MB).

23. Consider an IEEE 802.11a WLAN system in which OFDM baseband modulation scheme is used. The OFDM system has 52 subcarriers out of which eight subcarriers are used as pilot subcarriers. Find the data subcarriers. OFDM symbol duration including guard interval for ISI mitigation is 4 μs. If the system uses ¾ FEC code rate and 64 QAM carrier modulation scheme, find the number of data bits transmitted per OFDM symbol and approximate transmission data rate. (Ans: 44 subcarriers. Number of subcarriers = 44, number of data bits transmitted per OFDM symbol = 198, transmission data rate 49.5 Mbps ≈ 50 Mbps)

Objective type questions and answers

1. The Wi-Fi technology is specified in
   (a) IEEE 802.11 WLAN standards (b) IEEE 802.11a WLAN standards
   (c) IEEE 802.11b WLAN standards (d) IEEE 802.11g WLAN standards
2. The size of a file transferred in 8 s in the IEEE 802.11 WLAN system operating at 2 Mbps data transmission rate is
   (a) 2 MB (b) 4 MB
   (c) 16 MB (d) 32 MB
3. The data transfer time for a fixed size file ______ in the IEEE 802.11 WLAN system operating at 2 Mbps as compared to that of at 1 Mbps.
   (a) increases by ten times (b) increases by two times
   (c) decreases by two times (d) decreases by ten times
4. Quality of service in a mobile network is affected by
   (a) transmission errors and the quality of audio received
   (b) network connectivity, effective bandwidth availability, connection reliability, and data loss probability
   (c) atmospheric conditions and the number of simultaneous active mobile users
   (d) network load and the data lost per second
5. IEEE 802.15.1 WPAN standard uses ______ technique for separation of piconets
   (a) DSSS (b) OFDM (c) FHSS-TDMA (d) FHSS-CDMA
6. The WiMAX technology uses multicarrier OFDMA scheme in 2–11 GHz band, to achieve transmission data rates of
   (a) 11 Mbps (b) 54 Mbps (c) 155 Mbps (d) 2 Gbps
7. Bluetooth technology has been adopted as the IEEE _______ standards.
   (a) 802.11b (b) 802.15.1 (c) 802.15.3 (d) 802.16a
8. The ____ are installed as an add-on unit with the wireless terminals to provide wireless communications.
   (a) access points (b) wireless access interface cards
   (c) distribution systems (d) BSSs
9. The service set that does not contain access point is
   (a) independent BSS (b) infrastructure BSS
   (c) extended service Set ESS (d) none of the above
10. The blanket that have high concentration of users is referred as
    (a) access point (b) distribution spots
    (c) hot spots (d) none of the above

11. The bridge between the wireless LAN and the wired LAN is provided by
    (a) basic service set (BSS) (b) distribution system (DS)
    (c) access point (AP) (d) extended service set (ESS)
12. Wi-Fi networks can be configured in
    (a) ad-hoc mode (b) infrastructure mode
    (c) both a & b (d) neither a nor b
13. The interference between the MAC and wireless media is
    (a) physical layer (b) implementation layer
    (c) application layer (d) none of the above
14. Errors in physical layer header are detected using
    (a) frame check sequence (b) frame delimiter
    (c) sync bits (d) none of the above
15. The physical layer provides
    (a) frame exchange between MAC and PHY
    (b) transmits data frames over media
    (c) carrier sense indication back to the MAC
    (d) all the above
16. IEEE802.11a uses the modulation technique
    (a) DSSS (b) OFDM
    (c) both a & b (d) neither a nor b

Answers: 1. (c), 2. (a), 3. (b), 4. (b), 5. (d), 6. (c), 7. (b), 8. (b), 9. (a), 10. (c), 11. (c), 12. (c), 13. (a), 14. (a), 15. (d), 16. (b).

Open book questions

1. What is the difference between an access point and a portal?
2. What are the requirements of IEEE 802.11 wireless standards?
3. Describe how the ad-hoc networks are useful in day-today applications.
4. Why are multiple access points installed in a building?
5. Write about mobile ad-hoc network.
6. What are the interdependent layers of the IrDA protocol?
7. Tabulate the physical-layer characteristics of IEEE 802.11 family of standards.
8. What is key difference in the ad-hoc and infrastructure topologies of WLANS?
9. Explain the characteristics of HIPERLAN.

Further reading

G. Bianchi, "IEEE 802.11 saturation throughput analysis," IEEE Comm. Letters V2, N12, December 1998.

G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," IEEE Journal on Selected Areas in Communications, Vol. 18, No. 3, pp. 535–547, March 2000.

IEEE Standard for Wireless LAN Medium Access Control (MAC) and PHY Layer (PHY) specifications, IEEE 802.11b, 1999.

Prasant Mohapatra, J. Li, Chao Gui, "QoS in mobile ad-hoc networks", IEEE Wireless Communication, pp. 44–52, June 2003.

Ruopeng Wang, Jie Zhang and Xiangyi Zou, "Performance analysis and optimization of IEEE 802.11 DCF with constant contention window", ISECS International colloquium on CCCM, pp. 407–411, 2008.

X. Shugong, T. Saadawi, *Does the IEEE MAC protocol work well in multi-hop wireless ad-hoc networks?* IEEE Communications Magazine, pp. 130–137, June 2001.

Other Wireless Technologies 25

25.1 Introduction

The cellular network was a natural extension of the wired telephony network that became pervasive during the mid-twentieth century. As the need for mobility and the cost of laying new wires increased, the motivation for a personal connection independent of location of the network also increased. Coverage of large area is provided through (1–2 km) cells that cooperate with their neighbours to create a seemingly seamless network. Examples of standards are GSM, IS-136, IS-95. Cellular are standards basically aimed at facilitating voice communications throughout a metropolitan area. The IEEE 802.11 working group is formed to create a wireless local area network (WLAN) standard for data transfer.

The IEEE 802.15 working group is formed to create wireless personal area networks (WPANs) standard. WPANs are focused on a space around a person or object that typically extends up to 10 m in all directions. The focus of WPANs is on low cost, low power, short range, and very small size. This group has currently defined three classes of WPANs that are differentiated by data rate, battery drain, and quality of service (QOS).

- *IEEE 802.15.3* — High data rate WPAN is suitable for multimedia applications that require very high QOS.
- *IEEE 802.15.1* (Bluetooth) — Medium-rate WPANs will handle a variety of tasks ranging from cell phones to personal digital assistant (PDA) communications and have QOS suitable for voice communications.
- *IEEE 802.15.4* — The low-rate WPANs (LR-WPAN) are low-rate, low-power WPANs with low bandwidth requirements and relaxed QOS requirements.

The low data rate enables the LR-WPAN to consume very little power. The LR-WPAN is intended to serve a set of industrial, residential, and medical applications with very low power consumption at low cost. A technological standard created for control and sensor networks applications is the ZigBee. The ZigBee is based on the 802.15.4 standard and created by the ZigBee alliance. It operates in personal area networks (PANs) and device-to-device networks and provides connectivity between small packet devices. Primarily, ZigBee is used for control of lights, switches, thermostats, appliances, and so on.

WiMAX is an acronym that stands for **w**orldwide **i**nteroperability for **m**icrowave **acc**ess. WiMAX, also known as wireless metropolitan area networks (WMANs), provides broadband wireless connectivity across a large geographical area such as a large metropolitan city. It is based on the IEEE 802.16a standard. This chapter provides a detailed description of ZigBee, IEEE 802.15.4, WiMAX, and IEEE 802.16 technologies.

25.2 ZigBee and IEEE 802.15.4

ZigBee technology is a wireless networking protocol that is primarily developed for automation and control applications. It has a low power consumption, low data rate, and low cost.

ZigBee can be implemented in mesh networks larger than those possible with Bluetooth. ZigBee compliant wireless devices are expected to transmit 10–75 m, depending on the radio frequency (RF) environment. The data rate is 250 Kbps at 2.4 GHz, 40 Kbps at 915 MHz, and 20 Kbps at 868 MHz. IEEE and ZigBee alliance have been working closely to specify the entire protocol stack. The specifications of physical and data link layer of the protocol are according to IEEE 802.15.4. ZigBee alliance focuses the upper layers of protocol, that is, from network to application layer on security services, interoperable data networking, and wireless home and control applications (Figure 25.1(a)). Thus, ZigBee products bought from different manufactures are compatible to each other and can be used in any application to work together.

IEEE 802.15.4 is now detailing the specification of physical layer (PHY) and medium access control (MAC) by offering building blocks for different types of networking known as star, mesh, and cluster tree. Network routing schemes are designed to ensure power conservation and low latency through guaranteed time slots. A unique feature of ZigBee network layer is communication redundancy and eliminating single point of failure in mesh networks. Key features of PHY include energy and link quality detection, and clear channel assessment (CCA) for improved coexistence with other wireless networks.

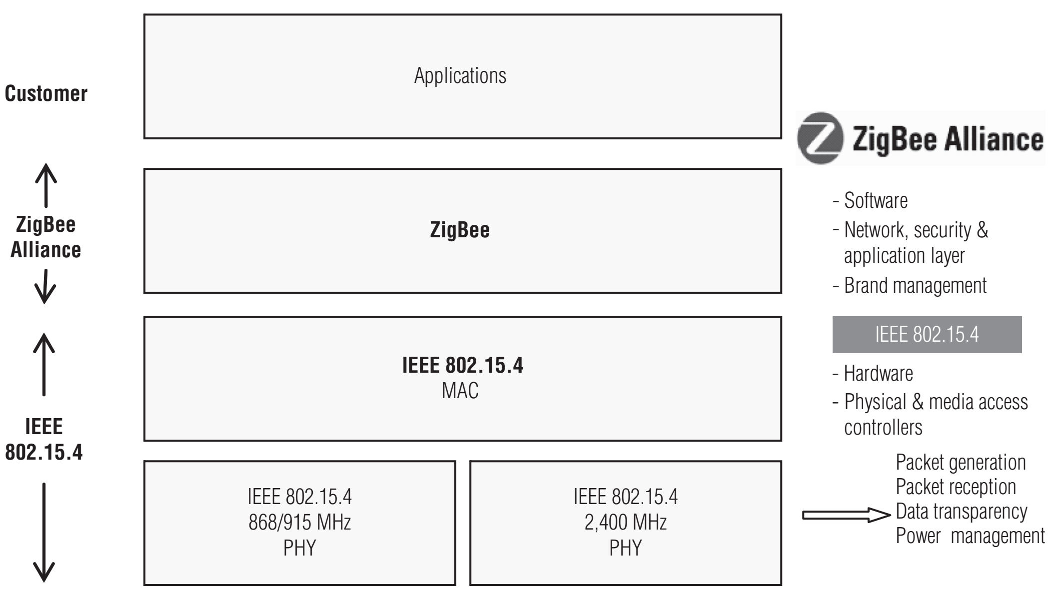


Figure 25.1(a) 802.15.4 & ZigBee in context

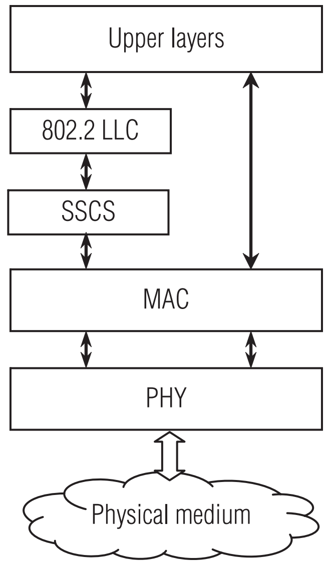


Figure 25.1(b) LR-WPAN device architecture

IEEE 802.15.4 was designed to address the need for a low-cost and low-power wireless solution and has become a solid foundation for monitoring and controlling networks.

25.2.1 Architecture

Figure 25.1(b) shows an LR-WPAN device. The device comprises a PHY, which contains the RF transceiver along with its low-level control mechanism, and a MAC sub-layer that provides access to the physical channel for all types of transfer.

The upper layers consist of a network layer which provides network configuration, manipulation, message routing, and the application layer which provides the intended function of a device. An IEEE 802.2 logical link control (LLC) can access the MAC sub-layer through the service specific convergence sub-layer (SSCS).

25.2.2 Physical layer

The PHY provides PHY data service and PHY management service, and also interfaces with the PHY management entity. The transmission and reception of PHY protocol data units in a radio channel are enabled by PHY data service.

The key functions of PHY are energy detection (ED), activation and deactivation of the radio transceiver, link quality indication (LQI), CCA, channel selection, and transmitting as well as receiving packets across the physical medium.

Two frequency options are given in the standard. Both the options are based on direct sequence spread spectrum. The data rate is 250 Kbps at 2.4 GHz, 40 Kbps at 915 MHz, and 20 Kbps at 868 MHz. The higher data rate at 2.4 GHz is attributed to a higher order modulation scheme. The advantage of using lower frequency is the longer range, as the propagation losses are less at lower frequencies. Higher rate means higher throughput, lower latency, or lower duty cycle. This information is summarized in Table 25.1.

Table 25.1 Frequency bands and data rates

		Spreading parameters		Data parameters		
PHY (MHz)	**Frequency band (MHz)**	**Chip rate (k chip/s)**	**Modulation**	**Bit rate (kb/s)**	**Symbol rate (k symbol/s)**	**Symbols**
868/915	868/868.6	300	BPSK	20	20	Binary
	902–928	600	BPSK	40	40	Binary
2450	2400–2483.5	2000	0-QPSK	250	62.5	16-ary Orthogonal

There is a single channel between 868 and 868.6 MHz, 10 channels between 902.0 and 928.0 MHz, and 16 channels between 2.4 and 2.4835 GHz as shown in Figure 25.2. Several channels in different frequency bands enable the ability to relocate within a spectrum. The standard also allows dynamic channel selection, a scan function that steps through a list of supported channels in search of beacon, receiver ED, LQI, and channel switching. Receiver sensitivities are –85 dBm for 2.4 GHz and –92 dBm for 868 or 915 MHz. The advantage of 6–8 dB comes from the advantage of lower rate. The achievable range is a function of receiver sensitivity and transmit power. The maximum transmit power shall confirm with local regulations. A compliant device shall have its nominal transmit power level indicated by the PHY parameter and PHY transmit power.

25.2.3 MAC layer

The MAC sub-layer provides the following two services:

- **MAC data service** – enables the transmission and reception of MAC protocol data units (MPDU) across the PHY data service.
- **MAC management service** – interfaces to the MAC sub-layer management entity (MLME) service access point (MLMESAP).

The features of MAC sub-layer are beacon management, channel access, frame validation, acknowledged frame delivery, association, and disassociation.

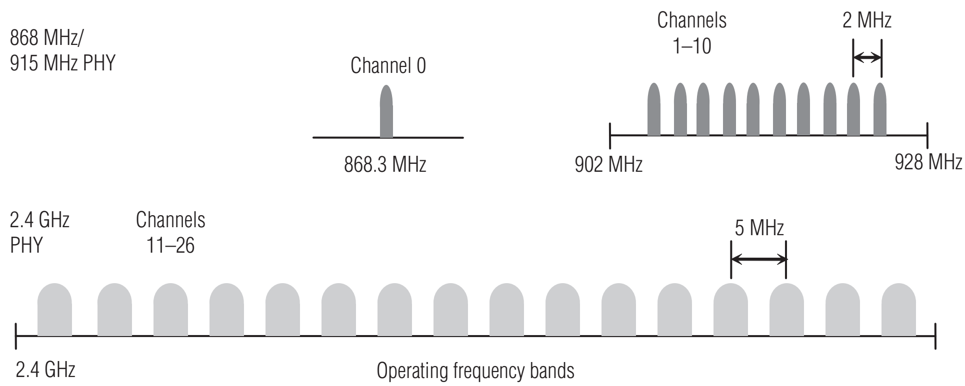


Figure 25.2 Operating frequency bands

Data transfer model

Three types of data transfer transactions exist: from a coordinator to a device, from a device to a coordinator, and between two peer devices. The mechanism for each of these transfers depends on whether the network supports the transmission of beacons. When a device wishes to transfer data in a non-beacon-enabled network, it simply transmits its data frame, using the unslotted CSMA-CA, to the coordinator. There is also an optional acknowledgement at the end as shown in Figure 25.3. When a device wishes to transfer data to a coordinator in a beacon-enabled network, it first listens for the network beacon. When the beacon is found, it synchronizes to the super frame structure. At the right time, it transmits its data frame, using slotted CSMA-CA, to the coordinator. There is an optional acknowledgement at the end as shown in Figure 25.4.

The applications transfers are completely controlled by the devices on a PAN rather than by the coordinator. This provides the energy-conservation feature of the ZigBee network. When a coordinator wishes to transfer data to a device in a beacon-enabled network, it indicates in the network beacon that the data message is pending. The device periodically listens to the network

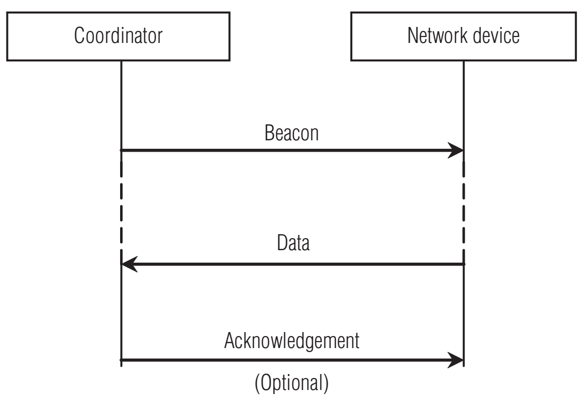


Figure 25.3 Communication to a coordinator in a beacon-enabled network

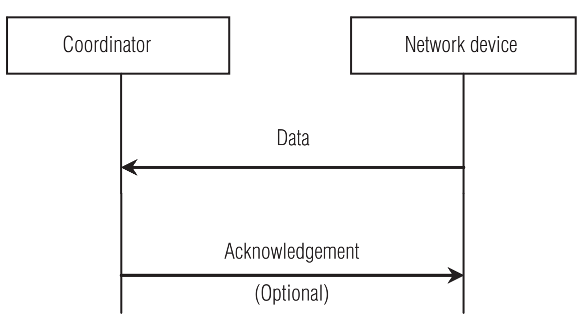


Figure 25.4 Communication to a coordinator in a non-beacon-enabled network

beacon, and if a message is pending, transmits a MAC command requesting this data, using slotted CSMA-CA. The coordinator optionally acknowledges the successful transmission of this packet.

The pending data frame is then sent using slotted CSMA-CA. The device acknowledges the successful reception of the data by transmitting an acknowledgement frame. Upon receiving the acknowledgement, the message is removed from the list of pending messages in the beacon as shown in Figure 25.5.

When a coordinator wishes to transfer data to a device in a non-beacon-enabled network, it stores the data for the appropriate device to make contact and to request data. A device may make contact by transmitting a MAC command requesting the data, using unslotted CSMA-CA, to its coordinator at *an application-defined rate*. The coordinator acknowledges this packet. If data are pending, the coordinator transmits the data frame using unslotted CSMA-CA. If data are not pending, the coordinator transmits a data frame with a zero-length payload to indicate that no data were pending. The device acknowledges this packet as shown in Figure 25.6.

In a peer-to-peer network, every device can communicate with any other device in its transmission radius. There are two options for this. In the first case, the node listens constantly and transmits its data using unslotted CSMA-CA. In the second case, the nodes synchronize with each other so that they can save power.

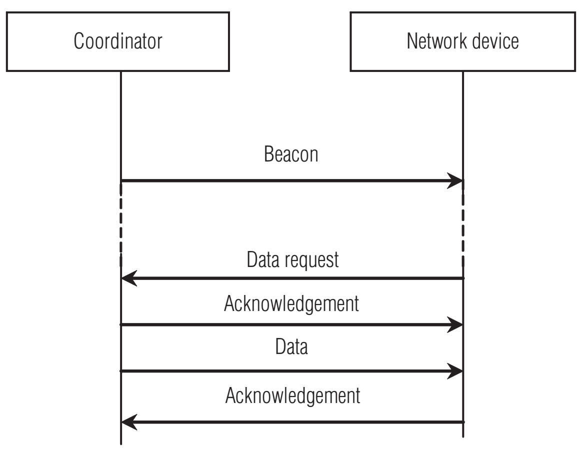


Figure 25.5 Communication from a coordinator in a beacon-enabled network

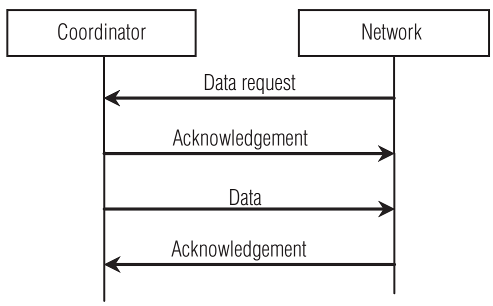


Figure 25.6 Communication from a coordinator in a non-beacon-enabled network

25.2.4 ZigBee

ZigBee is a specification for a suite of high-level communication protocols using small, low-power digital radios based on the IEEE 802.15.4-2003 standard for LR-WPANs, such as wireless light switches with lamps, electrical metres with in-home displays, and consumer electronics equipment via short-range radio that needs low rates of data transfer. The technology defined by the ZigBee specification is intended to be simpler and less expensive than other WPANs, such as Bluetooth. ZigBee is targeted at RF applications that require a low data rate, long battery life, and secure networking.

ZigBee is a low-cost, low-power, wireless mesh networking standard. Firstly, the low cost allows the technology to be widely deployed in wireless control and monitoring applications. Secondly, the low power-usage allows longer life with smaller batteries. Thirdly, the mesh networking provides high reliability and more extensive range. ZigBee is intended not to support powerline networking but to interface with it, at least for smart metering and smart appliance purposes.

ZigBee operates in the industrial, scientific, and medical (ISM) radio bands: 868 MHz in Europe, 915 MHz in the United States and Australia, and 2.4 GHz in most jurisdictions worldwide. The technology is intended to be simpler and less expensive than other WPANs such as Bluetooth. ZigBee chip vendors typically sell integrated radios and microcontrollers with flash memory ranging between 60 and 256 KB.

The ZigBee Smart Energy V2.0 specifications define an IP-based protocol to monitor, control, inform, and automate the delivery and use of energy and water. It is an enhancement of the ZigBee Smart Energy version 1 specifications, adding services for plug-in electric vehicle charging, installation, configuration and firmware download, prepaid services, user information and messaging, load control, demand response, and common information and application profile interfaces for wired and wireless networks.

Uses of ZigBee

ZigBee protocols are intended for use in embedded applications requiring low data rates and low power consumption. ZigBee's current focus is to define a general-purpose, inexpensive, self-organizing mesh network that can be used for industrial control, embedded sensing, medical data collection, smoke and intruder warning, building automation, home automation, and so on. The resulting network uses very small amount of power; individual devices must have a battery life of at least two years to pass ZigBee certification. Typical application areas include the following:

- Home entertainment and control: smart lighting, advanced temperature control, safety and security, movies and music
- Wireless sensor networks: starting with individual sensors like Telosb/Tmote and Iris from Memsic

ZigBee devices

There are three different types of ZigBee devices:

- ZigBee coordinator (ZC): The most capable device, the coordinator forms the root of the network tree and might bridge to other networks. There is exactly one ZC in each network since it is the device that started the network originally. It is able to store information about the network, including acting as the trust centre and repository for security keys.
- ZigBee router (ZR): Despite running an application function, a router can act as an intermediate router, passing on data from other devices.

- ZigBee end device (ZED): Contains just enough functionality to communicate to the parent node (either the coordinator or a router); it cannot rely on data from other devices. This relationship allows the node to be asleep a significant amount of time, thereby giving long battery life. A ZED requires the least amount of memory, and therefore can be less expensive to manufacture when compared to a ZR or ZC.

25.3 WiMAX and IEEE 802.16

WiMAX is an acronym that stands for **w**orldwide **i**nteroperability for **m**icrowave **acc**ess. WiMAX, also known as **w**ireless **m**etropolitan **a**rea **n**etworks (**WMANs**), provides broadband wireless connectivity across a large geographical area such as a large metropolitan city. It is based on the IEEE 802.16a standard.

The evolution of WiMAX began a few years ago when scientists and engineers felt the need of having a wireless Internet access and other broadband services which works well everywhere especially in the rural areas or in those areas where it is hard to establish wired infrastructure that is economically not feasible. The *IEEE 802.16, also known as IEEE wireless-MAN*, explored both licensed and unlicensed band of 2–66 GHz, which is the standard of fixed wireless broadband supporting mobile broadband application.

WiMAX forum was formed in June 2001 to coordinate and to develop the equipment that will be compatible and interoperable with other standards. After several years, in 2007, Mobile WiMAX equipment was developed with the standard IEEE 802.16e, providing mobility and roaming access. The IEEE 802.16e standard uses orthogonal frequency division multiple access (OFDMA) as air interface, and its main aim is to give better performance in non-line-of-sight (NLOS) environments.

Need of wireless MAN/WiMAX

WLANS and WPANS restrict the mobility of users to roam around a few hundreds of metres from the source of the RF signal (Figure 25.7). Also users are needed to stay within line-of-sight (LOS) antennas. Therefore, user mobility has remained largely confined to offices, homes, and hotspots except for voice communications and low-speed data over cellular networks. IEEE 802.16 and

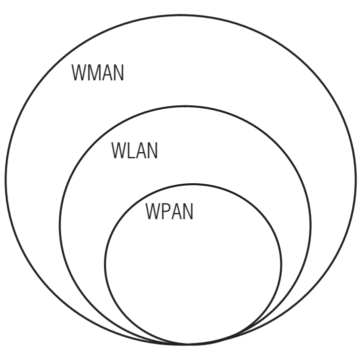


Figure 25.7 Wireless networks

Table 25.2 Features of wireless networks

Parameter	WPAN	WLAN	WMAN
Protocol	802.15	802.11	802.16
Standards	Bluetooth, IrDA, UWD	Wi-fi	Wi-Max
Frequency range	2.4 to 2.483 Ghz	5.15 to 5.35Ghz	10–66 Ghz
Speed	1–4 Mbps	1–54 Mbps	2–70 Mbps
Cell radius	1–10 ms	1–500 ms	1–50km
Modulation	FHSS	OFDM, DSSS	QPSK

WiMAX are designed as a complimentary technology to Wi-Fi and Bluetooth (Table 25.2). To prevent interference in unlicensed bands, user mobility has also been restricted in Wi-Fi networks due to the use of low transmitter power. In remote areas as well as areas with low user density, it may not be economically viable to implement mobile access and hotspots. In addition, the high cost of installing wired high-speed communication channels over long distances prevents high-speed Internet access at all places. WMANs are a group of technologies that provide wireless connectivity across a large geographical area such as a large metropolitan city. The network provides the access of wired networks beyond a single location without the expense of high-speed cable-based connections for extending user mobility throughout a metropolitan area. WMAN can provide high-speed connections, including Internet, to areas not serviced by any other method of connectivity.

25.3.1 WiMAX architecture

WiMAX architecture comprises several components but the basic two components are BS and SS. Other components are MS, ASN, CSN, and CSN-GW as shown in Figure 25.8. The WiMAX Forum's network working group has developed a network reference model according to the IEEE 802.16e-2005 air interface to make sure the objectives of WiMAX are achieved. To support fixed, nomadic, and mobile WiMAX network, the network reference model can logically be divided into three parts.

Mobile station (MS): It is for the end user to access the mobile network. It is a portable station that is able to move to wide areas and perform data and voice communication. It has

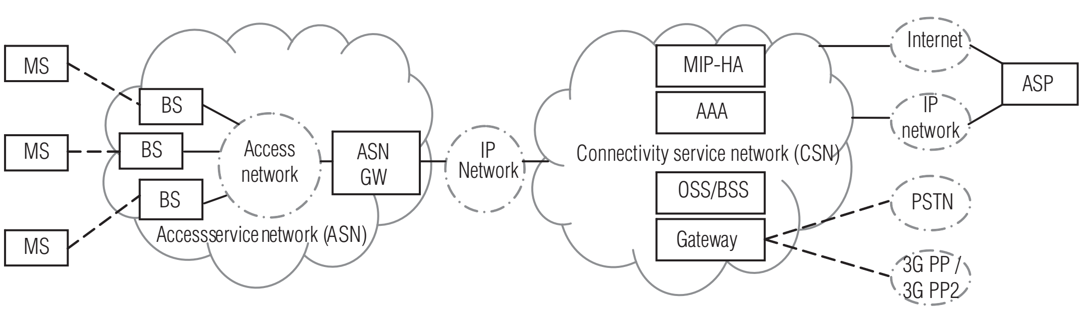


Figure 25.8 WiMAX network architecture based on IP

all the necessary user equipment such as an antenna, amplifier, transmitter, and receiver, and software needed to perform the wireless communication. GSM, frequency division multiple access (FDMA), time division multiple access (TDMA), CDMA, and W-CDMA devices are the examples of MS.

Access service network (ASN): It is owned by NAP, formed with one or several base stations and ASN gateways (ASN-GW), which create radio access networks. It provides all the access services with full mobility and efficient scalability. Its ASN-GW controls the access in the network and coordinates between data and networking elements.

Connectivity service network (CSN): It provides IP connectivity to the Internet or other public or corporate networks. It also applies per user policy management, address management, location management between the ASNs, and ensures QOS, roaming, and security.

25.3.2 Mechanism

WiMAX is capable of working in different frequency ranges but according to the IEEE 802.16, the frequency band is 10–66 GHz. A typical architecture of WiMAX includes a base station built on top of a high rise building and communicates on point-to-multi-point basis with subscriber stations which can be a business organization or a home. The base station is connected through customer premise equipment with the customer. This connection could be an LOS or an NLOS.

Line of sight (LOS)

In LOS connection, signal travels in a straight line which is free of obstacles, which means, a direct connection exists between a transmitter and a receiver. LOS requires its most of the Fresnel zone to be free from obstacles, but if the signal path is blocked by any means, the strength of the signal decreases significantly resulting in poor connectivity. There must be a direct link between a WiMAX base station and the receiver in LOS environment as shown in Figure 25.9.

The following are the features of LOS connections:

- Uses higher frequency between 10 and 66 GHz
- Huge coverage area
- Higher throughput
- Less interference

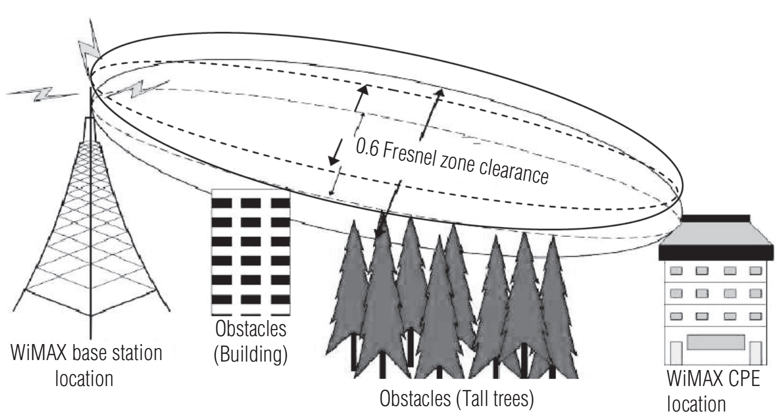


Figure 25.9 WiMAX in LOS condition

- Threat only comes from atmosphere and the characteristic of the frequency
- Requires most of its first Fresnel zone to be free of obstacles

Non-line of sight (NLOS)

In NLOS connection, signal experiences obstacles in its path and reaches to the receiver through several reflections, refractions, diffractions, absorptions, and scattering. These signals arrive at the receiver in different times with different attenuation and strength which makes it hard to detect the actual signal. WiMAX shows good performance in NLOS condition as it is based on orthogonal frequency division multiplexing (OFDM), which can handle delays caused in NLOS perfectly. WiMAX offers other following benefits which work well in NLOS condition:

- Frequency-selective fading can be overcome by applying adaptive equalization.
- Adaptive modulation and coding (AMC), AAS, and multiple-input multiple-output (MIMO) techniques help WiMAX to work efficiently in NLOS condition.
- Sub-channelization permits to transmit appropriate power on sub-channels.
- Based on the required data rate and channel condition, AMC provides the accurate modulation and code dynamically.
- AAS directs WiMAX BS to a subscriber station.
- MIMO helps to improve the signal strength and throughput in both stations.
- In NLOS condition, the speed is high but the coverage area would be lower than that of LOS condition. WiMAX in NLOS condition is shown in Figure 25.10.

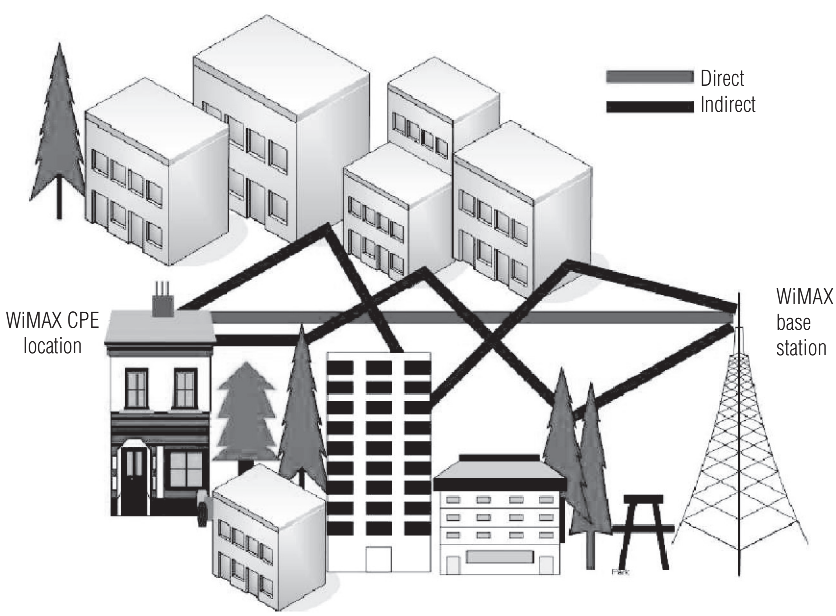


Figure 25.10 WiMAX in NLOS condition

25.3.3 IEEE 802.16 protocol layers

IEEE 802.16 standard WiMAX gives freedom in several things compared to other technologies. The focus is not only on transmitting tens of megabits of data to many miles of distance, but also on maintaining effective QOS and security. This chapter gives an overview of IEEE 802.16 protocol layers and OFDM features. WiMAX 802.16 is mainly based on the physical and data link layer in OSI reference model. Here, PHY can be single-carrier or multi-carrier (PHY) based and its data link layer is sub-divided into the following two layers:

- Logical link control (LLC)
- Medium access control (MAC)

MAC is further divided into three sub-layers:

- Convergence sub-layer (CS)
- Common part sub-layer (CPS)
- Security sub-layer (SS).

Physical layer (PHY)

PHY sets the connection between the communicating devices and is responsible for transmitting the bit sequence. It also defines the type of modulation and demodulation as well as the transmission power. WiMAX 802.16 PHY considers two types of transmission techniques: OFDM and OFDMA. Both of these techniques have frequency band below 11 GHz and use TDD and FDD as its duplexing technology. After implementing OFDM in IEEE 802.16d, OFDMA has been included in IEEE 802.16e to provide support in NLOS conditions and mobility. The earlier version uses 10–66 GHz, but the later version is expanded to use the lower bandwidth from 2 to 11 GHz, which also supports the 10–66 GHz frequency bands. There are some mandatory and some optional features associated with the PHY specification. From OSI 7 layer reference model, WiMAX only uses the PHY and MAC of data link layer. WiMAX physical and MAC layer architecture is shown in Figure 25.11.

25.3.4 OFDM PHY

The OFDM is developed to support high data rate and can handle multi-carrier signals. Its specialty is that, it can minimize the inter symbol interference (ISI) much more compared to

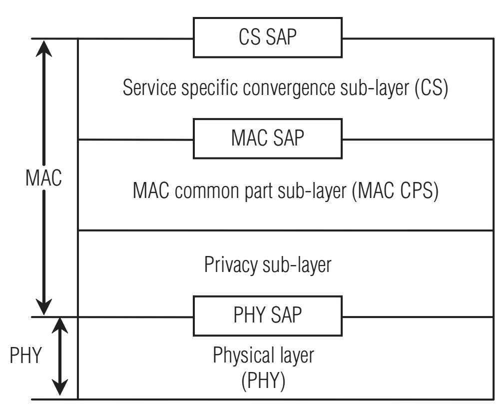


Figure 25.11 WiMAX physical and MAC layer architecture

other multiplexing schemes. This is the perfect choice for WiMAX as it can help to satisfy the requirements of efficient use of spectrum and to minimize the transmission cost. The OFDM also handles multi-path effect by converting serial data to several parallel data using fast Fourier transform (FFT) and inverse fast Fourier transform (IFFT).

The implementation of OFDM-PHY is different for two types of WiMAX. For fixed WiMAX, FFT size is fixed for OFDM-PHY and it is 256, but for mobile WiMAX the FFT size for OFDMA-PHY can be 128, 512, 1,024, and 2,048 bits. This helps to combat ISI and Doppler spread. Other difference between OFDM-PHY and OFDMA-PHY is OFDM splits a single high bit rate data into several low bit rate of data sub-stream in parallel, which are modulated by using IFFT; whereas OFDMA accepts several users' data and multiplexes those onto downlink (DL) sub-channel. Uplink multiple accesses are provided through uplink sub-channel.

OFDM-PHY

In OFDM-PHY, the FFT size is fixed and it is 256 bits in which 192 are for used data sub-carrier, 8 for pilot sub-carriers to perform synchronization and channel estimation, and 56 for null sub-carriers. For OFDM-PHY, the suitable symbol time is 64 μs, symbol duration is 72 μs, and guard time spacing is 15.625 kHz.

OFDMA-PHY

In mobile WiMAX, the FFT size can vary between 128 and 2048 bits, and to keep the sub-carrier spacing at 10.94 KHz, the FFT size should be adjusted. This will help minimize Doppler spreads. For OFDMA-PHY, the suitable symbol time is 91.4 μs, the symbol duration is 102.9 μs, and the number of symbols in 5 ms frames is 48.0.

Sub-channelization

WiMAX divides the available sub-carriers into several groups of sub-carriers and allocates to different users based on channel conditions and requirement of users. This process is called sub-channelization. Sub-channeling concentrates to transmit power to different smaller groups of sub-carrier to increase the system gain and to widen up the coverage area with less penetration losses that are caused by buildings and other obstacles. Mobile WiMAX's OFDMA-PHY permits sub-channelization in both uplink and DL channels. The BS allocates the minimum frequency and sub-channels for different users based on multiple access technique. That is why this kind of OFDM is called OFDMA. For mobile application, frequency diversity is provided by formation of distributed sub-carriers. Mobile WiMAX has several distributed carrier-based sub-channelization schemes. The mandatory one is called partial usage of sub-carrier. Another sub-channelization scheme based on unbroken sub-carrier is called AMC in which multi-user diversity got the highest priority. In this, allocation of sub-channels to users is done based on their frequency response. It is a fact that contiguous sub-channels are best suited for fixed and low mobility application, but it can give certain level of gain to the overall system capacity.

Mobile WiMAX's access method

Mobile WiMAX's access method is based on OFDMA which is also called multi-user-OFDM, especially designed for fourth generation wireless networks. It is a combination of FDMA, TDMA, and code division multiple access (CDMA), as it performs the same function like these access methods, that is, by dividing the available size to handle multiple users. OFDMA can be seen as an alternative to CDMA where each user gets different number of spreading code with different data rates. It resembles an alternative TDMA as low data rate users can send data with low transmission power, with constant and shorter delay. It can also be seen as a combination of TDMA and FDMA where the resources are divided according to time-frequency spaces and slots along with OFDM sub-carrier index. Different number of sub-carrier can be allotted to different number of users to maintain the data rate and error probability for each user. In a word, OFDMA is the best access method for multi-user environment.

25.4 Radio frequency identification (RFID)

The RFID stores product information in electronic tags that contain an antenna and a chip. RFID technology is similar to barcode labels but uses RF waves instead of laser rays to read the product code. RFIDs can respond to an RF signal and transmit their tag. They can store additional data, employ collision avoidance schemes, and comprise smart-card capabilities with simple processing power. They offer transmission rates of up to 115 Kbps and operate on many different ISM bands such as 27, 315, 418, 426, 433, 868, and 915 MHz.

> *The big difference between the bar code and RFID is that bar code is a LOS technology, whereas RFID does not require LOS. RFID tags can be read as long as they are within the range of a reader. Standard bar codes identify only the manufacturer and product, not the unique item.*

The size of the memory in a tag varies between 16 bits and hundreds of kilobits. RFID tags are initially programmed with a unique identification code obtained from electronic product code (EPC) global standards. EPC is a standardized numbering scheme so that it can be identified electronically.

EPC is either 96 or 64 bits long and is usually represented in hexadecimal notation. RFID tags are also commonly known as transponders. The transponder comprises a responder, an integrated circuit that contains non-volatile memory, and a simple microprocessor.

> *RFID offers transmission rates of up to 115 Kbps and operates on many different ISM bands such as 27, 315, 418, 426, 433, 868, and 915 MHz.*

25.4.1 Frequency bands and various types of RFID tags

The tags and readers use different transmission mechanisms in each HF (13.56 MHz) and UHF (400 to 900 MHz) band. There are two classes of tags (Class 0 tags are read-only and Class 1 tags are read/write). The various types of RFID tags are as follows:

- *Passive tags,* which are small, use the electromagnetic energy in the RF waves and do not require battery power. These can be produced in large quantities at low cost. The amount of data stored in a typical passive RFID tag is relatively small. Data transmission rates for the tags are also low.
- *Active tags* are equipped with a battery, and thus have a limited life due to the battery. They can transmit the signal farther away, and beacons transmit on a periodic basis.
- *Semi-active tags* use a built-in battery to power the circuit only when a reader first energizes the tag.
- *Sensory tags* can be equipped with various kinds of sensors to monitor and to record information.

25.5 Mobile ad-hoc networks (MANETs)

Wireless devices are creating a revolution in the way networked resources can interact with each other. Ad-hoc networks consist of a collection of wireless nodes. These nodes are connected with each other to dynamically establish an ad-hoc or on-the-fly network without the support of any centralized infrastructure. Such a network supports mobile computing anytime and anywhere, allowing the spontaneous formation of mobile networks for the period of usage. In such a network, each mobile host acts as a router that enables peer-to-peer as well as peer-to-remote wireless communications.

25.5.1 MANET topology

MANETs called as mobile ad-hoc networks are collections of mobile nodes dynamically establishing short-lived networks when fixed infrastructure is absent. Each mobile node is equipped with a wireless transmitter and a receiver with an appropriate antenna. All the mobile nodes are connected by wireless links that act as routers to all other mobile nodes. Nodes in MANETs are free to move and can be organized in an arbitrary manner. These features make MANETs very practical and easy to deploy in places where existing infrastructure is not capable enough to allow communication, for instance, in disaster zones or infeasible to deploy locations.

MANETs are considered to be the short-term temporary spontaneous wireless networks of mobile nodes communicating with each other without the intervention of any fixed infrastructure. It is an autonomous system of mobile nodes, mobile terminals, or MSs serving as routers interconnected by wireless links. Depending on the locations, antenna coverage patterns, transmit power levels, and co-channel interference levels, a wireless connectivity exists among participating mobile nodes at a given time, either in the form of random multi-hop transmissions or ad-hoc network. Network communications and management tasks are typically performed in a distributed manner. As the nodes move or adjust their transmission and reception parameters, MANET topology may change from time to time.

An ad-hoc mobile wireless network is a network without any base stations, that is, an infrastructure less network. Figure 25.12 depicts the formation and operation of a MANET. Data

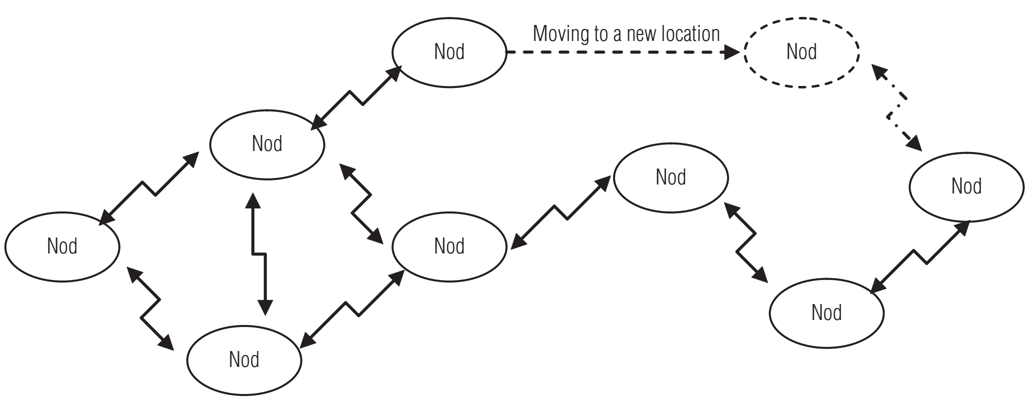


Figure 25.12 Formation of MANETs

packets are transmitted in a store-and-forward method from the source node to the destination node in peer-to-peer multi-hop intermediate nodes acting as routers. The network may either operate as standalone or as an extension of an infrastructure network with the help of few selected routers.

25.5.2 RF reader

A *reader* is a device that captures and processes the data received from the tags. A reader has separate transmitter and receiver circuits. HF tag readers can read 200 tags per second. For tags containing just an EPC, the actual rates will likely be between 500 and 800 tags per second. UHF specifications define the tag-to-reader data rate as twice that of the reader-to-tag. Tag-to-reader data rate can be up to 140.35 Kbps.

25.6 Summary

- The IEEE 802.15.4 standards define low-rate, low-power WPANs with low bandwidth requirements.
- ZigBee is a wireless technology developed as an open global standard to address the unique needs of low-cost, low-power wireless networks.
- IEEE 802.11 was concerned with features such as Ethernet matching speed, long range (100 m), complexity to handle seamless roaming, message forwarding, and data throughput of 2–11 Mbps.
- ZigBee technology is a low data rate, low power consumption, low cost, wireless networking protocol targeted towards automation and remote control applications.
- ZigBee can be implemented in mesh networks that cover larger area than is possible with Bluetooth.
- Network routing schemes are designed to ensure power conservation, and low latency through guaranteed time slots.
- Key features of PHY include energy and link quality detection, and CCA for improved coexistence with other wireless networks.
- IEEE 802.15.4 was designed to address the need for a low-cost and low-power wireless solution and has become a solid foundation for monitoring and controlling networks.
- The IEEE 802.16 defines the WMAN technology that is branded as WiMAX. The 802.16 includes two sets of standards: 802.16d for fixed WiMAX and 802.16e for mobile WiMAX. The WiMAX wireless broadband access standard provides the missing link for the "last mile" connection in metropolitan area networks where DSL, cable, and other broadband access methods are not available or too expensive. WiMAX also offers an alternative to satellite Internet services for rural areas and allows mobility of the customer equipment.
- The portable version of WiMAX, IEEE 802.16e utilizes OFDM and OFDMA, where the spectrum is divided into many sub-carriers. Each sub-carrier then uses QPSK or QAM for modulation.
- WiMAX eliminates the constraints of Wi-Fi. Unlike Wi-Fi, WiMAX is intended to work outdoors over long distances. WiMAX is a more complex technology and has to handle issues of importance such as QOS guarantee, carrier-class reliability, and NLOS.
- WiMAX is not intended to replace Wi-Fi. Instead, the two technologies complement each other.

Review questions

1. Explain in detail the architecture of ZigBee.
2. Write short notes on ZigBee devices.
3. What are the various functions of ZigBee and IEEE 802.15.4?
4. Explain WiMAX architecture.
5. Write short notes on LOS of WiMAX.
6. Compare the IEEE802.15 WPAN standards with respect to data rate, modulation scheme used, multiple access scheme, technology, and coverage.
7. Describe the need for WMAN technology.
8. Summarize IEEE 802.16 WMAN PHY specifications.
9. Compare the features of various wireless networks (WPAN, WLAN, WMAN).

Objective type questions and answers

1. ZigBee technology is _________ wireless networking protocol.
   (a) a low data rate (b) low power consumption
   (c) low cost (d) all
2. ZigBee compliant wireless devices are expected to transmit
   (a) 10–75 m (b) 75–100 m
   (c) 100–150 m (d) 150–200 m
3. The services provided by PHY are
   (a) PHY data service (b) PHY management service
   (c) both (a) and (b) (d) none
4. The features of the PHY are
   (a) energy detection (ED) (b) link quality indication (LQI)
   (c) channel selection (d) all
5. The services provided by MAC sub-layer are
   (a) MAC data service (b) MAC management service
   (c) both (a) and (b) (d) none
6. The features of MAC sub-layer are
   (a) channel access (b) GTS management
   (c) frame validation (d) all
7. In a peer-to-peer network, every device can communicate with any other device in its
   (a) transmission radius
   (b) outside transmission radius
   (c) inside and outside transmission radius
   (d) none
8. The device that forms the root of the network tree and bridge to other networks is
   (a) ZigBee coordinator (b) ZigBee router
   (c) ZigBee end device (d) none
9. The device that acts as an intermediate router, passing on data from other devices is
   (a) ZigBee coordinator (b) ZigBee router
   (c) ZigBee end device (d) none

10. The ZigBee device that requires the least amount of memory is
    (a) ZigBee coordinator (b) ZigBee router
    (c) ZigBee end device (d) none

Answers: 1. (d), 2. (a), 3. (c), 4. (d), 5. (c), 6. (d), 7. (a), 8. (a), 9. (b), 10. (c).

True/False

1. The 802.16e Standard is the version of WiMAX that will offer true mobility. (**T**rue/**F**alse)
2. Areas underserved by broadband access are not a potential application for WiMAX. (**T**rue/**F**alse)
3. WiMAX is sometimes labelled as the "wireless LAN" solution. (**T**rue/**F**alse)
4. Wireless ISPs type of wireless operator stands to benefit greatly with the availability of 802.16e. (**T**rue/**F**alse)
5. 3G wireless technology could actually pose a bigger threat to WiMAX. (**T**rue/**F**alse)

Answers: 1. T, 2. F, 3. F, 4. T, 5. T.

Open book questions

1. Mention the different task groups of IEEE 802 that have developed the wireless network standards for various wireless works.
2. List the benefits of WiMAX.
3. Describe the two forms of WiMAX services.
4. Write short notes on ZigBee and IEEE 802.15.4.
5. Write short notes on WiMAX and IEEE 802.16.
6. Compare the IEEE 802.15 WPAN standards.

Further reading

Bhushan, N., Lott, C., Black, P., Attar, R., Yu-Cheun Jou, Fan, M., Ghosh, D., and Au, J., "cdma2000 1xEV-DO Revision A: A Physical Layer and MAC Layer Overview," *IEEE Communications Magazine*, April 2005.

Eklund, C., Marks, R. B., Stanwood, K. L., Wang, S., "IEEE Standard 802.16: Technical Overview of the Wireless MAN Air Interface for Broadband Wireless Access," *IEEE Communications Magazine*, June 2002.

IEEE 802.11, "Wireless LAN Medium Access Control (MAC) and Physical (PHY) Layer Specifications," November 1997, 2007, 2008.

IEEE P802.16-REVd/D5-2004, "Air Interface for Fixed Broadband Wireless Access Systems," 2004.

Tschofenig, H., and M. Riegel, "An Introduction to 802.16(e)," SIEMENS, 2005.

WiMAX Forum, "Regulatory Positions and Goals of the WiMAX Forum," June 16, 2004.

WiMAX Forum, "Fixed, Nomadic, Portable and Mobile Applications for 802.16-2004 and 802.16e WiMAX Networks," November 2005.

Wireless Networking 26

26.1 Introduction

There are two ways to connect a computer to a network: wired and wireless. Wired network uses cables and network adapters. In wired networks, data travels as electrical signals in the case of a copper wire or coaxial cable and optical signal in the case of fibre optic cable. Two computers can be directly wired to each other using a crossover cable. To accommodate more computers in wired networks, network-connecting devices like hubs, switches, or routers are used. In wireless networks, wires are not used and the signal travels as an electromagnetic wave (EM wave or radio waves) or infrared light through the air. There are different techniques through which data is transferred between the sources to the destination known as routing techniques. These are based on (i) the type of traffic such as voice which needs a real time communication or data which may be bursty and (ii) the type of connection. Based on the type of connection, two types of networks are used in traffic routing of both wired and wireless networks. They are circuit switched and packet switched. Examples of wired and wireless networks are IEEE802.3 and IEEE802.11. IEEE802.3 is the wired local area network (LAN) and uses carrier sense multiple access with collision detection (CSMA/CD) as media access control (MAC) protocol. IEEE802.11 (or wireless fidelity (Wi-Fi)) is the wireless LAN and uses carrier sense multiple access with collision avoidance (CSMA/CA) as the MAC protocol.

This chapter begins with the discussion of various generations of wireless networks, differences in wireless and fixed telephone networks, and then describes the various traffic routing techniques such as circuit switching, packet switching, and various wide area networks (WANs) technologies (frame relay, asynchronous transfer mode (ATM), ISDN, etc.)

26.2 Various generations of wireless networks

Wireless communication technologies include various devices and systems with different standards, protocols, architectures, modulation, and coding techniques. The most important technologies are listed below:

- Global system for mobile communications (GSM)
- General packet radio service (GPRS)
- Wireless fidelity (Wi-Fi) IEEE 802.11
- Worldwide interoperability for microwave access (WiMAX) IEEE 802.16

Comprehensive description of wireless communication evolution and cellular system generations (1G to 4G) are given in Chapter 1. In this section, a brief review of various generations of wireless networks and comparison of various wireless technologies is presented. Table 26.1 gives an overview of the hybrid wireless technologies, standards, modulation techniques, data rate, bandwidth, and frequency of operation.

1G: The 1st generation (1G) communication technology employees *AMPS* standard using the *FDMA* technique. Although TDMA and FDMA techniques have been used in 1G technology, however, this technology could not handle high data rates and could support *voice services* only.

Table 26.1 General overview of hybrid wireless technologies

Technology	Standards	Multiple access and modulation coding	Data rate	Bandwidth	Frequency
1G	AMPS	FDMA, TDMA	10 kbps maximum	2.4 –3 KHz	400 – 800 MHz
2G	GSM	TDMA, CDMA	100 kbps maximum	9.6 – 14.4 KHz	850/900/ 1800/ 1900 MHz
2.5G	IS 95-B GPRS, IS 95C EDGE	8PSK	2Mbps	0.384 – 1MHz	850/950/ 1800/ 1900 MHz
3G PP2 3G	CDMA2000 IMT2000 UMTS	OFDM, OFDMA, SDMA HSPDA, WCDMA	42Mbps	1 – 2 MHz	
3G LTE	LTE	OFDMA for Downlink SC-FDMA for uplink Support FDD and TDD duplexing, DL/ UL Modulation QPSK,16QAM, 64 QAM	100 Mbps down link 50 Mbps uplink	1.4 – 20 MHz (1.4,3,5,10, 15,20) MHz	3.5 GHz
Wi-Fi	802.11	CCK in 802.11b OFDM with data modulation BPSK, QPSK, 16QAM, 64QAM, DSSS/FH, MiMo for 802.11 n	54 Mbps Up to 320 Mbps in 802.11n	20 MHz	2.412-2.484 5.15-5.25 GHz
WiMAX	802.16	OFDMA multiple access for uplink and down link, TDD duplexing, Data modulation QPSK, 16QAM, 64QAM	100 Mbps	5-20MHz (5,7,8.75, and 10MHz for Mobile	2-66 GHz in 802.16 2-11 GHz in 802.16a Mobile 3.4 GHz
DVB-S		QPSK	45 Mbps	Upto 540 MHz capacity	L band, C band
DVB-S2		QPSK, 8PSK, 16APSK, 32APSK	65 Mbps	Upto 780 MHz capacity	Ka, Ku band
DVB-RCS		MFTDMA	65 Mbps	Upto 500 MHz capacity	Ka band, Ku band

2G: The GSM technology, which is also known as the 2nd generation (2G) standard for mobile communication, is available all over the world allowing maximum bit rate of 14.4 Kbps using TMDA and CDMA techniques.

2.5G: The GSM technology standard rapidly evolved into the GPRS and enhanced data rates for GSM evolution (EDGE) standards, which became 2.5G technology. Meanwhile, the hybrid combination of 2.5G and 3G also came into existence using CDMA2000.

3G: The technology has been further developed to the universal mobile telecommunications system (UMTS) using wideband code division multiple access (WCDMA), which supports 14 Mbps with high-speed downlink packet access (HSDPA) and 42 Mbps with high-speed packet access (HSPA+).

Both CDMA2000 and WCDMA air interface systems are accepted as a part of the IMT-2000 family of 3G standards. User can access e-mail and Internet using HSPA.

LTE: The long-term evolution (LTE) technology is based on new orthogonal frequency division multiplexing (OFDM) air interface techniques instead of WCDMA and offers higher data rates. Orthogonal frequency division multiple access (OFDMA) and single carrier frequency division multiple access (SCFDMA) are now becoming popular due to better quality of service (QoS).

4G: In addition to OFDMA which is used in WiMAX, SCFDMA and multi-carrier codes division multiple access (MCCDMA) are gaining more popularity. This technology is called 4th generation (4G) and is purely based on packet switching instead of circuit switching, or a mixture of both as in 3rd generation (3G).

Wi-Fi: The Wi-Fi task group was created in 1997 and released the first set of specifications for Wi-Fi operating at 2.4 GHz. Wi-Fi is called the IEEE 802.11 standard developed by the IEEE standard committee working group 11. Three standards most commonly used by Wi-Fi are 802.11 a, b, and g. These standards define the physical (PHY) layer and the MAC layer. The *MAC* layer is the same for these three Wi-Fi IEEE 802.11 standards but the *PHY* layer differs among them.

WiMAX*:* The IEEE 802.16 family of standards is often referred to as WiMAX. WiMAX is a metropolitan area access technique with many encouraging features such as cost efficiency, flexibility, and fast networking, which provides wireless access as well as serves as a wireless expanding for wired network access. This facilitates the network access for remote or suburban areas. The coverage area of WiMAX spans 30–50 km. It provides high speed of data rates more than 100 Mbps in 20 MHz bandwidth.

The major difference between the Wi-Fi IEEE 802.11 and the WiMAX IEEE 802.16 standards is that WiMAX 802.16 uses both external Reed-Solomon block code concatenated and inner convolutional code.

Wi-Fi and WiMAX offer higher data rates as compared to other wireless technologies, including 3G. The Wi-Fi currently offers data rates up to 54 Mbps, which will increase, to 320 Mbps in the new IEEE 802.11n standards.

26.2.1 Satellite technologies

Very small aperture terminal (VSAT) networks offer satellite-based services that support broadband Internet, data, LAN, and voice and teleconference communications and can also provide powerful, value-added, dependable private and public network communications.

Other satellite technologies can be used as distribution networks such as digital video broadcasting – satellite (DVB-S) which only allows unidirectional links, and digital video broadcasting return channel through satellite (DVB-RCS) supports bidirectional links. Most current satellite links used for IP data distribution are based on single channel per carrier (SCPC) and DVB-S carrier. During the last few years, DVB-RCS was introduced to the market. In the near future, DVB-RCS compliant systems will become popular, while other technologies will evolve towards broader applications, first DVB-S2 and second on-board processor (OBP)-based systems. The DVB-RCS operates mainly ku bands with transmit frequency band 14–14.5 GHz and receive frequency band 10.7–12.2 GHz.

26.2.2 Wireless communication networks based on coverage

Wireless communication networks are classified into four different network types according to their range:

- Wireless personal area network (WPAN)
- Wireless local area network (WLAN) (e.g., Wi-Fi)
- Wireless metropolitan area network (WMAN) (e.g., WiMAX)
- Wireless wide area network (WWAN) (e.g., Satellite)

The classification of wireless networks is shown in Figure 26.1. The Wi-Fi and WiMAX are WLAN and WMAN network technologies, respectively, while satellite is considered as WWAN. Wireless communication systems such as Bluetooth, cellular, WLAN, and WMAN, are now in widespread use and have become essential in everyday life. Their popularity is extremely high due to the advantages of ubiquitous communication (i.e., anywhere, anytime with anyone).

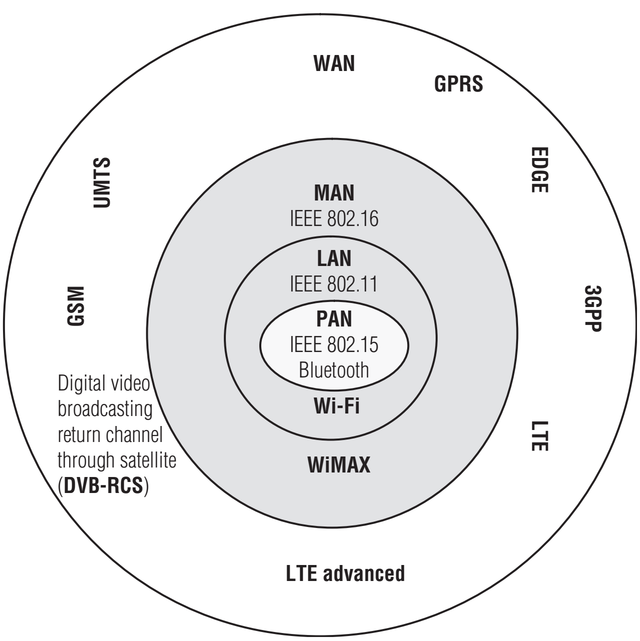


Figure 26.1 Classification of wireless network

26.3 Fixed network transmission hierarchy

Wireless networks rely heavily on landline connections. Several standard digital signalling (DS) formats form a transmission hierarchy that allows high data rate digital networks which carry a large number of voice channels to be interconnected throughout the world. These DS formats use time division multiplexing (TDM). The most basic DS format is called DS-0, which represents one duplex voice channel which is digitized into a 64 Kbps binary pulse code modulation (PCM) format. The second DS format is DS-1, which represents 24 full duplex DS-0 voice channels that are time division multiplexed into a 1.544 Mbps data stream (8 Kbps is used for control purposes).

Digital transmission hierarchy is the T (N) designation, which is used to denote transmission line compatibility for a particular DS format. DS-1 signalling is used for a T1 trunk, which is a popular point-to-point network signalling format used to connect base stations (BSs) to the mobile switching centre (MSC). T1 trunks digitize and distribute the 24 voice channels onto a simple four-wire full duplex circuit. In Europe, CEPT (European Conference of Postal and Telecommunications administrations) has defined a similar digital hierarchy. Level 0 represents a duplex 64 Kbps voice channel, whereas level 1 concentrates 30 channels into a 2.048 Mbps TDM data stream.

26.4 Differences in wireless and fixed telephone networks

Differences between wired and wireless networks are seen in terms of medium of communication (channel), speed, cost, ease, capacity, reliability, and so on. The main difference between a wired and wireless data communication networks is the existence of physical cabling. Wireless technology is very convenient, enabling a user to roam anywhere within the vicinity. Wireless networks have been one of the fastest growing segments of the telecommunications industry. They have become progressively faster over the years, and nowadays the speed at which they operate is almost indistinguishable from wired connections.

The performances of both networks are compared in terms of the following parameters:

- ***Mobility***: Wired networks make you immobile while wireless ones provide you with convenience of movement. Real-time information can be accessed from anywhere at any time.
- ***Cost***: Wired networks prove expensive when covering a large area because of the wiring and cabling while wireless networks do not involve this cost.
- ***Transmission speed***: Wired networks have better transmission speeds than wireless ones. In a wired network, a user does not have to share space with other users and thus gets dedicated speeds while in wireless networks, the same connection may be shared by multiple users.
- ***Dynamic versus static***:
  - Wireless networks are highly dynamic, that is, network configuration need to be changed every time a user changes coverage region, this is to be done in fraction of seconds. For example, dynamic network configuration in wireless cellular networks can be done by handoffs, roaming procedures.
  - Wired networks are virtually static. When a user changes the region (place) all the (wired) connectivity needs to be changed from one place to the other.
- ***Network capacity***: In wired networks, the network capacity can be increased by simply increasing the copper cables or optical fibre cables, whereas in wireless networks to increase the capacity extensive RF optimization needs to be done.

Table 26.2 Difference between wireless and fixed telephone networks

Parameter	Wired networks	Wireless networks
Communication channel/media	Optical fiber coaxial cable	Free space
Form of data	Current signal	RF or infrared signal
Reliability	High	Low
Congestion	Less	More
Speed	More	Less
Cost	More	Less
Security	More	Less
Mobility	No	Yes

- ***Reliability***: Wireless LANs suffer a few more reliability problems than wired LANs, due to interference from other home appliances including microwave ovens, cordless telephones, and garage door openers.

  Important differences of wireless and fixed telephone networks are summarized in Table 26.2

26.5 Traffic routing in wireless networks

Networking technologies can be distinguished on the basis of the method they use to determine the communication path between the devices over which data will flow. There are two approaches to transfer data: one is circuit switching and the other one is packet switching.

26.5.1 Circuit switching

Circuit switching is a technique in which a system seeks out the physical "copper" path from the source node to the destination node. Telephone system follows circuit switching. In circuit switching, an end to end path is set before any data is sent. That is why it takes several seconds between the end of dialling and the start of ringing for international calls. During this interval, the telephone system is actually hunting for a copper path, and the call request signal must propagate all the way to the destination and it should be acknowledged. In circuit switching, line is unused by others until the call is terminated and each caller may make one call at a time.

In circuit switching, once a connection is established it remains throughout the session.

Circuit-switched networks are based on TDM, in which numerous signals are combined for transmission on a single communication line or channel.

Figure 26.2 illustrates the circuit-switched network. From the figure we can observe that before communication can occur between two devices (A and B), a circuit is to be established between them. This is shown as a dotted line for the data flow from device A to device B, and a matching solid line from B back to A. Once connection is set up, all communication between these devices takes place over this circuit, even though there are other possible ways that data could conceivably be passed over the network of devices between them.

Advantages of circuit switching are as follows:

- Fixed delays
- Guaranteed continuous delivery

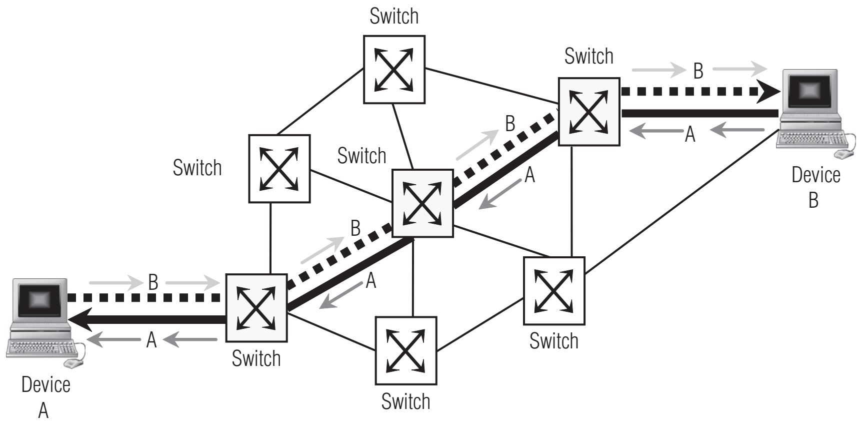


Figure 26.2 Circuit switching

Disadvantages of circuit switching are as follows:

- Circuits are not used when session is idle
- Inefficient for bursty traffic
- Circuit switching usually done using a fixed rate stream (e.g., 64 Kbps)
- Difficult to support variable data rates

26.5.2 Packet switching

In packet switching, no specific communication path is used for transfer of data. Instead, the data is divided into small units called packets and sent over the network. The packets can travel in different routes, combined or fragmented, as required to get them to their final destination. On the receiving end, the process is reversed where the data is read from the packets and re-assembled into the form of the original data. Circuit switching statically reserves the required bandwidth in advance, whereas packet switching acquires and releases bandwidth when it is needed. Such type of communication without a dedicated path between the transmitter and the receiver is termed as connectionless. Figure 26.3 illustrates the packet switching network.

When a message is broken into packets, a certain amount of control information is added to each packet to provide source and destination information and identification. Packet data format is shown in Figure 26.4.

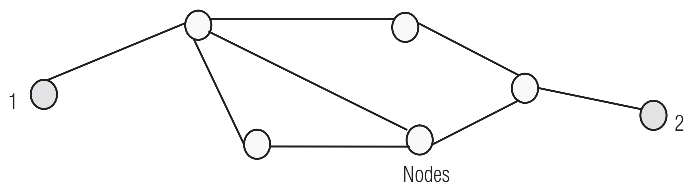


Figure 26.3 Packet switching

Figure 26.4 Packet data format

- Header – contains source address, destination address, packet sequence number, and other routing and billing information.
- Trailer – contains CRC sum, which is used for error detection at receiver.

All computer networks come under packet switching. They use the following protocols:

- X.25
- Frame relay
- ATM

Packet switching is also called packet radio when used by a wireless link. It provides excellent channel efficiency for data transmission, since the channel is utilized only when sending or receiving bursts of information.

Advantages of packet switching are as follows:

- Network supports many connections simultaneously
- Short messages not delayed by long messages
- More efficient than circuit switching

Disadvantages of packet switching are as follows:

- Performance drops when many users share the same network

26.5.3 Comparison of circuit and packet switching networks

Circuit switching is best suited for dedicated voice-only traffic, or for instances where data is continuously sent over long periods of time. Packet switching (or virtual switching) implements connectionless services for large number of data users, who remain virtually connected to the same physical network. The performance of the circuit and packet switched networks are compared and summarized in Table 26.3.

Table 26.3 Comparison of circuit and packet switching network

Parameter	Circuit switching	Packet switching
Dedicated "copper" path	Yes	No
Bandwidth available	Fixed	Dynamic
Each packet follows the same route	Yes	No
Store-and-forward transmission	No	Yes
Call setup	Required	Not required
Call charging	Per minute	Per packet

26.6 Wide area networks link connection technologies

A WAN is a computer network whose communications link cross metropolitan, regional, or national boundaries. WANs use routers and public communications links. The largest and most well-known example of a WAN is the *Internet*.

WANs exchange information across wide geographic areas. They operate at three-layer open system interconnect (OSI) model given below:

- PHY
- Data link
- Network

WAN connections typically function at the physical and data link layers of the OSI reference model, and are made over serial connections. WAN connections operate at a lower speed than LAN connections, and can be made as point-to-point, point-to-multipoint, and switched WAN connections.

There are several options available to implement WAN. The differences between them are *technology*, *speed,* and *cost*. First, we will shortly describe the differences between them. WAN connections can be either over a private infrastructure or over a public infrastructure. Private WAN connections include both dedicated and switched communication link options. A public WAN connection option, such as the Internet is now a sophisticated technology widely used in our daily communication. Figure 26.5 shows different ways of WAN link connection.

26.6.1 Leased line connection

Leased lines are digital, permanent, and dedicated connections. They are not suitable for a long distance connection due to the cost and time of pre-established line before successful transportation.

The main disadvantage of leased line is *fixed bandwidth*, whereas in many applications the traffic is variable, sometimes even empty. However, the dedicated capacity removes latency or

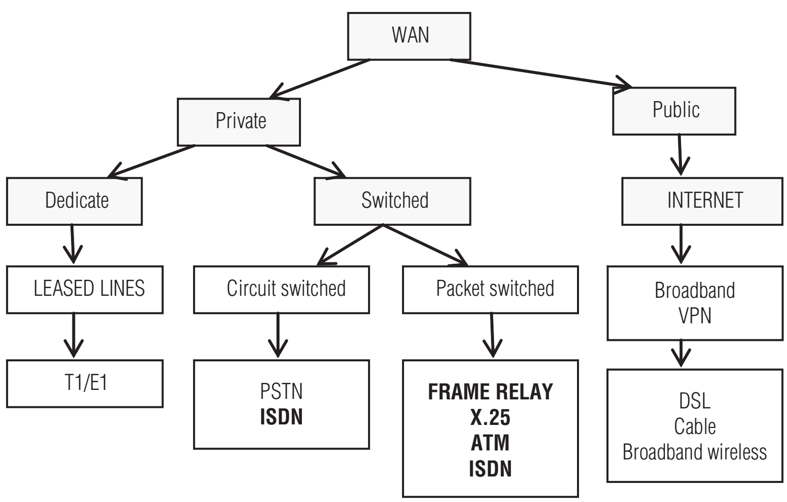


Figure 26.5 Different ways of WAN link connection

jitter between the endpoints. They are particularly suitable for applications such as video over IP (VoIP) where constant availability is essential. Examples of leased line connections are T1 and E1.

26.6.2 Internet connection

There are a number of ways of getting connected to the Internet. These include

- *Dial-up* using telephone networks
- Cable modem
- Cable television networks
- Integrated subscriber digital line (ISDN)
- Digital subscriber loop (DSL)
- Broadband (ADSL)
- *Wireless* such as Wi-Fi, WiMAX, and satellite
- Mobile devices (e.g., latest mobile phones)

Table 26.4 illustrates the comparison of various WAN link connections.

Table 26.4 Comparison of various WAN connections

Type	Description	Advantages	Disadvantages	Protocol used
Leased line	Point to point connection Between two computers or LANs	Most secure	expensive	PPP, HDLC, SDLC, HNAS
Circuit switching	A dedicated circuit path is created between endpoints. Best example is dialup connections	Less expensive	Call setup	PPP, ISDN
Packet switching	Devices transport packets via a shared single point to point or point to multipoint link across a carrier internetwork. Variable length packets are transmitted over permanent virtual circuits (PVCs) or switched virtual circuits (SVCs)	Available Dynamic Bandwidth	Shared media across link	X.25, Frame Relay
Cell relay	Similar to packet switching, but uses fixed length cells instead of variable length packets. Data is divided into fixed length cells and then transported across virtual circuits	Best for simulated use of voice and data	Overheated can be considerable	ATM
Internet	Connectionless packet switching using the internet as the WAN infrastructure uses network addressing to deliver packets. Because of security issues, VPN technology must be used	Least expensive globally available	Least secure	VPN, DSL, Cable Modem, Wireless

26.7 X.25 protocol

The first commercial product of packet switching technology was X.25 network.

X.25 defines how a packet-mode terminal can be connected to a packet network for the exchange of data.

X.25 protocol was adopted as a commonly used network protocol standard by the consultative committee for international telegraph and telephone (CCITT). The X.25 protocol allows computers on different public networks (such as CompuServe or a TCP/IP network) to communicate through an intermediate computer at the network layer level. X.25 defines the interface between data terminal equipment (DTE) and data circuit equipment (DCE). The DTE represents the end user, or host system. The DCE represents the boundary node of the packet switched public data network (PSPDN) and, in other words, the DCE is the point of access into the network (Figure 26.6).

X.25 architecture comprises first three layers [physical layer (X.21), frame layer (link access procedure, balanced (LAPB)), and packet layer protocol (PLP)] of the OSI seven-layered architecture. Figure 26.7 shows the hierarchy of X.25 protocols in the OSI model and Figure 26.8 shows the X.25 layers in relation to the OSI Layers.

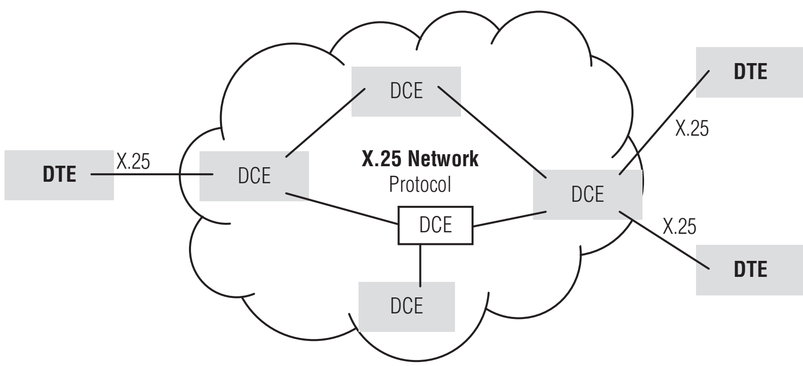


Figure 26.6 X.25 communication interface

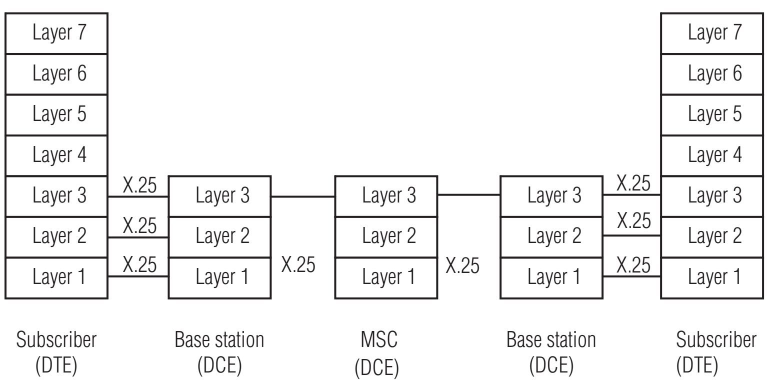


Figure 26.7 Hierarchy of X.25 in OSI model

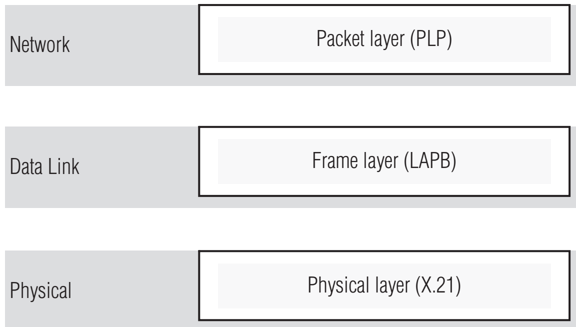


Figure 26.8 X.25 Layers in relation to the OSI layers

The Layer 1 (**physical layer (X.21)**) protocol deals with the electrical, mechanical, procedural, and functional interface between the subscriber (DTE) and the base station (DCE). The Layer 2 (**frame layer (LAPB)**) protocol defines the data link on the common air interface between the subscriber and the base station. The layer 3(PLP) provides connection between the BS and the MSC, and is called as PLP.

26.7.1 X.25 physical layer

The physical layer of the OSI reference model defines the physical interface between two adjacent devices. The physical link between a DTE and a DCE is a dedicated, synchronous, serial, point-to-point, and full duplex channel. The X.25 physical layer provides a subset of the functions that are defined by the physical layer of the OSI model. X.25 also requires a dedicated, synchronous, serial, point-to-point, and full-duplex circuit between the user (DTE) and the network (DCE).

26.7.2 X.25 link layer

The OSI data link layer provides error-free communication between two adjacent devices. Although the media connecting two devices may not be perfect, the data link layer is responsible for finding and correcting all bit errors on the line. The X.25 link layer provides the same functions of the OSI data link layer. The X.25 link layer is not a general protocol. For example, no mechanism exists in X.25 to turn the line around for half-duplex communication because the physical link must be full duplex.

26.7.3 X.25 Packet layer protocol

It is the network layer in X.25. This layer is responsible for establishing the connection, transferring the data, and terminating the connection between **two DTEs** (Figure 26.9). PLP is also responsible for creating the virtual circuits (VCs) and negotiating network services between two DTEs. VCs in X.25 are created at the network layer whereas in some other WANs such as frame relay and ATM it is created at the data link layers.

26.8 Frame relay

Frame relay is designated as a 2G X.25. It uses packet switching with VCs.

Frame relay operates only at PHY and data link layers.

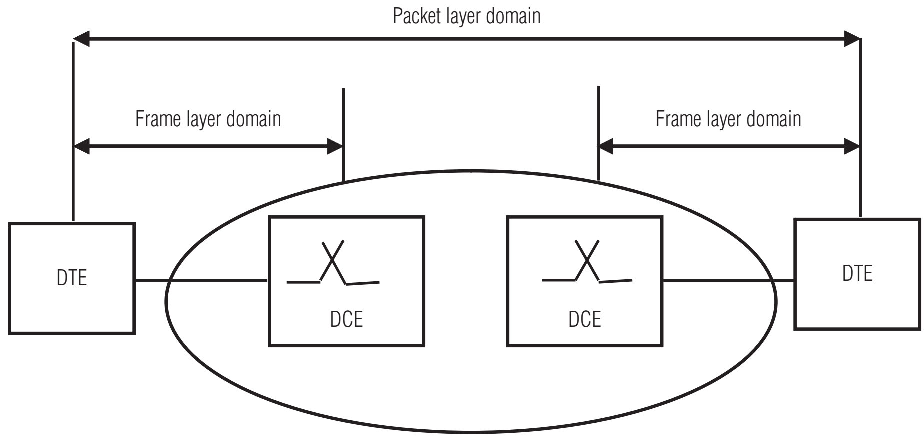


Figure 26.9 Frame layer and packet layer domains

The term VC is used most frequently to describe connections between two hosts in a packet-switching network. In this case, the two hosts can communicate as though they have a dedicated connection even though the packets might actually travel very different routes before arriving at their destination. An X.25 connection is an example of a VC. A permanent virtual circuit (PVC) is permanently available to the user just as though it were a dedicated or leased line continuously reserved for that user.

The frame relay is a VC-based packet-switching service with no error recovery and no flow control.

Whenever a frame relay switch detects an error in a packet, its only possible course of action is to discard the data. This results in a network with lower processing overheads and higher transmission rates than X.25, but requires intelligent end systems for data integrity. Frame relay is extensively used today to allow LANs on different corporate campuses to send data to each other at reasonably high speeds. As shown in Figure 26.10, frame relay interconnects LANs through IP routers, with each IP router in a different corporate campus. Frame relay offers a corporation

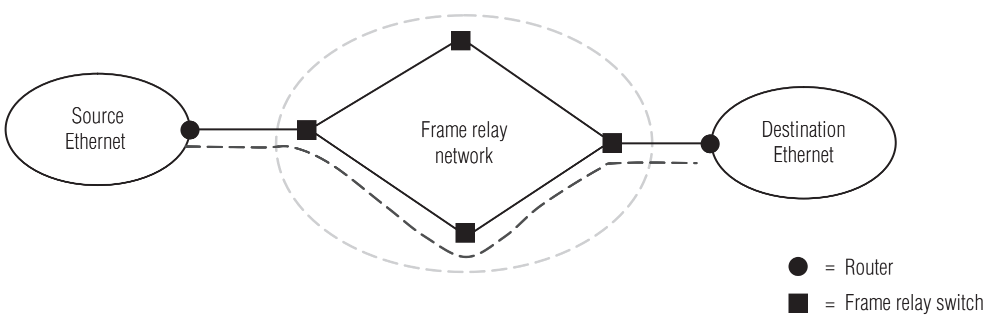


Figure 26.10 Public frame relay network interconnected two LANs through routers located on the Ethernets

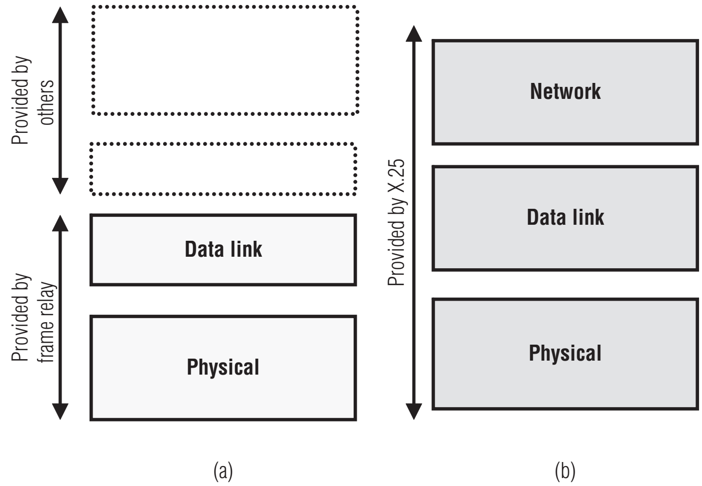


Figure 26.11 Comparing layers in frame relay and X.25

an alternative to sending its inter-campus IP traffic over the public Internet, for which the corporation may have reliability and security concerns.

The dotted line represents a VC. Frame relay networks can use either switched virtual circuits (SVCs) or PVCs. For router interconnection, PVC is often permanently established between each pair of routers.

Number of PVCs necessary to interconnect N routers is given by

$$N(N - 1)/2$$

Throughout our discussion, we shall assume that the frame relay network uses PVCs (which is the more common case).

26.8.1 Frame relay layers

Frame relay operates only at the physical and data link layers. Figure 26.11 illustrates the layers of frame relay and X.25.

Physical *layer*: This specifies the physical interface between the node (computer, terminal) and the link that connects the frame relay network. No specific protocol is used at this layer and it is left to the implementer to use whatever is available. Also, it supports any of the protocols recognized by ANSI (Figure 26.12).

***Data link layer*:** This employs a simplified version of High level Data Link Control (HDLC) called core Link Access Procedure/Protocol for Frame Mode services (LAPF) with no extensive error and flow control fields.

26.9 ATM

ATM is a standard for cell-based relay that carries voice, video, and data in small, fixed-size cells.

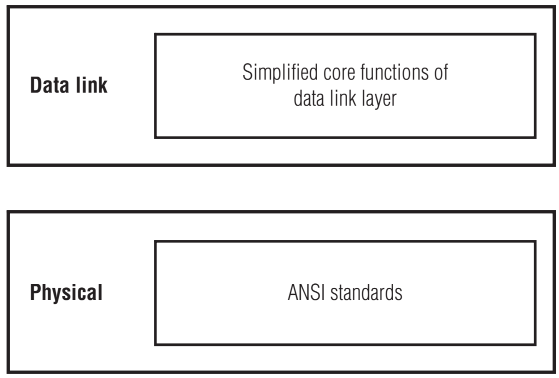


Figure 26.12 Physical and data link layers

ATM networks are connection-oriented networks that combine the benefits of circuit switching (guaranteed capacity and constant transmission delay) with those of packet switching (flexibility and efficiency for intermittent traffic). Traditional circuit-based networks use TDM, in which users are assigned a predetermined time slot; no other device can transmit during this time slot. If a station has a lot of data to send, it can transmit only during its time slot, even if the other time slots are empty. Conversely, if the station has nothing to transmit, the time slot is sent empty and is wasted. This arrangement is called synchronous transmission.

ATM is asynchronous, which means that time slots are available on demand. This allows for a more efficient use of available bandwidth. ATM uses small, fixed-sized cells (as opposed to the variable sized frames in frame relay), which have 53 bytes (Figure 26.13). Computers usually define things in powers of two or eight. The 53-byte cell size represents a compromise between the phone-standards folks and the data standards folks.

ATM networks use two devices. These are ATM switches and ATM endpoints. ATM switches accept cells from an endpoint or another switch, evaluate the cell header, and quickly forward the cell to another interface toward the destination. An ATM endpoint contains an ATM network interface adapter and is responsible for converting digital data into cells and back again. Examples of ATM endpoints include workstations, LAN switches, routers, and video coder decoders (codec's).

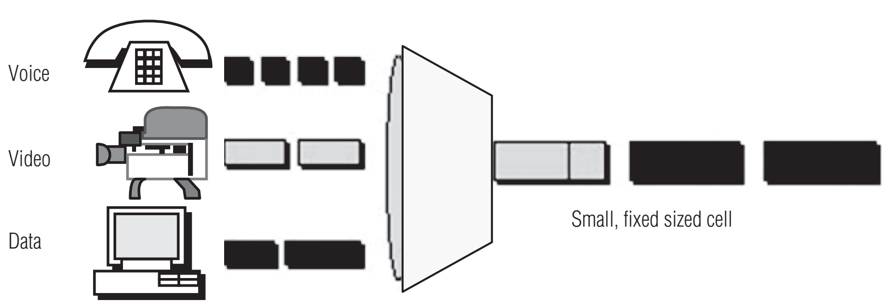


Figure 26.13 ATM cells

ATM networks can mark traffic after it is converted from its original data format. Some traffic, such as voice and video, must be transferred through the network at regular intervals with little variation in delay. Otherwise, the destination receives low-quality voice or video transmission. Data traffic is less sensitive to network delays and can be handled differently. To ensure the appropriate delivery for each of these traffic types, ATM devices employ QOS mechanisms that involve reserving bandwidth, shaping traffic to meet the reserved bandwidth, and policing traffic that exceeds the capacity that can be handled.

26.9.1 Data transmission speed

ATM transmits at speeds from a few Mbps to many Gbps. High-speed ATM circuits typically require optical fibre cables to transmit such high speeds. Speeds of these circuits are characterized as “optical carrier (OC)” class and are represented as OC-number. The number represents the multiple of the base OC-1, which is a standard circuit and which can carry 51.84 Mbps. Common circuit speeds are OC-3 (155.52 Mbps), OC-12 (622.08 Mbps), and OC-192 (9,953.28 Mbps or roughly 10 Gbps).

26.9.2 ATM advantages

Several organizations that have a large investment in client/server technology are moving to implement ATM. New demands to transmit voice and video data, as well as large database queries, require the bandwidth capabilities of ATM. The five main advantages of ATM are as follows:

- Accommodates high-speed telecommunication
- Dependable and flexible at geographic distances
- Handles data, voice, and video transmissions
- Provides potentially significant cost saving in network resources
- Encodes in fixed-length, 53-byte relay units of data called *cells*

26.10 Virtual private networks

The whole world is becoming a global village; this is due to the development in the communication world. This has also led to the expansion of the businesses across the globe. Companies throughout the world instead of communicating with the regional offices focused their attention towards developing their own private networks. Therefore, there is a need to develop a private network to communicate to far-off places with encrypted and secure communication mode. Virtual private network (VPN) is one such private network that companies use to communicate throughout the world.

A VPN is a secured connection between two devices over a shared, unsecured network. The security of the network is established by encryption method.

There are three types of VPNs:

- ***Remote access VPN*** – VPNs have been used for some time for mobile devices such as laptops to connect to their corporate headquarters over the Internet.
- ***Site-to-site intranet*** – when a gateway is established at various different locations but of the same company for secure communication lines then we term it as site-to-site intranet. It helps internal users communicate with each other.

- ***Site-to-site extranet*** – site-to-site extranet also has different communication gateways but the purpose is to communicate with the external business partners. The extranet provides enough business security, because only authenticated users can make use of it.

Site-to-site VPNs can be a very cost-effective way to connect relatively small locations to corporate headquarters over Internet services, such as broadband cable and DSL. VPNs are also used to some degree to authenticate users to local access points in a wireless environment. VPNs are addressed in a bit more depth in a later section.

26.11 Wireless data services

Wireless data services can provide the traveller with the required network access in many situations where wired access to the public network is impractical or inconvenient.

1G cellular system that provides data communications using circuit switching is inefficient for dedicated mobile data services such as Fax, e-mail, and short messaging. Also, voice filtering must be deactivated when data is transmitted over 1G cellular networks, and a dedicated data link must be established over the common air interface. Until recently, the demand for packet data services has been significantly less than the demand for voice services and, therefore, 1G subscriber equipment design has focused mainly on voice communications only.

26.11.1 Advanced radio data information systems

Advanced radio data information system (ARDIS) is a combined IBM's private packet radio data communication system and Motorola's international shared use radio network.

> *ARDIS provides 800 MHz two-way mobile data communications for short length radio messages in urban and in-building environments, and for users travelling at low speeds.*

Short ARDIS messages have low retry rates but high packet overhead, while long messages spread the overhead over the length of the packet but have a higher retry rate. ARDIS has been deployed to provide excellent in-building penetration, and large-scale spatial antenna diversity is used to receive messages from mobile users. When a mobile sends a packet, many BSs that are tuned to the transmission frequency attempt to detect and decode the transmission, in order to provide diversity reception for the case when multiple mobiles contend for the reverse link. In this manner, ARDIS BSs are able to insure detection of simultaneous transmissions, as long as the users are sufficiently separated in space.

26.11.2 RAM mobile data

RAM mobile data (RMD) was established in 1989 to build and operate a nationwide mobile data network. RMD network coverage is only slightly less expensive than that of ARDIS. RMD is a public, two-way data service based on the Mobitex protocol developed by Ericsson. RMD offers specific mobile data communication solutions built on mobile networks like Mobitex, GPRS, UMTS, and EDGE. These technologies were designed to make use of slandered, two-way voice, land mobile radio channels, with 12.5 or 25 kHz channel spacing. They include our track-and-trace vehicle tracking and journey registration application, an all-in-one solution for fleet and field service organizations.

RAM provides street level coverage for short and long messages for users moving in an urban environment. RAM has data and facsimile, Fax messages are transmitted as normal text to a gateway processor, which then converts the radio message to an appropriate format by merging it with a background page. Thus, the packet switched wireless transmission of a normal length message instead consists of a much larger fax image, even though the end user receives what appears to be a standard fax.

Mobitex is a packet switched network with packets reaching a maximum of 512 bytes long. Bit rates of 8 and 19.2 Kbps are offered.

26.12 Common channel signalling

Common channel signalling (CCS) is a digital communications technique that provides simultaneous transmission of user data, signalling data, and other related traffic throughout a network. CCS is signalling in which a group of voice and data channels share a separate channel that is used only for control signals (Figure 26.14).

Because control signals have a lower bandwidth requirement than voice signals, the same control channel can be used for carrying the control signals of multiple voice channels, hence the term *common channel*. In a multi-channel communications system, CCS is signalling in which one channel in each link is used for signalling to control, account for, and manage traffic on all channels of the link. Essentially, the CCS network is a highly robust sub-network that supports the operation of the primary communication network.

CCS is an out-of-band signalling technique that allows much faster communication between two nodes within the public switched telephone network (PSTN). Instead of being constrained to signalling data rates which are of the order of audio frequencies, CCS supports signalling data rates from 56 Kbps to many megabits per second. Thus, network-signalling data is carried in a seemingly parallel, out-of-band, signalling channel while only user data is carried on the PSTN. CCS provides a substantial increase in the number of users, which are served by trunked PSTN lines, but requires that a dedicated portion of the trunk time be used to provide a signalling channel used for network traffic. The most common CCS signalling methods in use today are ISDN and signalling system 7 (SS7).

Out-of-band signalling is signalling that does not take place over the same path as the conversation.

In 1G cellular systems, the SS7 family of protocols, as defined by the ISDN is used to provide CCS. Since network-signalling traffic is bursty and of short duration, the signalling channel may be operated in a connectionless fashion where packet data transfer techniques are efficiently

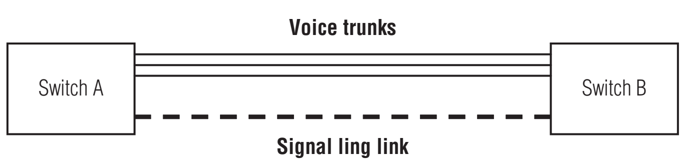


Figure 26.14 Common channel signaling

used. CCS generally uses a variable length packet sizes and a layered protocol structure. The expense of a parallel signalling channel is minor compared to the capacity improvement offered by CCS throughout the PSTN, and often the same physical network connection (i.e., a fibre optic cable) carries both the user traffic and the network signalling data.

For 2G wireless communications systems, CCS is used to pass user data and control/supervisory signals between the subscriber and the base station, between the base station and the MSC, and between the MSCs. Even though the concept of CCS implies dedicated, parallel channels, it is implemented in a TDM format for serial data transmissions.

> CCS is used in all modern telephone networks. Most recently, dedicated signalling channels have been used by cellular MSCs to provide global signalling interconnection, thereby enabling MSCs throughout the world to pass subscriber information.

26.12.1 Signalling system 7

In the mid-1970s the CCS network, separate from the voice network, was established; it utilizes the protocol called SS7. Within a communication network, the exchange of control information between network elements is performed by a signalling system. SS7 is a standard protocol approved by the International Telecommunication Union (ITU). Global billing, toll-free and 800 services, and international roaming for wireless calls are dependent on SS7. SS7 is the architecture for performing out-of-band signalling in support of the call establishment, billing, routing, and information exchange functions of the PSTN. It identifies the functions to be performed by a signalling system network and protocol to enable their performance. As a CCS standard, it is suitable for use with a wide range of circuit switched digital networks.

The SS7 protocol provides both error correction and retransmission capabilities to allow continued service in the event of signalling point or link failures. Compared to in-band signalling, out-of-band signalling provides faster call set-up times, more efficient use of voice circuits, and support for intelligent network (IN) services which require signalling to network elements without voice trunks (e.g., database systems), and improved control over fraudulent network usage.

> SS7 is used for trunk signalling in ISDN and widely used in today's public networks. SS7 is also used for SMS, prepaid, roaming, and other intelligent network functions.

There are three kinds of signalling points in the SS7 network:

- Service switching point (SSP)
- Signal transfer point (STP)
- Service control point (SCP)

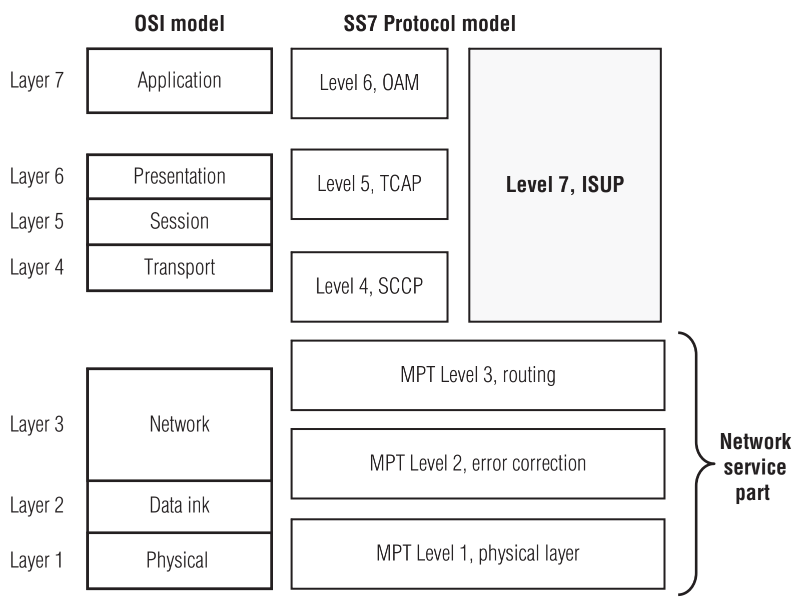
SSPs provide the SS7 functionality of a switch, they generally originate, terminate, or tandem calls. STPs may be either standalone or integrated STPs and are used to transfer signalling messages. They receive and route incoming signalling messages towards the proper destination. They also perform specialized routing functions. SCPs are databases that provide information necessary for advanced call processing capabilities. They allow service logic and additional routing information to be obtained to execute services.

> *SS7 is the global standard for telecommunications but it has the CCITT No.7. This standard covers procedures and protocols used by network elements in the PSTN to exchange data packets over a digital network.*

SS7 conforms to a layered model that parallels the OSI reference model. In SS7 terminology, these layers are called parts. The SS7 is responsible for the control of the fixed network as well as the mobile network. The inefficiencies of layered protocols are far outweighed by their flexibility in realization and management of complex functions. In a seven-layer network, a highly layered structure (transparent from layer to layer) is used to provide network communications. The seven layers are PHY, data link, network, transport, session, presentation, and application layers. Figure 26.15 illustrates the SS7 protocol model and the corresponding OSI layers.

The SS7 protocol model is an adaptation of the OSI model that is designed specifically for telephone networks. Layers 1 and 2 for OSI and SS7 are very much alike, but the upper layers have substantial differences. One difference is that SS7 "allows" level 7 integrated service digital network user part (ISUP) to communicate directly with layers 3, 4, 5, and 6.

The lowest three layers of the OSI model are handled in SS7 by the network service part (NSP) of the protocol, which in turn is made up of three message transfer parts (MTPs) and the signalling connection control pan (SCCP) of the SS7 protocol. SCCP provides an enhanced addressing capability that may be considered as a level close to level 4. Layers 4–6 in the OSI model do not exist in the SS7 protocol model. The transaction capabilities application part (TCAP) level and the operations maintenance and administration part (OMAP) level are considered the same as the application part (level 7) in OSI. The application service elements (ASEs) are at the same level as OMAP.



OMAP: operation maintenance administration part
ASE: application service element
TCAP: transaction capabilities application part
SCCP: signalling connection control part
MTP: message transfer part

Figure 26.15 OSI model and corresponding SS7 protocol architecture

Today, the network for supporting personal communication system (PCS) is narrowband ISDN, along with network intelligence based on advanced intelligent network (AIN) concepts, and a signalling system for mobility/location management and call control based on the SS7 network architecture.

Network services part of SS7

The SCCP in SS7 actually supports packet data network interconnections as well as connection-oriented networking to VC networks. The combination of MTP and SCCP corresponds to OSI layers 1, 2, and 3 and is known as NSP of SS7. This provides the reliable transfer and global routing and addressing of signalling messages through the network. The NSP provides ISDN nodes with a highly reliable and efficient means of exchanging signalling traffic using connectionless services. The NSP allows network nodes to communicate throughout the world without concern for the application or context of the signalling traffic.

Message transfer part of SS7

The function of the MTP is to ensure that signalling traffic can be transferred and delivered reliably between the end-users and the network. It is divided into three parts denoted as MTP1, MTP2, and MTP3. The SS7 physical layer is called MTP level 1. It defines the physical, electrical, and functional characteristics of a signalling digital link; physical channels may include copper wire, twisted pair, fibre, mobile radio, or satellite links, and are transparent to the higher layers. CCITT recommends that MTP level 1 uses 64 Kbps transmissions, whereas ANSI recommends 56 Kbps.

The data link layer is called MTP level 2. It provides means for a correct point-to-point transmission of a message through the signalling link. It implements flow control, message sequence validation, and error check. A single message signal unit (MSU) cannot have a packet length that exceeds 272 octets, and a standard 16 bit cyclic redundancy check (CRC) sum is included in each MSU for error detection.

A wide range of error detection and correction features are provided in MTP level 2. MTP level 2 also provides flow control data between two signalling points as a means of sensing link failure. If the receiving device does not respond to data transmissions, MTP level 2 uses a timer to detect link failure and notifies the higher levels of the SS7 protocol which take appropriate actions to reconnect the link. The network layer is called MTP level. Signalling network functions (MTP level 3) provide procedures that transfer messages between signalling nodes.

As in ISDN, there are two types of MTP level 3 functions: signalling message handling and signalling network management. Signalling message handling is used to provide routing, distribution, and traffic discrimination. Signalling network management allows the network to reconfigure in case of node failures, and has provisions to allocate alternate routing facilities in the case of congestion or blockage in parts of the network.

Discrimination is the process by which a signalling point determines whether a packet data message is intended for its use or not.

Signalling connection control part of SS7

This protocol provides functions for the transfer of messages that are not trunk related. It users are ISUP and TCAP. SCCP supports connection or connection-less oriented network services. One of the SCCP's most important features is an enhanced version of SS7 routing called global title translation (GTT) service. A global title is a sequence of digits, an 800 number, a calling card number, and so forth, that SCCP translates into a destination point code and subsystem number. SCCP provides enhancement to the addressing capabilities provided by the MTP. While the addressing capabilities of MTP are limited in nature, SCCP uses local addressing based on sub-system numbers (SSNs) to identify users at a signalling node. SCCP consists of four functional blocks. The SCCP connection oriented control block provides data transfer on signalling connections. The SCCP management block provides functions to handle congestion and failure conditions that cannot be handled at the MTP. The SCCP routing block routes and forwards messages received from MTP or other functional blocks.

SS7 user part

The SS7 user part provides call control and management functions and call set-up capabilities to the network. These are the higher layers in the SS7 reference model, and utilize the transport facilities provided by the MTP and the SCCP. The SS7 user part includes the ISUP, the TCAP, and the OMAP. The telephone user part (TUP) and data user part (DUP) are included in the ISUP.

Integrated services digital network user part

The ISUP supports both ISDN and non-ISDN calls in a digital network. ISUP is a protocol for call control and trunk maintenance procedure in both the telephone network and the ISDN. The ISUP transition provides the signalling functions for carrier and supplementary services for voice, data, and video in the ISDN environment. In the past, telephone requirements were lumped in the TUP, but this is now a subset of ISUP. ISUP uses the MTP for transfer of messages between different exchanges. ISUP message includes a routing label that indicates the source and destination of the message, a circuit identification code (CIC), and a message code that serves to define the format and function of each message. In addition to the basic bearer services in the ISDN environment, the facilities of user to user signalling, closed user groups, calling line identification, and call forwarding are provided.

Transaction capabilities application part

The TCAP provides a set of protocols and services used by an application process. The TCAP in SS7 refers to the application layer which invokes the services of the SCCP and the MTP in a hierarchical format. One application at a node is thus able to execute an application at another node and use these results. Thus, TCAP is concerned with remote operations.

Operation maintenance and administration part

The OMAP specifies network management functions and message related to operations and maintenance. The OMAP functions include monitoring, coordination, and control functions to ensure that trouble-free communications are possible. OMAP supports diagnostics and these are known throughout the global network to determine loading and specific sub-network behaviours.

Signalling traffic in SS7

Signalling traffic management is responsible for the management of the signalling link in the case of a failure. This means that it handles functions of signalling traffic routing in the case of failure. Call set-ups, inter-MSC handoffs, and location updates are the main activities that generate the maximum signalling traffic in a network, and these are all handled under SS7. Setting up of a call requires exchange of information about the location of the calling subscriber and information about the location of the called subscriber. Either or both of the calling and the called subscribers can be mobile, and whenever any of the mobile subscribers switches MSCs under a handoff condition, it adds to the amount of information exchanged. Table 26.5 shows the amount of signalling traffic that is generated for call set-up in GSM.

Table 26.5 Signaling load for call setup and handoffs in GSM

	Load in bytes
Call Originating from a mobile	
Information on the originating MSC and the terminating switch	120
Information on the originating MSC and the Associated VLR	550
Call terminating at a mobile	
Information on the switch and terminating MSC	120
Information on the terminating MSC and associated VLR	612
Information on the originating switch and HLR	126
Inter MSC handoffs	
Information on the new MSC and associated VLR	148
Information on the new MSC and old MSC	383

Table 26.6 Signaling load for location updating in GSM

Location Updating	Load in bytes
Information on the current MSC and Associated VLR	406
Information on the current VLR and HLR	55
Information on the new VLR and old VLR	406
Information on the new VLR and old VLR	213
Information on the old VLR and HLR	95
Information on the new VLR and HLR	182

Location update records are updated in the network whenever a subscriber moves to a new location. The traffic required by the location update process as a subscriber moves within and between visitor location register (VLR) areas is shown in Table 26.6. The MSC together with home location register (HLR) and VLR database provides the call routing and roaming capabilities of GSM.

Home location register (HLR) – used in cellular networks to store information such as current cellular phone location, billing, and cellular subscriber information.

Visitor location register (VLR) – used in cellular networks to store information on subscribers roaming outside the network. The VLR uses this information to communicate to the HLR database to identify the subscriber's location when roaming.

SS7 services

The implementation of SS7 involves a new set of protocols. The results for the end user will have a larger set of new potential services and network capabilities. The SS7 network is a packet switched network that is used solely for the purpose of connecting telephone calls. It provides two types of services: circuit related and non-circuit related services. There are three main types of services offered by the SS7 network: the Touchstar, 800 services, and alternate billing services. These services are briefly explained below.

Touchstar – Touchstar was established in 1986. It is also known as custom local area signalling services (CLASS) and is a group of switch controlled services that provide its users with certain call management capabilities. Touchstar is a collection of many advanced custom calling type service. Touchstar services include call return, call forwarding, repeat dialling, call block, call tracing, and caller ID.

800 services – 800 services were unusual. By definition, it was a toll-free long distance service and thus subject to deregulation and competition under the Bell system. The cost associated with the processing of calls is paid by the service subscriber. The service is offered under two plans known as the 800-NXX plan, and the 800-database plan. In the 800-NXX plan, the first six digits of an 800 call are used to select the interexchange carrier (IXC). In the 800 Database plan, the call is looked up in a database to determine the appropriate carrier and routing information.

Alternate billing service and line information database (ADB/LIDB) – ADB/LIDB is a transaction oriented database that contains line number and special billing number data that is needed by the operator. These services use the CCS network to enable the calling party to bill a call to a personal number (third party number, calling card, or collect etc.) from any number. LIDB provides subscriber or user information such as calling-card services including card validation and personal identification number (PIN) authentication and billing.

Performance of SS7

SS7 messages transported over Intelligent Peripheral networks must meet the rigid performance requirements imposed by both ITU SS7 standards and the user expectations. The performance of the signalling network is studied by connection set-up time (response time) or the end-to-end signalling information transfer time. The ITU standard specifies that the end-to-end call set-up delay cannot exceed 20–30 s after the ISUP initial address message is transmitted. The delays in the signalling point (SP) and the STP depend on the specific hardware configuration and switching software implementation.

Congestion control in SS7 networks – The reliable transmission is accomplished by the data link protocol and the congestion control mechanisms, which are provided by traffic, link, and route management of network protocol layer. With an increasing number of subscribers, it becomes important to avoid congestion in the signalling network under heavy traffic conditions. SS7 networking protocols provide several congestion control schemes, allowing traffic to avoid failed links and nodes.

Advantages of CCS over conventional signalling – CCS has several advantages over conventional signalling which have been outlined below:

- In CCS, high-speed signalling networks are used for transferring the call set-up messages resulting in smaller delay times when compared to conventional signalling methods, such as multi-frequency.
- CCS has shorter call set-up and tears down times that result in less call holding time, subsequently reducing the traffic on the network. In heavy traffic conditions, high trunking efficiency is obtained.
- CCS allows the transfer of additional information along with the signalling traffic providing facilities such as caller identification and voice or data identification.

26.13 Various networks for connecting to the Internet

There are a number of ways of connecting to the Internet as mentioned in Section 26.6.2. A few of these methods are described in the following sections.

26.13.1 Dial-up connection

The original telephone system is based on analogue technology. The phone calls were *frequency shifted* and *multiplexed* onto a single wire; that is, analogue devices would raise the pitch of a call and pack it onto a wire with other calls. At the receiving end, the calls would be separated by filtering and then readjusting the frequencies. As human beings tone of voice lies in the frequency range of 300 Hz to 3.3 kHz, the bandwidth required for voice communication over a landline telephone system is 4 kHz. Since voice is "analogue" in form, it changes smoothly and so this technology was perfect for carrying telephone conversations. However, the trick was how to transmit DIGITAL information over an ANALOGUE telephone line.

Modem: A device called "modem" was developed to allow digital data to be carried over the analogue telephone system. Modems are defined by their speed, that is, how many bits per second they can handle. A modern dial-up connection makes use of a 56 Kbps modem. A dial-up connection allows you to connect to the Internet via a local server using a standard 56 Kbps modem, your PC literally dials (hence the name dial-up) a phone number (provided by your Internet service provider (ISP)), and connects to the server and, therefore, the Internet.

Disadvantages of dial-up

- 56 Kbps is very slow by modern standards.
- Dial-up ties up your telephone line and so it cannot be used for anything else.
- It takes a while for the computer to set up a connection, that is many seconds.
- Internet service providers (ISPs) may charge you by the minute and so a slow link means you pay more.

26.13.2 ISDN

As telephone companies started moving to digital technologies, it was felt that there was a need to take services like ***voice, fax,*** and ***data*** and ***integrate*** them into a single unified service. ISDN allows multiplexing of devices over single ISDN line (Figs. 26.16 and 26.17).

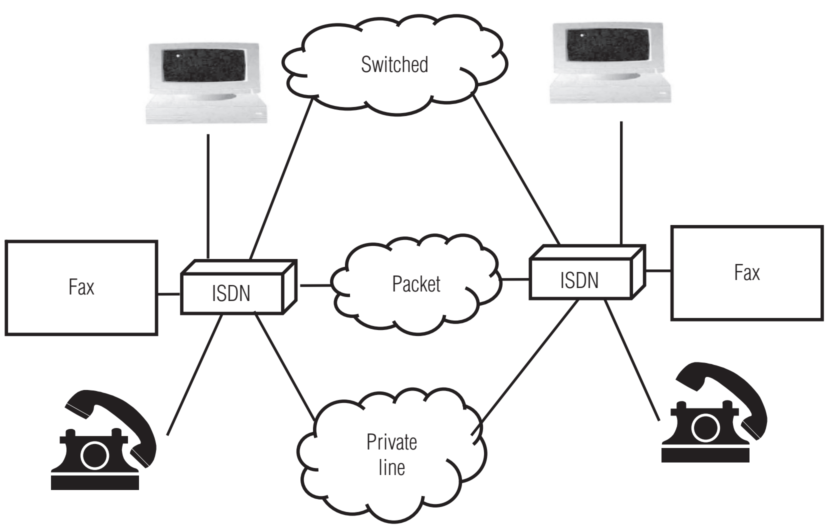


Figure 26.16 Multiplexing of devices over single line

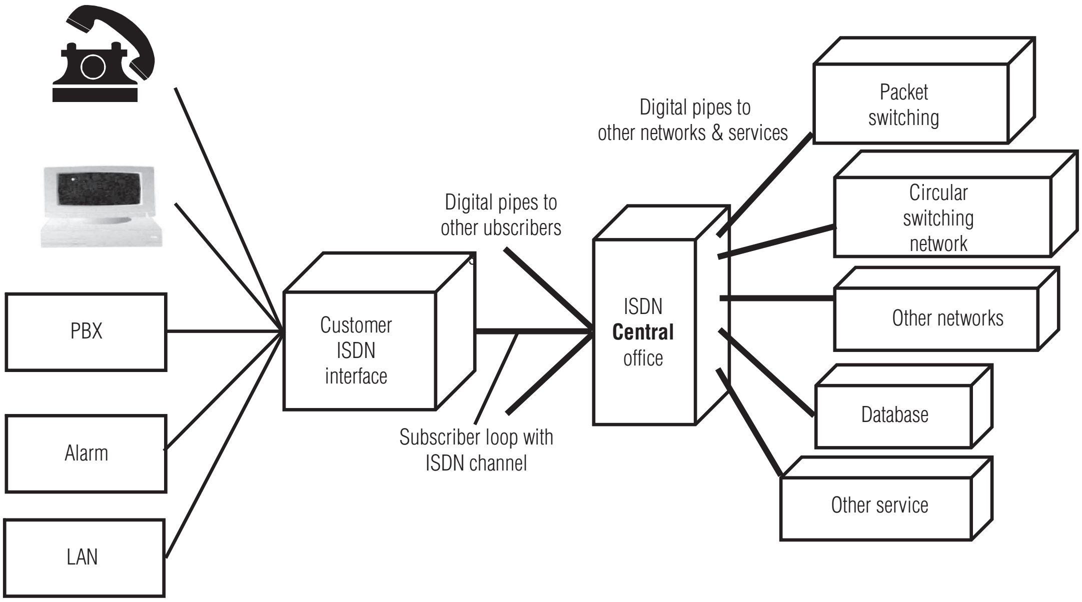


Figure 26.17 Services provided by ISDN

ISDN is one such unified service that can carry the voice, fax, and data simultaneously on a single line in a digital manner.

The ISDN was introduced in the 1970s, made official in 1984, and later refined in 1988 by the CCITT. Its connection speed is higher than that of analogue modem and is often used as a low cost alternative to T1 connections. ISDN carries all information in an end-to-end digital network and no analogue transmission services are used. As with other digital protocols, there must be a set of standards to allow digital networks to interconnect to analogue networks.

ISDN uses *TDM technique to provide a service (voice, data, video, etc.) that can be simultaneously provided over the same facility*. The signal to be transmitted is sampled at 8 kHz, that is, a sample every 125 μs. Each sample is 8 bits in length, so the data rate is 64 Kbps. This is the standard data rate with ISDN. The speed of ISDN is in between the speeds of narrow band modems and broadband ADSL. ISDN uses the full bandwidth in the voice channel to get a full 64 Kbps data rate. It can get this rate as it uses a fully digital path from home to ISP and no acoustic modems are involved (Figure 26.18). In the home or office, there is an encoder that produces a digital

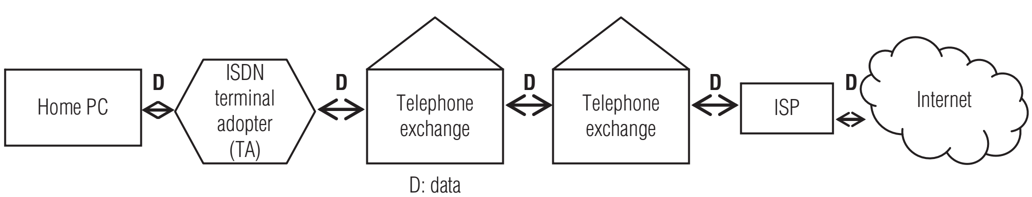


Figure 26.18 ISDN

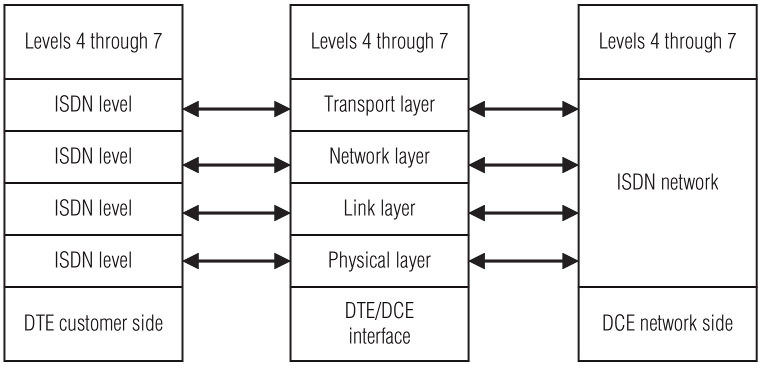


Figure 26.19 ISDN and OSI reference model

signal on the wire rather than an acoustic signal. This encoder is called an "ISDN modem" or "digital modem" or "terminal adapter." An ISDN requires extra hardware both at home and in the local exchange, telephone companies charge special rates for the use of ISDN.

ISDN and OSI reference model

ISDN incorporates the physical, data link, network, and transport layers of the OSI model. Similar to X.25, it uses LAPB and the data link layer to ensure the maximum detection of communication errors. Figure 26.19 shows ISDN in relation to the OSI-telecommunications reference model. ISDN is designed to be compatible with many existing digital networks, such as ATM, X.25, and T1 (T1 has a data rate of 1.54 Mbps). ISDN is divided into 64 Kbps channels.

ISDN data rate interfaces

ISDN provides two main types of services (Figure 26.20), namely

- Basic rate interface (BRI)
- Primary rate interface (PRI)

Basic rate interface

BRI connections offer three channels: two at 64 Kbps and one at 16 Kbps for a maximum throughput of 128 Kbps. The 64 Kbps channels are known as bearer or B-channels because they carry the data for the connection. ISDN BRI connections use the 16-Kbps signalling channel, which is also called the D-channel, to control the communications on the link.

Primary rate interface

- PRI connections offer 23 B-channels and one 64 Kbps D-channel for a bit rate of up to 1.544 Mbps. This service is known as T1-PRI and belongs to USA/JAPAN.
- European ISDN PRI (E1-PRI) service offers thirty 64 Kbps B-channels and one 64 Kbps D-channel yielding a total interface rate of 2.048 Mbps.
- In both ISDN BRI and PRI, a single D-channel is used for signalling information, and the B-channels are used to carry the data. *Because the control communications are conducted on a channel that is separate from the data transfer, ISDN is said to be out-of- band signalling.*

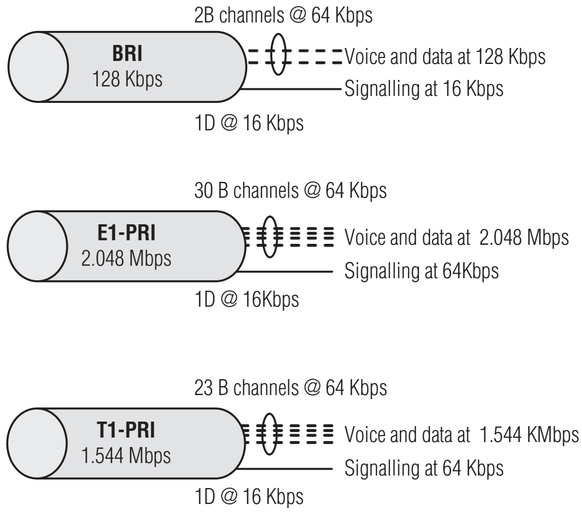


Figure 26.20 BRI and PRI channels

Benefits of ISDN

The following are the benefits of ISDN:

- Provides voice, data, and video services over one network.
- Also provides more bandwidth than a traditional 56 Kbps dial-up connection.
- Uses **bearer channels**, also called **B-channels**, as clear data paths.
  Each B-channel provides 64 Kbps of bandwidth. An ISDN connection with two B channels would provide a total usable bandwidth of 128 Kbps.
- Each ISDN B-channel can make a separate serial connection to any other site in the ISDN network.
- Layered protocol structure compatible with OSI.
- Network management services are offered via intelligent nodes.
- Provides switched and non-switched connection services.
- May provide video-conferencing through high-bandwidth capabilities.

26.13.3 Broadband ISDN

Increasing market demand for data rates substantially greater than those supported by ISDN has lead to the notion of Broadband ISDN (B-ISDN). With the proliferation of computer systems and video imaging, end-user applications are requiring much greater bandwidths than the standard 64 Kbps B-channel provided by ISDN. Recent work has defined ISDN interface standards that increase the end-user transmission bandwidth to several Mbps. It is based on ATM technology that allows packet switching rates up to 2.4 Gbps and total switching capacities as high as 100 Gbps.

B-ISDN is developed as an evolution of ISDN and hence follows the same principles. B-ISDN services are classified into *interactive and distribution* services. Interactive services involve the bidirectional flow of user information between two subscribers or between a subscriber and a service provider. Distribution services involve the unidirectional flow of user information from a service provider to a subscriber. 3G systems will use the B-ISDN to provide access to information networks, such as the Internet and other public and private databases.

26.13.4 Digital subscriber line

DSL is the generic name for a new technology that allows digital data to be sent over an ordinary copper telephone line at high speed. Although the transmitted information is in digital form, the transmission medium is usually an analogue carrier signal (or the combination of many analogue carrier signals) that is modulated by the digital information signal. It is much faster than ISDN. ISDN provides two voice/data channels, each at speeds of 64 Kbps, while DSL is predominantly a data service providing up to 1.5 Mbps downstream and up to 512 Kbps upstream. DSL offers two types of services:

- Asymmetric Digital Subscriber Line (ADSL)
- Symmetric Digital Subscriber Line (SDSL)

26.13.5 ADSL

Providing a high-bandwidth Internet connection to the user is expensive and needs special equipment at the exchanges and good quality telephone line. Telephone companies realised that most domestic customers would be downloading far more than they would be uploading. A new DSL technology has been developed to provide fast download speeds and slower upload speeds. This service is called ADSL. ADSL is also referred to as “Broadband.”

- ADSL connections are becoming more and more popular and can provide an excellent Internet connection. The connections work by splitting phone line into two separate channels, one for data (Internet) and one for voice (phone calls) which means user can talk on the phone and be connected to the Internet at the same time. ADSL connection offers three types of services based on different speed specifications (256 Kbps/128 Kbps, 512 Kbps/128 Kbps, 1 Mbps/256 Kbps).

Advantages of ADSL

- Faster downloads compared to dial-up or ISDN.
- No need for a second phone line – by allowing voice and data transfer at the same time (you can use the phone as normal while connected to the Internet).
- Because ADSL transfers data digitally it does not need to convert the data from digital to analogue and vice versa.
- ADSL connections are always on, which makes the usual long wait to connect a thing of the past.

Disadvantages of ADSL

- ***Distance sensitive***: ADSL connections are not available to everyone. You need to be within 3 miles of an ADSL enabled exchange.
- The hardware costs can be quite significant as you will need a special ADSL modem and ADSL filters to use the service. Most ISPs allow you to hire these items which can reduce the initial cost.
- Because ADSL connections are always on, you will need a firewall to protect your PC.

26.14 Summary

- ***Circuit switching***: A networking technology that provides a temporary, but dedicated, connection between two stations irrespective of the number of switching devices that the data are routed through.

- ***Packet switching***: A digital network technology that breaks up a message into small packets for transmission. Unlike circuit switching, which requires the establishment of a dedicated point-to-point connection, each packet in a packet-switched network contains a destination address. Thus, all packets in a single message do not have to travel the same path. The destination computer reassembles the packets into their proper sequence. Network protocols such as IP and IPX were designed for packet-based networks.
- The IP breaks data into small units or blocks of data called packets. A packet can also be called a segment or datagram.
- The first international standard for wide area packet switching networks was X.25, which was defined for analogue circuits and very susceptible to noise. Subsequent technologies, such as frame relay, were designed for today's almost error-free digital lines.
- WAN connections typically function at the physical and data link layers of the OSI reference model, and are made over serial connections.
- A WAN service that uses the connection-oriented model is frame relay. The service provider sets up PVCs through the network as required or requested by the customer. ATM is another networking technology that uses the connection-oriented VC approach.
- ATM uses a cell-switching technology that provides the bandwidth-sharing efficiency of packet switching with the guaranteed bandwidth of circuit switching.
- DSL is a family of technologies that provide digital data transmission over the wires of a local telephone network. DSL service is delivered simultaneously with regular telephone on the same telephone line. This is possible because DSL uses a higher frequency.
- ISDN systems are physically "in-band" and logically "out-of-band." This means that the same physical wires are used to multiplex both the voice traffic and the data traffic required to administer the system. "Out-of-band" means that various increments of time are dedicated for signalling and are not available for voice traffic.
- In contrast with ISDN, SS7 is both physically "out-of-band" and logically "out-of-band."
- An ISDN is an end-to-end digital network capable of simultaneous transmission of a range of services such as voice, data, video, and so on.

Review questions

1. What is the fundamental difference between wireless and fixed telephone networks? Give an example for the above two types.
2. Describe the circuit switching and packet switching methods of network communication and list the advantages and disadvantages of packet-switching networks.
3. What are the various traffic routing methods in wireless networks?
4. Describe the salient features of X.25 protocol.
5. What is frame relay and why it is needed?
6. Describe the PLP of X.25 network.
7. Explain the salient features of ATM technologies.
8. Differentiate between ISDN and SS7.

Objective type questions and answers

1. IEEE802.11 the wireless LAN uses
   (a) CSMA/CA (b) CSMA/CD (c) (a) and (b) (d) none of the above

2. Circuit switching networks are based on
   (a) FDM (b) TDM (c) SDM (d) none of the above
3. Protocols used by packet switching are
   (a) X.25 (b) frame relay (c) ATM (d) all the above
4. Data transmission speeds used in ATM are
   (a) OC-3 (b) OC-12 (c) OCM-109 (d) all the above

Answers: 1. (a), 2. (b), 3. (d), 4. (d)

True/False

1. In wireless networks, signals travel in the form of EM waves.
2. Wireless networks have better transmission speed.
3. In wired networks, capacity can be increased by optical fibres.
4. Bandwidth available in packet switching is dynamic.
5. Frame relay is designed as a 2G of X.25.

Answers: 1. (T), 2. (F), 3. (T), 4. (T), 5. (T).

Open book questions

1. Describe the ATM network technology.
2. What is the difference between wired and wireless networks? Mention the reasons for the development of wireless technology.
3. What are the basic advantages of ISDN over the dial-up network?
4. What are the significant differences between X.25 and frame relay.

Further reading

Haartsen, J. C., and Mattisson, S., “Bluetooth — A New Low-Power Radio Interface Providing Short Range Connectivity,” *Proceedings of IEEE*, 88(10), pp. 1651–1661, October 2000.

IEEE, *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, P. 802.11D6.2, July 1998.

ITU-T Recommendation I.363.1, ‘B-ISDN ATM Adaptation: Type 1 AAL’, August 1996.

ITU-T Recommendation I.363.3, ‘B-ISDN ATM Adaptation Layer Specification: Type 3/4 AAL’, August 1996.

ITU-T Recommendation I.363.2, ‘B-ISDN ATM Adaptation Layer Specification: Type 2 AAL’, September 1997.

Jain, R., *Wireless Local Area Network Recent Developments*. *Wireless Seminar Series*. Electrical and Computer Engineering Department, Ohio State University, 19 February 1998.

Jonsson, M., *HiperLan2—The Broadband Radio Transmission Technology Operating in the 5 GHz Frequency Band*. HIPERLAN-2 Global Forum White Paper.

JTC Technical Report on RF Channel Characterization and Deployment Model. Air Interface Standards, September 1994.

Muller, N. J., *Bluetooth Demystified*. New York: McGraw Hill, 2000.

Pahlavan, K., "Trends in Local Wireless Networks." *IEEE Communications Magazine*, Vol. 33, No. 3, pp. 88–95, March 1995.

Bluetooth 27

27.1 Introduction

Information transfer between devices has been cumbersome, mainly relying on cables. A new universal radio interface known as *Bluetooth* has been developed that enables wireless communication between electronic devices and communicates via short-range ad hoc radio connections.

The Bluetooth standard is based on a tiny microchip incorporating a radio transceiver that is built into digital devices. The transceiver takes the place of a connecting cable for devices such as cell phones, laptop and palmtop computers, portable printers and projectors, and network access points.

Bluetooth is a short-range (<10 m) and low bit rate (<1 Mbps) radio technology that connects portable devices such as cell phones, handheld devices, and notebook computers. The salient features of this technology are its low-cost, low-power attributes, and that it is based on ISM (industrial, scientific, and medical) radio bands.

> Bluetooth is also known as *wireless personal area network (WPAN) for* short-range and mobility applications around a room in the office or at home.

This chapter gives an overview of Bluetooth and its architecture and deals with radio specification, base band specification, and the protocols of Bluetooth technology in detail.

27.2 Bluetooth

Bluetooth technology is a short-range communications technology that is simple, secure, and can be used everywhere. Bluetooth technology was intended to hasten the convergence of voice and data in handheld devices, such as cellular telephones and portable computers. Research on the use of radio to link mobile phones and accessories was started by Ericsson Mobile Communications in 1994. It was not until the Bluetooth special interest group (SIG) was launched four years later, by Ericsson, IBM, Intel, Nokia, and Toshiba, that the concept started to broaden beyond mobile phones to include connections between PCs and other devices.

> Bluetooth is a wireless standard that enables devices to transmit data at up to 1 Mbps over a maximum distance of 10 m.

A complete Bluetooth system will require the following elements:

- A radio frequency (RF) portion for receiving and transmitting data
- A module with a baseband microprocessor
- Memory
- An interface to the host device (such as a mobile phone)

The following are the advantages of Bluetooth technology:

1. Wireless (no cables)
2. No set-up needed
3. Devices can be mobile
4. Industry-wide support
5. Easier to synchronize owing to it being omnidirectional and no line-of-sight (LOS) requirement

The following are the disadvantages of Bluetooth technology:

1. Short-range wireless radio technology – operating over a range of 10 m
2. Small throughput rates – data rate up to 1.0 Mbps
3. Mostly for personal use
4. Fairly expensive

27.2.1 Overview

Bluetooth devices operate in the 2.4 GHz ISM band (frequency range 2,400–2,483.5 MHz). This band consists of 79 channels each of 1 MHz bandwidth, with a lower guard band of 2 MHz and upper guard band of 3.5 MHz. Bluetooth uses a "fast frequency hopping (FH)" radio technique, changing its operating frequency 1,600 times a second. This technique enables it to carry on working even in areas of high interference, an important point, considering that it has to share its radio spectrum with many other devices, including microwave ovens and wireless local area networks (WLANs).

> A Bluetooth-enabled device communicates with another Bluetooth-enabled device over the radio medium to exchange information or to transfer data from one to the other.

The Bluetooth standard provides one asynchronous data channel at 723.2 Kbps. In this mode, also known as asynchronous connectionless (ACL), there is a reverse channel with a data rate of 57.6 Kbps. The specification also allows up to three synchronous channels each at a rate of 64 Kbps. This is also known as synchronous connection oriented (SCO) mode. It is mainly used for voice applications such as headsets, but can also be used for data transfer. These different modes result in an aggregate bit rate of approximately 1 Mbps. Routing of the asynchronous data is done via a packet switching protocol based on FH. There is also a circuit switching protocol for the synchronous data.

The following are some of the capabilities of Bluetooth:

- Make calls from a wireless headset connected remotely to a cell phone
- Eliminate cable linking computers to printers, keyboards, and the mouse
- Connect MP3 players to other machines to download music in a wireless fashion
- Call home from a remote location to turn applications on and off, set the alarm, and monitor activity

> The main aim of Bluetooth technology is to guarantee interoperability between different applications on devices in the same area that may run over different protocol stacks, and therefore to provide a solution for WPAN.

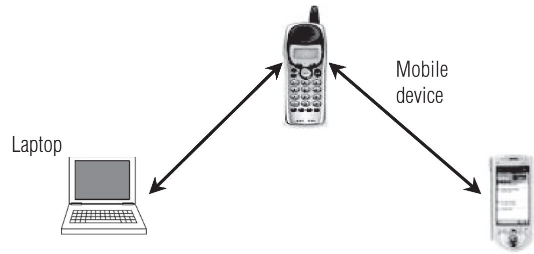


Figure 27.1 Bluetooth communication between portable devices

A Bluetooth WPAN involves up to eight devices, located within a 10 m radius personal operating space, that unite to exchange information or to share services (Figure 27.1). Because connectivity can be spontaneous according to immediate need, Bluetooth is also known as *ad-hoc networking*. Because a WPAN involves directly networking between different points, without the use of network infrastructure, it is also referred to as a "point-to-point network."

Laptop to mobile phone connectivity: A mobile phone can communicate with a computer using a phone's vendor specific cable, infrared (IR), or Bluetooth. The latter two depend on hardware capabilities of both the devices. Figure 27.2 presents a typical scheme for a laptop connecting to a server through the mobile phone and the cellular network. For laptop to mobile phone connectivity, the Bluetooth solution was selected for the following reasons:

- Compared to IR, it does not have the restriction of having the phone being in the LOS of the laptop's sensor.
- Compared to a cable, it does not require the phone to be physically attached to the laptop and restricted to the cable's length.
- With Bluetooth, the phone only needs to be in a range of a few metres (less than 10 m) from the laptop.

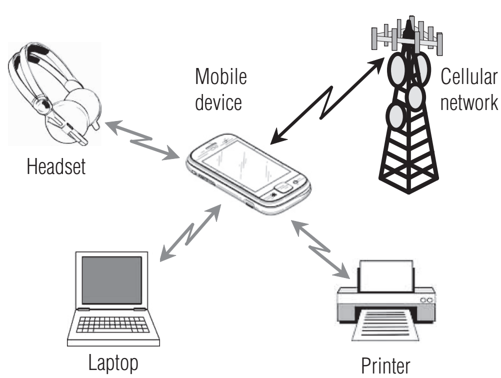


Figure 27.2 Typical schemes for a laptop connecting to a server through the mobile phone and the cellular network

27.2.2 Bluetooth architecture

The Bluetooth protocol architecture is shown in Figure 27.3. The figure provides an overview of the protocols that are used and supported in Bluetooth and the way they tie in together. The protocols are hierarchically shown with the "highest" protocols on top. In data communication contexts, this hierarchy means the following two things:

1. The higher protocols depend on the lower ones to exist. But the lower protocols can do without the higher, or they can support other protocols on the higher levels.
2. The higher protocols are usually "closer" to the user, insofar as they provide human-oriented services.

These protocols can be divided into the following four categories:

1. Core protocols
2. Cable replacement protocol
3. Telephony control protocol
4. Adopted protocols.

These four categories are detailed below.

1. Bluetooth core protocols

Baseband

The physical RF link between Bluetooth units forming a piconet can be enabled by using the baseband and link control layers. Usage of these layers, formed between the Bluetooth units, is

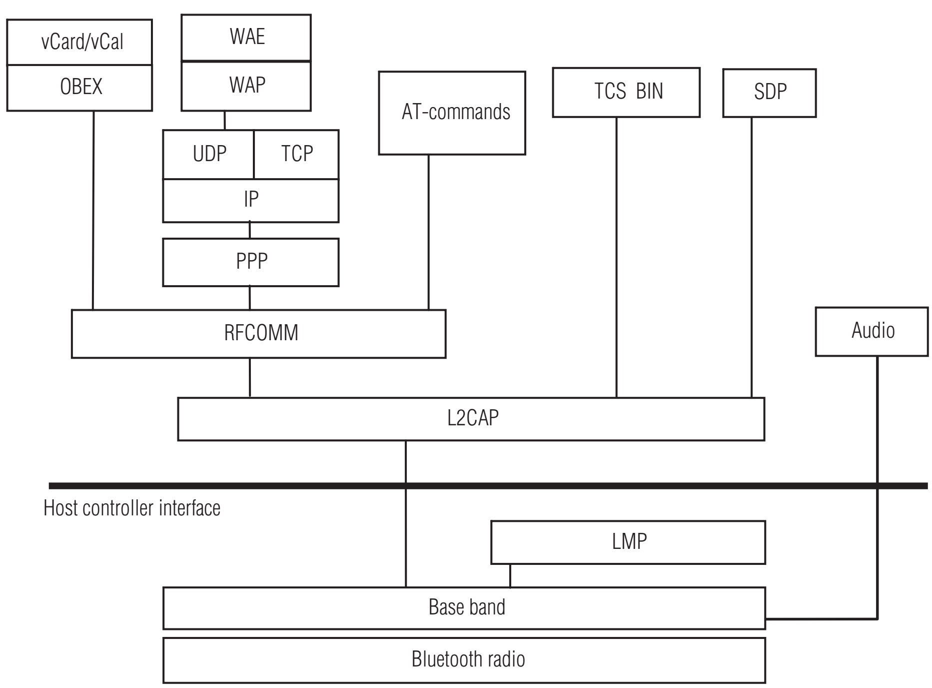


Figure 27.3 Bluetooth protocol architecture

synchronization and transmission of FH sequence. In addition, SCO and ACL are also managed. The above-mentioned layers are discussed in Section 27.2.4. The RF link can be multiplexed with ACL and SCO links for data and audio, respectively.

Audio

By using different usage models, audio transmission can be done between one or more Bluetooth units. Audio data are a straightforward set-up between Bluetooth units for direct transmission whenever a Bluetooth link is formed. It does not pass through logical link control and adaptation protocol (L2CAP) layer.

Host controller interface (HCI)

HCI manages a uniform interface method for accessing the Bluetooth hardware capabilities that contain a command interface to the baseband controller and link manager (LM), and access to hardware status, control, and event registers.

Link manager protocol (LMP)

LMP is reliable for link set-up between Bluetooth units. It manages the control and negotiation of packet sizes while in data transmission. The LMP is capable of managing power modems, power consumption, and state of a Bluetooth unit in a piconet. Finally, this layer utilizes generation, exchange and link control, and encryption keys for the purpose of encryption and authentication.

L2CAP

In the Bluetooth protocol, stack L2CAP is positioned over the baseband layer and beside the LMP. The L2CAP layer manages connection-oriented and connectionless data service to upper layers. The four main functions are as follows:

- L2CAP handles multiplexing of multiple packet resources. This layer determines the upper layer protocol to handle when a packet has been reassembled (e.g., service discovery protocol [SDP], radio frequency communication [RFCOMM], and TCS [telephony control protocol specification] binary).
- When the data packets exceed the maximum transmission, Bluetooth unit must be broken down before transmission.
- L2CAP handles the quality of service (QoS) requirement when both links are established during the normal operation.
- The L2CAP specification handles a group of feature that allows for mapping of groups on to a piconet

L2CAP implementation should not be ambitious and represents a low overhead since it should be compatible with less computational resources in piconet.

SDP

SDP defines how a Bluetooth client's application shall act to discover available Bluetooth server's services and their Bluetooth characteristics. This protocol briefs how a client can pursue for service based on specific attributes without knowing anything of the available services. Whenever the client enters an area where a Bluetooth server is being operated, the SDP discovers the new available services. There is another function for detecting when a service is no longer available.

2. Cable replacement protocol

This is the RFCOMM protocol, whose purpose is to emulate a serial port. The protocol covers applications that use serial ports of the kind used in PCs. Thus, RFCOMM emulates RS-232 control and data signals over the Bluetooth baseband. It provides transport capabilities for upper level services, such as OBEX.

3. Telephony control protocol

Telephony control – binary

It is a bit-oriented protocol that specifies call controlling signal for the development of audio and data calls between Bluetooth units, as well as signalling to ease. TCS binary provides functionality to interchange signalling data dissimilar to ongoing calls. Formation of voice or data call in a point-to-point configuration likewise in a point-to-multipoint configuration is covered. This protocol is based on ITU-T standard Q.931.

Telephony Control – AT Commands

A number of AT-commands are mapped for transmitting control signal for telephony control. For transmission they use RFCOMM and serial port emulation.

4. Adopted protocols

This section explains the protocols that are to be adopted to the Bluetooth protocol stack.

PPP

The IETF point-to-point protocol (PPP) in the Bluetooth technology is accomplished to run over RFCOMM to design point-to-point connections. It is a packet-oriented protocol that uses serial mechanism to convert the information packet stream to serial data stream.

TCP/UDP/IP

The TCP/UDP/IP standards provide the operation of Bluetooth units to communicate with the others connected to the Internet, which can act as a link to the Internet. The TCP/IP/PPP protocol arrangement is used for all the Internet Bridge strategies in Bluetooth 1.0 and OBEX in future versions. UDP/IP/PPP configuration is applicable for wireless application protocol (WAP) as transport.

OBEX protocol

IrDA object exchange protocol (IrOBEX) is an optional application layer protocol that supports IR communication to exchange a wide variety of data. OBEX uses a client–server model, which is independent of transport mechanism. It also defines a folder-listing object, which is used to browse the contents of folders. It uses RFCOMM as the main transport layer.

Content formats

vCard and vCalendar are transferred by OBEX. Some formats for transmitting vCard and vCalendar are mentioned in the Bluetooth specification, though they do not define the transport mechanisms.

WAP

WAP is a wireless protocol specification that works across a wide area wireless network technologies. Bluetooth, which can be used to provide a bearer for transporting data WAP client and server, can be used with regard to WAP.

The temporary networking capability of Bluetooth gives unique possibilities for a WAP client. The traditional forms of WAP communications involve server/proxy device using WAP protocols. If the WAP technology that supports server push is used over Bluetooth, it provides new possibilities for distributing information.

27.2.3 Radio layer and radio specification

At the physical layer (PHY), Bluetooth uses the IEEE 802.15.1 radio, which specifies an FH spread spectrum (FHSS) system and a hopping pattern controlled by the 48-bit media access control

(MAC) address of the master device. In some countries, the hopping pattern is reduced to cover just 23 channels to comply with specific local regulations. Its gross rate is up to 1 Mbps.

The IEEE 802.11 standard calls for three different PHY specifications: FHSS, direct sequence spread spectrum (DSSS), and IR. The transmit power for DS and FH devices is defined at a maximum of 1 W and the receiver sensitivity is set to –80 dBmW. Antenna gain is limited to 6 dB maximum. In this chapter, we focus on the 802.11b specification (DSSS) since it is in the same frequency band as Bluetooth and is the most commonly deployed one. All the three support 1–2 Mbps data rate. Both DSSS and FHSS use the 2.4 GHz ISM band (2.4–2.4835 GHz). The PHY provides three levels of functionality. These include the following:

- Frame exchange between the MAC and PHY under the control of the physical layer convergence procedure (PLCP) sub layer;
- Use of signal carrier and spread-spectrum (SS) modulation to transmit data frames over the media under the control of the physical medium dependent (PMD) sub layer; and
- Providing a carrier sense indication back to the MAC to verify activity on the media.

Bluetooth uses a technique called SSFH.

Radio specification

Bluetooth radio characteristics include low-power, short-range, and medium transmission speed. The low power consumption makes Bluetooth ideal for small, battery-powered devices like mobile phones and pocket PCs. Bluetooth is poised to capitalize on the emerging market of small mobile devices that is expected to grow.

Bluetooth's short range (< 10 m) is ideal for the concept of "personal operating space" and integrates the notion of using a device carried or worn on the body or otherwise located within immediate reach. Bluetooth's transmission speed of 780 Kbps works well for transferring small- to medium-sized files. Important specifications of the Bluetooth technology are illustrated in Table 27.1.

Table 27.1 Bluetooth technology specifications

Connection type	Spread spectrum (FH)
Multiple access scheme	FHSS-TDD-TDMA in piconet FHSS-CDMA in scatternet
Spectrum	2.4 GHz ISM band Open band (79 MHz of spectrum)
Number of RF channels	79
Channel bandwidth	1 MHz
Modulation	Gaussian frequency shift keying
Transmission power	1–100 mw
Aggregate data rate	1 Mbps
Range	30 ft
Supported stations	8 devices
Voice channels	3
Data security-authentication key	128 bit key
Data security-encryption key	8–128 bits (configurable)

Table 27.2 Bluetooth RF transmitter power classes

Class	Max RF power	Range (m)
1	100 mW (20 dBm)	100 m
2	2.5 mW (4 dBm)	50 m
3	1.0 mW (0 dBm)	10 m

Operating range

Three classes of RF transmitted power are defined from 0 to 20 dBm (1–100 mW), as shown in Table 27.2. Most Bluetooth devices have class 3 radios, although class 1 adapters are also available, providing a personal area network (PAN) range comparable with an IEEE 802.11b/g WLAN.

Most of the commercially available devices have a transmitting power of 1 mW and hence a range of 10 m. Transmit power control is mandatory for class 1 radios (optional for classes 2 and 3) and requires transmitting devices to dynamically adjust power to reduce interference. This also helps reduce power consumption and extend battery life for portable devices. To implement power control, a receiver signal strength indicator (RSSI) is used to determine whether a received signal is within a defined "golden receive power range," typically between 6 and 20 dBm above the receiver sensitivity level. If the received power is outside this power range, the receiver sends an LMP instruction to the transmitter to adjust its transmit power.

Frequency bands

The Bluetooth radio attains spectrum spreading by FH in 79 hops displaced by 1 MHz, starts at 2.402 GHz and achieves at 2.408 GHz. This frequency band range is reduced (temporarily) in few countries (i.e., France) and uses 23-hop system. In order to satisfy with out-of-band regulations in every country, guard band is used at lower and upper band edges.

Use of RF bands of the electromagnetic spectrum is regulated by governments in most countries, in a spectrum management process known as frequency allocation or spectrum allocation. RF ranges from 3 kHz to 300 GHz in the electromagnetic frequency spectrum as shown in Figure 27.4.

In the below figure, the abbreviations used are as follows:

VLF: very low frequency
LF: low frequency
MF: medium frequency
HF: high frequency

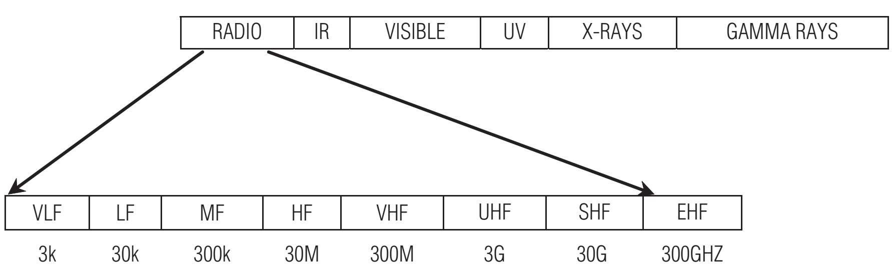


Figure 27.4 Frequency band designations

VHF: very high frequency
UHF: ultra-high frequency
SHF: super high frequency
EHF: extremely high frequency
UV: ultraviolet
IR: infrared

ISM bands

RF organizations, aware of the demand for local radio communications for individual users, have allocated certain frequency bands to be used in a more flexible way. The oldest and most commonly used frequencies are 900 MHz and 2.4 GHz, often called the ISM bands. The main characteristic of these bands is that they are unlicensed, which means that the frequencies are open for public use and require no registration or payment

The ISM radio bands are radio bands (portions of the radio spectrum) reserved internationally for the use of RF energy for ISM purposes other than communications. In general, communications equipment operating in these bands must tolerate any interference generated by ISM equipment, and users have no regulatory protection from ISM device operation.

RFs in the ISM bands have been used for communication purposes, although such devices may experience interference from non-communication sources. The 2.4 GHz band is proving to be increasingly attractive to industrial groups who are developing and promoting wireless network applications and products. WLAN devices use ISM bands as follows:

- Bluetooth 2,450 MHz band
- IEEE 802.11/Wi-Fi 2,450 and 5,800 MHz bands

Different countries have allocated various channels for Bluetooth operation. International Bluetooth frequency allocations are given in Table 27.3.

Modulation

Gaussian frequency shift keying (GFSK) is used as the modulation technique. In this technique, Binary 1 is represented by a positive frequency deviation and 0 by a negative frequency deviation. The radio receiver has to be designed in such a way to obtain a minimum bit error rate (BER) of 10^{-3}, that is, the radio should provide a link that ensures that there will not be more than one error for every 1,000 bits transmitted.

27.2.4 Baseband layer and baseband specification

The baseband is the PHY of the Bluetooth. It manages physical channels and links apart from other services like error correction, data whitening, hop selection, and Bluetooth security. The baseband layer lies on top of the Bluetooth radio layer in the Bluetooth stack. The baseband

Table 27.3 International Bluetooth frequency allocations

Area	Regulatory range	RF channels
U.S., most of Europe and most other countries	2.4–2.4835 GHz	$f = 2.4(2 + n)$ MHz, $n = 0,\ldots.78$
Japan	2.471–2.497 GHz	$f = 2.4(3 + n)$ MHz, $n = 0,\ldots.22$
Spain	2.445–2.475 GHz	$f = 2.4(9 + n)$ MHz, $n = 0,\ldots.22$
France	2.4465–2.4835 GHz	$f = 2.4(4 + n)$ MHz, $n = 0,\ldots.22$

protocol is implemented as a link controller that works with the LM for carrying out link level routines like link connection and power control. The baseband also manages asynchronous and synchronous links, handles packets, and does paging and inquiry to access and inquire Bluetooth devices in the area. Time division multiplexing is used to divide access to the channel across devices in the piconet (a set of devices form a small network). The baseband transceiver handles a time-division duplex (TDD) scheme. Therefore, the time is also slotted apart from different hopping frequencies (frequency division).

A collection of devices connected through Bluetooth technology in an ad hoc fashion is called piconet.

Baseband specification

In this section, we provide an overview of the key elements of the baseband specification of Bluetooth.

- ***FH***

  Bluetooth uses FH for multiple access with a carrier spacing of 1 MHz typically with up to 80 different frequencies being used for a total bandwidth of 80 MHz. Different logical channels (different hopping sequences) can simultaneously share the same 80 MHz bandwidth. Collisions will occur when devices in different piconets that are on different logical channels happen to use the same hop frequency at the same time. As the number of piconets in an area increases, the number of collisions increases and the performance degrades.

  In a piconet, the master and slave devices transmit alternatively. The master device starts its transmission in even-numbered time slots only and the slave device starts its transmission in odd-numbered time slots only.

  FH in Bluetooth serves the following two purposes:

  1. It provides resistance to interference and multipath effects.
  2. It provides a form of multiple access among colocated devices in different piconets.

  The FH scheme works as follows. The total bandwidth is divided into 79 (in almost all countries) physical channels, each with a bandwidth of 1 MHz. FH occurs by jumping from one physical channel to another in a pseudorandom sequence. The same hopping sequence is shared by all the devices on a single piconet and hence we will refer to this as an FH channel. The hop rate is 1,600 hops/s, so that each physical channel is occupied for a duration of 0.625 ms. Each 0.625 ms time period is referred to as a slot, and these are numbered sequentially.

  Bluetooth radios communicate using a TDD discipline. TDD is a link transmission technique in which data are transmitted in one direction at a time, with transmission alternating between the two directions. Because more than two devices share the piconet medium, the access technique is TDMA. Thus piconet access can be characterized as FH-TDD-TDMA.

- ***Physical links***

  Two types of links are established between master and slave: SCO and ACL. SCO links mainly carry voice transmission data and are symmetric links between the master and a single slave. To maintain the link, the master reserves transmit/receive time slots at regular intervals and, since the link is synchronous, SCO packets are not retransmitted in the event of a packet error. ACL links connect the master to all the slave devices in the piconet.

The master device can establish an ACL link to any slave using time slots that are not reserved for any active SCO links. Only one ACL link can exist at a time, but the link can be to a slave that already has an SCO link to the master. For most ACL packets, packet retransmission is applied in the event of a packet error. The Bluetooth baseband defines 13 packet types, including 4 specifically for high quality voice and voice data transmission. Each packet consists of a 68–72 bit access code, a 54 bit header, and a payload of up to 2,745 bits.

The access code is used during device discovery and to gain access to a specific piconet, and the header carries the slave address as well as information for acknowledgement, numbering and error checking of packets. The baseband controls the process of device discovery through an inquiry procedure, which enables a device to discover other devices in range and to determine their addresses and clock offsets, and a paging procedure that sets up the connection and synchronizes the slave device clock to the master. Once a connection is established, a device can be in one of the four states: active, sniff, hold, and park – in the order of decreasing power consumption.

- ***Active*** – In the active mode, the Bluetooth unit actively participates on the channel. The master schedules the transmission based on traffic demands to and from different slaves. In addition, it supports regular transmissions to keep slaves synchronized to the channel. Active slaves listen in the master-to-slave slots for packets. If an active slave is not addressed, it may sleep until the next new master transmission.
- ***Sniff*** – Devices synchronized to a piconet can enter power-saving modes in which device activity is lowered. In the sniff mode, a slave device listens to the piconet at reduced rate, thus reducing its duty cycle. The Sniff interval is programmable and depends on the application.
- ***Hold*** – Devices synchronized to a piconet can enter power-saving modes in which device activity is lowered. The master unit can put slave units into hold mode, where only an internal timer is running. Slave units can also demand to be put into hold mode. Data transfer restarts instantly when units transition out of hold mode.
- ***Park*** – In the park mode, a device is still synchronized to the piconet but does not participate in the traffic. Parked devices have given up their MAC (AM_ADDR) address and occasionally listen to the traffic of the master to re-synchronize and check on broadcast messages.

- ***Error correction***

The purpose of the forward error correction (FEC) scheme on the data payload is to reduce the number of retransmissions.

There are three kinds of error correction schemes used in the baseband protocol: 1/3 rate FEC, 2/3 rate FEC, and automatic repeat request (ARQ) scheme.

- In 1/3 rate FEC, every bit is repeated three times for redundancy.
- In 2/3 rate FEC, a generator polynomial is used to encode 10 bit code to a 15 bit code.

- In the ARQ scheme, DM, DH, and the data field of DV packets are retransmitted till an acknowledgement is received (or timeout is exceeded). Bluetooth uses fast, unnumbered acknowledgement in which it uses positive and negative acknowledgements by setting appropriate automatic repeat request number (ARQN) values. If the timeout value is exceeded, Bluetooth flushes the packet and proceeds with the next.

FEC is used on some SCO and ACL packets to correct errors and to reduce the number of ACL retransmissions.

27.2.5 Link Manager Protocol and LM specification

Link Manager Protocol (LMP) is used to set up and to manage baseband connections, including link configuration, authentication, and power management functions. This is achieved by exchanging protocol data units (PDUs) between the LM of two paired devices. PDUs include control of pairing, authentication, initiation of sniff, hold, and park modes, power increase or decrease requests, and selection of preferred packet coding and size to optimize data throughput.

The host drives a Bluetooth device through HCI commands, but it is the link manager (LM) that translates those commands into operations at the baseband level, managing the following operations:

1. Attaching slaves to piconets and allocating their active member addresses.
2. Breaking connections to detach slaves from a piconet
3. Configuring the link including master/slave switches
4. Establishing ACL and SCO links
5. Putting connections into low-power modes: hold, sniff, and park
6. Controlling test modes

A Bluetooth LM communicates with LMs on other Bluetooth devices using the LMP. Once the connection has been set-up, it can have up to three SCO connections created across it, or its mode can be changed, either to a low-power mode or to a test mode (these are useful for certification of Bluetooth devices by testing authorities and for a manufacture's production line testing of devices). The link can be configured at any time, including at mode changes, QoS changes, packet type changes, and any power level changes. Finally, information about an active link can be retrieved at any time. When the connection is no longer required, LMP can cause disconnection.

The LMP is used to set up and to control links. The three layers, that is, RF, link controller, and the LM, will be on the Bluetooth module attached to the device. The LM on one device exchanges messages with the LM on the other device. These messages, known as LMP messages are not propagated to higher layers. Link messages have higher priority than data. LMP messages are sent as single slot packets, with a header of 1 byte.

The functions of the LMP are as follows:

Authentication: When two devices have to communicate with each other, one has to verify the other device. So, one device is called the verifier and the other is called the claimant. The verifier sends a packet containing a random number, which is called a challenge. The claimant calculates the response, which is a function of the challenge, and sends the response along with its Bluetooth address (48-bit address) and a secret key. This is known as a challenge–response scheme, that is, you throw a challenge and check whether the other device can correctly respond to that challenge.

Encryption: To maintain confidentiality of data over the radio link, the data are encrypted. The master sends a key with which the data are encrypted to all the slaves, through an LMP message.

Clock offset request: To synchronize the clocks between the master and the slaves is a must for proper exchange of data. If the clock has to be offset, the LMP exchanges messages to ensure clock synchronization.

Timing accuracy information request: To ensure synchronization, the master can ask the slaves for timing accuracy information.

LMP version: It needs to be ensured that both devices use the same version of LMP. To achieve this, the version number of the LMP is exchanged.

Type of packets supported: Different Bluetooth-enabled devices may support different features, so an LMP features, request and response, are exchanged between the devices.

Switching between master–slave role: In a piconet, a device will act as a master and other devices as slaves. The master and a slave in a piconet can switch roles using the LMP messages. The master or the slave can initiate the switching operation

Name request: Each device can be given a user-friendly name having a maximum of 248 bits in ASCII format. A device can ask for the name through an LMP message and obtain the response.

Detach: Messages exchanged to close a connection.

Hold mode: Places an ACL link in hold for a specified time when there is no data to send. This feature is mainly to save power.

Park mode: In synchronization with the master, but not participating in data exchange.

Power control: Asks to transmit less power. This is useful particularly for class 1 devices, which are capable of transmitting 100 mW power.

QoS parameters exchange: In applications that require a good quality transmission link, QoS parameters can be specified. These parameters include the number of repetitions for broadcast packets, delay, and bandwidth allocation.

Request SCO link: Request an SCO link after the ACL link is established.

Multislot packet control: Controls the procedure when data are sent in consecutive packets.

Link supervision: Monitors link when device goes out of range (through a timeout mechanism).

Connection establishment: Establishes the connection after paging is successfully completed.

27.2.6 L2CAP

L2CAP enables transmission of data between upper and lower layers of the stack. It also enables support for many third-party upper-layer protocols such as TCP/IP. In addition, L2CAP provides group management by mapping upper-layer protocol groups to Bluetooth networks. It is also a factor in ensuring interoperability among Bluetooth units by providing application-specific protocols.

L2CAP runs above the baseband and carries out the data link layer functionality. L2CAP layer is only for ACL links. L2CAP data packets can be up to 64 kb long. L2CAP protocol runs on the host such as laptop, cellular phone, or other wireless devices. L2CAP does not do any checksum calculation. When L2CAP messages are exchanged between two devices, it assumes that an ACL link is already established between the two devices. It also assumes that packets are delivered in sequence. Note that L2CAP does not support SCO links for voice communication. Also, it does not support multicasting.

Other protocols interfacing with the L2CAP include SDP, RFCOMM, TCS, and IrOBEX.

- ***SDP*** provides service discovery specific to Bluetooth. That is, one device can determine the services available in another connected device by implementing the SDP.
- ***RFCOMM*** is a transport protocol that provides serial data transfer. In other words, it enables legacy software applications to operate on a Bluetooth device.
- ***TCS*** is for voice and data call control. It provides group management capabilities and allows for signalling unrelated to an ongoing call.
- ***OBEX*** is a session protocol, and for Bluetooth devices, only connection-oriented OBEX is supported. Three application profiles have been developed using OBEX: synchronization (for phonebooks, calendars, messaging, and so on), file transfer between connected devices, and object push for business card support.

27.2.7 L2CAP functions and security

This layer is only for ACL links or in other words, for data applications. The functions of L2CAP layer are as follows:

Protocol multiplexing*:* A number of other protocols can be running above L2CAP in the protocol stack seen in Figure 27.3. A packet received by L2CAP has to be passed on to the correct higher layer. This is protocol multiplexing.

Segmentation and reassembly*:* Baseband packets are limited in size as we saw in packet format. Large L2CAP packets are segmented into small baseband packets and sent to the baseband. Similarly, the small packets received from the baseband are reassembled and sent to higher layers.

QoS*:* QoS parameters such as delay can be specified, and this layer ensures that the QoS constraints are honoured.

The L2CAP layer sends connection request and QoS request messages from the application programs through the higher layers. It receives the responses for these requests from the lower layers. The responses can be connection indication, connection confirmation, connect confirmation negative, connect confirmation pending, disconnection indication (from remote) and disconnect confirmation, timeout indication, and QoS violation indication.

Security

Bluetooth is mainly intended for short-range connectivity between personal devices. Some basic security elements are included to prevent unauthorized usage and eavesdropping. At connection establishment, an authentication process is carried out to verify the identities of the units involved. The authentication process uses a conventional challenge response routine illustrated in Figure 27.5.

The claimant (right) transmits its claimed 48-bit address to the verifier (left). The verifier returns a challenge in the form of a 128-bit random number (AU_RAND). The AU_RAND, the claimant address, and a 128-bit common secret link key form the inputs to a computational secure hash function *E1* based on SAFER+, which produces a 32-bit signed response (SRES). The SRES produced by the claimant is sent to the verifier, which compares this result with its own SRES. Only if the two calculated SRES numbers are the same then challenger will continue with connection establishment. The authentication can be unidirectional or bidirectional.

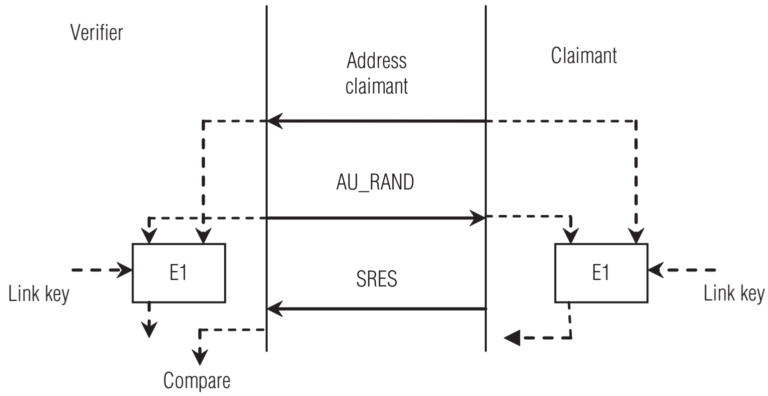


Figure 27.5 Bluetooth authentication procedure

27.3 Bluetooth network topology

The Bluetooth technology provides both a point-to-point connection and a point-to-multipoint connection. In point-to-multipoint connections, the channel is shared among several Bluetooth units. In point-to-point connections, only two units share the connection. If several piconets overlap a physical area, and members of the various piconets communicate with each other, this new, larger network is known as a *scatternet*.

In a PAN, a set of devices form a small network called a *piconet*. In a piconet, there will be one master and one or more slaves. All the slaves tune to the master. The master decides the hop frequency sequence and all the slaves tune to these frequencies to establish communication links. Any device can be a master or slave. The master–slave terminology is only for the protocols. The device capabilities are not defined by this terminology. It is also possible for a master and slave to switch roles, that is, a slave can become a master. A piconet can have maximum number of seven slaves that can actively communicate with the master. In addition to these active slaves, a piconet can contain many slaves in parked mode. These parked devices are synchronized with the master, but they are not active on the channel. The communication between the master and the slave uses TDD.

The three types of network configurations for Bluetooth devices are as following:

1. ***Single point-to-point (piconet):*** In this topology, the network consists of one master and one slave device (Figure 27.6(a)).
2. ***Multipoint (piconet):*** Such a topology combines one master device and up to seven slave devices in an ad hoc network (Figure 27.6(b))
3. ***Scatternet:*** A Scatternet is a group of piconets linked via a slave device in one piconet that plays a master role in the other piconet (Figure 27.6(c)).

27.4 Bluetooth versus Wi-Fi WLAN

Bluetooth and Wi-Fi are both wireless network standards that provide connections by radio waves. The main difference between these two technologies is that Bluetooth is used to replace cables and wires while Wi-Fi is used to provide wireless, high-speed access to the Internet.

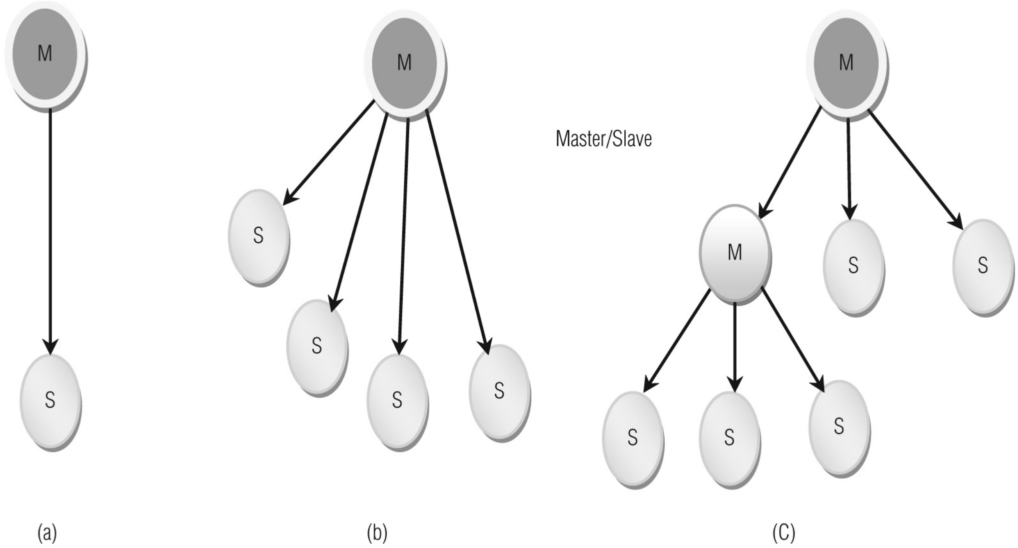


Figure 27.6 (a) Piconet (point-to-point); (b) Piconet (multipoint); and (c) Scatternet

The dominant WLAN technology, IEEE802.11b (or "Wi-Fi"), offers much higher data rates than Bluetooth, delivering usable data rates up to 6 Mbps. More recent WLAN standards (802.11a and 802.11b) are even more impressive, offering useful throughputs in excess of 25 Mbps; WLANs offer much greater range, operating over distances in the order of 50 m indoors as opposed to a typical range of around 10 m for Bluetooth (for the widely deployed Class 2 products).

While these comparisons are technically correct, they completely miss the key point about Bluetooth that it is designed for a myriad of wireless PAN applications, not as a competitor to 802.11b. Moreover, the low cost and size of the Bluetooth radio means that it is being included in equipment by default, and over time it is likely to become significantly more widespread than 802.11b WLAN. Whereas, in the short term, the use of WLANs will be dominated by business users; the Bluetooth user base will consist of people from all market segments, from true mass market right through to the high-end corporate. This is what is special about Bluetooth, that is, the sheer quantity of enabled devices. The comparison of Bluetooth and Wi-Fi technologies is given in Table 27.4.

The federal communications commission (FCC) qualifies devices that radiate RF energy as either intentional or unintentional radiators. Both these radiators are found to be in plenty in the crowded 2.4 GHz ISM bands that are shared by Wi-Fi and Bluetooth.

Table 27.4 Comparison of Bluetooth and Wi-Fi technologies

Technology	Application	Range (m)	Data rate	Connection type
Bluetooth	Cable replacement Ad hoc PAN	10–100	< 1Mbps	FHSS
802.11.b WLAN	High speed PAN	> 100	11 Mbps	DSSS

Wi-Fi and Bluetooth occupy a section of the 2.4 GHz ISM band that is 83 MHz-wide. Bluetooth uses FHSS and is allowed to hop between 79 different 1 MHz-wide channels in this band. Wi-Fi uses DSSS instead of FHSS. Its carrier does not hop or change frequency and remains centered on one channel that is 22 MHz-wide. While there is room for 11 overlapping channels in this 83 MHz-wide band, there is only room for three non-overlapping channels. Thus there can be no more than three different Wi-Fi networks operating in close proximity to one another.

When a Bluetooth radio and a Wi-Fi radio are operating in the same area, the single 22 MHz-wide Wi-Fi channel occupies the same frequency space as 22 of the 79 Bluetooth channels that are 1 MHz wide. When a Bluetooth transmission occurs on a frequency that lies within the frequency space occupied by a simultaneous Wi-Fi transmission, some level of interference can occur depending on the strength of each signal.

27.5 Introduction to WLL technology

In principle, *wireless local loop* (WLL) is a simple concept to grasp: it is the use of radio to provide a telephone connection to the home. In practice, it is more complex to explain because wireless comes in a range of guises, including mobility, because WLL is proposed for a range of environments and the range of possible telecommunications services' delivery is widening.

Traditionally, the provision of voice and data communication services to the end subscribers, over the subscriber loop or local loop, has been provided by wired systems. For residential subscribers, twisted pair cable has been and continues to be the standard means of connection. For business and government subscribers, twisted pair cable, coaxial cable, and optical fiber cable are in use.

WLL is a system that provides a wireless connection between subscribers and the local telephone station. The other names of WLL are *radio in the loop (RITL) or fixed-radio access (FRA)*. WLL systems are suggested for voice, data, Internet access, TV, and other new applications

WLL is the use of radio to provide a telephone connection to the home.

Figure 27.7 shows the role of WLL in the world, where a house is connected to a switch via a traditional local loop (copper cable buried in the ground), then through a distribution node onto a trunked cable going back to the switch. Historically, the local loop was copper cable buried in the ground. WLL replaces the local loop section with a radio path rather than a copper cable. It

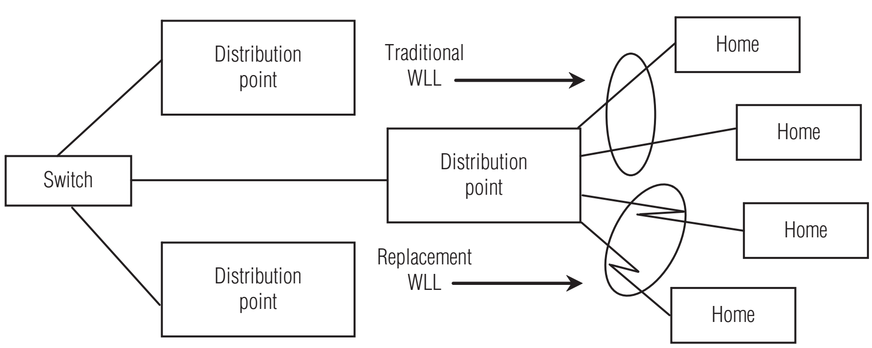


Figure 27.7 Role of WLL

is concerned only with the connection from the distribution point to the house; all other parts of the network are left unaffected.

In a WLL system, the distribution point is connected to a radio transmitter. A radio receiver is mounted on the side of the house, in much the same manner as a satellite receiver dish and a cable is run down the side of the house to a socket inside the house. The socket is identical to the one into which users currently plug their home telephones. Hence, apart from a small receiver on the side of their house, the home subscriber does not notice any difference.

Modern WLL systems use CDMA technology. A WLL service provider serves one or more cells. Each cell includes a base-station antenna mounted on top of a tall building or tower. Individual subscribers have a fixed antenna mounted on a building or a high pole that has an unobstructed LOS to the base-station antenna. From the base station, there is a wired or wireless link to a switching centre. The switching centre is typically a telephone exchange, which provides connections to the local and long-distance landline telephone networks. An Internet service provider (ISP) may be connected to the switching centre by a high-speed data link.

The WLL has a number of advantages over a wired approach to subscriber loop support, such as the following:

1. WLL systems are less expensive than wired systems.
2. WLL systems can be installed rapidly.
3. Subscriber radio units are installed only for those willing subscribers who want the service at a given time. With a wired system, typically a cable is laid out in anticipation of serving every potential subscriber in a local area.
4. A large geographical area is still not covered with landline telephone service or not covered for high-speed data transmission applications.
5. WLL has become cost-competitive with wired local loops, and new requirements are preferred to be met with WLL approach.
6. Cellular systems are quite expensive and do not provide sufficient facilities to act as a realistic alternative to broadband WLL.
7. A major advantage of WLL over the cellular mobile system is that the fixed subscriber can use a directional antenna pointed at the base-station antenna, providing improved signal quality in both directions.

WLL has been allocated a frequency band of 2–40 GHz, especially the unused frequency bands available above 25 GHz. Note that these frequencies are considerably higher than those used for cellular systems. At these frequencies, often referred to as millimetre wave frequencies, propagation characteristics are quite different from those in the MHz ranges.

WLL applications include local multipoint distribution service (LMDS) and multichannel multipoint distribution service (MMDS). LMDS is a relatively new WLL service capable of providing video, telephony, and high-data rates of the order of several Mbps, within the short range from the base station, requiring a relatively large number of base stations to service a given area.

MMDS signals operate at lower band of millimetre range and can operate in considerably larger cells within a radius of 50 km, but subscriber antennas must be in the LOS. MMDS can be used to support two way services as well as an alternative for broadband services such as Internet access. But it offers much less bandwidth than LMDS. LMDS is useful for larger companies

Table 27.5 Comparisons of WLL, mobile, wireless and wireline technologies

WLL	Mobile wireless	Wireline
Good LOS component	Mainly diffuse components	No diffuse components
Rician fading	Rayleigh fading	No fading
Narrowbeam directed antennas	Omnidirectional antennas	Expensive wires
High channel reuse	Less channel reuse	Reuse limited by wiring
Simple design, constant channel	Expensive DSPs, power control	Expensive to build and maintain
Low in-premises mobility only, easy access	High mobility allowed, easy access	Low in-premises mobility, wiring of distant areas is cumbersome
Weather conditions effects	Not very reliable	Very reliable

requiring greater bandwidths, whereas MMDS is likely to be used by residential subscribers and small businesses.

Fixed wireless terminal *(FWT):* This is a service provided using FWT and a telephone instrument. The FWT will be wall mounted and provided with an indoor or outdoor kind of aerial depending on the site of the premises, keeping in view the strength of the radio signal to be transmitted and received. WLL and FWTs are general terms for radio-based telecommunications technologies, and the respective devices that can be implemented using a number of different wireless and radio technologies are as given in Table 27.5.

27.6 Summary

- *Bluetooth* refers not only to a technology but also to a standard and a specification.
- Bluetooth is a short-range wireless communication technology that connects portable devices such as cell phones, handheld devices, and notebook computers. The technology has a range of up to 10 m and wirelessly transfers data at rates of up to 720 Kbps.
- The main aim of Bluetooth technology is to guarantee interoperability between different applications on devices in the same area that may run over different protocol stacks, and therefore to provide a solution for WPAN.
- Bluetooth is designed to be a small form-factor, low-cost and low-power radio communication technology. Bluetooth technology supports raw data transfer speed of 1 Mbps in the 2.4 GHz band (2.400–2.483 GHz) and communication at a range of up to 10 m.
- Bluetooth is one step above IrDA and one step below Wi-Fi WLAN. It is a step above IrDA because it allows simple devices, close together, to communicate together, simply, and they do not need to be in LOS. Also, its data capabilities (e.g., communications speed and distance) are much worse than those available with Wi-Fi.
- Bluetooth supports the ad hoc networking between different mobile wireless devices for spontaneous networking and immediate communication. Two supported network types are piconet and scatternet. Piconet is a network consisting of one master and up to seven slaves. This means that generally one Bluetooth device can be connected to up to seven other Bluetooth devices at the same time. Scatternet is a network formed by several piconets.
- WLL is the use of radio to provide a telephone connection to the home.

Review questions

1. In general terms, what are the application areas supported by Bluetooth?
2. What is the relationship between master and slave in a piconet?
3. How is it possible to combine FH and TDD?
4. List and briefly define the types of links that can be established between a master and a slave.
5. What error correction schemes are used in Bluetooth baseband?
6. List and briefly define Bluetooth baseband logical channels.
7. What security services are provided by Bluetooth?
8. List and briefly define L2CAP logical channels.

Objective type questions and answers

1. What are the transmitting range and the bit rate of a Bluetooth standard, respectively?
   (a) <10 m, <1 Mbps (b) >20 m, <1 Mbps
   (c) >10 m, <1 Mbps (d) >1.0 m, <1 Mbps
2. The frequency range of Bluetooth operating in ISM band is
   (a) 2,400–2,583.5 MHz (b) 2,400–2,483.5 MHz
   (c) 2,500–2,583.5 MHz (d) 2,400–2,493.5 MHz
3. The function of a Bluetooth radio (physical) protocol is
   (a) provides physical links among Bluetooth devices
   (b) provides transport service of packets on physical links
   (c) set-up and management of physical links
   (d) data transmission
4. Which protocol provides a transport service of packets on physical links?
   (a) PHY (b) LMP (c) baseband layer (d) SDP
5. Set-up and management of physical links are done by
   (a) PHY (b) SDP (c) baseband protocol (d) LMP
6. L2CAP services provides
   (a) Physical links among Bluetooth devices
   (b) Transport service of packets on physical links
   (c) Data transmission
   (d) Set-up and management of physical links
7. Which protocol is used to find the type of service that is available in the network?
   (a) PHY (b) SDP (c) baseband protocol (d) LMP
8. Which multiple access control scheme is used in Bluetooth technology?
   (a) FH-CDMA (b) FH-TDMA (c) FH-FDMA (d) none
9. What transmission power is used in Bluetooth technology?
   (a) 1–100 mW (b) 10–100 mW (c) 1–10 mW (d) 20–100 mW
10. Which modulation scheme is used in Bluetooth?
    (a) Quadrature phase shift keying (b) Gaussian phase shift keying
    (c) GFSK (d) none
11. Number of voice channels and supported stations in a Bluetooth technology, respectively, are
    (a) 4, 4 devices (b) 4, 6 devices (c) 4, 8 devices (d) 6, 6 devices
12. The radio receiver's minimum BER in Bluetooth technology is
    (a) 0.01 per cent (b) 0.2per cent (c) 0.02 per cent (d) 0.1 per cent

13. Data packet size of an L2CAP can be ________ long.
    (a) 64 kb (b) 32 kb (c) 16 kb (d) 8 kb
14. The IEEE 802.15.1 Bluetooth system has a typical frequency-hop rate of ________ hops per second.
    (a) 2.5 (b) 1,600
    (c) 3,200 (d) one access point only
15. In FH scheme, bandwidth is divided into ________ physical channels.
16. The data encryption is possible in Bluetooth by sending a key by Master, with which the data is encrypted to all the Slaves through ________.
17. Authentication of two communication devices can be achieved by ________.

Answers: 1. (a), 2. (b), 3. (a), 4. (c), 5. (d), 6. (c), 7. (b), 8. (a), 9. (a), 10. (c), 11. (c), 12. (d), 13. (a), 14. (b), 15. 79, 16. LMP message, 17. Challenge response scheme.

Open book questions

1. What is Bluetooth?
2. What are the advantages and disadvantages of Bluetooth?
3. What are the operations of LMP?
4. What are the differences between Wi-Fi (802.11b) and the Bluetooth wireless technology?
5. Differentiate between LMDS and MMDS applications of WLL technology.

Further reading

ETSI RES, "High Performance Radio Local Area Network (HIPERLAN) Type 1, Functional Specifications," ETS 300 652, 1996.
Freeman, R., *Radio System Design for Telecommunications (1–100 GHz)*. New York: Wiley & Sons, 1987.
IEEE 802.11, "Wireless LAN MAC and Physical Layer Specification," June 1997.
"Personal Handy Phone Standard (PHS)," CRC STD-28, 1993.
Steele, R. (Ed.), *Mobile Radio Communications*, New York: Wiley & Sons, 1992.
Tuttleby, W. (Ed.), *Cordless Telecommunications World-Wide*. Berlin, Germany: Springer-Verlag, 1996.
Withers, D., *Radio Spectrum Management*. London: IEE Press, 1987.

GPRS 28

28.1 Introduction

A number of mobile subscribers are discriminating day by day with the increasing demand for value-added data services like e-mail, Internet access, and receiving messages. Global system for mobile communications (GSM) of 2G is mainly used for circuit-switched communications. The general packet radio service (GPRS) is an extension of the GSM system, and uses the same channels, the same modulation, and the same network backbone as the existing GSM network. But the major difference is that GPRS is a packet-switching-based data service. In packet switching, the data are divided into packets with each data packet transmitted separately and then reassembled at the receiving end. GPRS supports the world's leading packet-based Internet communication protocols: Internet protocol (IP) and X.25. GPRS enables the existing IP or X.25 application to operate over a GSM cellular connection.

In GPRS, data can be exchanged directly in the form of a packet to the Internet or other networks. Even though, GPRS increases the capacity requirements on the radio and base-station subsystems, high mobile data speeds were not available to individual mobile users until enhanced data rates for global evolution (EDGE) or universal mobile telephone system (UMTS) was introduced. EDGE in GSM enhances existing GPRS/GSM infrastructure and increases speeds up to 384 Kbps. GPRS is well suited for non-real time Internet usage such as the retrieval of e-mail, faxes, and asymmetric web browsing, where the user downloads much more data than it uploads on the Internet.

This chapter provides an overview of GPRS, the GPRS architecture, different entities, and the major functionalities of the interfaces. In addition to the architecture, attach and detach procedures, packet data protocol (PDP) context procedures, and combined RA/LA update procedures are discussed.

28.2 GPRS architecture

The main features of the GPRS architecture are its flexibility, scalability, interoperability, and roaming capability. The network architecture of a GPRS system is shown in Figure 28.1.

In the core network, the existing mobile switching centres (MSCs) are designed to support circuit-switched traffic and cannot process packetized traffic. The packet-switched traffic is supported by enabling GPRS on a GSM network. This requires the addition of two core modules (network nodes) called as GPRS support nodes (GSN).

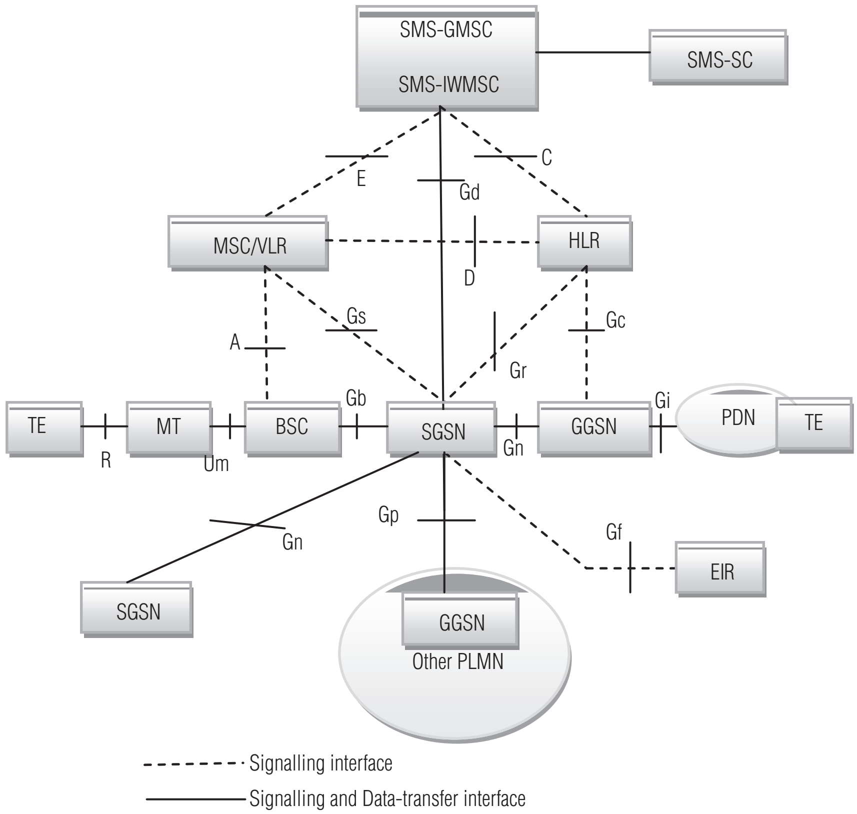


Fig 28.1 Typical GPRS network architecture

1. Serving GPRS support node (SGSN)
2. Gateway GPRS support node (GGSN)

The SGSN is essentially a packet-switched MSC. It delivers packets to mobile stations within its service area. SGSNs send queries to home location registers (HLR) to obtain profile data of GPRS subscribers. SGSNs detect new GPRS mobile stations in a given service area, process the registration of new mobile subscribers, and keep a record of their location inside a given area. So the SGSN performs mobility management functions such as mobile subscriber attach/detach and location management. The SGSN is connected to the base-station subsystem via a frame relay connection to the packet control unit (PCU) in the base station controller (BSC).

The GGSN is used as an interface to external IP networks such as the public Internet, other mobile service providers GPRS services, or business intranets. As the word gateway in its name suggests, the GGSN acts as a gateway between the GPRS network and IP-based public switch data network (PSDN). GGSNs also connect to other GPRS networks to facilitate GPRS roaming. The GGSN is similar to a GMSC (in GSM) since it provides a gateway between the GPRS network and the public packet data network (PDN) or other GPRS networks. It provides authentication and location management functions in addition to firewall functions on the Gi interface to the PDN.

All GSNs are connected via an IP-based GPRS backbone network. Within this backbone, the GSNs encapsulate the PDN packets and transmit (tunnel) them using the GPRS tunnelling protocol GTP. There are two types of GPRS backbones:

1. Intra-public land mobile network (PLMN) backbone networks connect GSNs of the same PLMN and private IP-based networks of the GPRS network provider.
2. Inter-PLMN backbone networks connect GSNs of different PLMNs. A roaming agreement between two GPRS network providers is necessary to install such a backbone.

The MSC/visitor location register (VLR) may be extended with functions and register entries that allow efficient coordination between packet-switched (GPRS) and circuit-switched (conventional GSM) services. Moreover, paging requests of circuit-switched GSM calls can be performed via the SGSN. For this purpose, the Gs interface connects the databases of SGSN and MSC/VLR. To exchange messages in the short message service (SMS) via GPRS, the Gd interface is defined. It interconnects the SMS gateway MSC (SMS-GMSC) with the SGSN. In addition to adding GPRS network nodes (SGSN, GGSN), some other nodes need to be added to a GSM network to implement a GPRS service. These include the following:

- The addition of PCUs often hosted in the GSM base-station subsystems
- Mobility management to locate the GPRS mobile station
- A new air interface for packet traffic
- New security features such as Ciphering
- New GPRS-specific signalling systems

GPRS attempts to reuse existing GSM network elements as much as possible, but to effectively build a packet-based mobile wireless network, some new network elements, interfaces, and protocols that handle packet traffic are required. Table 28.1 describes the network element modifications of GSM required for GPRS.

Table 28.1 Modifications required in GSM networks to support GPRS

GSM network element	Modification or upgrade required for GPRS
Subscriber terminal	A totally new subscriber terminal is required to access (Mobile Phone) GPRS services. These new terminals will be backward compatible with GSM for voice calls.
Base station	A software upgrade is required in the existing Base Transceiver Station (BTS).
BSC	The BSC will also require a software upgrade and the installation of a new piece of hardware called a PCU. The PCU directs the data traffic to the GPRS network and can be a separate hardware element associated with the BSC.
Core network	The deployment of GPRS requires the installation of new core network elements called the SGSN and the GGSN.
Databases (i.e. HLR, VLR)	All the databases involved in the network will require software upgrades to handle the new call models and functions introduced by GPRS.

Interfaces in GPRS architecture

- The MS and BSS communicate via the ***Um*** interface.
- The BSS and the SGSN are connected by the ***Gb*** interface using frame relay.
- Within the same GPRS network, SGSNs/GGSNs are connected through the ***Gn*** interface.
- When SGSN and GGSN are in different GPRS networks, they are interconnected with ***Gp*** interface.
- The GGSN connects to external networks through ***Gi*** interface.
- The MSC/VLR communicates with the BSS using the existing GSM ***A*** interface. Interface ***A*** is used for both signaling and voice transmission in GSM.
- The MSC/VLR communicates with the SGSN using the ***Gs*** interface (optional). SGSN may send/receive location information to/from MSC/VLR. SGSN may receive the paging request.
- The HLR connects to the SGSN via the ***Gr*** interface, and to the GGSN via the ***Gc*** interface.
- Both ***Gr*** and ***Gc*** follow GSM MAP protocol.
- The HLR and VLR are connected through the existing GSM ***D*** interface.
- Interfaces ***A, Gs, Gr, Gc,*** and ***D*** are used for signalling, without involving user data transmission in GPRS.
- Interfaces ***Um, Gb, Gn, Gp,*** and ***Gi*** are used for both signalling and transmission in GPRS.

28.2.1 Benefits of GPRS

GPRS makes use of the same communications transmission network as that of GSM. It has been designed for the transmission of data used for applications such as multimedia messaging service (MMS) picture messaging, mobile Internet browsing, and machine-to-machine data communications. The main benefits of GPRS are as follows:

- Allocates radio resources only when there are data to send and it reduces dependence on traditional circuit-switched network elements.
- The increased functionality of GPRS decreases the incremental cost to provide data services and in turn, increases the use of data services among consumer and business wireless users.
- Improves the quality of data services as measured in terms of reliability, response time (lower latency), and features supported.
- Offers new and improved data services to residential and business markets to aid retention and loyalty.
- Increases revenues from data services.
- Provides opportunity to increase subscriber numbers – there are more mobile phones in general use than there are PCs in people's homes. This means that the potential market for GPRS is high and that new Internet users are more likely to upgrade to a GPRS handset rather than making a larger investment in a PC.
- Offers innovative tariffs based on new dimensions such as the number of kilobytes or megabytes.
- Return on investment – investment in GPRS will be double since the new network infrastructure pieces will be used as part of the UMTS network requirements as well as GPRS.
- GPRS provides an upgrade path and test bed for UMTS.
- Control of large content portals.
- Access to the key member of the value chain – the customer.
- Cost effectiveness through spectrum efficiency – with packet-switching, radio resources are used only when users are actually sending or receiving data. This efficient use of scarce radio resources means that the large numbers of GPRS users can potentially share the same

bandwidth and be served from a single cell. GPRS spectrum efficiency means that there is less need to build in idle capacity that is only used in peak hours. GPRS therefore allows network operators to maximize the use of their network resources in a dynamic and flexible way.

28.3 GPRS procedures

Before data can be transferred between the MS and the external data network, some preparation is necessary to enable the transfer of IP packets through the GPRS network. Figure 28.2 illustrates the GPRS procedure. There are three important steps involved are described as follows:

1. The MS must be attached in the GPRS network. The procedure is called as the "GPRS attach procedure." This is a logical procedure between the MS and the SGSN that takes note of the position (i.e., the "routing area [RA]") of the MS. Storing and updating the position of the MS is particularly important for down link (DL) transmissions because this information enables the GPRS network to locate the MS.
2. A connection between the MS and the GGSN must be set up; that is, the activation of a PDP context. After this procedure, each node in the GPRS network knows how it has to forward the IP packets of this MS.
3. The path between the MS and the external data network is prepared. So IP packets can be sent through the GPRS network towards the destination address.

28.3.1 Attach and detach procedures

A subscriber requests a GPRS attach procedure when he registers with the GPRS network. This is the case when he switches on his mobile device or he explicitly activates GPRS while already GSM is attached. The result of this procedure is that the current SGSN (i.e., SLR) knows that the MS has activated GPRS. If the MS was already registered with another SGSN, the new SGSN updates the HLR such that the HLR knows the current SGSN identity. The HLR then sends the GPRS-specific data of the MS to this current SGSN. This GPRS-specific data correspond to the different PDP

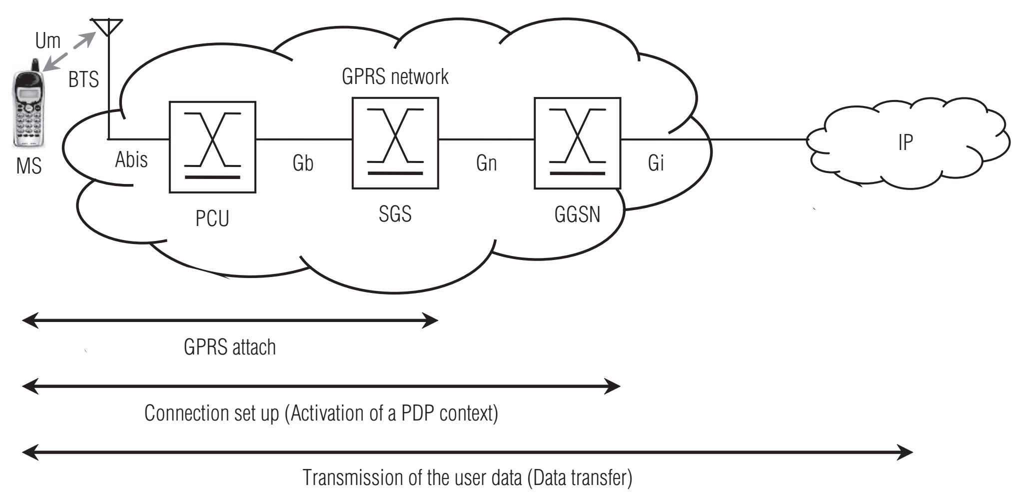


Figure 28.2 GPRS procedure

contexts that are set up for this MS. A PDP context describes the GPRS data connection of an MS. The information that describes such a context is sent from the HLR and consists of:

- Access point name (APN): A logical name for the desired data network – for example, intranet. tfk.de
- Quality of service (QoS): The priorities, delays, reliability, throughputs – for the desired application (speech, video, Internet surfing, download, etc.)
- PDP: The protocol used between the MS and the external data network, usually IP

Several PDP contexts can be set up for an MS such that the MS can reach several IP networks (i.e., several APNs) with different QoS. Several APNs can also be given to the same data network for different services with different QoS.

GPRS attach

A GPRS mobile station is not reachable or not known by the network until the mobile station performs the attach procedure and switches into the ready mode. To attach to the network (actually the SGSN), the mobile station provides its identity and indicates which type of attach procedure is to be performed:

- **GPRS attach:** Needs the mobile station's P_TMSI (packet temporary mobile subscriber identity) and the RAI (routing area identity) where the mobile is located.
- **IMSI (international mobile subscriber identity) attach:** Specific to GSM, but may be performed via GPRS if a TMSI is not already assigned to the mobile station.
- **IMSI/GPRS attach** will be possible in a later release of GPRS.

In the ready state, both the MS and the SGSN have established mobility management contexts for the subscriber's IMSI, which is the primary key to the GPRS subscription data stored in the HLR. The mobile station may send and receive data protocol data units (PDUs) that are nothing more than packets.

The mobile station may also activate or deactivate PDP contexts (data addresses) with the network. A mobile station may have many active PDP contexts simultaneously. The mobile station listens to the GPRS packet common control channel (PCCCH) and may also use discontinuous reception (DRX). DRX means that the mobile station will only use resources when data are present to receive (in the form of packets). Other times, the always-on condition is not using any radio resource. The mobility management remains in the ready mode until the ready timer expires and the mobile station then moves to the standby mode. Figure 28.3 shows the process of conducting the GPRS attach as the mobile station initiates the request.

Detach procedures

The GPRS detach procedure is used to detach the IMSI for GPRS services only. Independent of the network operation mode, this procedure is used by all kind of GPRS MSs. A combined GPRS detach procedure is used by GPRS MSs operation mode to detach the IMSI for GPRS and non-GPRS services or for non-GPRS services only. In the case of a network failure condition, it indicates the MS that a re-attach with successive activation of previously active PDP contexts shall be performed.

After completions of a GPRS detach procedure or combined GPRS detach procedure for GPRS and non-GPRS services, the GMM context is released. The GPRS detach procedure shall be invoked by the MS if the MS is switched off, the SIM card is removed from the MS, or if the GPRS or non-GPRS capability of the MS is disabled. The procedure may be invoked by the network to

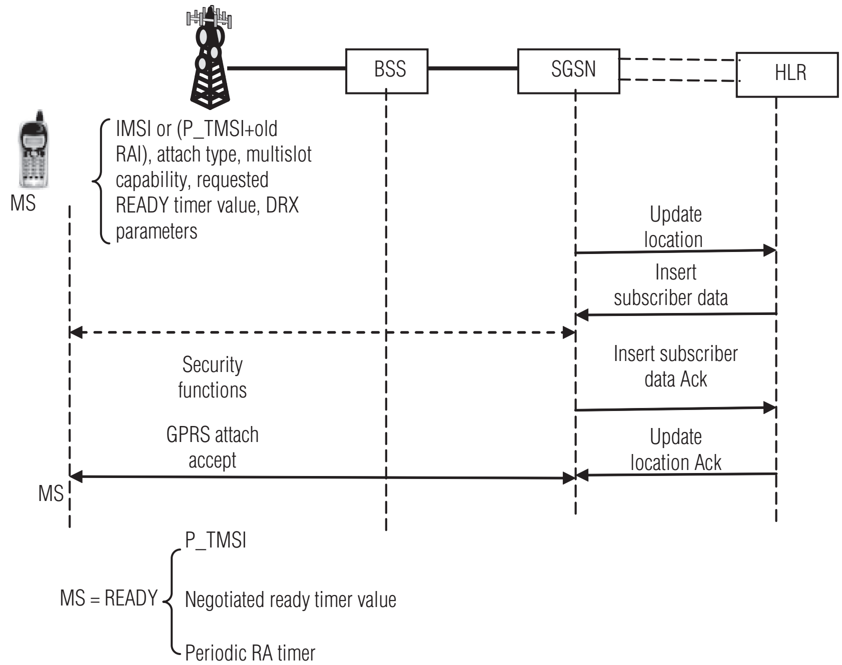


Figure.28.3 GPRS attach procedure to move to the ready mode

detach the IMSI for GPRS services. The GPRS detach procedure causes the MS to be marked as inactive in the network for GPRS services, non-GPRS services, or both services.
There are two types of detach procedures:

1. Mobile station–initiated GPRS detach
To move from the ready state to the idle state, the mobile station initiates a GPRS detach procedure. The result of the GPRS detach function is that the SGSN may delete the MM and the PDP contexts; the PDP contexts are actually deleted in the GGSN. The mobile station detaches by sending a detach request (detach type or a switch off) message to the SGSN. The detach type may include such things as detach for GPRS purposes only, IMSI detach (a GSM function), or both.

If GPRS detaches, the active PDP contexts in the GGSN regarding this particular mobile station are deactivated by the SGSN sending a Delete PDP Context Request message to the GGSN. The GGSN acknowledges with a Delete PDP Context Response. If the switch off indicates that the detach request is not due to a switch-off situation, the SGSN sends a GPRS Detach Accept message to the mobile station. Figure 28.4 shows the mobile station–initiated detach procedures, where the steps indicate the type of detach initiated.

2. Network-initiated GPRS detach
When it becomes necessary for the network to detach a mobile station, the SGSN informs the mobile station that it has been detached by sending a Detach Request (Attach Indicator) to the mobile station. The Attach Indicator indicates if the mobile station is requested to make a new GPRS attach and performs PDP context activation procedures for the previously activated PDP contexts. If so, the GPRS attach procedure is initiated when the GPRS detach procedure is

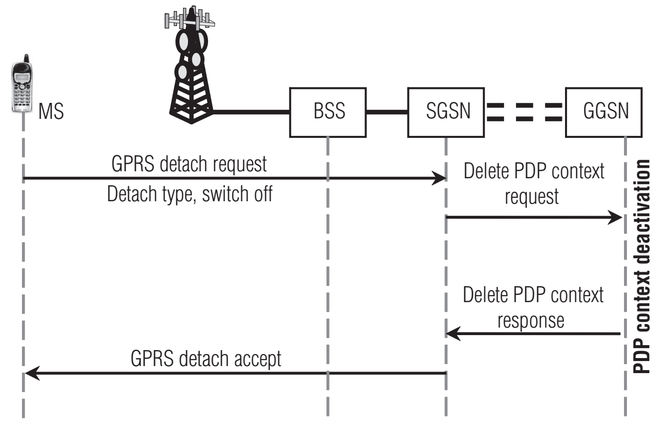


Fig 28.4 Illustration of mobile station-initiated detach procedure

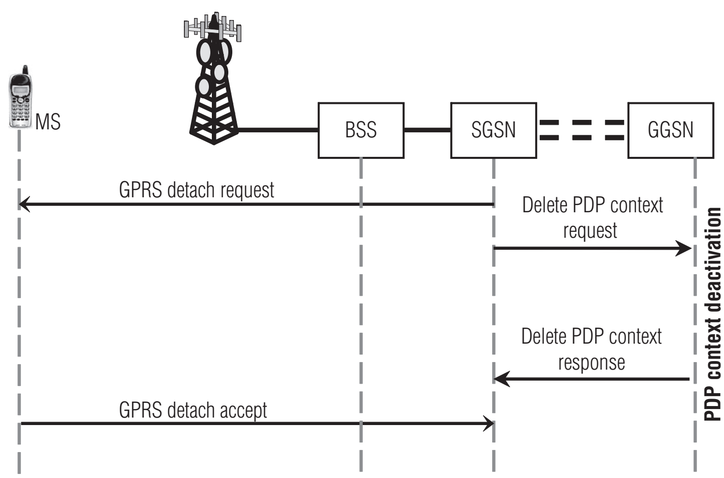


Fig 28.5 Illustration of network-initiated GPRS detach

complete. Figure 28.5 shows the network-initiated detach procedure. The active PDP contexts in the GGSN are deactivated by the SGSN. The mobile station sends a GPRS Detach Accept message back to the SGSN anytime after the Detach Request message.

28.4 PDP context procedure

After the GPRS Attach procedure (i.e., terminal is attached to SGSN and it was authorized, which means that it can access GPRS services and the identity was checked – authentication) is over, the terminal is pre-prepared for the data communication. In the most common situation, the terminal wants to communicate with the server in the Internet or in another external network (called PDN) accessed through operator's backbone. Some data services (MMS) are also provided by the operator itself. Mobile device must establish the data connection, and the PDP address (currently IPv4 or IPv6) from the accessed network rank must be assigned to the terminal.

Hence, when a terminal wants to start communication, it initiates the establishment of PDP context. In other words, after a GPRS attach, when the terminal is attached to an SGSN, it must activate a PDP address (in this case, an IP address) when it wishes to begin a packet data communication. This is usually done on application request, but in some situations, the user could decide to be "on-line" for the whole time and then the data connection is established during the device boot sequence (e.g., registration into IMS).

Before data can be sent or received, a PDP context (a data address) must be activated (created for the MS). A GPRS subscription contains several PDP addresses and an individual PDP context is maintained in the MS, SGSN, and GGSN for every PDP address. It is possible to inquire/set the following parameters in PDP:

- Requested QoS (peak bit rate, mean bit rate, delay requirements, reliability level expected)
- Data compression or no data compression
- Whether or not to use TCP/IP header compression
- PDP address and the type requested

Each PDP context can be either active or inactive. There are three PDP context functions available: activate, deactivate, and modify. The MS is responsible for activation and deactivation, GGSN is responsible for activation (for incoming packets) and deactivation, and SGSN is responsible for modification. An MS in standby or ready state can initiate activation or deactivation at anytime, to activate the PDP context in the MS, SGSN, or GGSN.

28.4.1 GPRS context activation scenario

A mobile station is attached itself to an SGSN in the GPRS PLMN. The mobile station has been assigned a TLLI that the wireless network knows. However, the external network nodes (IP or X.25) do not know of the mobile station. Therefore, the mobile station must initiate a PDP context with the GGSN. Both the SGSN and the GGSN are identified by IP addresses. A many-to-many relationship exists between the SGSN and the GGSN. Multiple tunnels (used for secure data transfer between the SGSN and the GGSN) may exist between a pair of GGSNs, each with a specific *tunnel identifier* (TID). Four steps are involved in the activation process, as shown in Figure 28.6.

- The mobile station sends a PDP context activation request to the SGSN.
- The SGSN chooses the GGSN based on information provided by the mobile station and other configurations and requests the GGSN to create a context for the mobile station. The SGSN will select a GGSN that serves the particular type of context needed (such as one for IP network access and one for X.25 access).
- The GGSN replies to the SGSN with the TID information. It also updates its tables wherein it maps the TID and the SGSN IP addresses with the particular mobile associated with them.
- The SGSN sends a message to the mobile station informing it that a context has been activated for that particular mobile. The SGSN also updates its links with the TID and the GGSN IP address with which it has established the tunnel for the mobile.

28.4.2 Mobile-initiated PDP data protocol context

Activation

The PDP context activation aims to establish a PDP context between the mobile station and the network, as shown in Figure 28.7. It may be performed automatically or manually depending

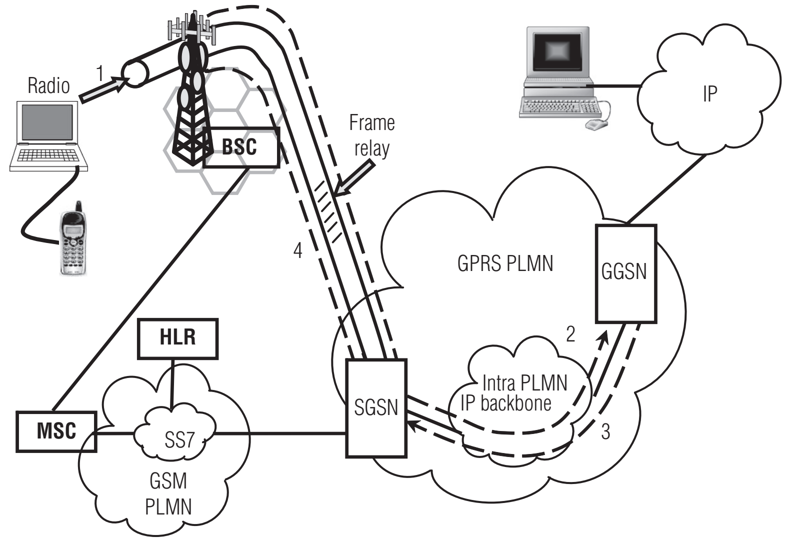


Fig 28.6 Illustrates GPRS context activation scenario

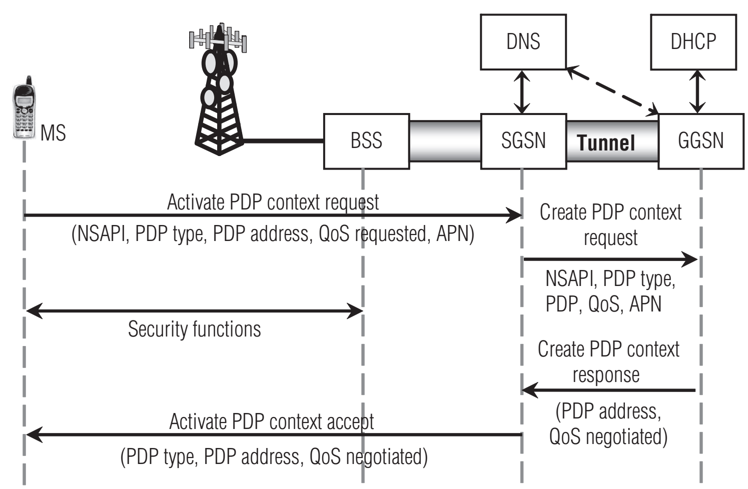


Fig 28.7 The mobile-initiated PDP activation

upon the manufacturer's implementation and configuration. The mobile station first sends an Activate PDP Context Request message that contains the following:

- NSAPI (Netscape server application programming interface)
- PDP type
- PDP address, whether it is a static or dynamic address (IP address)

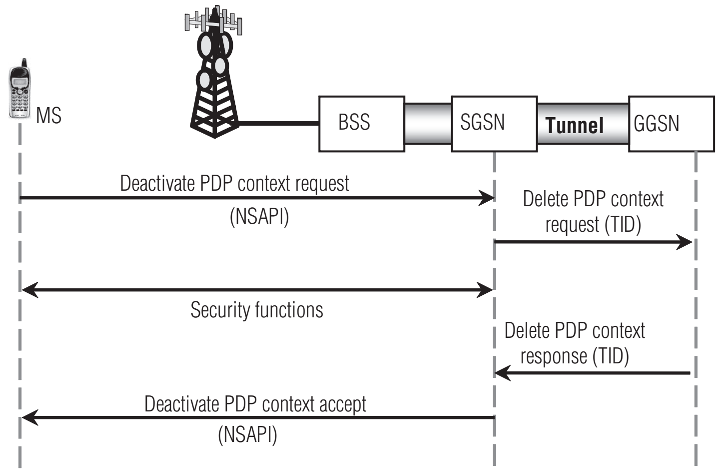


Fig 28.8 Mobile-initiated PDP context deactivation

- Requested QoS (best effort is all that is currently available, but will get to specific QoS in the future)
- APN (optional) to select a certain GGSN, either the IP address or logical name is used
- PDP configuration options

The mobile station only exchanges messages with the SGSN, which acts as a relay to the GGSN. The SGSN performs the following:

- Check the subscription data that was stored in the SGSN during the GPRS attach to determine if the mobile station is able to activate the PDP address.
- Insert the NSAPI along with the GGSN address in its PDP context.
- Return an Activate PDP Context Accept message to the mobile station.
- Become ready to route PDP packets (PDUs) between the GGSN and the mobile station.

Deactivation

To initiate this procedure, the mobile station sends a Deactivate PDP Context Request (NSAPI) message to the SGSN and security functions may be executed, as shown in Figure 28.8. The SGSN sends a Delete PDP Context Request (TID) message to the GGSN, which removes the PDP context and returns a Delete PDP Context Response (TID) message to the SGSN. If mobile stations were using a dynamic PDP address, the GGSN would release this address and make it available for subsequent activation by other mobile stations. The SGSN returns a Deactivate PDP Context Accept (NSAPI) message to the mobile station. At GPRS detach, all PDP contexts for the mobile station are implicitly deactivated.

28.4.3 Network-initiated PDP context

Activation

The network may also initiate a PDP context, if data arrives of a mobile user who has not already established an address. Figure 28.9 shows this network-initiated context activation. When

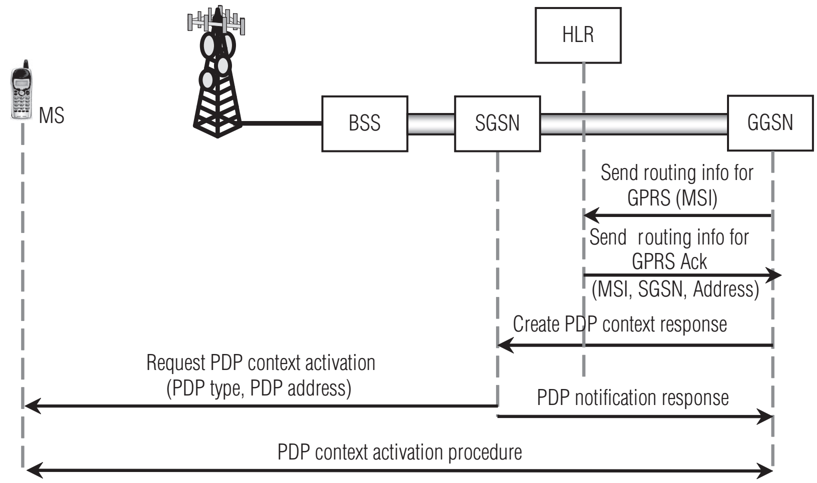


Fig 28.9 Network-initiated PDP context

receiving a PDP PDU, the GGSN determines if the network-initiated PDP context activation procedure has been initiated. The GGSN may send a Send Routing Information for GPRS (IMSI) message to the HLR (via the SGSN). The HLR returns a Send Routing Information for GPRS ACK (the information contained includes the IMSI, SGSN address, and cause) message to the GGSN.

- If a request can be served, the HLR includes the IP address of the serving SGSN.
- If a request cannot be served, the HLR only includes cause to indicate the reason for the negative response (cannot find network address and so on).

If the SGSN address is present and the cause is not present or is equal to a No Paging Response, the GGSN sends a PDU Notification Request message to the SGSN indicated by the HLR, which acknowledges it by sending a PDU Notification Response message.

The SGSN sends a Request PDP Context Activation (PDP type, PDP address) message to request the mobile station to activate the indicated PDP context, using the PDP context activation procedure (same as the mobile station-initiated).

Deactivation

When the GGSN initiates the PDP context deactivate procedure as shown in Figure 28.10, it sends a Delete PDP Context Request (TID) message to the SGSN, which sends a Deactivate PDP Context Request message (NSAPI) to the mobile station. The mobile station removes the PDP context and returns a Deactivate PDP Context Accept (NSAPI) to the SGSN. The SGSN returns a Delete PDP Context Response (TID) message to the GGSN. The SGSN may not wait for a response from the mobile station before sending the Delete PDP Context Response message.

28.5 Combined RA/LA update procedures–billing

The MS detects that a new cell has been entered by comparing the broadcasted cell identity with the cell identity stored in its MM context. The MS detects that a new RA has been entered by periodically comparing the RAI stored in its MM context with that received from the new cell.

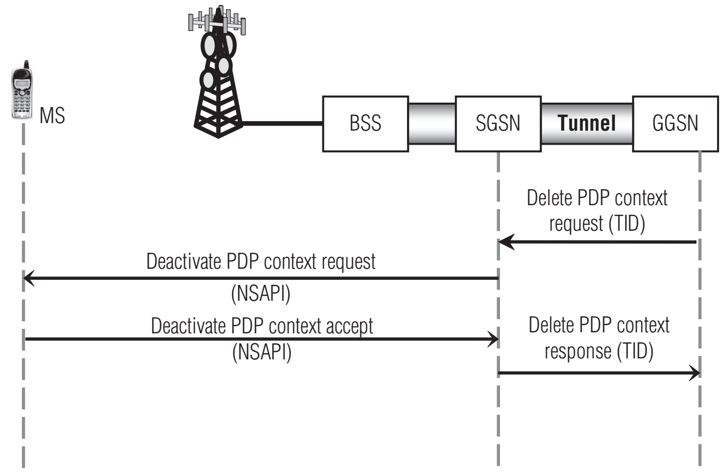


Fig 28.10 Network-initiated PDP context deactivation

When the MS camps on a new cell, possibly in a new RA, this indicates one of the following three possible scenarios:

- A cell update is required.
- An RA update is required.
- A combined RA and location area (LA) update is required.

In all three scenarios, the MS stores the new cell identity in its MM context. A cell update takes place only when the MS is in the MM state READY. The SGSN becomes aware of the cell update because the BSS adds the cell identity to the BSSGP-PDU sent over the Gb interface. An RA update takes place when a GPRS-attached MS detects that it has entered a new RA, when the periodic RA update timer has expired, or when a suspended MS is not resumed by the BSS. The SGSN detects that it is an intra-SGSN RA update by noticing that it also handles the old RA.

In this case, the SGSN has the necessary information about the MS and there is no need to inform the GGSNs or the HLR about the new MS location. A periodic RA update is always an intra-SGSN RA update. A combined RA/LA update takes place in network operation mode I when the MS enters a new RA or when a GPRS-attached MS performs IMSI attach. The MS sends an RA Update Request indicating that an LA update may also need to be performed, in which case the SGSN forwards the LA update to the VLR. This concerns only CS idle mode, since no combined RA/LA updates are performed during a CS connection. In network mode of operation II and III, whenever an MS determines that it shall perform both an LA update and an RA update, the MS shall perform the LA update first.

The MS sends the RA or LA Update Request message to the new SGSN. This message is not ciphered so that the new SGSN can process the message. For both GPRS and UMTS, the update type can be: RA update, periodic RA update, combined RA/LA update, or combined RA/LA update with IMSI attach. The “follow on request” parameter is used in UMTS to indicate if the connection should be kept for pending uplink traffic. This parameter does not exist in GPRS. In GPRS, before the BS passes the message to the SGSN, it adds the cell global identity information (including cell, RA and LA identities). In UMTS, the RNC adds the RAI information (including RA and LA identities). The SGSN and GGSN register all possible aspects of a GPRS user's behaviour and generate billing information accordingly. This information is gathered in so-called charging data records (CDR) and is delivered to a billing gateway. The GPRS service charging can be based on the following parameters:

- **Volume:** The amount of bytes transferred, that is, downloaded and uploaded.
- **Duration:** The duration of a PDP context session.
- **Time:** Date, time of day, and day of the week (enabling lower tariffs at off-peak hours).
- **Final destination:** A subscriber could be charged for access to the specific network, such as through a proxy server.
- **Location**: The current location of the subscriber.
- **QoS:** Pay more for higher network priority.
- **SMS:** The SGSN will produce specific CDRs for SMS.
- **Served IMSI/subscriber:** Different subscriber classes (different tariffs for frequent users, businesses, or private users).
- **Reverse charging:** The receiving subscriber is not charged for the received data; instead, the sending party is charged.
- **Free of charge:** Specified data to be free of charge.
- **Flat rate:** A fixed monthly fee.
- **Bearer service:** Charging based on different bearer services (for an operator who has several networks, such as GSM900 and GSM1800, and who wants to promote the usage of one of the networks). Or, perhaps the bearer service would be good for areas where it would be cheaper for the operator to offer services from a wireless LAN rather than from the GSM network.

28.6 Enhanced data rates for GSM evolution

EDGE is a new technology that offers much higher data rates over the air interface. EDGE stands for enhanced data rates for GSM evolution, and can be combined with both high-speed circuit-switched data (HSCSD) and GPRS to realize new coding schemes that, along with time slot bundling, can theoretically provide data rates around three-times higher than normal GPRS.

EDGE can be integrated into existing random selection for service and can be thought of as an upgrade for the modulation of bits onto the air interface. The term modulation refers to how the bits are transformed into radio waves and sent across the air interface. The so-called Gaussian minimum shift keying (GMSK) modulation used in GSM sends the bits one at a time; with EDGE, however, the so-called "8PSK" modulation sends the bits three at a time. Hence, EDGE offers roughly triple the bit rate of GSM.

A comparison of GSM and EDGE modulation is given in Figure 28.11. With GSM, a single bit (0 or 1) is represented by a higher or a lower frequency. With EDGE, a group of three bits (001, 010, 011, …, 111) is represented by one of eight different phase values, transmitted with a constant frequency.

The gross data rate over the radio interface with EDGE is 69.2 Kbps, and several new coding schemes are introduced that offer net bit rates of up to 59.2 Kbps. If a subscriber has all the eight time slots of a carrier, the maximal theoretical data rate with EDGE is then

$$59.2 \text{ Kbps} \times 8 \text{ time slots} = 473.6 \text{ Kbps}.$$

As EDGE offers such high data rates, this standard is also considered as belonging to the third generation. At the time of writing, several operators in the United States are implementing EDGE

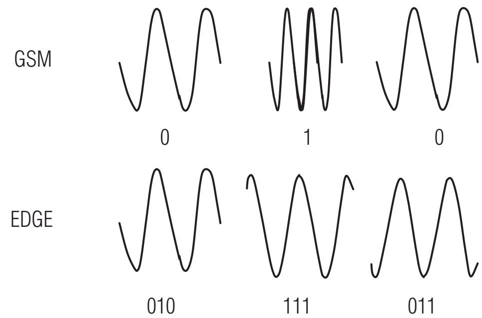


Figure 28.11 Comparison of GSM and EDGE modulation

instead of UMTS or CDMA2000. In the network, new modules must be implemented in the base stations, such that the new modulation is supported. Some other software updates must be done (e.g., in the PCU and in the SGSN) to support the additional signalling requirements. Another important remark regarding the implementation of the network, and especially of the BSS, is that more bandwidth (around three-times more) is necessary in the network, and especially on the Abis interface, because of the higher transmission rates over the radio interface. It is very common in GSM to install some chains of BTSs that are connected to the BSC, but because of the higher information load the length of these chains may need to be shortened; otherwise the link between the BSC and the first BTS has insufficient bandwidth to transport the volume information.

New mobiles supporting the new modulation are also necessary. Apart from the new modulation, EDGE uses the same radio interface as GSM. Furthermore, EDGE can transport either circuit- or packet-switched applications. For this purpose, two different applications are distinguished: enhanced circuit-switched data (ECSD) and enhanced general packet radio service (E-GPRS).

28.7 Summary

- GPRS is a packet-based wireless communication service designed for continuous connection to the Internet for portable terminals such as 2.5G cell phones and laptops.
- GPRS is an extension of the GSM system, and uses the same channels, the same modulation, and the same network backbone as the existing GSM network.
- The main benefits of the GPRS architecture are its flexibility, scalability, interoperability, and roaming capability.
- A subscriber requests a GPRS attach procedure when he registers with the GPRS network.
- The GPRS detach procedure is used to detach the IMSI for GPRS services only.
- The PDP context activation aims to establish a PDP context between the mobile station and the network
- The mobile station only exchanges messages with the SGSN, which acts as a relay to the GGSN.
- The network may also initiate a PDP context if data arrive for a mobile user who has not established an address already.
- The GMSK modulation used in GSM sends the bits one at a time; with EDGE, however, the “8PSK” modulation sends the bits three at a time.

Review questions

1. What is EDGE? How is it different from GPRS?
2. Explain about GPRS architecture.
3. Explain GPRS Attach and Detach procedures with neat diagrams.
4. Write about PDP context procedures.
5. Explain in detail about combined RA/LA procedures.
6. What is the role of SGSN in GPRS networks?
7. How are data transferred to external PDN through GPRS network?
8. What are the main interfaces involved in relation to signalling plane of GPRS networks?
9. Write the benefits of GPRS.
10. Explain deactivation procedure in network-initiated PDP context.

Objective type questions and answers

1. The main benefits of the GPRS architecture are its
   (a) flexibility (b) scalability (c) interoperability (d) all the above
2. Modulation technique used in GSM is
   (a) 8PSK (b) BPSK (c) GMSK (d) none
3. Modulation technique used in EDGE is
   (a) 8PSK (b) BPSK (c) GMSK (d) none
4. The PDP context activation aims to establish a PDP context between
   (a) mobiles (b) networks (c) mobile and network (d) none
5. ________ is an extension of the GSM system.
6. EDGE stands for ________.
7. ________ are responsible for the delivery and routing of data packets between the MSs and the external PDNs.
8. The BSS and the SGSN are connected by the ________ using frame relay.
9. ________ is used for both signalling and voice transmission in GSM in GPRS architecture.
10. The gross data rate over the radio interface with EDGE is ________.
11. The mobile station exchanges messages only with the ________, which acts as a relay to the GGSN.
12. EDGE uses the same radio interface as ________.
13. If a subscriber has all eight time slots of a carrier, the maximal theoretical data rate with EDGE is ________.
14. In network-initiated PDP context, the deactivation procedure is initiated by ________.
15. The mobile station sends a PDP context activation request to the ________.

Answers: 1. (d), 2. (c), 3. (a), 4. (c), 5. GPRS, 6. EDGE, 7. GSNs, 8. Gb interface, 9. Aninterface, 10. 69.2Kbps, 11. SGSN, 12. GSM, 13. 473.6Kbps, 14. GGSN, 15. SGSN.

Open book questions

1. Mention the benefits of GPRS.
2. What is the role of SGSN in GPRS?
3. Write about GGSN in GPRS network.
4. Write any three applications of GPRS.

Further reading

Bryan Strange, Lawrence Harte, *Introduction GPRS and EDGE*, ALTHOS Publications, 2005.

Bates, R. J., *GPRS: General Packet Radio Service*, Mc GRAW HILL Telecommunications, 2002.

Munsuri, A. A., Arzuaga Canals, J. M and Aiartzaguena, M. Z., *GPRS Technology and Applications*, PAC March 2010, World Magazine.

Shengyao Chen, *GPRS Attach/Detach*, UMTS Network WS 2003/2004, Technical University, Berlin.

Mobile IP and Wireless Application Protocol 29

29.1 Introduction

Mobile internet protocol (IP) is the underlying technology for support of various mobile data and wireless networking applications. With the advent of packet-based mobile data applications and the increase of wireless computing, there is a corresponding need for the ability for seamless communication between the mobile-node (MN) device and the packet data network (PDN) such as the Internet. To accomplish this, mobile IP established the visited network as a foreign node and the home network as the home node.

Mobile IP uses a tunnelling protocol to allow messages from the PDN to be directed to the MN's IP address. This is accomplished by way of routing messages to the foreign node for delivery via tunnelling the original IP address inside a packet destined for the temporary IP address assigned to the MN by the foreign node. The home agent (HA) and foreign agent (FA) continuously advertise their services on the network through an agent discovery process, enabling the HA to recognize when a new FA is acquired and allowing the MN to register a new care-of address (CoA).

Another open standard, besides mobile IP, providing mobile users of wireless terminals access to telephony and information services is wireless application protocol (WAP). It is an application communication protocol used to access services and information from handheld devices such as mobile phones. WAP is an enabling technology based on the Internet client–server architecture model, for transmission and presentation of information from the World Wide Web (WWW) and other applications utilizing the IP to a mobile phone or other wireless terminal. This chapter deals with introduction of mobile IP, WAP, and its application layer.

29.2 Mobile IP introduction

Mobile IP is an open standard, defined by the Internet Engineering Task Force (IETF), which allows users to keep the same IP address, stay connected, and maintain ongoing applications while roaming between IP networks. Mobile IP is scalable for the Internet because it is based on IP. Any media that can support IP can support mobile IP.

The number of wireless devices for voice or data is projected to surpass the number of fixed devices. Mobile data communication will likely emerge as the technology supporting most communication including voice and video. Mobile data communication will be persistent in cellular systems such as 3G and in wireless LAN, and will extend into satellite communication. Though mobility may be enabled by link-layer technologies, data crossing networks or different link layers are still a problem. The solution to this problem is a standards-based protocol, mobile IP.

In IP networks, routing is based on stationary IP addresses, similar to how a postal letter is delivered to the fixed address on the envelope. A device on a network is reachable through normal IP routing by the IP address that is assigned on the network.

The problem occurs when a device roams away from its home network and is no longer reachable using normal IP routing. This results in the active sessions of the device being terminated. Mobile IP was created to enable users to keep the same IP address while travelling to a different network (which may even be on a different wireless operator), thus, ensuring that a roaming individual could continue communication without sessions or connections being dropped.

Mobile IP functions

An important difference between mobile IP and mail forwarding is one that represents the classic distinction between people and computers: people are smart and computers are not. To this end, mobile IP includes a host of special functions that are used to set up and manage datagram forwarding. To see how these support functions work, we can describe the general operation of mobile IP as a simplified series of the following steps:

- ***Agent communication:*** The MN finds an agent on its local network by engaging in the agent discovery process. It listens for agent advertisement messages sent out by agents and from this it can determine where it is located. If it does not hear these messages it can ask for one using an agent solicitation message.
- ***Network location determination:*** The MN determines whether it is on its home network or a foreign network by looking at the information in the agent advertisement message. If it is on its home network, it functions using regular IP. To show how the rest of the process works, let us consider that the device just moved to a foreign network. The remaining steps are as follows:
- ***CoA acquisition:*** The device obtains a temporary address called a CoA. This either comes from the agent advertisement message from the FA, or through some other means. This address is used only as the destination point for forwarding datagrams.
- ***Agent registration:*** The MN informs the HA on its home network of its presence on the foreign network and enables datagram forwarding, by registering with the HA. This may be done either directly between the node and the HA, or indirectly using the FA as a conduit.
- ***Datagram forwarding:*** The HA captures datagrams intended for the MN and forwards them. It may send them either directly to the node or indirectly to the FA for delivery, depending on the type of CoA in use.
    - Datagram forwarding continues until the current agent registration expires. The device can then renew it. If it moves again, it repeats the process to get a new CoA and then registers its new location with the HA. When the MN returns back to its home network, it de-registers to cancel datagram forwarding and resumes normal IP operation.

29.2.1 Operation of mobile IP

Mobile IP has the following functional entities:

- ***Mobile node:*** This is the mobile device, the one moving around the Internetwork.
- ***Home agent (HA):*** This is a router on the home network that is responsible for catching datagrams intended for the MN and forwarding them to it when it is traveling. It also implements other support functions necessary to run the protocol.
- ***Foreign agent (FA):*** This is a router on the network to which the MN is currently attached. It serves as a "home away from home" for the MN, normally acting as its default router as well as implementing mobile IP functions. Depending on the mode of operation, it may

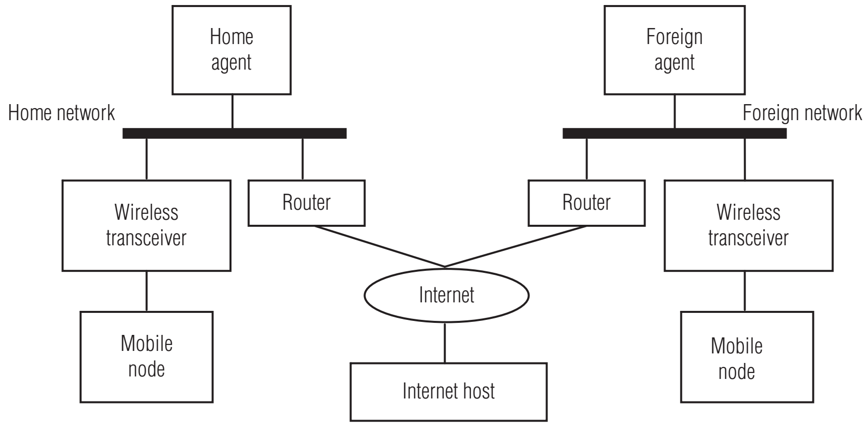


Figure 29.1 Mobile IP topology

receive forwarded datagrams from the HA and forward them to the MN. It also supports the sharing of mobility information to make mobile IP operate. The FA may not be required in some mobile IP implementations but is usually considered part of how the protocol operates.

The topology of a mobile IP system is shown in Figure 29.1.

In mobile IP the following scenario shows how a datagram moves from one point to another within the mobile IP framework:

- The Internet host sends a datagram to the MN by using the MN's home address (normal IP routing process).
- If the MN is on its home network, the datagram is delivered through the normal IP process to the MN. Otherwise, the HA receives the datagram.
- If the MN is on a foreign network, the HA forwards the datagram to the FA. The HA must encapsulate the datagram in an outer datagram so that the FA's IP address appears in the outer IP header.
- The FA delivers the datagram to the MN.
- Datagram from the MN to the Internet host are sent by using normal IP routing procedures. If the MN is on a foreign network, the packets are delivered to the FA. The FA forwards the datagram to the Internet host.
- In situations with ingress filtering present, the source address must be topologically correct for the subnet that the datagram is coming from, or a router cannot forward the datagram. If this scenario exists on links between the MN and the correspondent node, the FA needs to provide reverse tunnelling support. Then, the FA can deliver every datagram that the MN sends to its HA. The HA then forwards the datagram through the path that the datagram would have taken had the MN resided on the home network. This process guarantees that the source address is correct for all links that the datagram must traverse.

29.2.2 Co-located address

The IP devices have only a single address, since most of us use only a single e-mail address. Our travelling consultant, however, needs to have two addresses, a normal one and the other that is used while he is away. Mobile-IP-equipped notebook that our consultant carries needs to have two addresses as well:

- ***Home address:*** The "normal", permanent IP address assigned to the MN. This is the address used by the device on its home network, and the one to which datagrams intended for the MN are always sent.
- ***CoA:*** A secondary, temporary address used by a MN while it is "travelling" away from its home network. It is a normal 32-bit IP address in most respects, but is used only by mobile IP for forwarding IP datagrams and for administrative functions. Higher layers never use it, nor do regular IP devices when creating datagrams.

Mobile IP care-of address types

The CoA is a slightly tricky concept. There are two different types, which correspond to two distinctly different methods of forwarding datagrams from the HA router.

Foreign agent care-of address

This is a CoA provided by a FA in its agent advertisement message. It is, in fact, the IP address of the FA itself. When this type of CoA is used, all datagrams captured by the HA are not relayed directly to the MN, but indirectly to the FA, which is responsible for final delivery. Since in this arrangement the MN has no distinct IP address valid on the foreign network, this is typically done using a layer two technology.

Co-located care-of address

This is a CoA assigned directly to the MN using some means external to mobile IP. For example, it may be assigned on the foreign network manually, or automatically using DHCP. In this situation, the CoA is used to forward traffic from the HA directly to the MN.

29.2.3 Mobile IP registration

Once a MN has completed agent discovery, it knows whether it is on its home network or on a foreign network. If on its home network it communicates as a regular IP device, but if on a foreign network it must activate the mobile IP. This requires that it communicate with its HA so information and instructions can be exchanged between the two. This process is called HA registration, or more simply, just registration. A mobile IP HA registration is shown in Figure 29.2.

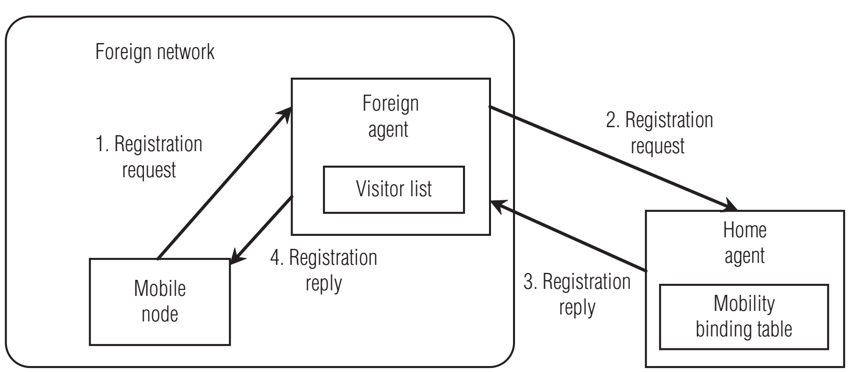


Figure 29.2 Mobile IP home agent registration

The main purpose of registration is to actually start mobile IP working. The MN must contact the HA and tell it that it is on a foreign network and request that datagram forwarding be turned on. It also must let the HA know its CoA so the HA knows where to send the forwarded datagrams. The HA, in turn, needs to communicate various types of information back to the MN when registration is performed. Note that the FA is not really involved in registration, except perhaps to relay messages, as we will see in the chapter.

MN registration events

Successful registration establishes what is called in the standard a mobility binding between a HA and a MN. For the duration of the registration, the MN's regular home address is tied to its current CoA and the HA will encapsulate and forward datagrams addressed to the home address over to the CoA. The MN is supposed to manage its registration and handle various events using several actions as follows:

- ***Registration:*** The MN initiates a registration when it first detects it has moved from its home network to a foreign network.
- ***Deregistration:*** When the MN returns home, it should inform its HA to cancel forwarding, a process called deregistration.
- ***Re-registration:*** If the MN moves from one foreign network to another, or if its CoAs changes, it must update its registration with the HA. It also must do so if it's current registration is about to expire, even if it remains stationary on one foreign network.

Each registration is established only for a specific length of time, which is why regular re-registration is required whether the device moves or not. Registrations are time-limited to ensure that they do not become stale. If, for example, a node forgets to de-register when it returns home, the datagram forwarding will eventually stop when the registration expires.

New registration request and registration reply messages

To perform registration, two new message types have been defined in mobile IP: the registration request and the registration reply. Each of these does what you would expect from its name. Interestingly, these are not ICMP messages like the one used in agent discovery; they are user datagram protocol (UDP) messages. Thus, technically speaking, registration is performed at a higher layer than the rest of mobile IP communication. Agents listen for registration requests on well-known UDP port number 434, and respond back to MNs using whatever ephemeral port the node used to send the message.

Registration procedures

Depending on the type of CoA used by the MN and other specifics, which we will discuss shortly, two different procedures are defined for registration. The first is the direct registration method, which has the following two steps:

- MN sends registration request to HA.
- HA sends registration reply back to MN.

In some cases, however, a slightly more complex process is required, where the FA conveys messages between the HA and the MN. In this situation, the process has the following four steps and also shown in Figure 29.3.

- MN sends registration request to FA.
- FA processes registration request and forwards to HA.

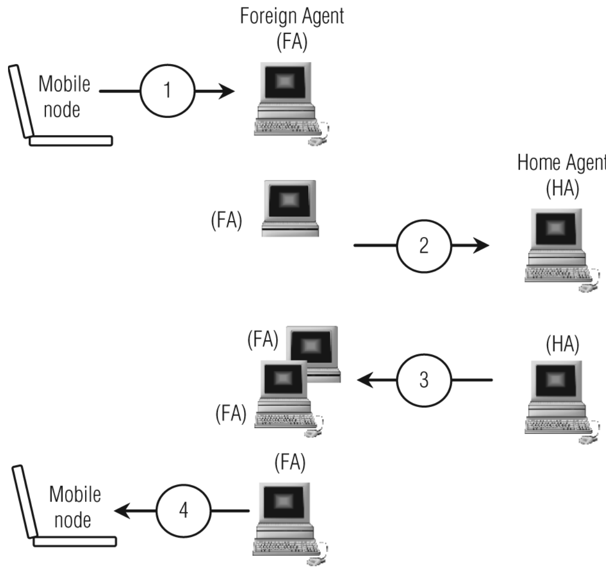


Figure 29.3 Mobile IP registration process

- HA sends registration reply to FA.
- FA processes registration reply and sends back to MN.

The first, simpler method is normally used when a MN is using a co-located CoA. In that situation, the node can easily communicate directly with the HA, and the MN is also set up to directly receive information and datagrams from the HA. When there is no FA, this is obviously the method that *must* be used. It is also obviously the method used when a MN is de-registering with its HA after it arrives back on the home network.

The second method is required when a MN is using a foreign CoA. Recall that in this situation, the MN does not have its own unique IP address at all; it is using a shared address given by the FA, which precludes direct communication between the node and the HA.

Note that the FA really is just a “middleman”; the exchange is still really between the HA and the MN. However, the FA can deny registrations if the request violates whatever rules are in place for using the foreign network. It is for this reason that some FAs may require that they be the conduit for registrations even if the MN has a co-located CoA. However, if the FA cannot contact the HA the registration will not be able to proceed.

The description above is really a highly simplified explanation of the basics of registration. The mobile IP standard specifies many more details on exactly how agents and nodes perform registration, including particulars on when requests and replies are sent, how to handle various special conditions such as invalid requests, rules for how HAs maintain a table of mobility bindings, and much more. The standard covers the definition of extensions to the regular registration messages to support authentication, which is required for secure communications. It also includes the ability to have a MN maintain more than one concurrent binding, when needed.

29.2.4 Mobile IP tunnelling

Once a MN on a foreign network has completed a successful registration with its HA, the mobile IP datagram forwarding process described in the general operation topic will be fully "activated". The HA will intercept datagrams intended for the MN as they are routed to their home network, and forward them to the MN. This is done by encapsulating the datagrams and then sending them to the node's CoA.

The encapsulation process creates a logical construct, called a tunnel, between the device that encapsulates and the one that decapsulates. This is the same idea of a tunnel used in discussions of virtual private networks (VPNs), IPsec tunnel mode, or the various other tunnelling protocols used for security. The tunnel represents a medium over which datagrams are forwarded across an arbitrary internetwork, with the details of the encapsulated datagram (i.e., the original IP headers) temporarily hidden and is shown in Figure 29.4.

In mobile IP, the start of the tunnel is the HA, which does the encapsulation. The end of the tunnel depends on what sort of CoA is being used:

- ***Foreign agent care-of address:*** The FA is the end of the tunnel. It receives encapsulated messages from the HA, strips off the outer IP header and then delivers the datagram to the MN. This is generally done using layer two, because the MN and FA are on the same local network, and of course, the MN does not have its own IP address on that network (it is using that of the FA.)
- ***Co-located care-of address:*** The MN itself is the end of the tunnel and strips off the outer header.

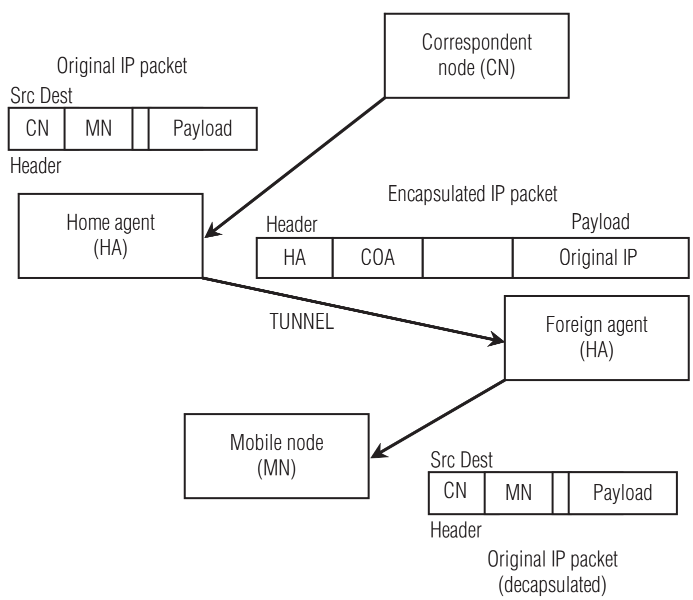


Figure 29.4 Mobile IP tunnelling

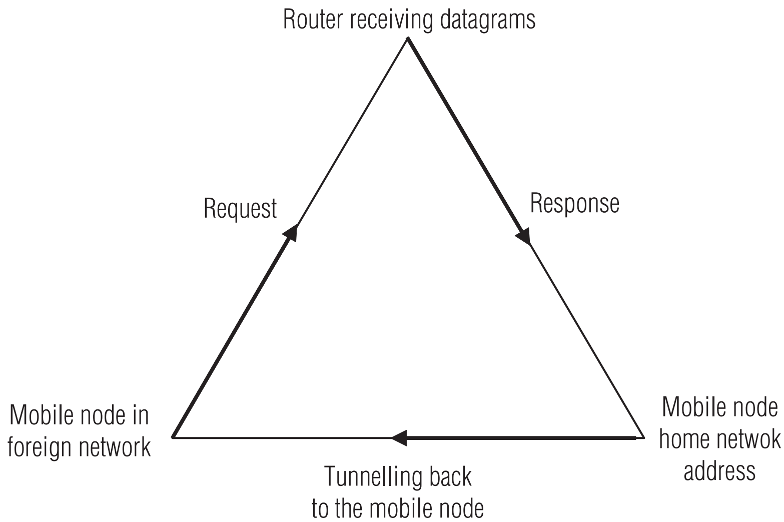


Figure 29.5 Mobile IP tunnelling

Mobile IP conventional tunnelling

Normally, the tunnel described above is used only for datagrams that have been sent to the MN and captured by the HA. When the MN wants to send a datagram, it does not tunnel it back to the HA; this would be needlessly inefficient. Instead it just sends out the datagram directly using whatever router it can find on its current network, which may or may not be a FA. When it does this, it uses its own home address as the source address for any requests it sends. As a result, any response to those requests will go back to the home network. This sets up a "triangle" of sorts for these kinds of transactions:

- The MN sends a request from the foreign network to some third-party device somewhere on the internetwork.
- The third-party device responds back to the MN. However, this sends the reply back to the MN's home address on its home network.
- The HA intercepts the response on the home network and tunnels it back to the MN.

The reverse transaction would be pretty much the same, just in the reverse order. In that case the third-party (Internet) device would send a request to MN, which would be received and forwarded by the HA. The MN would reply back directly to the Internet host. Mobile IP tunnelling is shown in Figure 29.5.

Mobile IP reverse tunnelling

There may be situations where it is not feasible or desired to have the MN send datagrams directly to the internetwork using a router on the foreign network as we just saw. In this case, an optional feature called reverse tunnelling may be deployed, if it is supported by the MN, the HA and if relevant, the FA.

When this is done, a reverse tunnel to complement the normal one is set up between the MN and the HA, or between the FA and the HA, depending on CoA type. All transmissions from the MN are tunnelled back to the home network where the HA transmits them over the internetwork, resulting in a more symmetric operation rather than the "triangle" just described. This is basically what described earlier as being "needlessly inefficient", because it means each communication requires four steps. Thus, it is used only when necessary.

One situation where reverse tunnelling may be required is if the network where the MN is located has implemented certain security measures that prohibit the node from sending datagrams using its normal IP address. In particular, a network may be set up to disallow outgoing datagrams with a source address that does not match its network prefix. This is often done to prevent "spoofing" (i.e., impersonating another's IP address.)

29.2.5 Internet protocol version 4 and Internet protocol version 6

IPv4

IPv4 is an abbreviation of "Internet protocol version four". The IP is in the third layer of the OSI network model. This is also known as the network layer. The network layer is the first layer in the OSI, which is software based. The network layer or the third layer of the OSI model deals with finding, routing, and switching for end-to-end communications that are not directly connected to each other using one physical link, for example, an Ethernet cable. The IP is the most dominant protocol on the Internet today and usually runs on upper-layer protocols such as the transmission control protocol (TCP) and the UDP. IP networking is an example of connectionless networking service (CLNS). IPv4 address consists of 32 bits, 4 bytes that are a combination between zeros and ones. The address consists of two parts: the network part and the host address part.

The current mobile IPv4 protocol is completely transparent to the transport and higher layers and does not require any changes to existing Internet hosts and routers. The mobile IP protocol allows the MNs to retain their IP address regardless of their point of attachment to the network. This can be fulfilled by allowing the MN to use two IP addresses. The first one, called home address, is static and is mainly used to identify higher layer connections, for example, TCP. The second IP address that can be used by an MN is the CoA. While the mobile is roaming among different networks, the CoA changes. The reason for this is that the CoA has to identify the mobile's new point of attachment with respect to the network topology. In mobile IPv4, the CoA management is achieved by an entity called FA. The MN, using its home address, is appearing to be able to receive data on its home network through an HA. In the situation that the mobile roams into a foreign region, it will need to obtain a new CoA via the FA.

This new CoA will be registered with its HA. At the moment that the HA receives a packet that has to be sent to the mobile, it delivers it from the home network to the mobile's CoA. The delivery can take place only if the packet is redirected or tunnelled, such that the CoA appears as the destination IP address. The HA tunnels the packet to the FA. After receiving the packet, the FA will have to apply the reverse transformation to decapsulate it, such that the packet will appear to have the mobile's home address as the destination IP address. After decapsulation, the packet is sent to the MN. Due to the fact that the packet arrives at the MN, being addressed to its home address, it will be processed properly by the upper protocol layers, for example, TCP. The IP packets sent by the MN are delivered to their proper destination by standard IP routing procedures.

IPv6

Internet protocol version 6 (IPv6) is designed to succeed the IPv4 and it is a version of IP. In addition to functions provided by IPv4, the primary changes from IPv4 to IPv6 include expanded addressing and routing, auto configuration, authentication, and confidentiality capabilities. The

IPv6 addresses are 128 bits long. Auto configuration refers to dividing the IP address into two distinct parts: network identifier and a unique host identifier, using the host's IEEE-802 48-bit MAC address. IPv6 establishes three important security services: packet authentication, packet integrity, and packet confidentiality.

All security functions are achieved using IPv6's optional extension header. It simplifies aspects of address assignment (stateless address auto configuration), network renumbering and router announcements when changing Internet connectivity providers. The IPv6 subnet size has been standardized by fixing the size of the host identifier portion of an address to 64 bits to facilitate an automatic mechanism for forming the host identifier from link-layer media addressing information (MAC address). Network security is also integrated into the design of the IPv6 architecture, and the IPv6 specification mandates support for Internet Protocol Security (IPsec) as a fundamental interoperability requirement.

Mobile IPv6 is an IETF standard that has added the roaming capabilities of MNs in IPv6 network. The major benefit of this standard is that the MNs (as IPv6 nodes) change their point-of-attachment to the IPv6 Internet without changing their IP address. This allows mobile devices to move from one network to another and still maintain existing connections. Although mobile IPv6 is mainly targeted for mobile devices, it is equally applicable for wired environments. The need for mobile IPv6 is necessary because the MNs in fixed IPv6 network cannot maintain the previously connected link (i.e., using the address assigned from the previously connected link) when changing location.

To accomplish the need for mobility, connections to mobile IPv6 nodes are made (i.e., without user interaction) with a specific address that is always assigned to the MN, and through which the MN is always reachable. Mobile IPv6 is expected to be used in IP over WLAN, WiMAX, or BWA.

Mobile IPv6 operation

When a MN is away from home, it sends information about its current location to the HA. A node that wants to communicate with a MN uses the home address of the MN to send packets. The HA intercepts these packets, and using a table, tunnels the packets to the MN's CoA.

Mobile IPv6 uses CoA as source address in foreign links. Moreover, to support natural route optimization, the correspondent node uses IPv6 routing header than the IP encapsulation. The following discussion makes mobile IPv6's understanding more clearly by highlighting the benefit of mobile IPv6 over mobile IPv4.

- Route optimization is a built-in feature for mobile IPv6. In mobile IPv4, this feature was available via an optional set of extensions that was not supported by all nodes.
- There is no requirement of FAs in mobile IPv6. As mentioned previously, neighbour discovery and address auto-configuration features enable MNs to function in any location without the services of any special router in that location.
- There is no access filtering problem in mobile IPv6 (In mobile IPv4 this happens because the correspondent node puts its home address as the source address of the packet). In mobile IPv6, the correspondent node puts the CoA as the source address and having a home address destination option, allows the use of the CoA to be transparent over the IP layer.

29.3 Wireless application protocol

Another open standard, besides mobile IP, providing mobile users of wireless terminals access to telephony and information services is the WAP. However, WAP is not actually a protocol as the terms hypertext transfer protocol (HTTP) and IP suggest. Instead, multiple protocols and complete network architecture are involved in WAP for delivery of wireless content. The architecture specified by the WAP is based on the layers that follow OSI model. WAP is designed to work with all wireless network technologies such as GSM, CDMA, and TDMA.

In near future, the WAP applications may extend the function of telephone, for example, the user will be allowed to answer the phone message with an e-mail. The news feeds, stock quotes, and weather forecasts were featured in early WAP application. These developments are setting hopes in wireless industry that in e-commerce applications like online banking, the WAP devices will become popular. Significant backlash against the hype and optimism surrounding WAP has certainly occurred as a result of the uncertainty about its future.

The WAP is a universal, open standard developed by the WAP forum to provide mobile users of wireless phones and other wireless terminals such as pagers and personal digital assistants (PDAs) access to telephony and information services, including the Internet and the Web. WAP is designed to work with all wireless network technologies (e.g., GSM, CDMA, and TDMA). WAP is based on existing Internet standards, such as IP, XML, HTML, and HTTP, as much as possible. It also includes security facilities. Ericsson, Motorola, Nokia, and Phone.com established the WAP forum in 1997, which now has several hundred members.

The use of mobile phones and terminals for data services are the significant limitations of the devices and the networks that connect them. The devices have limited processors, memory, and battery life. The user interface is also limited, and the displays small. The wireless networks are characterized by relatively low bandwidth, high latency, and unpredictable availability and stability compared to wired connections. WAP is designed to deal with these challenges.

The WAP specification includes the following:

- A programming model based on the WWW programming model
- A markup language, the wireless markup language, adhering to Extendable Markup Language (XML)
- A specification of a small browser suitable for a mobile, wireless terminal
- A lightweight communications protocol stack
- A framework for wireless telephony applications (WTAs)

29.3.1 WAP architecture

WAP is designed in a layered fashion so that it can be extensible, flexible, and scalable. The WAP stack is divided into five layers. The WAP architecture is shown in Figure 29.6. The WAP layered architecture enables other services and applications to utilize the features of the WAP stack through a set of well-defined interfaces. External applications may access the session, transaction, security, and transport layers directly.

The WAP programming model is based on three elements: the client, the gateway (GW), and the original server. HTTP is used between the GW and the original server to transfer content. The GW acts as a proxy server for the wireless domain. Its processor(s) provide services that free from the limited capabilities of the handheld, mobile, wireless terminals. For example, the GW provides DNS services, converts between WAP stack and the WWW stack (HTTP and TCP/IP), encodes information from the Web into a more compact form that minimizes wireless communication, and, in the other direction, decodes the compacted form into standard Web communication conventions. The GW also caches frequently requested information.

Using WAP, a mobile user can browse Web content on an ordinary Web server. The Web server provides content in the form of HTML-coded pages that are transmitted using the standard Web

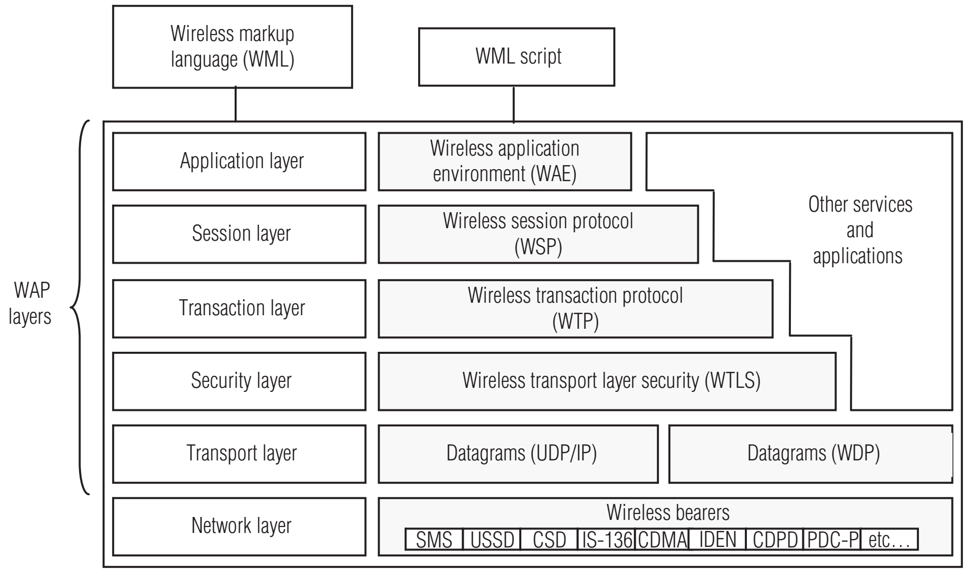


Figure 29.6 WAP architecture

protocol stack (HTTP/TCP/IP). The HTML content must go through an HTML filter, which may either be co-located with the WAP proxy or in a separate physical module. The filter translates the HTML content into WML content. If the filter is separate from the proxy, HTTP/TCP/IP is used to deliver the WML to the proxy.

The proxy converts the WML to a more compact form known as binary WML and delivers it to the mobile user over a wireless network using the WAP stack. If the Web server is capable of directly generating WML content, then the WML is delivered using HTTP/TCP/IP to the proxy, which converts the WML to binary WML and then delivers it to the MN using WAP. The layers in the WAP architecture are defined below.

- ***Application layer:*** This layer contains content developer programming languages, WML, WML script, device specifications, etc. which are of great interest to content developers.
- ***Session layer:*** WAP forum has designed wireless session protocol (WSP) to provide fast reconnection and connection suspension.
- ***Transaction layer*:** Like UDP, the wireless transaction protocol (WTP) runs at the top of the datagram services. The WTP which is part of the standard suit of TCP/IP protocols provides simplified protocol suitable for low-bandwidth wireless stations.
- ***Security layer:*** According to the transport layer security (TLS) protocol standard, wireless transport layer security (WTLS) incorporates security features. These features include data integrity checks, privacy, service denial, and authentication services.
- ***Transport layer:*** The WAP is made bearer-independent by adapting the transport layer of the underlying bearer which is the feature provided by wireless datagram protocol (WDP). The consistent data format is presented to the higher layers of the WAP stack by WDP. This offers the advantage to the bearer-independent application developers.

All these layers provide a well-defined interface to the layer above it, which means the internal working of each layer is transparent and visible to the layers above it. Through layered architecture the other applications and the services can utilize the features provided by the WAP-stack also.

29.3.2 WML scripts

WML script is the client-side scripting language of WML. A scripting language is similar to a programming language, but is of lighter weight. With WML script, the wireless device can do some of the processing and computation. This reduces the number of requests and responses to/from the server.

WML script components

WML script is very similar to Java script. Almost WML script components have similar meaning as they have in Java script. WML script program components are summarized as follows:

WML script operators

WML script supports the following type of operators:

- Arithmetic operators
- Comparison operators
- Logical (or relational) operators
- Assignment operators
- Conditional (or ternary) operators

WML script control statements

Control statements are used for controlling the sequence and iterations in a program (Table 29.1).

WML script functions

The user-defined functions are declared in a separate file having the extension .wmls. Functions are declared as follows:

```
function name (parameters)
{
control statements;
return var;
}
```

The functions used are stored in a separate file with the extension .wmls. The functions are called as the filename followed by a hash, followed by the function name.

Example: `maths.wmls#squar()`

In the example, `maths.wmls` is the file name and `squar()` is the function name.

Table 29.1 WML control statements

Statement	Description
if-else	Conditional branching
for	Making self-incremented fixed iteration loop
while	Making variable iteration loop
break	Terminates a loop
continue	Quit the current iteration of a loop

WML scripts standard libraries

In total, there are six standard libraries. Here is an overview of them:

- ***Lang***: The Lang library provides functions related to the WML script language core.
  Example function: `abs(), abort(), characterSet(), float(), isFloat(), isInt(), max(), isMax(), min(), minInt(), maxInt(), parseFloat(), parseInt(), random(), seed()`
- ***Float***: The Float library contains functions that help us perform floating-point arithmetic operations.
  Example function: `sqrt(), round(), pow(), ceil(), floor(), int(), maxFloat(), minFloat()`
- ***String***: The String library provides a number of functions that help us manipulate strings.
  Example function: `length(), charAt(), find(), replace(), trim(), compare(), format(), isEmpty(), squeeze(), toString(), elementAt(), elements(), insertAt(), removeAt(), replaceAt()`
- ***URL***: The URL library contains functions that help us manipulate URLs.
  Example function: `getPath(), getReferer(), getHost(), getBase(), escapeString(), isValid(), loadString(), resolve(), unescapeString(), getFragment()`
- ***WMLBrowser***: The WMLBrowser library provides a group of functions to control the WML browser or to get information from it.
  Example function: `go(), prev(), next(), getCurrentCard(), refresh(), getVar(), setVar()`
- ***Dialogs***: The Dialogs library contains the user interface functions.
  Example function: `prompt(), confirm(), alert()`

29.3.3 WAP services

WAP is looking for the online services that we are accustomed to today and that can be of interest in the wireless community as well. The key issue in successfully launching these services is *usefulness*. If the usefulness factor is not high enough, then the majority of users will just ignore the service. We also need to remember that the majority of the public is not very familiar even with the basic Internet services available today. However, some examples of useful mobile services are in the following fields:

Banking

- Accessing account statements
- Paying bills
- Transferring money between accounts

Finance

- Retrieving stock and share prices
- Buying and selling stocks and shares
- Looking up interest rates
- Looking up currency exchange rates

Shopping

- Buying everyday commodities
- Browsing and buying books
- Buying CDs

Ticketing

- Booking or buying airline tickets
- Buying concert tickets
- Booking theatre tickets

Entertainment

- Retrieving restaurant details
- Looking up clubs
- Finding out what is playing in what cinemas
- Playing solitaire games
- Playing interactive games

Weather

- Retrieving local weather forecasts
- Looking up weather at other locations

Advanced phonebook management

- Updating a personal phonebook
- Downloading a corporate phonebook

WAP also opens new possibilities for service and content providers, since they do not necessarily have to come to an agreement with a specific operator about providing services to their customers. This offers several benefits:

- You only need to create *a service once*, and it is then accessible on a broad range of wireless networks.
- You can build and address new market segments by creating innovative mobile value-added services.
- You can keep existing customers by adapting current Internet services to WAP.
- Creating a WAP service is no harder than creating an Internet service today, since WML and WML script are based on well-known Internet technology.
- You can continue to use standard tools like ASP or common gateway interface (CGI) to generate content dynamically.
- You can continue to utilize existing investments in databases and hardware that are the basis of existing Internet services.

29.3.4 Wireless session protocol

The WSP, a protocol in the WAP suite, provides the wireless application environment a consistent interface with the following two services:

- Connection-mode session services over a transaction service, to operate above the transaction layer protocol. This mode will be used for long-lived connections. A session state is maintained. There is reliability for data sent over a connection-mode session.
- Non-confirmed, connectionless services over a datagram transport service, which operates above secure or non-secure datagram service. This service is suitable when applications do not need reliable delivery of data and do not care about confirmation. It can be used without actually having established a session.

Basic functionality

At present, the protocols of the WSP family provide HTTP/1.1 functionality and semantics in a compact encoding, long-lived session state with session suspend and resume capabilities, a common facility for reliable and unreliable data push as well as a protocol feature negotiation. These protocols are optimized to be used in low-bandwidth bearer networks with relative long latency to connect a WAP client to a HTTP server.

WSP is part of a system offering Web capabilities to cell phones. The protocol offers connection-oriented services that are lacking in the transport layer of the WAP stack. Each session is given an ID so that several sessions can be maintained between the same client and server without the data being muddled between sessions.

Unlike most session management protocols, WSP allows a session to be suspended and resumed in the same place. This functionality requires the server to mark a save point in the session activities. WSP has a long timeout delay to account for typical long waits in establishing connections over low-bandwidth networks.

WSP communications are carried out over HTTP. Messaging is text-based and negotiates allowable HTTP extensions for the session at the point of session establishment.

29.3.5 Wireless transaction protocol

The WTP, a protocol in the WAP suite, operates efficiently over either secure or non-secure wireless datagram networks. It provides three different kinds of transaction services: unreliable one-way, reliable one-way, and reliable two-way transactions. This layer also includes optional user-to-user reliability by triggering the confirmation of each received message. To reduce the number of messages sent, the feature of delaying acknowledgements can be used.

WTP manages transactions by conveying requests and responses between a user agent (such as a WAP browser) and an application server for such activities as browsing and e-commerce transactions. WTP provides a reliable transport service but dispenses with much of the overhead of TCP, resulting in a light-weight protocol that is suitable for implementation in “thin” clients (e.g., MNs) and suitable for use over low-bandwidth wireless links.

WTP includes the following features:

- Three classes of transaction service
- Optional user-to-user reliability: WTP user triggers the confirmation of each received message
- Optional out-of-band data on acknowledgments
- PDU concatenation and delayed acknowledgment to reduce the number of messages sent
- Asynchronous transactions

WTP is transaction oriented rather than connection oriented. With WTP, there is no explicit connection setup or teardown, but rather a reliable connectionless service.

WTP transaction classes: WTP provides three transaction classes that may be invoked by WSP or another higher layer protocol.

- Class 0: Unreliable invoke message with no result message
- Class 1: Reliable invoke message with no result message
- Class 2: Unreliable invoke message with one reliable result message

Class 0: provides an unreliable datagram service, which can be used for an unreliable push operation. Data from a WTP user are encapsulated by WTP (e.g., the initiator or the client) in an invoke PDU and transmitted to the target WTP (e.g., the responder or the server), with no acknowledgment. The responder WTP delivers the data to the target WTP user.

Class 1: provides a reliable datagram service, which can be used for a reliable push operation. Data from an initiator are encapsulated in an invoke PDU and transmitted to the responder. The responder delivers the data to the target WTP user and acknowledges receipt of the data by sending back an ACK PDU to the WTP entity on the initiator side, which confirms the transaction to the source WTP user. The responder WTP maintains state information for some time after the ACK has been sent to handle possible retransmission of the ACK, if it gets lost and/or the initiator retransmits the invoke PDU.

Class 2: provides a request/response transaction service and supports the execution of multiple transactions during one WSP session. Data from an initiator are encapsulated in an invoke PDU and transmitted to the responder, which delivers the data to the target WTP user. The target WTP user prepares response data, which are handed down to the local WTP entity. The responder WTP entity sends these data back in a result PDU. If there is a delay in generating the response data beyond a timer threshold, the responder may send an ACK PDU before sending the result PDU. This prevents the initiator from unnecessarily retransmitting the invoke message.

29.3.6 Wireless datagram protocol

The WDP, a protocol in WAP architecture, covers the transmission layer protocols in an Internet model. As a general transport service, WDP offers to the upper layers an invisible interface independent of the underlying network technology used. In consequence of the interface common to transport protocols, the upper layer protocols of the WAP architecture can operate independent of the underlying wireless network. By letting only the transport layer deal with physical network-dependent issues, global interoperability can be acquired using mediating GWs. WDP architecture is shown in Figure 29.7.

WDP offers a consistent service at the transport service access point to the upper layer protocol of WAP. This consistency of service allows for applications to operate transparently over different available bearer services. The varying heights of each of the bearer services shown in Figure 29.7 illustrates the difference in functions provided by the bearers and thus the difference in WDP necessary to operate over those bearers to maintain the same service offering at the transport service access point is accomplished by a bearer adaptation. WDP can be mapped onto different bearers, with different characteristics. To optimize the protocol with respect to

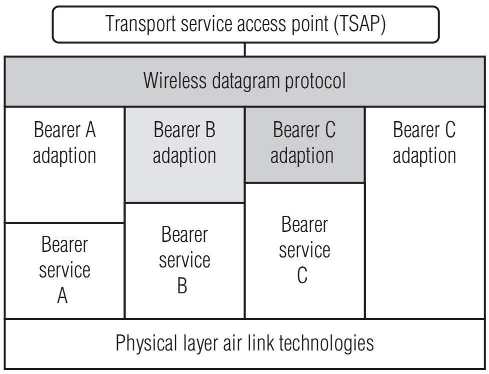


Figure 29.7 Wireless datagram protocol

memory usage and radio transmission efficiency, the protocol performance over each bearer may vary. However, the WDP service and service primitives will remain the same, providing a consistent interface to the higher layers.

The WDP layer operates above the data capable bearer services supported by the various network types. As a general datagram service, WDP offers a consistent service to the upper layer protocol (such as security, transaction, and session) of WAP and communicate transparently over one of the available bearer services. WDP supports several simultaneous communication instances from a higher layer over a single underlying WDP bearer service. The port number identifies the higher layer entity above WDP. This may be another protocol layer such as the WTP or the WSP or an application such as electronic mail. By reusing the elements of the underlying bearers, WDP can be implemented to support multiple bearers and yet be optimized for efficient operation within the limited resources of a mobile device.

29.4 Application layer

WAPs application layer is the wireless application environment (WAE). WAE directly supports WAP application development with WML instead of HTML, and WML script instead of Java script. WAE also includes the wireless telephony application interface (WTAI or WTA) that provides a programming interface to telephones for initiating calls, sending text messages, and other networking capability. The WAE is a general-purpose application environment based on a combination of WWW and mobile telephony technologies. The primary objective of the WAE effort is to establish an interoperable environment that will allow operators and service providers to build applications and services that can reach a wide variety of different wireless platforms in an efficient and useful manner. WAE includes a micro-browser environment containing the following functionality:

- ***Addressing model*:** Syntax suitable for naming resources stored on servers. WAP uses the same addressing model as the one used on the Internet, that is, uniform resource locators (URL).
- ***WML*:** A lightweight mark-up language designed to meet the constraints of a wireless environment with low bandwidth and small handheld devices. The WML is WAP's analogy to HTML used on the WWW. WML is based on XML.
- ***WML script*:** WML script is a light-weight scripting language. It is based on European Computer Manufacturers Association (ECMA) Script, the similar scripting language that Java script is based on. It can be used for enhancing services written in WML in the way that it, to some extent, adds intelligence to the services, for example, procedural logic, loops, conditional expressions, and computational functions.
- ***WTA or WTAI*:** A framework and programming interface for telephony services. The WTA environment provides a means to create telephony services using WAP.

29.4.1 WAP model

The WAP programming model illustrated in Figure 29.8 describes the following:

- The WAP client (the handheld device or WAP terminal)
- The WAP GW
- The Web server

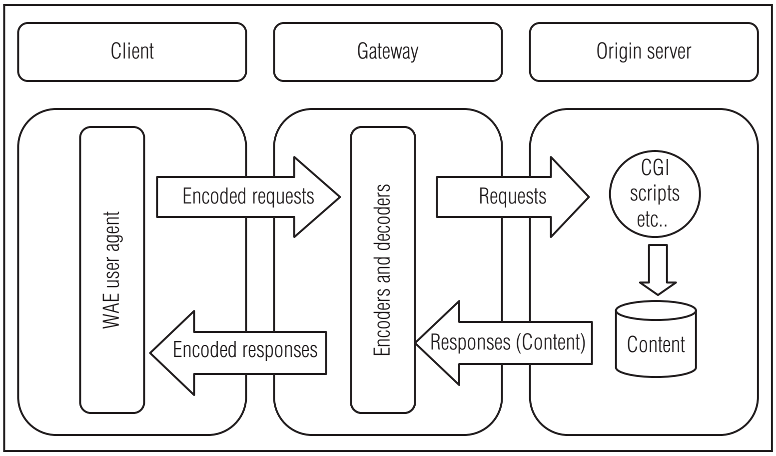


Figure 29.8 WAP programming model

WAP client

The client, also known as WAE user agent, is a component of the WAP terminal. It consists of a micro browser and the WAP stack to handle the execution of all requests and responses going through the WAP layered structure.

For example, this includes

- Session establishment
- Connectionless or connection-oriented data transport
- Setting up a secure environment including
  - Applying encryption and authentication
  - Encoding of outgoing requests
- Decoding of incoming responses to minimize bandwidth

Origin server

The client's micro browser requests WML pages. These WML pages are stored on the origin server, which might be a Web server, connected via the Internet or intranet. WML pages may also be stored in an application server installed in the GW itself. A WML page consists of a WML deck. One WML deck is divided into one or more WML cards. A WML card can be regarded as a unit of interaction. Services let the user navigate back and forth between cards from one or several WML pages. WML, especially designed for WAP terminals, provides a smaller set of mark-up tags than HTML. It is based on HTTP 1.1. WML decks may also contain WML scripts, another way of coding Web pages.

29.4.2 Mobile location-based service

In this age of significant telecommunications competition, mobile network operators continuously seek new and innovative ways to create differentiation and to increase profits. One of the best ways to accomplish this is through the delivery of highly personalized services. One of the most powerful ways to personalize mobile services is based on location. One of the most obvious technologies behind location-based service (LBS) is positioning, with the most widely recognized system being the global positioning system (GPS). There are however, other means of positioning in addition to GPS. These other technologies are network-based positioning and typically rely on various means of triangulation of the signal from cell sites serving a mobile phone. In addition, the serving cell site can be used as a fix for location of the user.

Geographical data are an important aspect of any location system. Geographic information systems (GIS) provide the tools to provision and administer base map data such as man-made structures (such as streets and buildings) and terrain (such as mountains and rivers). GIS is also used to manage point-of-interest data such as location of gas stations, restaurants, and nightclubs. Finally, GIS information also includes information about the radio frequency characteristics of the mobile network. This allows the system to determine the serving cell site of the user. It is not enough to be able to position the mobile user and know the map data around that position. There must be a location management function to process positioning and GIS data on behalf of LBS applications. The location management function acts as a GW and mediator between positioning equipment and LBS infrastructure.

Location-based information

Many people are familiar with wireless Internet, but many do not realize the value and potential to make information services highly personalized. One of the best ways to personalize information services is to enable them to be location based. An example would be someone using their WAP-based phone to search for a restaurant. The LBS application would interact with other location technology components to determine the user’s location and provide a list of restaurants within a certain proximity to the mobile user.

Location-based billing

The ability to have preferential billing is provided by this type of application. Through location-based billing, the user can establish personal zones such as a home zone or work zone. Through arrangements with the serving wireless carrier, the user could perhaps enjoy flat-rate calling while in the home area and special rates while in other defined zones. This type of application can be especially useful when used in conjunction with other mobile applications such as prepaid wireless.

Tracking

This is a rather large category that contains everything from the difficult fleet applications to enabling mobile commerce. Fleet applications typically entail tracking vehicles for purposes of the owning company knowing the whereabouts of the vehicle and/or operator. Tracking is also an enable of mobile commerce services. A mobile user could be tracked and provided information that he or she has pre-determined his or her desires, such as notification of a sale on men’s suits at a store close to the user’s current proximity.

29.4.3 WAP gateway

The GW works as a WAP proxy. It receives requests from the client, transforms an HTTP message, and sends it (i.e., based on the URL) to the addressed Web server. When the Web server returns the response, the GW transforms it again into a bit-coded output. It then sends it to the mobile network, which directs it to the WAP client. This method allows the data content and applications to be hosted in standard Web servers using traditional Web technology.

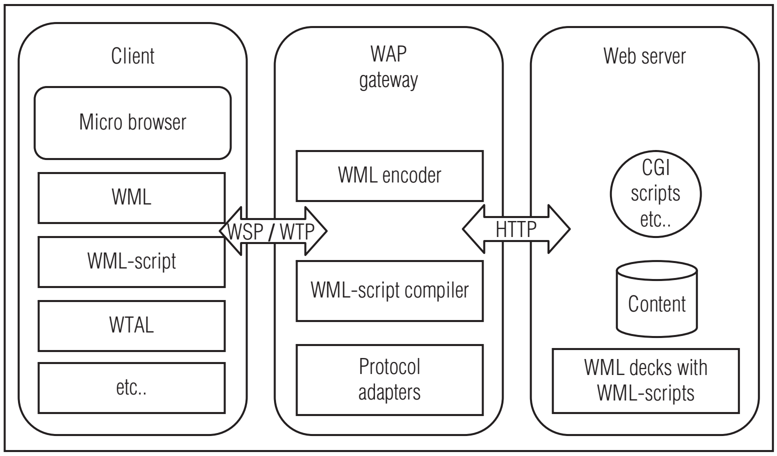


Figure 29.9 Usage of a WAP gateway

The WAP GW is also able to work as a WAP application server. This solution does not need a Web server. It can be used to support end-to-end security configurations. This could be for applications which require higher access control or a guarantee of service, like those used for WTA.

The WAP GW decreases the response time to the WAP terminal by aggregating data from different Web servers and caching frequently used information. It also divides large HTTP responses into smaller transmission units before sending the responses to the client.

The WAP GW can also interface with subscriber databases and use information to dynamically customize WML pages for a certain group of users.

The WAP GW provides transition between the Internet and different non-voice mobile services. These might be services such as short message services (SMS), circuit-switched data (CSD), or general packet radio services (GPRS).

Usage of a WAP GW

This type of configuration, illustrated in Figure 29.9, consists of the following two connections:

- The first connection is between the client and a WAP GW.
- The second connection is between the WAP GW and the Web server.

While the first connection uses the WSP and the WTP, the second connection uses the traditional HTTP, which runs above TCP. The proxy GW maintains the two connections as one logical connection between the client and the Web server.

Client–server flow

The client–server flow is as follows:

- A client asking for a particular service from an origin server submits a request to this server using the WML user agent.
- The WML user agent uses the URL addressing scheme operation to find the origin server. URLs are used to address applications on a Web server, for example, CGI.

- The client's request is transmitted to the GW using WSP/WTP.
- The GW does the bit-encoding, transforms the request into an HTTP message, and sends it to the Web server (addressed by the URL).
- The origin server replies to the request by returning a single deck, if it is stored in textual format in the origin server.
- The deck is transmitted to the GW.
- On its way through the GW, the textual format of each deck has to be converted into a format that is better suited for over-the-air transmission to WAP terminals. This conversion is done by the GW using the WML encoder to convert the textual format into a binary format.
- The GW sends the encoded content to the client via the WAP stack layers over the wireless network.
- At the client's WAP terminal, the data are received and can be displayed and interpreted, for example, by a micro browser.

29.4.4 WAP user agent profile

Existing mark-up languages and content written in those mark-up languages presume that devices have similar display sizes, memory capacities, and software capabilities. Content is also largely oblivious to the available network bandwidth and perceived network latency. Moreover, users may have content presentation preferences that also cannot be transferred to the server for consideration.

Recently, work has begun within the WWW consortium (W3C) to define mechanisms for describing and transmitting information about the capabilities of Web clients and the display preferences of Web users. The composite capabilities/preferences profile (CC/PP) note defines a high-level structured framework for describing this information using the resource description framework. CC/PP profiles are structured into named "components" each containing a collection of attribute-value pairs, or properties. Each component may optionally provide a default description block containing either a set of default values for attributes of that component or a URI that refers to a document containing those default values.

End-to-end architecture of user agent profile

This specification provides for the end-to-end specification, delivery, and processing of composite capability information from the device. The information is collected on the client device, encoded into an efficient binary form, transmitted, and cached within a WSP session, optionally enhanced with information provided with a particular request, optionally combined with other information available over the network, and made available to the origin server. Over the Internet, this specification assumes the use of the CC/PP, CC/PP exchange protocol over HTTP, and HTTP with the HTTP extension framework. The user agent profile end-to-end systems are shown in Figure 29.10.

The end-to-end system consists of five logical components:

- A client device capable of requesting and rendering WAP content.
- A wireless network employing WAP 1.1 or later protocols.
- A WAP-capable GW capable of translating WAP requests into corresponding requests over the Internet and translating responses from the Internet into corresponding responses over the WAPs.
- The Internet or an intranet using TCP/IP-based protocols and possibly having one or more protocol GWs and Web/HTTP proxies.

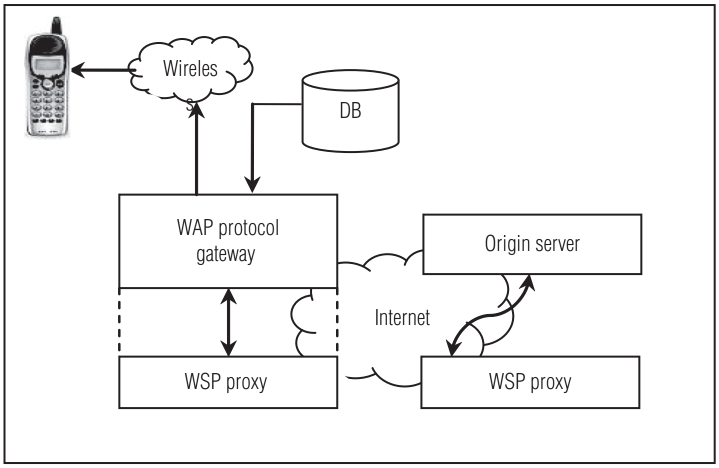


Figure 29.10 User agent profile end-to-end systems

- An origin (Web) server that can generate requested content. Though this specification refers to five end-to-end system components, actual configurations may physically deploy those components in many forms. For example, the latter three components (i.e., WAP GW, Internet/intranet, and origin server) might easily be merged into a single server-side system connected to the WAP network. Moreover, the WAP GW may itself be distributed, with different hosts serving as endpoints for different layers of the WAP stack.

29.4.5 Caching model

The WAP user agent caching model tailors the HTTP caching model to support WAP handsets with limited functions. For cached resources that will not be changed during user retrievals, the resources can be efficiently accessed by the WAP handsets without revalidation. A time-sensitive cached resource is set to "must-revalidate." If this cached resource is stale when the user tries to go back in the history, the user agent revalidates this cached source. In general, navigation and processing within a single cached resource does not require revalidation, except for the first fetch. Examples include function calls within a single WML script compilation unit and intra-deck navigation within a single WML deck.

The HTTP caching model is sensitive to time synchronization. Since WAP follows this model, a reliable time-of-day clock should be maintained in the WAP GW. If a WAP user agent does not have access to a time-of-day clock, it should exchange the time-of-day request and response message with the WAP GW and synchronize with the clock value returned from the WAP GW.

Another important issue for caching is security. The private information in the user agent cache is protected from unintended or malicious access. WAP GWs implementing a caching function must obey all security related considerations defined in HTTP.

Several studies indicate that the hit rate of caching can be more than 50 per cent, which significantly reduces the utilized network resources. Complementary techniques, such as push and pre-fetching, can be used to speed up web access. Pre-fetching is based on the fact that after retrieving a page from the origin server, the WAP handset, as well as the air link of the wireless network, is idle, providing time that can be used to pre-fetch the next page so that when the user proceeds to retrieve the information, the page is available immediately. In this way, transmission delay can be reduced. However, if the pre-fetched data are not used by the user, the network resources used by pre-fetching are wasted. Furthermore, it is difficult to change the user if the operator does not know whether the pre-fetched data are actually used.

29.4.6 Wireless bearers for WAP

The WAP stack is built on top of wireless bearer services. The WAP bearers are as follows:

- Short message service (SMS)
- Circuit-switched data (CSD)
- Unstructured supplementary services data (USSD)
- General packet radio service (GPRS)

Short message service

Given its limited length of 160 characters per short message, SMS may not be an adequate bearer for WAP because of the weight protocol of the protocol. The overhead of the WAP that would be required to be transmitted in an SMS message would mean that even for the simplest of transactions several SMS messages may in fact have to be sent. This means that using SMS as a bearer can be a time consuming and expensive exercise. Only one network operator – SBC of the United States – is known to be developing WAP services based on SMS.

Circuit switched data

Most of the trial WAP-based services use CSD as the underlying bearer. Since CSD has relatively few users currently, WAP could kick-start the usage of and traffic generated by this bearer.

However, CSD lacks immediacy – a dial-up connection taking about 10 s is required to connect the WAP client to the WAP GW, and this is the best case scenario when there is a complete end-to-end digital call. In the case of the need for analogue modem handshaking (because the WAP phone does not support the digital protocol V.110 or the WAP GW does not have a digital direct connection such as ISDN into the mobile network), the connect time is increased to about 30 s.

Unstructured supplementary services data

USSD is a means of transmitting information or instructions over a GSM network. USSD has some similarities with SMS since both use the GSM network's signalling path. Unlike SMS, USSD is not a store and forward service and is session-oriented such that when a user accesses a USSD service, a session is established and the radio connection stays open until the user, application, or time out releases it. This has more in common with CSD than SMS. USSD text messages can be up to 182 characters in length.

USSD has some advantages and disadvantages as a tool for deploying services on mobile networks:

- Turnaround response times for interactive applications are shorter for USSD than SMS because of the session-based feature of USSD, and because it is NOT a store and forward service. According to Nokia, USSD can be up to seven times faster than SMS to carry out the same two-way transaction.
- Users do not need to access any particular phone menu to access services with USSD – they can enter the USSD command direct from the initial mobile phone screen.
- Because USSD commands are routed back to the home mobile network's home location register, services based on USSD work just as well and in exactly the same way when users are roaming.

- USSD works on all existing GSM mobile phones.
- Both SIM application toolkit and the WAP support USSD.
- USSD Stage 2 has been incorporated into the GSM standard. Whereas USSD was previously a one-way bearer useful for administrative purposes such as service access. Stage 2 is more advanced and interactive. By sending in a USSD2 command, the user can receive an information services menu. As such, USSD Stage 2 provides WAP-like features on EXISTING phones.
- USSD strings are typically complicated for the user to remember, involving the use of the "*" and "#" characters to denote the start and finish of the USSD string. However, USSD strings for regularly used services can be stored in the phonebook, reducing the need to remember and re-enter them.

As such, USSD could be an ideal bearer for WAP on GSM networks.

General packet radio service

The GPRS is a new packet-based bearer that is being introduced on many GSM and TDMA mobile networks from the year 2000 onwards. It is an exciting new bearer because it is immediate (there is no dial-up connection), relatively fast (up to 177.2 kbps in the very best theoretical extreme), and supports virtual connectivity, allowing relevant information to be sent from the network as and when it is generated.

At the time of writing in early August 1999, there has been no confirmation from any handset vendors that mobile terminated GPRS traffic (i.e., direct receipt of GPRS packets on the mobile phone) will be supported by the initial GPRS terminals. Availability or not of GPRS MT is a central question with critical impact on the GPRS business case such as application migration from other non-voice bearers.

There are two efficient means of delivering proactively sending ("pushing") content to a mobile phone: (i) by the SMS, which is of course one of the WAP bearers or (ii) by the user maintaining more or less a permanent GPRS (mobile originated) session with the content server.

However, the mobile terminated IP traffic may allow the information to reach the terminal which was not requested. The internet sources originating such unrequested information may not be chargeable. But in a worst scenario, the mobile user may have to pay for such information. This is the major reason that the mobile vendors are not supporting GPRS mobile terminate in their GPRS terminal. However, by originating the session by themselves from their handset, users agree to pay for the delivery of content from that service. Users can make their request via a WAP session, which need not be blocked. As such, a WAP session initiated from the WAP micro browser will be the only way that GPRS users can receive information onto their mobile terminals.

Except the early WAP-enabled phones all others support GPRS, and WAP and GPRS could well be synergistic and be used widely together. For the interactive, menu-based information exchanges that WAP anticipates, CSD is not immediate because of the need to set up a call. Early prototypes of WAP services based on CSD were therefore not usable. On the other hand, SMS is immediate but always stores and forwards, such that even when a subscriber has just requested information from their micro browser, the SMS centre resources are used in the information transfer. As such, GPRS and WAP are ideal for each other.

WAP incorporates two different connection modes – WSP connection mode or WSP connectionless protocol. It is similar to the two GPRS point-to-point services – - connection oriented and connectionless services.

The predominant bearer for WAP-based services will depend on delays in the availability of WAP handsets and GPRS terminals. If WAP terminals are delayed until the year 2000, most WAP

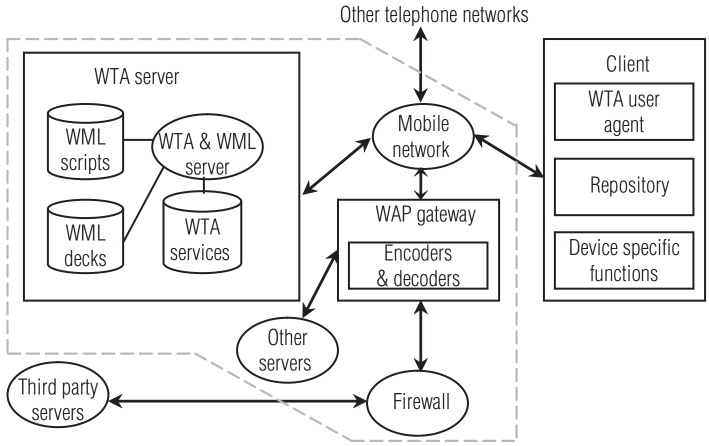


Figure 29.11 Wireless telephony networks

terminals will support GPRS as well. If the first WAP terminals support SMS and CSD, but not GPRS, then SMS could become the predominant initial WAP bearer.

WAP certainly will be important for the development of GPRS-based applications. Because the bearer level is separated from the application layer in the WAP stack, WAP provides the ideal, defined, and standardized means to port the same application to different bearers. As such, many application developers will use WAP to facilitate the migration of their applications across bearers once GPRS-based WAP are supported.

29.4.7 Wireless telephony application

The WAP WTAI (WTA application programming interface) features provide the means to create telephony applications, using a WTA user agent with the appropriate WTAI function libraries. A typical example is to setup a mobile originated call using the WTAI functions accessible from either a WML deck/card or WML script. The application model for WTA is shown in Figure 29.11 and is based on a WTA user agent, executing WML and WML script. The WTA user agent uses the WTAI function libraries to make function calls related to network services. The WTA user agent is able to receive WTA events from the mobile network and pushed content, like WML decks and WTA events, from the WTA server. WTA events and WTAI functions make it possible to interact and handle resources, for call control, and much more, in the mobile network. The WTA server can invoke applications dynamically using content push with WML and WML script.

WTA provides tools for building telephony applications. It is primarily designed for network operators, carriers, and equipment vendors. Network security and reliability is a major consideration. Extensions are added to the standard WML/WML script browser to support an additional WTA application programming interface (WTAI). WTAI includes the following functions: call control, network text messaging, phonebook interface, indicator control, and event processing.

i-Mode

WAP uses a special language known as WML for communication between a special protocol conversion device called a WAP GW and content on the Internet. The WAP GW converts between WML and HTML; this allows delivery of WAP-based content to a WAP-capable mobile device.

As explained above, the MSC must utilize the public-switched telecommunications network to connect to the GW. Similarly, in most of the networks today, the connection between the MSC and the GW is circuit switched. Since mobile network operators deploy next generation packet-data technologies such as GPRS, the connection between the MSC and the WAP GW will be upgraded to leverage the faster packet connection facilitated by the GPRS network. The i-mode utilizes an overlay packet network for direct communications (no GW needed) to the content providers on the Internet, which is in contrast to WAP.

i-Mode is the packet-based service for mobile phones offered by NTT DOCOMO, Japan's leader in wireless technology. i-Mode is built on a firm foundation of advanced technology. The use of packet transmissions offers continuous access, while the use of a subset of HTML makes content creation easy and provides simple conversion of existing websites. NTT DOCOMO's packet-switched technology was specifically designed to provide users with "always-on" network access, eliminating the need to log on or off. In turn, i-mode the service developed for this packet-switched technology provides the most efficient wireless access possible as no dedicated radio channel is required. The result is lower costs for customers as they are billed according to the volume of data sent and received, rather than by time spent connected. i-Mode offers another significant benefit, that is, mobile voice and data service in a single convenient package. This voice/data flexibility is unprecedented. People can download information about events, restaurants, etc., and then make reservations, or place calls to numbers searched in the i-mode directory.

SyncML

SyncML is the leading open industry standard for universal synchronization of remote data and personal information across multiple networks, platforms, and devices. The SyncML initiative recently consolidated into the open mobile alliance (OMA), contributing their technical work to the OMA technical working groups: device management working group and data synchronization working group. SyncML's goal is to have networked data that support synchronization with any mobile device, and mobile devices that support synchronization with any networked data. SyncML is intended to work on transport protocols as diverse as HTTP, WSP (part of WAP), and OBEX, and with data formats ranging from personal data (e.g., vCard and vCalendar) to relational data and XML documents.

The ability of applications running on handheld devices, on desktop computers, and in networks to update a single body of information, and to coordinate the updates intelligently is the key to the popularity of the paradigm of ubiquitous computing. The process of making two or more sets of data identical is called data synchronization.

To synchronize personal information on your handheld device and your PC you probably use the proprietary technology that comes with the device – and use a different technology to synchronize data on your phone with data on your laptop. Each technology can synchronize one or a few applications, and is limited to a particular type of device or network connectivity. Users waste huge amounts of time and money configuring and installing all these different applications. Given the large and growing diversity of applications and devices we use today, we need a standard synchronization technology. The SyncML initiative is such a standard, which promises to enable users to buy devices that synchronize with a wide range of data and devices, and to reduce the effort and costs expended by device manufacturers, service providers, and application developers as well.

SyncML is an open standard that drives data mobility by establishing a common language for communication among devices, applications, and networks. The goal is to enable smooth, efficient synchronization of remote data and personal information across devices, platforms, and multiple networks, both fixed and mobile. This universal language enables devices and applications to synchronize data like email, contact information, calendar, and to-do lists over the network, so that information is consistent, up-to-date, and accessible, no matter where it is stored: on a mobile phone, a PDA, a laptop, a PC, or a server.

Benefits of SyncML

SyncML represents a common synchronization standard that benefits all kinds of users in the mobility industry:

- ***Device manufacturers:*** Because storage on mobile devices is limited, manufacturers can support only one synchronization standard. Adopting a common standard enables the device to interoperate with a wide range of applications, services, and network transmission technologies.
- ***Service providers:*** If multiple synchronization technologies are widely used, service providers must install and configure multiple server infrastructures, one for each synchronization technology, and must update and maintain each to preserve compatibility and performance. Using a common synchronization standard reduces maintenance effort and operating costs.
- ***Application developers:*** Developers who adopt SyncML can create applications that connect to broad varieties of devices and networked data. Applications that do not need to support multiple synchronization technologies are less complex and less costly to maintain. Deployment is easier, too – a feature likely to appeal to service providers deciding which of similar applications to offer.
- ***End users:*** If platforms and applications use a standard synchronization technology, users will be able to synchronize a wide range of data and devices without spending all those hours configuring multiple, incompatible connections.

29.5 Summary

- Mobile IP uses a tunnelling protocol to allow messages from the PDN to be directed to the MN's IP address.
- The HA and FA continuously advertise their services on the network through an agent discovery process, enabling the HA to recognize when a new FA is acquired and allowing the MN to register a new CoA.
- A mobile host (MH) is an Internet device, such as a notebook or a PDA, that could potentially be moving from one domain to another and should thus be equipped with mobile IP.
- The CoA is an IP address for the FA. This FA will receive datagrams at this CoA, intended for the MN, and then forwards them across the foreign network to the MN.
- Agent discovery is a type of mobility-detection mechanism. This process consists of the MAs broadcasting messages (agent advertisement) of their presence using one of the ICMP message types called router discovery.
- Tunnelling is used by HA to send IP datagrams to the MH when the MH is in a foreign domain. Tunnelling implies the use of double protocol headers, in this case, an IP over an IP.
- To encapsulate an IP datagram using IP in IP encapsulation, an outer IP header is inserted before the datagram's existing IP header.
- The WAP is a universal, open standard developed by the WAP forum to provide mobile users of wireless phones and other wireless terminals such as pagers and PDAs access to telephony and information services, including the Internet and the Web.

- A WAP GW is a server through which all wireless (WAP) data are transferred from wireless devices (using WAP requests) to content sites (in WML format) and back again.
- The user agent profile (UAProf) specification extends WAP 1.1 to enable the end-to-end flow of a UAProf, also referred to as capability and preference information, between the WAP client, the intermediate network points, and the origin server.
- WAP uses a special language called WML for communication between a special protocol conversion device called a WAP GW and content on the Internet.

Review questions

1. What are mobile IP components? Define each of them.
2. Write short notes on co-located addresses.
3. Explain WAP architecture.
4. Explain WML scripts and its components.
5. Write short notes on wireless transaction protocol.

Objective questions and answers

1. If the MH has a permanent address in the home domain, it is called
   (a) mobile address (b) home address (c) node address (d) none
2. An Internet host device with which MH is communicating during an instance of a mobile IP connection is called
   (a) correspondent host (b) mobile host (c) mobile HA (d) none
3. DHCP stands for
   (a) dynamic host configuration protocol (b) dual host configuration protocol
   (c) dual host common protocol (d) dynamic host common protocol
4. Tunnelling is used by HA
   (a) to stop IP datagrams from the MH (b) to send IP datagrams to the MH
   (c) both (a) and (b) (d) none
5. The WAP specification includes
   (a) a programming model based on the WWW programming model
   (b) a mark-up language, the wireless markup language, adhering to XML
   (c) a specification of a small browser suitable for a mobile, wireless terminal
   (d) all
6. The WDP allows WAP to be bearer-independent by adapting the
   (a) application layer (b) session layer (c) transport layer (d) security layer
7. WML script supports
   (a) arithmetic operators (b) comparison operators
   (c) logical (or relational) operators (d) all
8. The WDP covers the following transmission layer protocols
   (a) application layer protocols (b) session layer protocols
   (c) transport layer protocols (d) transmission layer protocols
9. A WAP GW is a server through which all wireless (WAP) data are transferred from
   (a) wireless devices to content sites (b) content sites to wireless devices
   (c) both (a) and (b) (d) none

10. The end-to-end system consists of
    (a) a client device capable of requesting and rendering WAP content
    (b) a wireless network employing WAP 1.1 or later protocols
    (c) an origin (Web) server that can generate requested content
    (d) all

Answers: 1. (b), 2. (a), 3. (a), 4. (c), 5. (d), 6. (c), 7. (d), 8. (d), 9. (c), 10. (d).

Open book questions

1. Define the terms registration, de-registration, and CoA.
2. Write short notes on WSP.

Further reading

Aftab Ahmad, *Wireless and Mobile Data Networks*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

Rappaport, T. S., *Wireless Communications*, Virginia Polytechnic University Mobile & Portable Research Group, Pearson Education.

Stallings, W., *Wireless Communications and Networks*, Newyork, Mac Millan Technical publishing, Pearson Education.

Mobile Data Networks 30

30.1 Introduction

The increase and development of cellular voice systems over the past several years has exposed the capabilities and the effectiveness of wireless communications and thus, has paved the way for wide-area wireless data applications as well. The demand for such applications is currently experiencing a significant increase and therefore, there is a strong call for advanced and efficient mobile data technologies. This chapter deals with these mobile data technologies and aims to exhibit their potential. It provides details of the most important mobile packet data services and technologies, including cellular digital packet data (CDPD), short messaging service (SMS) in Global System for Mobile (GSM) communication technology, and the emerging general packet radio service (GPRS). For each technology, this chapter outlines its main technical characteristics, discusses its architectural aspects, and explains the medium access protocol, the services provided, and the mobile routing scheme.

30.2 Data-oriented CDPD network

CDPD is a mobile data technology that permits subordinate data operation on the spectrum assigned to the advanced mobile phone service (AMPS). It was first introduced by IBM as a packet-switching overlay to the existing analogue cellular voice network and frequencies. Later, a CDPD system specification was created by a consortium of cellular carriers including AirTouch, McCaw Cellular, Southwestern Bell Mobile Systems, NYNEX, Ameritech, GTE, Bell Atlantic Mobile, and Contel Cellular.

CDPD technology is being deployed by a number of cellular companies in the United States, including Bell Atlantic, Ameritech, GTE, and McCaw Cellular. This equipment is provided by a variety of manufacturers. CDPD systems are designed to take advantage of the idle voice channels of an analogue cellular network, such as AMPS. These idle channels are used to transmit short data messages and establish a packet-switching service. In order to utilize these idle channels, CDPD implements a hopping procedure among the available cellular frequencies. The air interface operates at a raw data rate of 19.2 Kbps and provides forward error correction to combat the interference and fading of the cellular channels.

The Wireless Data Forum is an industry association that handles the shaping of the CDPD technology and supports the growth of the commercial market place. This forum aims to help the operators, equipment providers, and billing system developers through the complicated issues they face. Among other things, the forum tries to help all CDPD operators to develop interoperated roaming and invoicing. According to the Wireless Data Forum, by the end of the third quarter of 1998, CDPD was available in 195 markets in the United States: 118 metropolitan statistical areas (MSAs), 41 rural statistical areas (RSAs), and 36 international markets and was available to 53 per cent of the U.S. population.

30.2.1 System description

The primary elements of a CDPD network are the end systems (ESs) and the intermediate systems (ISs), as shown in Figure 30.1. The ESs represent the actual physical and logical end nodes that exchange information, while the ISs represent the CDPD infrastructure that stores, forwards, and routes the information. There are two kinds of ESs: The mobile end system (M-ES), which is a device used by a subscriber to access the CDPD network over the wireless interface, and the fixed end system (F-ES), which is a common host, server, or gateway that supports or provides access to data and applications. By definition, the location of an F-ES is fixed, whereas the location of an M-ES may change. Typically, each M-ES consists of a mobile terminal (personal computer, personal digital assistant, or other standard device), and an additional device, the radio modem, that attaches to the mobile terminal and manages the radio links and protocols. These devices usually communicate over standard serial protocols, such as the serial line Internet protocol (SLIP), or the point-to-point protocol (PPP).

On the other hand, there are two kinds of ISs: a "Generic" IS, which is simply a (IP) router that has no knowledge of CDPD and mobility issues, and a mobile data intermediate system (MD-IS), which is a specialized IS that routes messages based on its knowledge of the current location of an M-ES. More specifically, an MD-IS is a set of hardware components and software functions that provide switching, accounting, registration, authentication, encryption, and mobility

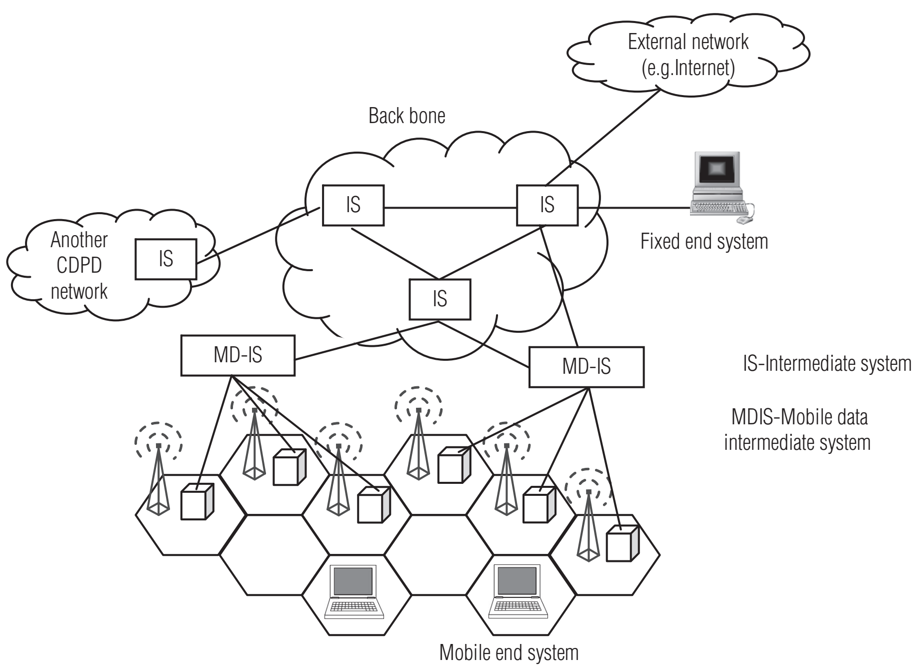


Figure 30.1 CDPD network architecture

management functions. The mobility management software allows the switching system to track M-ESs regardless of their location in the network, and allows M-ESs to use a single network address. The CDPD mobility management software follows the mobile-IP model, established by the Internet Engineering Task Force (IETF). Besides the ESs and the ISs, there is also another element called the mobile data base station (MDBS), which is analogous to the AMPS base station (BS). An MDBS is a combination of a computer, power amplifiers, and a radio transceiver. It performs no networking functions but it is a link-layer relay; it sends and receives information from the M-ESs and relays it back to the MD-IS. It also controls the radio interface and manages the radio communications and monitors the activity on the voice network (to ensure that data and voice do not interfere with each other). The MDBS creates an air link comprising two RF channels for forward and reverse communications with multiple M-ESs.

CDPD interfacing

Figure 30.2 shows the standardized interfaces used across the CDPD network. The M-ESs are connected to the CDPD network through the A-interface (the air interface), while the F-ESs are connected through the E-interface. The E-interface (external to the CDPD provider) is also used to interconnect with external networks. Finally, the I-interface (internal to the CDPD provider) is used in the backbone, between the various ISs, and at the interconnection points with other CDPD networks.

The CDPD backbone provides connectionless transport services, also called “datagram” services. This means that the network individually routes packets, based on the destination address of the packet and on the knowledge of the current network topology. For routing of packets, CDPD supports both the Internet protocol (IP) and the connectionless network protocol (CLNP), which is an OSI standard protocol. The CDPD system is designed to facilitate the interoperability between the networks of different service providers. Especially in CDPD, where there may be many different operators in a given geographical area (since there may also be many different cellular operators, each one deploying its own CDPD infrastructure), interoperability is a main concern. Interoperability means that all the subscriber equipment, components, functions, and processes of the CDPD system work together within all the cooperating CDPD networks.

Interoperability allows for seamless roaming inside a large geographical area and provides the means for ever-present CDPD service. In fact, the CDPD network is envisioned as an internetwork composed of multiple administrative domains, each one operated by a different service provider as shown in Figure 30.3.

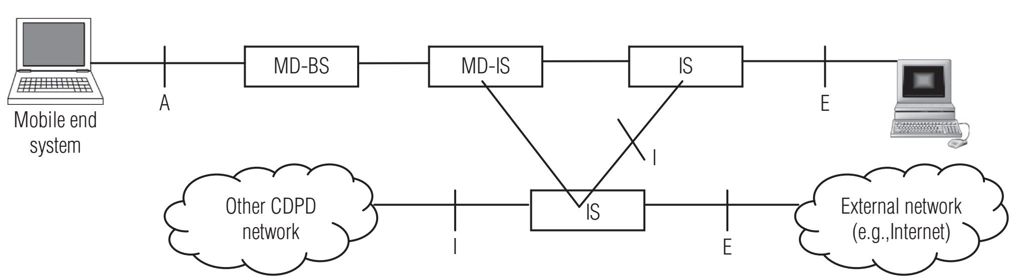


Figure 30.2 Standardized interfaces within CDPD architecture

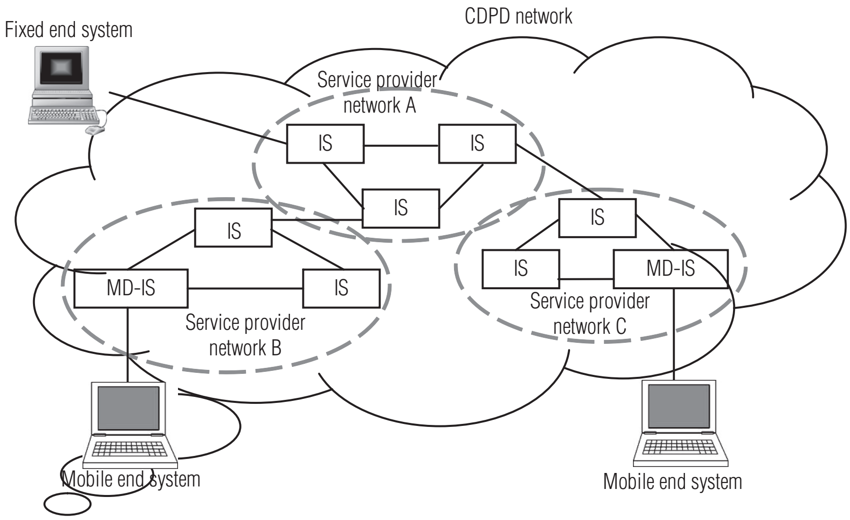


Figure 30.3 CDPD service with interoperability between service providers

30.2.2 Protocol architecture

The protocol architecture of the CDPD air interface is illustrated in Figure 30.4.

Physical layer

The physical layer (PHY) in CDPD corresponds to a functional entity that accepts a sequence of bits from the medium access control (MAC) layer and transforms them into a modulated

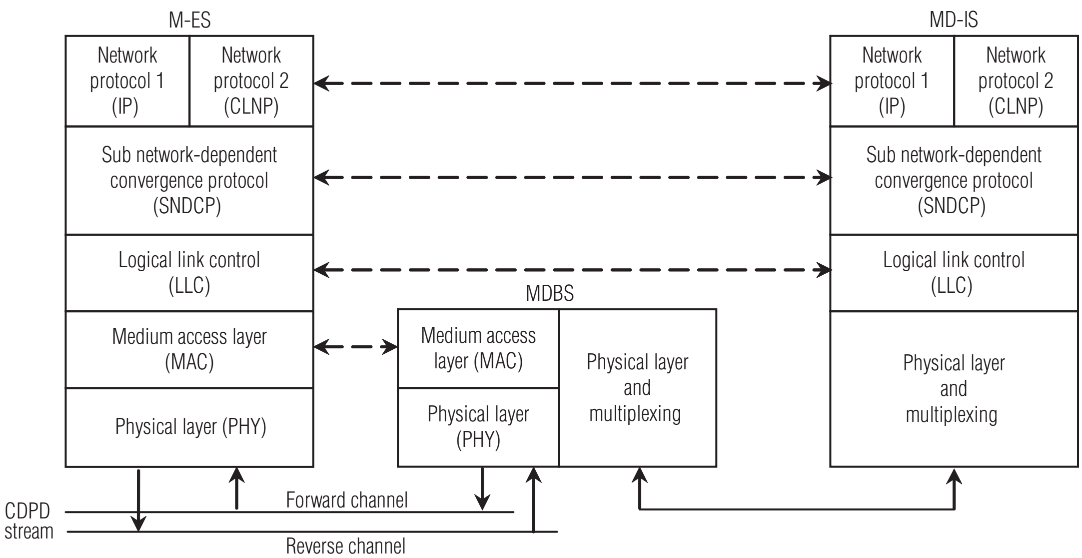


Figure 30.4 Protocol architecture of CDPD air interface

waveform for transmission onto a physical 30-kHz RF channel. As illustrated in Figure 30.4, communication between an MDBS and an M-ES takes place over a pair of such RF channels (having a fixed frequency separation).

There are two channels in CDPD channel stream:

- The first channel, called the forward channel, accommodates transmissions in the direction from the MDBS to the M-ESs and is either dedicated to CDPD use or shared with the voice cellular network. In any case, transmission on the forward channel is continuous as long as it is in use for CDPD.
- The second channel, called the reverse channel, accommodates transmissions in the direction from the M-ESs to the MDBS and is shared among all M-ESs communicating with the same MDBS. A pair of associated reverse and forward channels forms a CDPD channel stream. The PHY interfaces with another entity, the radio resource management entity (RRME).

Through this control interface, the RRME can perform the following functions:

- Tune the physical layer to a specific RF channel pair.
- Set the transmission power level to the desired value.
- Measure the received signal level of a RF channel and estimate its potential to offer acceptable communication.
- Suspend and resume the operation of the physical layer in cases where power saving facilities is required.

A frequency greater than the central carrier frequency represents a logical '1', while a logical '0' is represented by a frequency less than the central carrier frequency. The modulation rate on both the forward and reverse RF channels is 19.2 Kbps.

Medium access control layer

The medium access control (MAC) layer models a functional entity logically operating between the physical and logical link control (LLC) layers. The MAC layer within an M-ES cooperates with the corresponding MAC layer within the MDBS. The purpose of this layer is to convey information, namely, link protocol data units (LPDUs) between peer LLC entities across the CDPD air interface.

For this purpose, the MAC layer provides the following services:

- Encapsulates LPDUs into frame structures to ensure LPDU delimiting, frame synchronization, and data transparency.
- Encodes LPDUs to provide error protection against mobile channel impairments.
- Detects and corrects bit errors within received frames.
- Arbitrates access to the shared reverse channel.
- Synchronizes with the forward channel transmissions to make feasible for the reception of data as well as control information transmitted in every CDPD cell.

Logical link control layer

The purpose of the LLC layer is to convey information between network layer entities across the CDPD air interface. The protocol applied in this layer is called the mobile data link protocol (MDLP). As illustrated in Figure 30.4, the MDLP in an M-ES communicates with a peer MDLP located in its serving MD-IS. Hence, it is seen that the functionality of an MDBS is restricted to the physical and MAC layers. Above the MAC layer, an MDBS is completely transparent.

The primary service offered by the MDLP to the upper layer sub-network dependent convergence protocol (SNDCP) is the provision and control of one or more logical data link connections on a CDPD channel stream. Above the LLC layer, these data link connections are treated as individual bit pipes that may be used to convey messages back and forth between an MD-IS and one or more M-ESs. Within each data link connection, one or more network traffic flows may be accommodated through facilities provided by the SNDCP. Discrimination between data link connections is made by means of an address label contained in each message (frame). This address label is called the temporary equipment identifier (TEI) and is a pure LLC layer concept, that is, it is used internally by the LLC layer and is not necessarily known by other functional layers.

Sub-network-dependent convergence protocol

Functionally, the SNDCP lies between the data link layer and the network layer. The latter is assumed to be sub-network independent, that is, it is built to work over virtually any data link and therefore, it does not take into account the specific features of the MDLP. For this reason, the services assumed by the network protocol(s) may not map directly into the services provided by the MDLP. In this case, the SNDCP is operated to provide the required cooperation.

More specifically, the SNDCP provides the following functions:

***Segmentation*:** Network protocol data units (NPDUs) are segmented and reassembled wherever needed, in order to be accommodated within the limited length of the data link frames. With this segmentation, the maximum size of an NPDU can be 2,048 bytes; while the maximum size of user data supported by MDLP is considerably smaller (default value is 130 bytes).

***Encryption*:** To provide user data confidentiality over the CDPD air interface, the NPDUs are encrypted after being segmented. The secret keys used for the encryption and decryption are obtained by means of a security management entity (SME) that operates on top of SNDCP as a network layer entity.

***Multiplexing*:** The SNDCP provides the means for multiplexing a number of network-layer traffic streams within the same data link connection. Note that this facility is not provided by the MDLP. This makes feasible the simultaneous utilization of various network-layer entities on top of SNDCP. For example, as illustrated in Figure 30.4, two (or more) network protocols may simultaneously operate on top of SNDCP. Each one is discriminated by its own network layer protocol identifier (NLPI).

***Header Compression*:** The SNDCP compresses and recovers redundant network-control information to increase data link performance and efficiency.

***Quality of Service (QoS)*:** Two data transport modes are provided by the SNDCP: the acknowledged mode, which transfers NPDUs within the data link control procedures and the unacknowledged mode, which transfers NPDUs outside the data link control. The transport service mode utilized depends on the QoS parameter requested by the network layer.

30.2.3 Channel hopping

Since CDPD was added to the voice system after the latter was already operational, its design was subject to the constraint that no changes should be necessary to the existing voice system. For this reason, CDPD was designed to be completely transparent to the underlying voice system. Consequently, when the voice system selects a new channel for voice transmission, it is not aware of the existence of CDPD and it may select to use the channel currently used by CDPD. To avoid collisions in such cases, CDPD transmissions should obstruct the currently used channel as soon as possible and hop to another idle channel. CDPD monitors the transmit signal of the underlying voice system by sensing the power that enters into its transmitting antenna.

As soon as a power ramp-up is detected, which indicates the initiation of voice traffic, a channel hopping procedure begins. First, the MDBS sends a special signal that closes down the channel. This should be accomplished within 40 ms after the start of power ramp-up, because it takes so long before any voice is transmitted. While closing the CDPD channel, the MDBS may also announce the new CDPD channel where it will hop (if it is already known). Afterward, the MDBS finds a new idle voice channel and starts transmitting an identification signal in this channel. In the case where the CDPD channel was closed without an announcement of a new channel number, mobile terminals must hunt around among a designated set of potential CDPD channels in order to find the new one. In this way, CDPD can occupy any idle capacity in a cell, without interfering with the voice system. However, nothing in the design prevents having dedicated CDPD channels. Actually, as CDPD grows in popularity, providers are more likely to reserve channels exclusively for it.

30.2.4 Channel access

As stated above, all M-ESs that communicate with the same MDBS (i.e., they are in the same cell) share a common transmission channel, called the "reverse" channel. On the other hand, the MDBS uses the "forward" channel to transmit information to the M-ESs. The reverse and the forward channels are different RF channels and can be used simultaneously. An M-ES can access the reverse channel using a slotted non-persistent digital sense multiple access with collision detection (DSMA/CD) algorithm. This algorithm is similar to the carrier sense multiple access with collision detection (CSMA/CD) used in Ethernet. However, in CDPD, because M-ESs cannot sense the status of the reverse channel directly (since they employ different reception and transmission frequency bands), a different collision detection scheme is applied.

The DSMA/CD algorithm makes use of the busy/idle flag and the decode status flag, which are periodically transmitted on the forward channel. The busy/idle flag is a 5-bit sequence that is transmitted once every 60 bits, that is, once every one micro-slot period. This flag provides periodic binary information with one micro slot resolution indicating whether the reverse channel is busy or idle. On the other hand, the decode status flag is a 5-bit sequence that indicates whether or not the MDBS has decoded successfully a data block previously transmitted by an M-ES.

An M-ES wishing to transmit first senses the busy/idle flag (actually, a locally stored version of it, which is updated once every a micro slot period). If the reverse channel is found busy, the M-ES defers for a random number of micro slots and then repeats the sensing of the busy/idle flag again. Because the M-ES does not persist in continuously sensing the busy/idle flag, the access scheme is referred to as non-persistent. Once the reverse channel is found idle, the M-ES may initiate transmission. Note that a transmission may be initiated only at a micro slot boundary, which is why the access scheme is termed slotted. As soon as the MDBS detects a transmission start on the reverse channel, it sets the busy/idle flag in order to prevent further transmissions.

After an M-ES gains access to the reverse channel, it transmits its data as a sequence of fixed-length blocks. As mentioned before, the decode status flag provides "real-time" information regarding the successful reception of these blocks. The M-ES checks this flag and continues transmission if the MDBS has encountered no decoding errors. In the opposite case, it ceases transmission and attempts to regain access to the reverse channel after an appropriate exponential back off retransmission delay. This delay is increased exponentially by a factor of two on every subsequent retransmission attempt, hence the name exponential back off.

30.3 GPRS and higher data rates

GPRS is a GSM phase 2 bearer service that provides wireless packet data access to mobile GSM users. Its introduction is one of the key steps in the evolution of today's GSM networks to the third generation. The main feature of GPRS is that it reserves radio resources only when there is such a need and that these radio resources are shared by all mobile stations (MSs) in a cell. Therefore, as is the case for all packet data services, effective resource utilization is provided for bursty data applications, such as telemetry, train control systems, interactive data access, toll road charging systems, and Internet browsing using the World Wide Web. Data rates of up to 115 Kbps would be supported.

The main objective of GPRS is to offer a mobile packet interface to standard data networks (such as TCP/IP, X.25, and CLNP). This interface should be embedded in the conventional GSM network architecture, which was originally designed for circuit-switched integrated services. For this purpose, GPRS introduces some new functional elements to the general GSM architecture and it modifies some mobility management functions. Some cooperation still exists between the elements of the standard GSM services and GPRS. In comparison with CDPD, it is important to note that GPRS provides a data overlay within the standard GSM infrastructure, while CDPD provides a packet data overlay but with a totally separate infrastructure.

30.3.1 System description

The GPRS reference model is depicted in Figure 30.5. GPRS allows the subscriber to send and receive data in an end-to-end packet transfer mode, without using any network resources in circuit-switched mode.

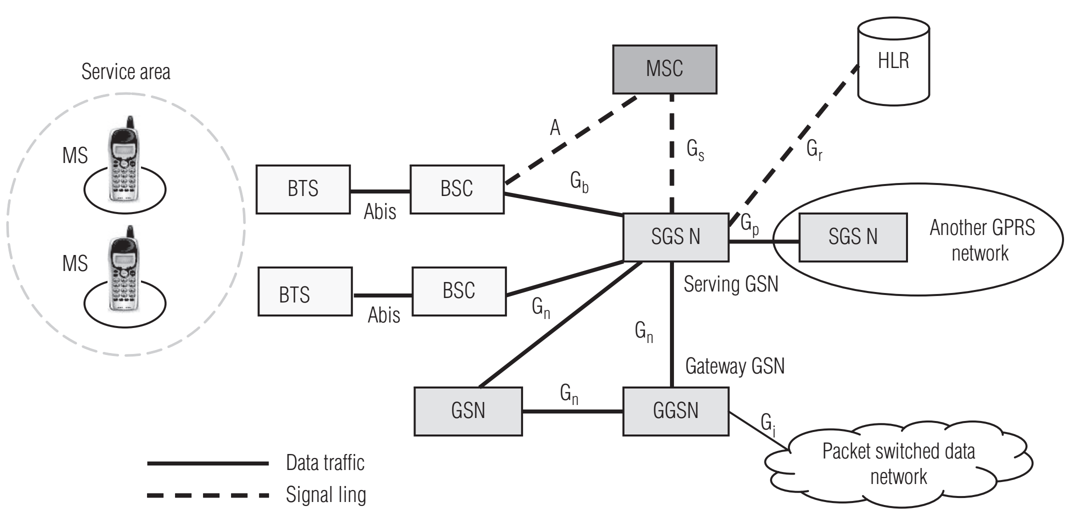


Figure 30.5 GPRS system architecture

This allows for autonomous operation of GPRS and best fits the bursty traffic characteristics. Packet routing and transfer within the public land mobile network (PLMN) is supported by definition of a new logical network node called a GPRS support node (GSN). The GSN is basically a packet router with additional mobility management features and connects with various network elements through standardized interfaces.

The GSN that acts as a physical interface to the external packet data networks (PDNs) is called as the gateway GSN (GGSN), whereas the GSN that connects with a base station controller (BSC) and directly handles packet delivery to and from the MSs is called as the serving GSN (SGSN). Each SGSN is responsible for the delivery of packets to the MSs within its service area. The general packet routing procedure is performed as illustrated in Figure 30.6. In the case where a mobile originates traffic for a fixed host, the SGSN encapsulates the data packets and routes them to the appropriate GGSN, where they are forwarded to the correct packet switched data network (PSDN). Specific routing policies are applied inside this PSDN to send the packets to the corresponding host. On the other hand, packets coming from a corresponding host are first routed to the GGSN through the PSDN based on the examination of the destination address. The GGSN checks the routing context associated with this destination address and determines the address of the SGSN currently serving the addressed MS. Subsequently, the original data packet is encapsulated into another packet (this procedure is called tunnelling), which is forwarded to the SGSN and ultimately delivered to the correct MS.

The GPRS backbone network is a private IP network. The IP addresses used in this backbone are selected by the GPRS operator and they are not known outside the PLMN. Routing between GPRS mobile terminals is usually accommodated through inter-PLMN packet networks, as illustrated in Figure 30.6. Within GPRS, two different encapsulation schemes are used:

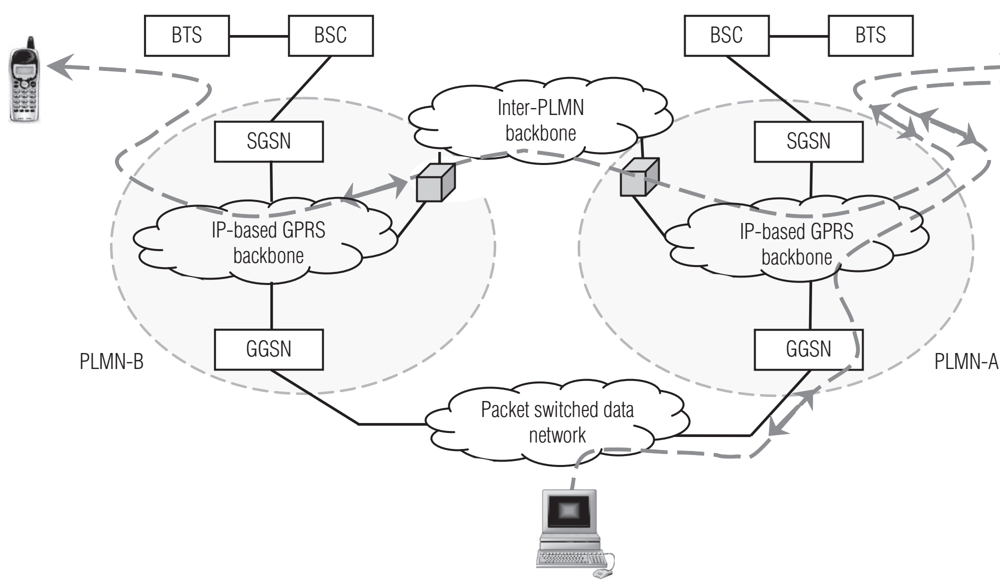


Figure 30.6 Typical routing within GPRS

- First, between the GSNs all packets are encapsulated by means of a GPRS tunnelling protocol (GTP) in order to enable the usage of different packet data protocols, even if these protocols are not supported by all SGNSs.
- Second, encapsulation between the MS and SGSN is performed to decouple the logical link management from the network-layer protocols.

30.3.2 Protocol architecture

The GPRS service is based on the layered architecture shown in Figure 30.7(a). At the top of the protocol layer, there may be a number of network protocols, which are treated as prospective users of the GPRS service. The SNDCP provides a multiplexing layer that aids the transmission of multiple network layer messages onto a single logical link connection. Furthermore, the SNDCP includes ciphering, segmentation, and compression facilities. Its functionality is very similar to the functionality of the SNDCP layer in CDPD, explained previously. The LLC layer, at the bottom of the SNDCP layer, provides a logical link between the MS and SGSN. Protocol functionality is based on LAPD (Link access procedures on the D channel is the second layer protocol on the ISDN protocol stack in the D channel, where **D channel** (data) is a telecommunications term, which refers to the ISDN channel in which the control and signalling information is carried) as used within the GSM signalling plane, but additional features are supported, for example, point-to-multipoint transmission. This link control protocol is sometimes called link access procedure on the G-channel (LAPG). The radio link control/medium access control (RLC/MAC) layer arbitrates access to the shared medium between the MSs and the network, and also provides a reliable link between an MS and a BS. It is worth noting that RLC is specific to the radio technology used in the wireless interface, while LLC is independent of the wireless interface characteristics. The protocol data units at the LLC layer are segmented into one or more RLC frames, which are handed over to the MAC sub-layer. In turn, each MAC frame is translated into four fixed-length blocks which, after bit-interleaving, are transmitted on four consecutive TDMA frames (using the same time slot in each frame).

Apart from the efficient multiplexing of data and signalling information, the functionality of the RLC/MAC layer includes contention resolution, QoS control, framing, and error handling. The MAC protocol for the GPRS radio interface is essentially a slotted ALOHA reservation protocol (ALOHA with an additional constraint in which time is divided into discrete time intervals) and operates between the MS and the base station transceiver (BTS). Before an MS is capable of using the GPRS service, it must attach to this service. Effectively, this attachment corresponds to the establishment of a logical link between the MS and its SGSN. As a result, a temporary logical link identity (TLLI) is assigned to the MS. A logical condition after this assignment is illustrated in Figure 30.7(b). As explained previously, a logical link can accommodate more than one traffic flow through the multiplexing facilities provided by the SNDCP layer. MSs roaming within the service area of a given SGSN are communicating with this SGSN through the reserved GPRS resources and via separate logical links. Each link has its own unique TLLI. Additionally, within each logical link, a number of different traffic flows (e.g., different network protocols) are multiplexed with the aid of the SNDCP protocol. After attachment, one or more routing contexts for one or more network protocols can be negotiated with the SGSN. In order to verify that a given MS is allowed to use a network protocol, the home location register (HLR) is queried. Among other things, the subscription profile found in the HLR includes the matching GGSN address. If access is permitted, the GGSN is requested to update the routing context (i.e., the serving SGSN address and tunnelling information) accordingly.

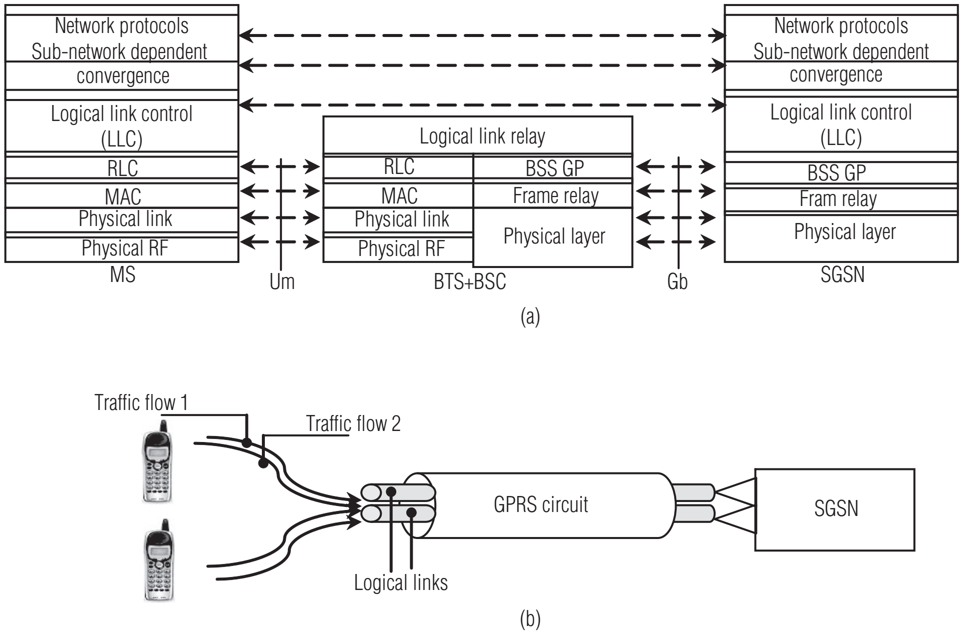


Figure 30.7 (a) GPRS protocol architecture; (b) schematic presentation of traffic and logical link multiplexing in GPRS

As explained earlier, this information is used to support the routing procedure in a mobile environment. During the GPRS session, the location of an MS is being tracked. When in ready state (i.e., during active communication), the MS informs the SGSN about every cell change. However, when in standby state (i.e., when waiting for an event to trigger transmission/reception), the MS requests location updates upon a routing area (RA) change. The RA consists of an operator-defined group of cells. If an RA update takes place and the new RA is handled by another SGSN, then the new SGSN asks the old SGSN to send the mobility management information of the MS. Subsequently, the GGSN and the HLR are updated with the new routing context, and the old SGSN deletes the corresponding information of this MS.

30.3.3 Channel access

When a network operator decides to offer GPRS-based services within a cell, one or several physical channels from the pool of available channels are dedicated to packet mode transfer. Each of these so-called packet data channels (PDCHs) is mapped onto one physical time slot. In order to support a flexible adaptation to different traffic requirements, allocation of PDCHs is based on demand. Prior to packet transmission, an MS initiates a random access request, that is, it sends a short request on a control uplink channel, called the packet random access channel (PRACH). Together with the access request, the MS indicates the number of GPRS slots required for the forthcoming transaction.

This access request is handled by the BSC, which provides the radio resource management functionality. Upon correct reception of the access request, a control downlink channel (the packet access grant channel) is used to identify the reserved slots and the timing advance. If an MS receives no response to an access request, a retransmission procedure takes place after a random back off time up to a maximum number of access attempts. After the transmission in the reserved time slots is completed, the BTS acknowledges the reception status of the transmitted blocks. It sends either a negative acknowledgement (NACK) to specify the erroneous blocks, or a positive ACK to specify no errors.

A NACK message lists the erroneous blocks that should be retransmitted and also includes an appropriate channel reservation for the retransmission to take place with minimal delay. If after transmission of data, the MS does not receive an ACK within a certain time period, a data recovery procedure is initiated by attempting a new random access. In the case of mobile-terminated traffic, the network sends a page to a specific MS through the packet paging channel (PPCH). If the cell wherein this MS is located is known, this page message may include either a direct reservation of uplink slots for uplink transmission, or an indication of downlink slots for data reception. In this case, the MS may immediately start data reception/transmission on the pre-reserved slots. On the other hand, if the accurate location of the MS is not known, no slots are reserved for immediate transmission. Instead, a single slot is reserved for a paging response that precedes the channel reservation and therefore, the collision-sensitive random access is avoided. Of course, if a page without reservation is sent, the MS initiates the usual random access procedure and asks for the reservation of one block to be able to identify itself after access is granted.

30.4 Short messaging service in GSM

The proliferation of GSM enabled the introduction of the SMS, which has become extremely popular in Europe. It is similar to the peer-to-peer instant messaging services on the Internet. Users of SMS [PEE00a], [PEE00b] can exchange alphanumeric messages up to 160 characters (mapped into 140 bytes) within seconds of submission of the message. The service is available wherever GSM exists and makes it a very attractive wide area data service.

30.4.1 What is SMS?

SMS was developed as part of GSM Phase 2 specifications. It operates over all GSM networks making use of the GSM infrastructure completely. It uses the same network entities (with the addition of a SMS centre—SMSC), the same PHY layer, and intelligently reuses the logical channels of the GSM system to transmit the very short alphanumeric messages.

Service description

SMS has both an almost instant delivery service if the destination MS is active and a store and forward service if the MS is inactive. Two types of services are specified: In the cell broadcast service, the message is transmitted to all MSs that are active in a cell and that are subscribed to the service. This is an unconfirmed one-way service used to send weather forecasts, stock quotes, and so on. In the *point-to-point* (PTP) service, an MS may send a message to another MS using a handset keypad, a personal digital assistant (PDA), or a laptop connected to the handset, or by calling a paging centre. Recently, SMS messages can be transmitted via dial-up to the service centre and the Internet as well [PEE00a].

A short message (SM) can have a specified priority level, future delivery time, expiration time, or it might be one of the several short predefined messages. A sender may request acknowledgment of message receipt. A recipient can manually acknowledge a message or may have predefined messages for acknowledgement.

An SM will be delivered and acknowledged whether a call is in progress because of the way logical channels in GSM are used for SMS.

30.4.2 Overview of SMS operation

The SMS makes use of the GSM infrastructure, protocols, and the physical layer to manage the delivery of messages. Note that the service has a store-and-forward nature. As a result, each message is treated individually. Each message is maintained and transmitted by the SMS centre (SMSC). The SMSC sorts and routes the messages appropriately. The short messages (SMs) are transmitted through the GSM infrastructure using signalling system-7 (SS-7).

Figure 30.8 shows the reference architecture and the layered protocol architecture for SMS. There are two cases of SMs: A mobile-originated SM and a mobile-terminated SM. An SM originating from an MS has to be first delivered to a service centre. Before that, it reaches a message service centre (MSC) for processing. A dedicated function in the MSC called the SMS-interworking MSC (SMS-IWMSC) allows the forwarding of the SM to the SMSC using a global SMSC ID. An SM that terminates at the MS is forwarded by the SMSC to the SMS-gateway MSC (SMS-GMSC) function in an MSC. As in the case of GSM, it either queries the HLR or sends it to the SMS-GMSC function at the home MSC of the recipient. Subsequently, the SM is forwarded to the appropriate MSC that has the responsibility of finally delivering the message to the MS. This delivery is performed by querying the VLR for details about the location of the MS, the BSC controlling the BTS, providing coverage to the MS, and so on. There are four layers in SMS: the application layer (AL), the transfer layer (TL), the relay layer (RL), and the link layer (LL). The AL can generate and display the alphanumeric message.

The SMS-TL services the SMS-AL to exchange SMs and receive confirmation of receipt of SMs. It can obtain a delivery report or status of the SM sent in either direction. The RL relays the SMS PDUs through the LL. There are six PDU types in SMS that convey the SM from the SMSC to the MS and vice versa, convey a failure cause, and convey status reports and commands. Over the air, the SMs are transmitted in time slots that are freed up in the control channels. If the MS is in an idle state, the SMs are sent over the SDCCH at 184 bits within approximately 240 ms. If the MS is

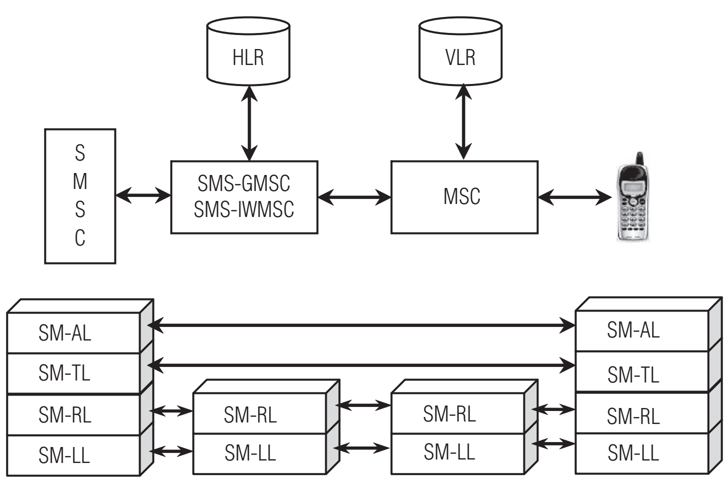


Figure 30.8 Reference and layered protocol architecture for SMS

in the active state (i.e., it is handling a call), the SDCCH is used for call setup and maintenance. In that case, the SACCH has to be used for delivering the SM. This occurs at around 168 bits every 480 ms and is much slower. Failures can occur if there is a state change when the SM is in transit. The SM will have to be transmitted later.

In the case of cell broadcast, a cell broadcast entity and a cell broadcast centre are used to send the weather forecast or other broadcast SMs to multiple BSCs for delivery. The broadcast contains the data and identities of MSs that are to receive the message. The cell broadcast is on the cell broadcast channel (CBCH).

30.5 Summary

- CDPD is a mobile data technology that permits subordinate data operation on the spectrum assigned to the AMPS.
- CDPD systems are designed to take advantage of the idle voice channels of an analogue cellular network, such as AMPS.
- The physical layer (PHY) in CDPD corresponds to a functional entity that accepts a sequence of bits from the MAC layer and transforms them into a modulated waveform for transmission onto a physical 30-kHz RF channel.
- The LLC layer is used to convey information between the network layer entities across the CDPD air interface.
- The primary service offered by the MDLP to the upper layer (SNDCP) is the provision and the control of one or more logical data link connections on a CDPD channel stream.
- NPDUs are segmented and reassembled wherever needed, in order to be accommodated within the limited length of the data link frames.
- The SNDCP provides the means for multiplexing a number of network-layer traffic streams within the same data link connection.
- The SNDCP compresses and recovers redundant network-control information to increase data link performance and efficiency.
- An M-ES can access the reverse channel using a slotted non-persistent DSMA/CD algorithm.
- GPRS is a GSM Phase 2 bearer service that provides wireless packet data access to mobile GSM users.
- The main objective of GPRS is to offer a mobile packet interface to standard data networks (such as TCP/IP, X.25, and CLNP).
- The SNDCP provides a multiplexing layer that aids the transmission of multiple network layer messages onto a single logical link connection.
- The SMS makes use of the GSM infrastructure, protocols, and the PHY to manage the delivery of messages.

Review questions

1. Explain the system description of data-oriented CDPD network.
2. Write short notes on channel hopping.
3. Briefly explain the system description of GPRS.
4. Explain short messaging service in GSM.
5. Explain mobile application protocol.

Objective questions and answers

1. CDPD was first introduced by
   (a) Sony (b) IBM (c) Apple (d) Nokia
2. The location of an F-ES is
   (a) fixed (b) dynamic (c) both (a) and (b) (d) none
3. An MDBS consists of
   (a) computer (b) power amplifiers (c) a radio transceiver (d) all
4. The M-ESs are connected to the CDPD network through the
   (a) A-Interface (b) E-Interface (c) both (a) and (b) (d) none
5. The F-ESs are connected to the CDPD network through the
   (a) A-Interface (b) E-Interface (c) both (a) and (b) (d) none
6. A frequency greater than the central carrier frequency represents
   (a) a logical 1 (b) a logical 0 (c) a logical –1 (d) none
7. The modulation rate on both the forward and reverse RF channels is
   (a) 18.2 Mbps (b) 19.2 Kbps (c) 30.2 Mbps (d) 31.2 Kbps
8. The main objective of GPRS is to offer a ________ to standard data networks.
9. The GSN node that acts as a physical interface to the external packet data networks (PDNs) is called as ________.
10. The logical link control (LLC) layer, at the bottom of the SNDCP layer, provides a logical link between ________ and ________.

Answers: 1. (b), 2. (a), 3. (d), 4. (a), 5. (b), 6. (a), 7. (b), 8. Mobile Packet Interface, 9. GGSN, 10. MS and SGSN.

Open book questions

1. Explain the channel access of data-oriented CDPD network.
2. Write short notes on service description of SMS in GSM.

Further reading

Aftab Ahmad, *Wireless and Mobile Data Networks*. Hoboken, NJ: John Wiley & Sons, 2005.

Garg, V., *Wireless Communications and Networking*. San Francisco: Morgan Kaufmann Publishers, Elsevier, 2007.

Pahlavan, K., and Krishnamurthy, P., *Principles of Wireless Networks*. Englewood Cliffs, NJ: Prentice Hall, 2002.

Appendices

Appendix A: The Decibel (dB)

The unit of power is watt (W). One watt is equal to one ampere of current flowing at one volt. But microwave engineers often use a relative unit known as decibel (dB) to express the ratio of two powers. It is used for stating the gain or loss of one device (P1) in relation to another (P2). The formula for calculating gain or loss in dB is derived as follows:

1 bel is a ratio of 10:1 between two power levels. Therefore, a power ratio of 200:20 is 1 bel (10:1), 200:40 is 5 bels (5:1), and 200:10 is 2 bels (20:1).

$$\text{bels} = \log(\text{P1}/\text{P2})$$

$$\text{decibels (dB)} = 10 \times \log(\text{P1}/\text{P2}) = 10 \times \log(P_{out}/P_{in})$$

The dB by itself is not an absolute number, but a ratio.
In general, a dimensionless quantity P in decibels (denoted P dB) is defined by

$$P\ \text{dB} = 10 \log_{10} P$$

P usually represents a ratio of powers, where the denominator is the reference, and log10 is simply written as log. Characters are added to the "dB" to denote the reference quantity. For example, dBm is decibels relative to a milliwatt. Therefore, if P is in watts

$$P\ \text{dBW} = 10\log(P/1) \text{ or } P\ \text{dBm} = 10 \log(P/0.001)$$

For amplifiers, a common reference unit is the dBm, with 0 dBm being equal to 1 milliwatt (mW). Thus, an amplifier with an output of 30 dBm puts out 1 W.

Therefore, we can represent any quantity in terms of dBW, dBm, or dBμ as

$$\text{Power (dB or dBW)} = 10 \times \log_{10} \text{(power in watts)}$$

$$\text{Power (dBm)} = 10 \times \log_{10} \text{(power in milli (m)watts)}$$

$$\text{Power (dB}\mu\text{)} = 10 \times \log_{10} \text{(power in micro (}\mu\text{)watts)}$$

dBW is decibels relative to 1 W and dBm is decibels relative to 1 mW.

A.1 Relationship between power and voltage

Since power = V^2/R

$$\text{Power in dB} = 10 \times \log10\ (\text{power in watts}) = 10 \times \log_{10}(V^2/R) = 20 \times \log_{10}(\text{voltage}/\sqrt{R})$$

dB representation for antennas:

dBd is used to represent gain (a ratio) relative to a dipole antenna:

Power (in maximum direction) of a dipole antenna = P_{dip}

Power (in maximum direction) of some other antenna = P_{ant}

Gain of that antenna relative to a dipole = P_{ant}/P_{dip}

$$\text{Gain (dBd)} = 10 \times \log_{10}(P_{ant}/P_{dip})$$

dBi is the unit that is used to measure the gain of an antenna. It states the gain of an antenna as referenced to an isotropic source. The greater the dBi value, the higher the gain and, as such, the more acute the angle of coverage. Gain of antenna with respect to isotropic antenna given in dBi = 10 log(power max/power isotropic).

A.2 Common dB Factors

Example:

1 W = 0 dB
2 W = 0 + 3 dB = 3 dB
20 W = 3 dB + 10 = 13 dB

Linear function	dB function
1	0 dB
* 2	+3 dB
/ 2	–3 dB
* 10	+10 dB
* 100	+20 dB
/ 10	–10 dB
/ 100	–20 dB

Comparison of milliWatts and decibel change (relative to 1 mW)

Formula for converting dBm to mW:

$$\text{dBm} = \log_{10}(\text{mW}) * 10$$

Formula for converting mW back to dBm:

$$\text{mW} = 10^{(\text{dBm}/10)}$$

The differences between the values can become extremely large or small and more difficult to deal with. It is easier to say that a 100 mW signal decreased by 70 dB than to say that it decreased to 0.00001 mW.

$$10 \log(100/0.00001) = 70 \text{ dB}$$

Table A.1 gives the power in milliWatts and corresponding decibel change, and Table A.2 gives the 10's and 3's rule of RF power.

Table A.1 Power in milliWatts and corresponding decibel change

dB change	–40	–30	–20	–10	0	+10	+20	+30	+40
mW	0.001	0.001	0.01	0.1	1	10	100	1000	10000

Table A.2 The 10's and 3's rules of RF power

Rule (dB)	Explanation	% of power lost/ gained	Current power level	Example
–3	Half of the watt value	50% lost	Half of original	100 mW – 3 dB = 50 mW
+3	Double the watt value	100% gained	Double the original	100 mW + 3 dB = 50 mW
–10	Decrease watt value to one-tenth of original	90% lost	One-tenth of original	300 mW – 10 dB = 30 mW
+10	Increase the watt value by 10-fold	1000% gained	Ten times the original	10 mW + 10 dB = 100 mW

Appendix B: Frequencies for Communication

The fundamental principle of wireless communication is electromagnetic wave transmission between a transmitter and a receiver. Signals are characterized by their frequencies in use. A radio signal is characterized by wavelength and frequency. Multiple signals or noises of the same frequency will cause interference at the receiver. To avoid interference, various wireless technologies use distinct frequency bands with well-controlled signal power which are portions of the so-called frequency spectrum. As a scarce public resource, the frequency spectrum is strictly regulated by governments of countries around the world.

Figure B.1 shows various bands of electromagnetic frequency spectrum. The frequency spectrum can be divided into the following categories: very low frequency (VLF), low frequency (LF), medium frequency (MF), high frequency (HF), very high frequency (VHF), ultra-high frequency (UHF), super-high frequency (SHF), extremely high frequency (EHF), infrared, visible light, ultraviolet, X-ray, gamma-ray, and cosmic ray, each of which represents a frequency band. Table B.1 gives the frequency range, wavelength, and applications of the various bands of the frequency spectrum.

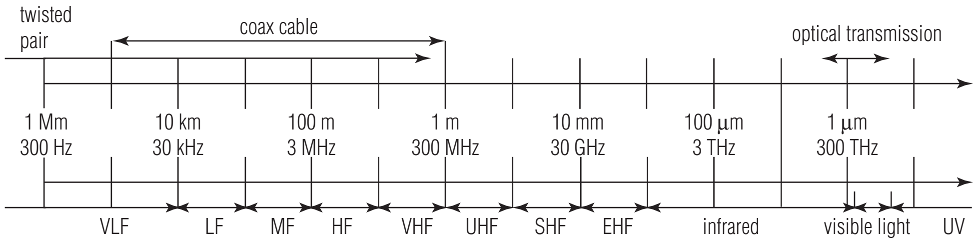


Figure B.1 Electromagnetic frequency spectrum

Table B.1 Applications of various frequency bands

Frequency	Wavelength (Free space)	Designation	Applications
<3 Hz	>100 mm		Geophysical prospecting
3–30 Hz	10–100 mm	ELF	Detection of buried metals
30–300 Hz	1–10 mm	SLF	Power transmission, submarine communications
0.3–3 KHz	0.1–1 mm	ULF	Telephone, audio
3–30 KHz	10–100 km	VLF	Navigation, positioning, naval communications
30–300 KHz	1–10 km	LF	Navigation, radio beacons
0.3–3 MHz	0.1–1 km	MF	AM broadcasting
3–30 MHz	10–100 m	HF	Short wave, citizens' band
30–300 MHz	1–10 m	VHF	TV, FM, police
54–72			TV channels 2–4
76–88			TV channels 5–6
88–108			FM radio
174–216			TV channels 7–13
0.3–3 GHz	10–100 cm	UHF	Radar, TV, GPS, cellular phone
470–890 MHz			TV channels 14–83
915 MHz			Microwave ovens (Europe)
800–2500 MHz		"money band"	PCS cellular phones, analog at 900 Mz , GSM/CDMA at 1900
1–2 GHz			
2–4.5			L-band, GPS system
2–4			Microwave ovens (US)
			S-band
3–30 GHz	1–10 cm	SHF	Radar, satellite communications
4–8			C-band
8–12			X-band (police radar at 11 GHz)
12–18			K_u-band (dBs, Primestar at 14 GHz)
			K-band (police radar at 22 GHz)

(*Continued*)

Table B.1 Applications of various frequency bands (*Continued*)

Frequency	Wavelength (Free space)	Designation	Applications
18–27			
30–300 GHz	0.1–1 cm	EHF	Radar, remote sensing
27–40			K_a-band (police radar at 35 GHz)
40–60			U-band
60–80			V-band
80–100			W-band
0.3–1 THz	0.3–1 mm	Millimetre	Astronomy, meteorology
10^{12}–10^{14} Hz	3-300 μm	Infrared	Heating, night vision, optical communications
3.95×10^{14} to 7.7×10^{14} Hz	390–760 nm	Visible light	Vision, astronomy, optical communications
	625–760		Red
	600–625		Orange
	577–600		Yellow
	492–577		Green
	455–492		Blue
	390–455		Violet
10^{15}–10^{18} Hz	0.3–300 nm	Ultraviolet	Sterilization
10^{16}–10^{21} Hz		X-rays	Medical diagnosis
10^{18}–10^{22} Hz		Gamma rays	Cancer therapy, astrophysics
$>10^{22}$ Hz		Cosmic rays	Astrophysics

Appendix C: Fundamentals of Antenna Radiation

The electric and the magnetic fields of an electromagnetic wave are interdependent while propagating through a space. During wave propagation, the time-varying magnetic field generates a time-varying electric field and vice versa, and they propagate through the free space at the velocity of light. Figure C.1 shows a travelling plane wave, wherein the electric field (E) and the magnetic field (H) are perpendicular to each other. Together they propagate in a direction transverse to the plane. Such a wave where the electrical and the magnetic fields are perpendicular to a plane and the direction of propagation is transverse to that plane is called a transverse electromagnetic (TEM) wave. Maxwell's equations govern the principles of guiding and the propagation of electromagnetic energy and provide the foundations of all electromagnetic phenomena and their applications.

Maxwell's equations relate the fields (**E** and **H**) and their sources (**r** and **J**) to each other. Maxwell's equations in differential form are

$$\nabla \times H = \dot{D} + J \tag{C.1}$$

$$\nabla \times E = -\dot{B} \tag{C.2}$$

$$\nabla \cdot D = \rho \tag{C.3}$$

$$\nabla \cdot B = 0 \tag{C.4}$$

Equation (C.4) indicates the presence of a vector potential such that

$$B = \nabla \times A \tag{C.5}$$

(since $\nabla \cdot \nabla \times A = 0$ for any vector A)
Substituting Equation (C.5) into Equation (C.2) yields

$$\nabla \times (E + \dot{A}) = 0 \tag{C.6}$$

Equation (C.6) indicates that $E + \dot{A}$ can be expressed as gradient of a scalar, since curl of gradient of a scalar is zero. Substituting $E + \dot{A}$ equals to $-\nabla V$, where V is scalar potential yields

$$E = -\nabla V - \dot{A} \tag{C.7}$$

Thus, Equations (C.2) and (C.4) suggest the existence of magnetic and electric potentials.

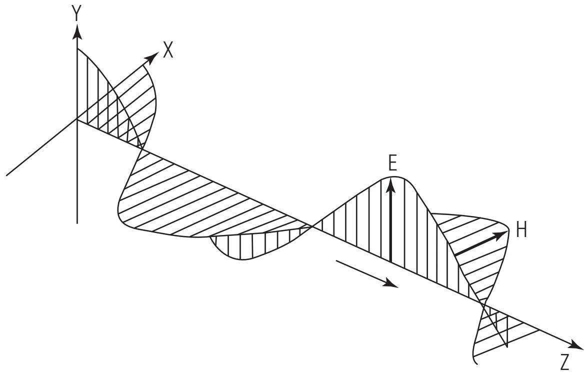


Figure C.1 Plane travelling wave with *E* and *H* vectors perpendicular to each other and on a plane perpendicular to the direction of propagation

The established Lorentz condition is given as

$$\nabla \cdot A = -\mu\varepsilon\dot{V} \tag{C.8}$$

Substituting Equations (C.5), (C.7), and (C.8) in Equation (C.1) yields

$$\frac{1}{\mu}\nabla \times \nabla \times A = -\varepsilon\nabla\dot{V} - \varepsilon\ddot{A} + J$$

$$\Rightarrow \nabla^2 A - \mu\varepsilon\ddot{A} = -\mu J \tag{C.9}$$

Substituting Equations (C.7) and (C.8) in Equation (C.3) yields

$$\nabla^2 V - \mu\varepsilon\ddot{V} = -\frac{\rho}{\varepsilon} \tag{C.10}$$

Solutions to Equations (C.9) and (C.10) will be of the form

$$A(r) = \frac{\mu}{4\pi}\int J(r')\frac{e^{-jBR}}{R}\mathrm{d}V' \tag{C.11}$$

and

$$V(r) = \frac{1}{4\pi\varepsilon}\int \rho(r')\frac{e^{-jBR}}{R}\mathrm{d}V' \tag{C.12}$$

To find the radiation from antenna, consider an infinitesimal linear current element $I\mathrm{d}l$ flowing through an infinitesimal length $\mathrm{d}l$ which is very much less than the wavelength of excited wave. From the knowledge of radiation from this element, electromagnetic field of any antenna can be obtained by proper integration. Figure C.2 shows the radiation from an infinitesimal element.

The magnetic vector potential A at any point P due to this finite current element can be expressed as

$$A(r) = \frac{\mu}{4\pi}\int I\frac{e^{-jBR}}{R}\mathrm{d}l \tag{C.13}$$

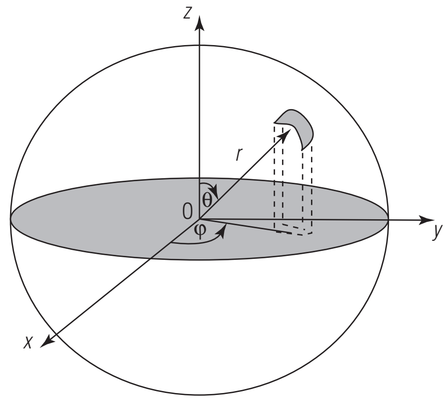


Figure C.2 Radiation from an infinitesimal element

The above equation is obtained by substituting $J\mathrm{d}v = I\mathrm{d}l$ (this is valid for a linear current source). As the current element is very small, Equation (C.13) can be slightly modified with $\mathrm{d}l = \Delta l\,\hat{z}$ to obtain magnetic potential at point P as

$$A(P) = \frac{\mu I \Delta l\, e^{-j\beta r}}{4\pi r}\hat{z} \tag{C.14}$$

In spherical coordinates

$$A_r = A_z \cos\theta$$

$$= \frac{\mu I \Delta l\, e^{-j\beta r}}{4\pi r}\cos\theta \tag{C.15}$$

$$A_\theta = -A_z \sin\theta$$

$$= -\frac{\mu I \Delta l\, e^{-j\beta r}}{4\pi r}\sin\theta \tag{C.16}$$

$$A_\varphi = 0$$

Now as $H = {}^{1}\!/_{\mu}\nabla \times \overline{A}$

$$H_\varphi = j\beta\, I\Delta l \sin\theta\left[1 + \frac{1}{j\beta r}\right]\frac{e^{-j\beta r}}{4\pi r} \tag{C.17}$$

$$H_\theta = H_r = 0$$

From Equation (C.7)

$$E = -j\omega \bar{A} - \frac{j}{\omega\mu\varepsilon} \nabla \nabla . \bar{A} \tag{C.18}$$

It follows that

$$E_r = 2\eta\, I\Delta l \cos\theta \left[\frac{1}{r} + \frac{1}{j\beta r^2}\right] \frac{e^{-j\beta r}}{4\pi r} \tag{C.19}$$

$$E_\theta = j\eta\beta\ I\Delta l\ \sin\theta \left[1 + \frac{1}{j\beta r} - \frac{1}{\beta^2 r^2}\right] \frac{e^{-j\beta r}}{4\pi r} \tag{C.20}$$

$$E_\varphi = 0 \tag{C.21}$$

On the basis of the preceding discussion, we can divide the antenna radiation field into three regions, which are functions of wavelength λ and dimension of antenna D.
The definitions of these three regions are listed in the following.

Reactive Near-Field Region: Regions where R is less than wavelength of EM wave is the region where the reactive field (stored energy – standing waves) is dominant. In this region only $1/r^3$ terms of the Equations (C.17), (C.19), and (C.20) will dominate and remaining can be ignored. The limit for this distance R is

$$R_1 = 0.62\sqrt{\frac{D^3}{\lambda}}$$

Near-Field Region: This is the region between the reactive near field and the far field, where the radiation fields predominate and the field distribution is dependent on the distance from the antenna. In this region only $1/r^2$ terms of the Equations (C.17), (C.19), and (C.20) will dominate and remaining can be ignored. The limit for this distance R is

$$R_2 = 2\frac{D^2}{\lambda}$$

***Far-Field Region*:** In this region, the field distribution is essentially independent of the distance from the antenna. In this region, fields with the $1/r$ term will prevail and remain. This region is defined by a sphere with radius $R3 > R2$.

Appendix D: Weiner Filter

A class of linear optimum discrete time filters are collectively known as "Wiener Filters". Wiener filter theory is formulated for the general case of complex valued stochastic process with the filter specified in terms of its impulse response. This is because, in many practical applications (Radar, Sonar, Communication etc.) the observables are measured in baseband form. The term baseband is used to designate a band of frequencies representing the original signal as delivered by source of information. Consider a block diagram built around a linear siscrete time filter. The filter input consists of a time series $u(0)$, $u(1)$, $u(2)$, etc., and let the impulse response of the filter be w_0, w_1, w_2, etc. Filter produces an output $y(n)$ at some discrete time n. Estimation error denoted by $e(n)$ is defined as the difference between desired response $d(n)$ and filter output $y(n)$. The requirement is to make estimation error $e(n)$ as small as possible in some statistical sense.

The following two restrictions have been so far placed on the filter:

- Filter is linear – which makes mathematical analysis easy to handle.
- Filter operates in discrete time – which makes it possible for the filter to be implemented using digital hardware or software.

Final details of filter specification depend on the following two choices that have to be made:

1. Whether the impulse response of the filter has finite or infinite duration.
2. Type of statistical criterion used for optimization.

   Choice 1 is dictated by practical considerations.

   Choice 2 is influenced by mathematical tractability.

What criterion we need to choose for statistical optimization?

For this, we may consider optimizing filter design by minimizing a cost function or index of performance selected from the following possibilities.

1. Mean square value of estimation error.
2. Expectation of absolute value of estimation error.
3. Expectation of third or higher powers of absolute value of estimation error.

Option 1 has clear advantage over the other two because it leads to tractable mathematics.

The essence of filtering problem is to design a linear discrete time filter with output $y(n)$ that provides estimates of desired response $d(n)$ with a given set of input samples $u(0)$, $u(1)$, $u(2)$, etc. with a minimum mean square value of estimation error $e(n)$. (Estimation error $e(n)$ is the difference between desired response $d(n)$ and actual response $y(n)$.)

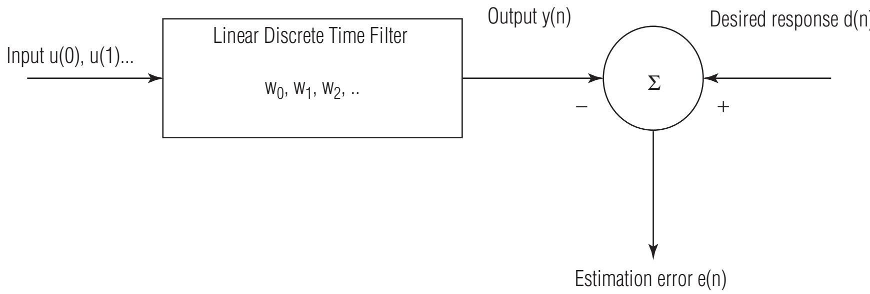


Figure D.1 Block diagram of a linear discrete time filter

To develop mathematical solution for the problem stated earlier, we follow two different approaches that are complementary.

Principle of Orthogonality: Filter input is denoted by time series $u(0)$, $u(1)$, $u(2)$, etc. impulse response is denoted by w_0, w_1, w_2, etc., both of which are assumed to have complex values and infinite duration. Filter output at discrete time n is defined by linear convolution sum

$$y(n) = \sum_{k=0}^{\infty} w_k * u(n-k), \quad n = 0, 1, 2, 3, \ldots \tag{D.1}$$

(* (asterisk) denotes complex conjugation).

$w_k{}^* u(n-k)$ represents the scalar version of an inner product of filter coefficients w_k and filter input $u(n-k)$.

The purpose of filter is to produce an estimate of desired response $d(n)$. We assume that the filter input and desired response $d(n)$ are single realizations of jointly wide sense stationary stochastic processes both with zero mean. If the mean is non-zero, we simply subtract them from the input $u(n)$ and desired response $d(n)$ before filtering.

The estimation of $d(n)$ is naturally accompanied by an error $e(n)$ which is defined as

$$e(n)=d(n) - y(n) \qquad \text{(D.2).}$$

$e(n)$ is sample value of random variable.

To optimize filter design, we choose to minimize the mean square value of $e(n)$.We thus define the cost function as mean-square error (J).

$$J = E[e(n)e * (n)] = E[Ie(n)I^2] \tag{D.3}$$

E denotes the statistical expectation operator. Our requirement is to get the operating conditions under which "J" attains its minimum values. For complex input data, the filter coefficients are also complex. kth filter coefficient w_k is

$$w_k = a_k + jb_k \quad (k = 0, 1, 2, \ldots) \tag{D.4}$$

Correspondingly we may define gradient operator ,the kth element of which is written in terms of first order partial derivatives with respect to real part a_k and imaginary part b_k as

$$\nabla_k = \frac{\partial}{\partial a_k} + j\frac{\partial}{\partial b_k} \quad (k = 0, 1, 2, \ldots) \tag{D.5}$$

Applying ∇ operator to cost function (J), we get a multidimensional complex gradient vector (∇J), the kth element of which is

$$\Delta_k J = \frac{\partial J}{\partial a_k} + j\frac{\partial J}{\partial b_k} \quad (k = 0, 1, 2, \ldots) \tag{D.6}$$

For the cost function (J) to attain a minimum value all the elements of gradient vector ∇J must be simultaneously equal to zero. That is,

$$\nabla_k J = 0. \quad K = 0, 1, 2, 3, \ldots \tag{D.7}$$

Under this set of conditions, the filter is said to be optimum in the mean square error sense. Substituting Equation (D.3) in Equation (D.6) we get,

$$\nabla_k J = E\left[\frac{\partial e(n)}{\partial a_k}e^*(n) + \frac{\partial e^*(n)}{\partial a_k}e(n) + \frac{\partial e(n)}{\partial b_k}je^*(n) + \frac{\partial e^*(n)}{\partial b_k}je(n)\right] \tag{D.8}$$

Using Equations (D.2) and (D.4), we obtain partial derivatives

$$\frac{\partial e(n)}{\partial a_k} = -u(n-k),$$

$$\frac{\partial e(n)}{\partial b_k} = -ju(n-k),$$

$$\frac{\partial e^*(n)}{\partial a_k} = -u^*(n-k), \tag{D.9}$$

$$\frac{\partial e^*(n)}{\partial b_k} == -ju^*(n-k).$$

Substituting Equation (D.9) in Equation (D.8), we get

$$\nabla_k J = -2E\left[u(n-k)\ e^*(n)\right] \tag{D.10}$$

We are now ready to specify the operating conditions required for minimizing the cost function J. Let e_0 denote the special value of estimation error that results when the filter operates in its optimum condition.

$$E[u(n-k)e^*(n)] = 0 \quad k = 0, 1, 2, \ldots \tag{D.11}$$

Equation (D.11) states that,

"The necessary and sufficient condition for cost function J to attain a minimum value is for the corresponding value of estimation error $e_0(n)$ to be orthogonal to each input sample that enter into an estimation of desired response at time n. This statement constitutes the principle of orthogonality."

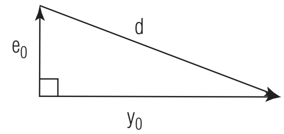


Figure D.2 Geometric interpretation of the relation between desired response, filter output, and the estimation error

There is a corollary to the principle of orthogonality that we may derive by examining the correlation between filter output $y(n)$ and estimation error $e(n)$. From Equation (D.1), we may express the correlation as,

$$E\left[y(n)e^*(n)\right] = E\left[\sum_{k=0}^{\infty} w_k^* u(n-k)e^*(n)\right] = \sum_{k=0}^{\infty} w_k^* E\left[u(n-k)e^*(n)\right] \tag{D.12}$$

Let $y_0(n)$ denote the output produced by the filter optimized in the mean square error sense, with corresponding estimation error $e_0(n)$. Using the principle of orthogonality from Equation (D.11), we get the result

$$E[y_0(n)e_0^*(n)] = 0 \tag{D.13}$$

Let $\hat{d}(n|U_n)$ denote the estimate of desired response, that is optimized in mean square error sense, given input data that span the space U_n up to and including time n.

$$\hat{d}(n \mid U_n) = y_0(n) \tag{D.14}$$

Note: The estimate $\hat{d}(n|U_n)$ has zero mean because the tap inputs are assumed to have zero mean. This condition matches the assumed zero mean of desired response $d(n)$.

Figure D.2 gives the geometric interpretation of relationship between desired response, estimate at filter output, and the estimation error.

Minimum Mean Square Error (MMSE): Equation (D.2) becomes

$$e_0(n) = d(n) - y_0(n) = d(n) - \hat{d}(n \mid U_n) \tag{D.15}$$

Let $J_{\min}$ denote the minimum mean square error,

$$J_{\min} = E[|e_0(n)^2|] \tag{D.16}$$

$$\sigma_d^2 = \sigma_{\hat{d}}^2 + J_{\min} \tag{D.17}$$

σ_d^2 = variance of desired response

$\sigma_{\hat{d}}^2$ = variance of estimate $d(n|U_n)$

$$J_{\min} = \sigma_d^2 - \sigma_{\hat{d}}^2 \tag{D.18}$$

It is convenient to normalize the expression in Equation (D.18), in such a way that mean square error always lies between zero and unity. We do this by dividing both sides of Equation (D.18) by σ_d^2,

$$J_{\min} / \sigma_d^2 = 1 - \sigma_{\hat{d}}^2/\sigma_d^2$$

Let $J_{\min}/\sigma_d^2 = \epsilon$.

ϵ is called the normalized mean square error,

$$\epsilon = 1 - \sigma_{\hat{d}}^2/\sigma_d^2$$

We note the following

1. The ratio ϵ can never be negative.
2. The ratio $\sigma_{\hat{d}}^2/\sigma_d^2$ is always positive.

$$\epsilon \text{ range is } 0 \le \epsilon \le 1.$$

If $\epsilon = 0$, the optimum filter operates perfectly, which means there is complete agreement between the estimate $\hat{d}(n|U_n)$ and the desired response $d(n)$. If $\epsilon = 1$, then there is no agreement between these quantities.

Appendix E: Mathematical Formulas

$$\sin(A + B) = \sin A \cos B + \cos A \sin B$$

$$\sin(A - B) = \sin A \cos B - \cos A \sin B$$

$$\cos(A + B) = \cos A \cos B - \sin A \sin B$$

$$\cos(A - B) = \cos A \cos B + \sin A \sin B$$

$$\tan(A + B) = \frac{\tan A + \tan B}{1 - \tan A \tan B}$$

$$\tan(A - B) = \frac{\tan A - \tan B}{1 + \tan A \tan B}$$

$$\sin^2 A + \cos^2 A = 1$$

$$1 + \tan^2 A = \sec^2 A$$

$$1 + \cot^2 A = \operatorname{cosec}^2 A$$

$$2 \cos A \sin B = \sin(A + B) - \sin(A - B)$$

$$2 \cos A \cos B = \cos(A + B) + \cos(A - B)$$

$$2 \sin A \sin B = \cos(A - B) - \cos(A + B)$$

$$\sin A + \sin B = 2 \sin \frac{A + B}{2} \cos \frac{A - B}{2}$$

$$\sin A - \sin B = 2 \cos \frac{A + B}{2} \sin \frac{A - B}{2}$$

$$\cos A + \cos B = 2\cos\frac{A+B}{2}\cos\frac{A-B}{2}$$

$$\cos A - \cos B = 2\sin\frac{A+B}{2}\sin\frac{B-A}{2}$$

$$2 \sin A \cos B = \sin(A + B) + \sin(A - B)$$

$$2 \cos A \sin B = \sin(A + B) - \sin(A - B)$$

$$2 \cos A \cos B = \cos(A + B) + \cos(A - B)$$

$$2 \sin A \sin B = \cos(A - B) - \cos(A + B)$$

$$\sin\frac{A}{2} = \pm\sqrt{\frac{1-\cos A}{2}}$$

$$\sin 2A = 2 \sin A \cos A$$

$$\cos 2A = \cos^2 A - \sin^2 A = 2\cos^2 A - 1 = 1 - 2\sin^2 A$$

$$\tan 2A = \frac{2\tan A}{1-\tan^2 A}$$

$$\sinh A = \frac{e^A - e^{-A}}{2}$$

$$\cosh A = \frac{e^A + e^{-A}}{2}$$

$$\tanh A = \frac{\sinh A}{\cosh A} = \frac{e^A - e^{-A}}{e^A + e^{-A}}$$

$$\coth A = \frac{\cosh A}{\sinh A} = \frac{e^A + e^{-A}}{e^A - e^{-A}}$$

$$\cos\frac{A}{2} = \pm\sqrt{\frac{1+\cos A}{2}}$$

$$\tan\frac{A}{2} = \pm\sqrt{\frac{1-\cos A}{1+\cos A}} = \frac{\sin A}{1+\cos A} = \frac{1-\cos A}{\sin A}$$

$$e^{jA} = \cos A + j\sin A$$

$$e^{-jA} = \cos A - j\sin A$$

$$\sinh(A+B) = \sinh A\cosh B + \cosh A\sinh B$$

$$\cosh(A+B) = \cosh A\cosh B + \sinh A\sinh B$$

$$\sinh(A-B) = \sinh A\cosh B - \cosh A\sinh B$$

$$\cosh(A-B) = \cosh A\cosh B - \sinh A\sinh B$$

$$\tanh(A+B) = \frac{\tanh A + \tanh B}{1+\tanh A\tanh B}$$

$$\tanh(A-B) = \frac{\tanh A - \tanh B}{1-\tanh A\tanh B}$$

$$\log_a xy = \log_a x + \log_a y$$

$$\log_a \frac{x}{y} = \log_a x - \log_a y$$

$$\log_a x^y = y\log_a x$$

$$\log_a x = \log_b x \times \log_a b = \frac{\log_b x}{\log_b a}$$

$$\ln x = \log_{10} x \times \ln 10 = 2.302585 \times \log_{10} x$$

$$\log_{10} x = \ln x \times \log_{10} e = 0.434294 \times \ln x$$

$$e = 2.718281828$$

$$e^A = 1 + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \frac{A^4}{4!} + \ldots$$

$$e^{-A} = 1 - A + \frac{A^2}{2!} - \frac{A^3}{3!} + \frac{A^4}{4!} - \ldots$$

$$\sinh A = \frac{e^A - e^{-A}}{2} = A + \frac{A^3}{3!} + \frac{A^5}{5!} + \frac{A^7}{7!} + \ldots$$

$$\cosh A = \frac{e^A + e^{-A}}{2} = 1 + \frac{A^2}{2!} + \frac{A^4}{4!} + \frac{A^6}{6!} + \ldots$$

$$\sin A = \frac{e^{jA} - e^{-jA}}{2j} = A - \frac{A^3}{3!} + \frac{A^5}{5!} - \frac{A^7}{7!} + \ldots$$

$$\cos A = \frac{e^{jA} + e^{-jA}}{2} = 1 - \frac{A^2}{2!} + \frac{A^4}{4!} - \frac{A^6}{6!} + \ldots$$

$$\tan A = j\frac{e^{-jA} - e^{jA}}{e^{jA} + e^{-jA}} = A + \frac{A^3}{3} + 2\frac{A^5}{15} + 17\frac{A^7}{315} + \cdots$$

Appendix F: Computer Networks

A computer network connects several geographically spread computers so that they can exchange information by message passing, and one computer can access resources available to other computers. If computers are connected by a *network*, then existing information on them can be shared. A computer network can communicate over a wired or wireless medium to share resources.

F.1 Computer Network Types

There are 3three types of computer networks, namely local area networks (LANs), metropolitan area networks (MANs), and wide area networks (WANs). The main difference among these classifications is their area of coverage.

LAN: It is a group of computers and network devices connected together, usually within the same building. LANs are communication networks connecting computers by wire, or wireless link. LANs allow users to share software, hardware, and data. By definition, the connections must be high speed and relatively inexpensive (e.g., token ring or Ethernet).

MAN: It is a data network designed for a town or city. In terms of geographic breadth, MAN is larger than LAN, but smaller than WAN. MAN is usually characterized by very high-speed connections using fibre optic cable or other digital media.

WAN: It is a computer network that covers a broader area, that is, any network whose communication links cross the metropolitan, regional, or national boundaries. The technology is high speed and relatively expensive. The Internet is an example of a worldwide public WAN. WAN uses routers and public communications links.

F.2 Computer Network Architectures

Computer network architecture specifies the design of a computer network, for example, a computer network consisting of multiple connected nodes that communicates over a wired or a wireless medium.

Examples of network architectures are client–server and peer-to-peer networks.

Client/Server Networks: It is a server-based network that allows a network administrator to control the access of the users to different resources to enhance the security. A client–server network also enables the users to share expensive equipment, such as laser printers and mass storage. Each instance of the client software can send requests to a server.

Specific types of servers include web servers, application servers, file servers, terminal servers, and mail servers.

Peer-to-Peer Networks: This architecture relies primarily on the computing power and bandwidth of the participants in the network rather than concentrating it in a relatively low number of servers. They are typically used for connecting nodes via largely ad hoc connections.

Unlike server-based networks, a peer-to-peer network does not use any dedicated server. In a peer-to-peer network, all the computers in a network handle security and administration for themselves.

Appendix G: OSI Model

G.1 Computer Network Standards

Computer networks connect terminals, devices, and computers from many different manufacturers across many types of networks, such as metropolitan areas, local areas, and wide areas. The media of communication can be either wired or wireless. For the different devices on various types of networks to be able to communicate, the network must use similar techniques of moving data through the network from one application to another. For example, an IBM server with UNIX as operating system cannot communicate directly with a Dell PC running Windows XP. Some form of translation must occur for these two types to communicate.

To ease the problems of *incompatibility* and to ensure that hardware and software components can be integrated into any network, organizations such as American National Standard Institute (ANSI) and IEEE propose, develop, and approve network standards. A network standard defines guidelines that specify the way computers access the medium to which they are attached, the type(s) of medium used, the speed used on different types of networks, and the type of the physical cable and/or the wireless technology used.

A standard that outlines characteristics of how two network devices communicate is called a protocol.

A protocol would define packet format, coding scheme, error handling, and sequencing techniques. Software and hardware manufacturers design their products to meet the guidelines specified in a particular standard. The following sections discuss some of the more widely used network standards and protocols for both wired and wireless networks. These network standards and protocols work together to move data through a network. Thus, as data moves through the network from one application to another, it may use one or more of these standards.

G.2 OSI Model

We have stated that protocols allow incompatible systems to communicate. A protocol that would allow any two different types of systems to communicate regardless of their underlying architecture is called an open system. The international organization for standardization (ISO) developed the open systems interconnection model (OSI) reference model to describe how information is transferred from one system to another. When information is converted into electrical or light signals, it is transferred along a piece of wire or radio waves. It is very

Table G.1 OSI model

	Data unit	Layer	Function
User support layer		Application	Network process to application
		Presentation	Data representation and encryption
	Data	Session	Inter host communication
	Segments	Transport	End-to-end connections and reliability (TCP)
Network support layer	Packets	Network	Path determination and logical addressing (IP)
	Frames	Data link	Physical addressing (MAC & LLC)
	Bits	Physical	Media, signal and binary transmission

important to define a set of rules for their transmission. This model covers all aspects of network communication and is known as OSI model or seven layer model (Table G.1).

Figure G.1 shows the data flow from Node A to Node B. Each layer performs a specific function and communicates with the layer directly above and below it. Higher layers deal more with user services, and the lower layers deal more with the actual transmission of data.

Physical layer

Physical layer is required for transmitting a bit stream over a physical medium. It deals with mechanical and electrical specification of the interface and transmission medium. It also defines functions of the physical devices and interface for the transmission to occur.

Data link layer

The data link layer transforms the raw bit stream received from the physical layer into reliable information and is responsible for node to node delivery. It makes the physical layer appear error-free to the upper layers. Some of the responsibilities of data layer are as follows:

1. It deals with the physical address of both the source and the destination.
2. When two or more devices are connected to the same link, data link layer protocols determine which device will have control over the link at any given time.
3. This layer also deals with error control mechanism, that is, it detects and re-transmits the damaged or lost *frame*. It uses a mechanism to prevent duplication of frames.
4. It also deals with the rate at which the sender should send the data to the receiver so that a fast sender does not overwhelm a slow receiver, this mechanism is called *flow control*.

Network layer

The network layer is responsible for source to destination delivery of packets across multiple networks, whereas data link layer's delivery is a hop-to-hop delivery. Network layer assures that each packet gets from its point of origin to its destination.

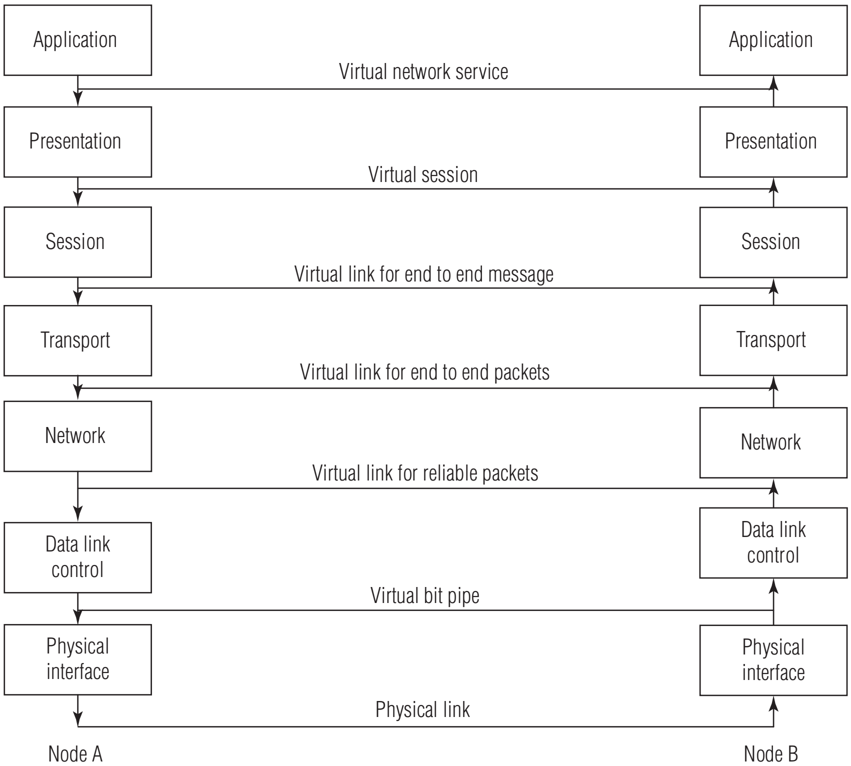


Figure G.1 Data flow from Node A to Node B

The specific responsibilities are as follows:

1. It deals with the logical addressing of sender and receiver.
2. It deals with the routing of data between different links and networks.
3. This layer also deals with the overall congestion control mechanism of the network.
4. It divides the outgoing message into packets and assembles the incoming packets into message.

Transport layer

The transport layer is responsible for process to process delivery of the entire message, that is, this layer ensures that the whole message arrives intact and in order over error control and flow control process to process level. Some responsibilities are as follows:

1. It deals with segmentation and re-assembly of the message with proper sequence number in case the message is very large.
2. It deals with connection control, that is, whether an end to end connection should be connectionless or connection oriented.
3. This layer also deals with flow control of the data, but here the flow control is performed process to process rather than across a single link.
4. This layer also deals with error control so that the receiving transport layer receives, the message without any error or duplication; error correction is usually achieved by re-transmission.

Session layer

The job of the session layer is to establish, maintain, and synchronize the interaction between the communicating systems. Specific responsibilities of session layer are as follows:

1. It allows two systems to start a communication (dialog control) with each other. The communication between two systems is either in half or full duplex mode.
2. This layer allows addition of check points, that is, synchronization points into a stream of data, so that in case of crash during the transmission of data, data can be re-transmitted from the check point last inserted, instead of re-transmitting from the beginning.

Presentation layer

The presentation layer is concerned with the syntax and semantics of information exchange between two communicating systems. Some responsibilities are as follows:

1. It deals with protocol conversion.
2. It carries out data compression to reduce the bandwidth of the data to be transmitted.
3. It translates data between the format that the network requires and the format that the computer expects.
4. For security and privacy purpose, it carries out encryption of data at the sender's end and decryption at the receiver's end.

Application layer

This layer enables the user, whether human or software, to access the network.

Appendix H: Erlang B-Table for 1 to 50 Channels, 0.7%–40%

The task of teletraffic theory is to design systems as cost effective as possible with a predefined grade of service when we know the future traffic demand and the capacity of system elements. Erlang B was developed by A. K. Erlang, for whom the ***erlang*** a unit of traffic density in a telecommunications system was named.

Definition of Erlang: If the call intensity is 5 calls per minute, and the mean service time is 3 minutes then the offered traffic is equal to 15 erlang.

Erlang-B formula is used for the traffic engineering. Erlang B can determine the number of trunks, or lines, needed to handle a calling load during a one-hour period. However, the formula assumes that lost calls are cleared,that is, if callers get a busy signal, they will never retry. The Erlang B formula used to predict the probability that a call will be blocked is:

$$\text{Erlang B loss probability} = B = \frac{a^n / n!}{\sum_{i=0}^{n} \frac{a^i}{i!}}$$

where, n is the number of trunks in full availability group, a is the traffic offered to group in Erlangs.

	Loss probability (*E*)										
n	0.007	0.008	0.009	0.01	0.02	0.03	0.05	0.1	0.2	0.4	*n*
1	.00705	.00806	.00908	.01010	.02041	.03093	.05263	.11111	.25000	.66667	1
2	.12600	.13532	.14416	.15259	.22347	.28155	.38132	.59543	1.0000	2.0000	2
3	.39664	.41757	.43711	.45549	.60221	.71513	.89940	1.2708	1.9299	3.4798	3
4	.77729	.81029	.84085	.86942	1.0923	1.2589	1.5246	2.0454	2.9452	5.0210	4
5	1.2362	1.2810	1.3223	1.3608	1.6571	1.8752	2.2185	2.8811	4.0104	6.5955	5
6	1.7531	1.8093	1.8610	1.9090	2.2759	2.5431	2.9603	3.7584	5.1086	8.1907	6
7	2.3149	2.3820	2.4437	2.5009	2.9354	3.2497	3.7378	4.6662	6.2302	9.7998	7
8	2.9125	2.9902	3.0615	3.1276	3.6271	3.9865	4.5430	5.5971	7.3692	11.419	8
9	3.5395	3.6274	3.7080	3.7825	4.3447	4.7479	5.3702	6.5464	8.5217	13.045	9
10	4.1911	4.2889	4.3784	4.4612	5.0840	5.5294	6.2157	7.5106	9.6850	14.677	10
11	4.8637	4.9709	5.0691	5.1599	5.8415	6.3280	7.0764	8.4871	10.857	16.314	11
12	5.5543	5.6708	5.7774	5.8760	6.6147	7.1410	7.9501	9.4740	12.036	17.954	12
13	6.2607	6.3863	6.5011	6.6072	7.4015	7.9667	8.8349	10.470	13.222	19.598	13
14	6.9811	7.1155	7.2382	7.3517	8.2003	8.8035	9.7295	11.473	14.413	21.243	14
15	7.7139	7.8568	7.9874	8.1080	9.0096	9.6500	10.633	12.484	15.608	22.891	15
16	8.4579	8.6092	8.7474	8.8750	9.8284	10.505	11.544	13.500	16.807	24.541	16
17	9.2119	9.3714	9.5171	9.6516	10.656	11.368	12.461	14.522	18.010	26.192	17
18	9.9751	10.143	10.296	10.437	11.491	12.238	13.385	15.548	19.216	27.844	18
19	10.747	10.922	11.082	11.230	12.333	13.115	14.315	16.579	20.424	29.498	19
20	11.526	11.709	11.876	12.031	13.182	13.997	15.249	17.613	21.635	31.152	20
21	12.312	12.503	12.677	12.838	14.036	14.885	16.189	18.651	22.848	32.808	21
22	13.105	13.303	13.484	13.651	14.896	15.778	17.132	19.692	24.064	34.464	22
23	13.904	14.110	14.297	14.470	15.761	16.675	18.080	20.737	25.281	36.121	23
24	14.709	14.922	15.116	15.295	16.631	17.577	19.031	21.784	26.499	37.779	24
25	15.519	15.739	15.939	16.125	17.505	18.483	19.985	22.833	27.720	39.437	25
26	16.334	16.561	16.768	16.959	18.383	19.392	20.943	23.885	28.941	41.096	26
27	17.153	17.387	17.601	17.797	19.265	20.305	21.904	24.939	30.164	42.755	27
28	17.977	18.218	18.438	18.640	20.150	21.221	22.867	25.995	31.388	44.414	28
29	18.805	19.053	19.279	19.487	21.039	22.140	23.833	27.053	32.614	46.074	29
30	19.637	19.891	20.123	20.337	21.932	23.062	24.802	28.113	33.840	47.735	30
31	20.473	20.734	20.972	21.191	22.827	23.987	25.773	29.174	35.067	49.395	31
32	21.312	21.580	21.823	22.048	23.725	24.914	26.746	30.237	36.295	51.056	32
33	22.155	22.429	22.678	22.909	24.626	25.844	27.721	31.301	37.524	52.718	33
34	23.001	23.281	23.536	23.772	25.529	26.776	28.698	32.367	38.754	54.379	34
35	23.849	24.136	24.397	24.638	26.435	27.711	29.677	33.434	39.985	56.041	35
36	24.701	24.994	25.261	25.507	27.343	28.647	30.657	34.503	41.216	57.703	36

(*Continued*)

n	Loss probability (*E*) 0.007	0.008	0.009	0.01	0.02	0.03	0.05	0.1	0.2	0.4	n
37	25.556	25.854	26.127	26.378	28.254	29.585	31.640	35.572	42.448	59.365	**37**
38	26.413	26.718	26.996	27.252	29.166	30.526	32.624	36.643	43.680	61.028	**38**
39	27.272	27.583	27.867	28.129	30.081	31.468	33.609	37.715	44.913	62.690	**39**
40	28.134	28.451	28.741	29.007	30.997	32.412	34.596	38.787	46.147	64.353	**40**
41	28.999	29.322	29.616	29.888	31.916	33.357	35.584	39.861	47.381	66.016	**41**
42	29.866	30.194	30.494	30.771	32.836	34.305	36.574	40.936	48.616	67.679	**42**
43	30.734	31.069	31.374	31.656	33.758	35.253	37.565	42.011	49.851	69.342	**43**
44	31.605	31.946	32.256	32.543	34.682	36.203	38.557	43.088	51.086	71.006	**44**
45	32.478	32.824	33.140	33.432	35.607	37.155	39.550	44.165	52.322	72.669	**45**
46	33.353	33.705	34.026	34.322	36.534	38.108	40.545	45.243	53.559	74.333	**46**
47	34.230	34.587	34.913	35.215	37.462	39.062	41.540	46.322	54.796	75.997	**47**
48	35.108	35.471	35.803	36.109	38.392	40.018	42.537	47.401	56.033	77.660	**48**
49	35.988	36.357	36.694	37.004	39.323	40.975	43.534	48.481	57.270	79.324	**49**
50	36.870	37.245	37.586	37.901	40.255	41.933	44.533	49.562	58.508	80.988	**50**
	0.007	**0.008**	**0.009**	**0.01**	**0.02**	**0.03**	**0.05**	**0.1**	**0.2**	**0.4**	

Erlang B-Table for 1 to 50 channels, 0.001%–0.6%

n	Loss probability (*E*) 0.00001	0.00005	0.0001	0.0005	0.001	0.002	0.003	0.004	0.005	0.006	n
1	.00001	.00005	.00010	.00050	.00100	.00200	.00301	.00402	.00503	.00604	**1**
2	.00448	.01005	.01425	.03213	.04576	.06534	.08064	.09373	.10540	.11608	**2**
3	.03980	.06849	.08683	.15170	.19384	.24872	.28851	.32099	.34900	.37395	**3**
4	.12855	.19554	.23471	.36236	.43927	.53503	.60209	.65568	.70120	.74124	**4**
5	.27584	.38851	.45195	.64857	.76212	.89986	.99446	1.0692	1.1320	1.1870	**5**
6	.47596	.63923	.72826	.99567	1.1459	1.3252	1.4468	1.5421	1.6218	1.6912	**6**
7	.72378	.93919	1.0541	1.3922	1.5786	1.7984	1.9463	2.0614	2.1575	2.2408	**7**
8	1.0133	1.2816	1.4219	1.8298	2.0513	2.3106	2.4837	2.6181	2.7299	2.8266	**8**
9	1.3391	1.6595	1.8256	2.3016	2.5575	2.8549	3.0526	3.2057	3.3326	3.4422	**9**
10	1.6970	2.0689	2.2601	2.8028	3.0920	3.4265	3.6480	3.8190	3.9607	4.0829	**10**
11	2.0849	2.5059	2.7216	3.3294	3.6511	4.0215	4.2661	4.4545	4.6104	4.7447	**11**
12	2.4958	2.9671	3.2072	3.8781	4.2314	4.6368	4.9038	5.1092	5.2789	5.4250	**12**
13	2.9294	3.4500	3.7136	4.4465	4.8306	5.2700	5.5588	5.7807	5.9638	6.1214	**13**
14	3.3834	3.9523	4.2388	5.0324	5.4464	5.9190	6.2291	6.4670	6.6632	6.8320	**14**
15	3.8559	4.4721	4.7812	5.6339	6.0772	6.5822	6.9130	7.1665	7.3755	7.5552	**15**

(*Continued*)

	Loss probability (*E*)										
n	**0.00001**	**0.00005**	**0.0001**	**0.0005**	**0.001**	**0.002**	**0.003**	**0.004**	**0.005**	**0.006**	***n***
16	4.3453	5.0079	5.3390	6.2496	6.7215	7.2582	7.6091	7.8780	8.0995	8.2898	**16**
17	4.8502	5.5583	5.9110	6.8782	7.3781	7.9457	8.3164	8.6003	8.8340	9.0347	**17**
18	5.3693	6.1220	6.4959	7.5186	8.0459	8.6437	9.0339	9.3324	9.5780	9.7889	**18**
19	5.9016	6.6980	7.0927	8.1698	8.7239	9.3515	9.7606	10.073	10.331	10.552	**19**
20	6.4460	7.2854	7.7005	8.8310	9.4115	10.068	10.496	10.823	11.092	11.322	**20**
21	7.0017	7.8834	8.3186	9.5014	10.108	10.793	11.239	11.580	11.860	12.100	**21**
22	7.5680	8.4926	8.9462	10.180	10.812	11.525	11.989	12.344	12.635	12.885	**22**
23	8.1443	9.1095	9.5826	10.868	11.524	12.265	12.746	13.114	13.416	13.676	**23**
24	8.7298	9.7351	10.227	11.562	12.243	13.011	13.510	13.891	14.204	14.472	**24**
25	9.3240	10.369	10.880	12.264	12.969	13.763	14.279	14.673	14.997	15.274	**25**
26	9.9265	11.010	11.540	12.972	13.701	14.522	15.054	15.461	15.795	16.081	**26**
27	10.537	11.659	12.207	13.686	14.439	15.285	15.835	16.254	16.598	16.893	**27**
28	11.154	12.314	12.880	14.406	15.182	16.054	16.620	17.051	17.406	17.709	**28**
29	11.779	12.976	13.560	15.132	15.930	16.828	17.410	17.853	18.218	18.530	**29**
30	12.417	13.644	14.246	15.863	16.684	17.606	18.204	18.660	19.034	19.355	**30**
31	13.054	14.318	14.937	16.599	17.442	18.389	19.002	19.470	19.854	20.183	**31**
32	13.697	14.998	15.633	17.340	18.205	19.176	19.805	20.284	20.678	21.015	**32**
33	14.346	15.682	16.335	18.085	18.972	19.966	20.611	21.102	21.505	21.850	**33**
34	15.001	16.372	17.041	18.835	19.743	20.761	21.421	21.923	22.336	22.689	**34**
35	15.660	17.067	17.752	19.589	20.517	21.559	22.234	22.748	23.169	23.531	**35**
36	16.325	17.766	18.468	20.347	21.296	22.361	23.050	23.575	24.006	24.376	**36**
37	16.995	18.470	19.188	21.108	22.078	23.166	23.870	24.406	24.846	25.223	**37**
38	17.669	19.178	19.911	21.873	22.864	23.974	24.692	25.240	25.689	26.074	**38**
39	18.348	19.890	20.640	22.642	23.652	24.785	25.518	26.076	26.534	26.926	**39**
40	19.031	20.606	21.372	23.414	24.444	25.599	26.346	26.915	27.382	27.782	**40**
41	19.718	21.326	22.107	24.189	25.239	26.416	27.177	27.756	28.232	28.640	**41**
42	20.409	22.049	22.846	24.967	26.037	27.235	28.010	28.600	29.085	29.500	**42**
43	21.104	22.776	23.587	25.748	26.837	28.057	28.846	29.447	29.940	30.362	**43**
44	21.803	23.507	24.333	26.532	27.641	28.882	29.684	30.295	30.797	31.227	**44**
45	22.505	24.240	25.081	27.319	28.447	29.708	30.525	31.146	31.656	32.093	**45**
46	23.211	24.977	25.833	28.109	29.255	30.538	31.367	31.999	32.517	32.962	**46**
47	23.921	25.717	26.587	28.901	30.066	31.369	32.212	32.854	33.381	33.832	**47**
48	24.633	26.460	27.344	29.696	30.879	32.203	33.059	33.711	34.246	34.704	**48**
49	25.349	27.206	28.104	30.493	31.694	33.039	33.908	34.570	35.113	35.578	**49**
50	26.067	27.954	28.867	31.292	32.512	33.876	34.759	35.431	35.982	36.454	**50**
	0.00001	**0.00005**	**0.0001**	**0.0005**	**0.001**	**0.002**	**0.003**	**0.004**	**0.005**	**0.006**	

List of Abbreviations

1G	First Generation
2G	Second Generation
2.5G	2.5 Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
4G	Fourth Generation
ABS	Analysis by Synthesis
AC	Alerting Channel
ACCH	Associated Control Channels
ACES	ASIA Cellular Satellite
ACI	Adjacent Channel Interference
ACK	Acknowledge Signal
ACL	Asynchronous Connection–Less
ADB	Alternate Billing Service
ADSL	Asymmetric Digital Subscriber Line
AFD	Average Fade Duration
AGC	Automatic Gain Control
AGCH	Access Granted Channel
AIN	Advanced Intelligent Network
AKA	Authentication and Key Agreement
AL	Application Layer
ALT	Automatic Link Transfers
AM	Amplitude Modulation
AMC	Adaptive Modulation and Coding
AMPS	Advanced Mobile Phone Service
AMTS	Automated Mobile Telephone System
ANSI	American National Standard Institute
AOC	Advice of Charge
AP	Access Point
APC	Automatic Power Control
APN	Access Point Name
ARA	Alerting/Registration Area
ARDIS	Advance Radio Data Information Systems
ARQ	Automatic Repeat Request
ARQN	Automatic Repeat Request Number
ASEs	Application Service Elements
ASK	Amplitude Shift Keying
ASN	Access Service Network
ATC	Adaptive Transform Coding
AUC	Authentication Centre
AWGN	Additive White Gaussian Noise
BCA	Borrowing Channel Algorithm
BCC	Blocked Call Cleared
BCCH	Broadcast Control Channel
BCD	Blocked Call Delayed
BCH	Blocked Call Held
BER	Bit Error Rate
BFSK	Binary Frequency Shift Keying
B-ISDN	Broadband ISDN
BLER	Block Error Rate
BoD	Bandwidth on Demand
BPSK	Binary Phase Shift Keying
BRI	Basic Rate Interface
BS	Base Station
BSA	Base Station Antenna
BSC	Base Station Controller
BSIC	Base Site Identity Code

BSS	Base Station Systems
BSS	Basic Service Set
BSSID	BSS Identifier
BTS	Base Transceiver Station
BTS	Base Transceiver Stations
CAI	Common Air Interface
CBCH	Cell Broadcast Channel
CC	Country Code
CC/PP	Capabilities/Preferences Profile
CCA	Clear Channel Assessment
CCCH	Common Control Channel
CCI	Co-Channel Interference
CCITT	Consultative Committee for International Telegraph and Telephone
CCK	Complementary Code Keying
CCS	Common Channel Signalling
CDM	Code Division Multiplexing
CDMA	Code Division Multiple Access
CDMA2000	Code Division Multiple Access 2000
CDPD	Cellular Digital Packet Data
CDR	Charging Data Records
CEPT	Conference of European Posts And Telegraphs
CF	Contention Free
CGI	Common Gateway Interface
CI	Cell Identifier
CIC	Circuit Identification Code
CIR	Carrier To Interference Ratio
CLASS	Custom Local Area Signaling Services
CLNP	Connection Less Network Protocol
CLNs	Connection Less Networking Service
CM	Connection Management
CO	Central Office
CPE	Customer Premise Equipment
CPS	Common Part Sub-Layer
CRC	Cyclic Redundancy Check
CS	Convergence Sub-Layer
CSD	Circuit Switched Data
CSMA	Carrier Sense Multiple Accessing
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
CSMA/CD	Carrier Sense Multiple Accessing With Collision Detection
CSN	Connectivity Service Network
CTS	Clear To Send
CUGs	Closed User Groups
D-AMPS	Digital AMPS
DARPA	Defence Advanced Research Projects Agency
DBPSK	Differential Binary Phase Shift Keying
DCA	Dynamic Channel Algorithm
DCCH	Dedicated Control Channel
DCE	Data Circuit Equipment
DCF	Distributed Coordination Network
DCS1800	Digital Cellular System 1800
DDE	Data Directed Estimation
DECT	Digital Enhanced Cordless Telephone
DFE	Decision Feedback Equalizer
DID	Direct Inward Dialling
DLL	Data Link Layer
DM	Delta Modulation
DPSK	Differential Phase Shift Keying
DQPSK	Differential Quadrature Phase Shift Keying
DRX	Discontinuous Reception
DS	Direct Sequence
DS-CDMA	Direct Sequence Code Division Multiple Access
DSL	Digital Subscriber Lines/Digital Subscriber Loop
DSMA/CD	Digital Sense Multiple Access with Collision Detection
DSP	Digital Signal Processing
DSS	Distribution System Services
DSSS	Direct Sequence Spread Spectrum
DTE	Data Terminal Equipment
DUP	Data User Part
DVBC	Digital Video Broad Casting Cable

DVB-RCS	Digital Video Broadcasting Return Channel through Satellite
DVB-S	Digital Video Broadcasting–Satellite
EC	Echo Canceller
ECMA	European Computer Manufacturers Association
ED	Energy Detection
EDGE	Enhanced Data Rates for GSM Evolution
EFR	Enhanced Full Rate
EGPRS	Enhanced GPRS
EIA	Electronic Industries Association
EM	Electro Magnetic
EOC	Embedded Operations Channel
EPC	Electronic Product Code
ESs	End Systems
ESS	Extended Service Set
ETSI	European Telecommunications Standards Institute
EUL	Enhanced Uplink
EURO-COST	European Co-Operative for Scientific And Technical Research
EV-DO	Evolution-Data Optimized
EV-DV	Evolution for Integrated Data and Voice
FA	Foreign Agent
FAC	Final Assembly Code
FACCH	Fast Associated Control Channels
FB	Front-to-Back
FBCA	Forcible-Borrowing Channel Algorithm
FC	Fast Channel
FCA	Fixed Channel Algorithm
FCA	Fixed Channel Assignment
FCC	Federal Communication Commission
FCCH	Frequency Correction Channel
FDD	Frequency Division Duplex
FDM	Frequency Division Multiplexing
FDMA	Frequency Division Multiple Access
FEC	Forward Error Correction
F-ES	Fixed End System
FFT	Fast Fourier Transform
FH	Frequency Hopped
FH-CDMA	Frequency Hopped CDMA
FHSS	Frequency Hopping Spread Spectrum
FIR	Finite Impulse Response
FL	Forward Link
FM	Frequency Modulation
FOCC	Forward Control Channel
FR	Full Rate
FRA	Fixed Radio Access
FSK	Frequency Shift Keying
FVC	Forward Voice Channels
FWT	Fixed Wireless Terminal
GEO	Geostationary Earth Orbits
GFSK	Gaussian Frequency Shift Keying
GGSN	Gateway GPRS Support Node/ Gateway GSN
GIS	Geographical Information Systems
GMM	GPRS Mobility Management
GMPCS	Global Mobile Personal Communications by Satellites
GMSC	Gateway MSC Server
GMSK	Gaussian Minimum Shift Keying
GMSS	Geostationary Mobile Satellite Standard
GOS	Grade of Service
GPRS	General Packet Radio Service
GPS	Global Positioning System
GSM	Global System for Mobile
GSN	GPRS Support Node
GSU	Globalstar Subscriber User
GTP	GPRS Tunnelling Protocol
GTT	Global Title Translation
GW	Gateway
GWL	Gateway Link
HA	Home Agent
HCA	Hybrid Channel Algorithm
HCI	Host Controller Interface

HCL	Host Controller Interface
HDR	High Data Rate
HDSL	High Speed Digital Subscriber Lines
HEO	Highly Elliptical Orbit
HIPERLAN	High Performance Radio LAN
HLR	Home Location Register
HPBW	Half Power Beamwidth
HPT	High Power Transmitter
HSCSD	High Speed Switched Circuit Data
HSDPA	High Speed Downlink Packet Access
HSOPA	High Speed OFDM Packet Access
HSPA	High Speed Packet Access
HSUPA	High Speed Uplink Packet Access
HTTP	Hyper Text Transfer Protocol
IBSS	Independent BSS
IC	Incoming Call
ID	Identification
IDEN	Integrated Digital Enhanced Network
IDFT	Inverse Discrete Fourier Transform
IETF	Internet Engineering Task Force
IFFT	Inverse Fast Fourier Transform
IFS	Inter Frame Space
IMEI	International Mobile Station Equipment Identity
IMSI	International Mobile Subscriber Identity
IMT	International Mobile Telecommunications
IN	Intelligent Network
IP	Internet Protocol
IPv4	Internet Protocol Version 4
IPv6	Internet Protocol Version 6
IR	Infrared
IrDA	Infrared Data Association
IrDA-SIR	IrDA Serial Infrared Data Link Standard
IrOBEX	IrDA Object Exchange Protocol
IS	Interim Standard
ISDN	Integrated Services Digital Network
ISI	Inter Symbol Interference
ISLs	Inter Satellite Links
ISM	Industrial Scientific and Medical
ISP	Internet Service Provider
ISs	Intermediate Systems
ISUP	ISDN User Part
ISUs	Iridium Subscriber Unit's
ITU	International Telecommunications Union
ITU-T	International Telecommunication Union Technical
IWF	Inter Working Function
IXC	Interexchange Carrier
JDC	Japanese Digital Cellular
Ki	Subscriber Authentication Key
L2CAP	Logical Link Control and Adaption Protocol
L2TP	Layer 2 Tunnelling Protocol
LA	Location Area
LAC	Local Area Code
LAI	Local Area Identity
LAN	Local Area Network
LAPD	Link Access Procedure on the D-Channels
LAPG	Link Access Procedure on the G-Channel
LCR	Level Crossing Rate
LEO	Low Earth Orbits
LFSR	Linear Feedback Shift Register Generator
LIDB	Line Information Database
LIFA	Linear Inverted-F Antenna
LL	Link Layer
LLC	Logical Link Control
LM	Link Manager
LMDS	Local Multipoint Distribution Service
LMP	Link Manager Protocol
LMS	Least Mean Square
LMSI	Local Mobile Subscriber Identity
LoS	Line-of-Sight

LPAS	Linear Prediction Based Analysis by Synthesis
LPC	Linear Predictive Coding
LPDA	Log Periodic Dipole Array
LPDU	Link Protocol Data Units
LPF	Low Pass Filter
LPT	Low Power Transmitter
LQI	Link Quality Indication
LR-WPAN	Low Rate WPAN
LTE	Long Term Evolution
MAC	Medium Access Control
MAHO	Mobile Assisted Handoff
MAN	Metropolitan Area Network
MAP	Mobile Application Part
MC	Multi Carrier
MCC	Mobile Country Code
MC-CDMA	Multi Carrier Code Division Multiple Access
MCHO	Mobile Controlled Handoff
MCS	Multiple Modulation Coding Schemes
MCSS	Multi Carrier Spread Spectrum
MDBS	Mobile Data Base Station
MD-IS	Mobile Data Intermediate System
MDLP	Mobile Data Link Protocol
ME	Mobile Equipment
ME	Mobile Equipment
MEG	Mean Effective Gain
MEO	Medium Earth Orbits
M-ES	Mobile End System
MFSK	M-Ary or Multiple FSK
MGCF	Media Gateway Control Function
MGW	Media Gateway
MIMO	Multi Input Multi Output/ Multiple Input Multiple Output
MIPS	Millions of Instructions per Second
MISO	Multiple Input Single Output
MLME	MAC Sub Layer Management Entity
MLMESAP	MAC Sub Layer Management Entity Service Access Point
MLS	Maximum Length Sequences
MLSE	Maximum Likelihood Sequence Estimation
MM	Mobility Management
MMDS	Multichannel Multipoint Distribution Service
MMS	Multi Media Service
MNC	Mobile Network Code
MNO	Mobile Network Operator
MNs	Mobile Nodes
MOS	Mean Opinion Scores
MPDU	MAC Protocol Data Units
MS	Mobile Station
MSAs	Metropolitan Statistical Areas
MSC	Mobile Switching Centre
MSC	Mobile Switching Centre/ Mobile Services Switching Centre
MSE	Mean Square Error
MSISDN	Mobile Station International Service Digital Network
MSK	Minimum Shift Keying
MSRN	Mobile Station Roaming Number
MSS	Mobile Satellite Services
MTP	Message Transfer Protocol
MTS	Mobile Telephone Service
MTSO	Mobile Telephone Switching Centre
NACK	Negative ACK
NAMTS	Nippon Advanced Mobile Telephone System
NAT-PT	Network Address Translation Protocol Translation
NAV	Network Allocation Vector
NB	Narrow Band
NCC	Network Control Centres
NCHO	Network Controlled Handoff
NDC	National Destination Code
NF	Non Forwarder
NLMS	Normalized Least Mean Square
NLoS	Non-Line-of-Sight
NLPI	Network Layer Protocol Identifier
NMC	Network Management Centre
NMT	Nordic Mobile Telephone

NOP	Network and Operations Plan
NPDU	Network Protocol Data Units
NRZ	Non Return to Zero
NSP	Network Service Part
NSS	Network Switching Subsystem/ Network Subsystem
NWG	Network Working Group
OBP	On-Board Processor
OC	Originating Call
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
OMA	Open Mobile Alliance
OMAP	Operations Maintenance and Administration Part
OMC	Operations and Maintenance Centre
OQPSK	Offset Quadrature Phase Shift Keying
OSI	Open Systems Interconnect
OSS	Operation and Maintenance/ Support Subsystem
OVSF	Orthogonal Variable Spreading Factor
PACS	Personal Access Communication System
PAM	Pulse Amplitude Modulation
PAN	Personal Area Network
PAPR	Peak to Average Power Ratio
PBX	Private Branch Exchange
PCCCH	Packet Common Control Channel
PCF	Point Coordination Function
PCH	Paging Channel
PCM	Pulse Code Modulation
PCM	Pulse Code Modulation
PCS	Personal Communication System/Personal Communication Service
PCU	Packet Control Unit
PD	Protocol Discriminator
PDA	Personal Digital Assistants
PDC	Personal Digital Cellular
PDCHs	Packet Data Channels
PDF	Probability Density Function
PDN	Public Data Network
PDNs	Packet Data Networks
PDU	Protocol Data Units
PER	Packet Error Rate
PEV	Plug In Electric Vehicle
PF	Point Coordinator
PHS	Personal Handy Phone System
PHY	Physical Layer
PIFA	Planer Inverted-F Antenna
PIMs	Passive Inter-Modulation Products
PIN	Personal Identity Number
PL	Path Loss
PLCP	Physical Layer Convergence Procedure
PLL	Phase Locked Loop
PLME	Physical Layer Management Entity
PLMN	Public Land Mobile Network
PLP	Packet Layer Protocol
PMD	Physical Medium Dependent
PPCH	Packet Paging Channel
PPDU	PHY Protocol Data Units/PLCP Protocol Data Unit
PPM	Pulse Position Modulation
PPP	Point-to-Point Protocol
PRACH	Packet Random Access Channel
PRC	Priority Request Channel
PRI	Primary Rate Interface
PRMA	Packet Reservation Multiple Access
PRN	Pseudo Random Noise
PSDN	Packet Switched Data Network/ Public Switch Data Network
PSDU	PLCP Service Data Unit
PSF	PLCP Signalling Field
PSK	Phase Shift Keying
PSPDN	Packet Switched Public Data Network
PSTN	Public Switched Telephone Network
PTP	Point To Point
PTT	Push To Talk
PUSC	Partial Usage of Sub-Layer
PVC	Permanent Virtual Circuit

QAM	Quadrature Amplitude Modulation
QHA	Quadrifilar Helix Antenna
QI	Quality Indicator
Qos	Quality of Service
QPSK	Quadrature Phase Shift Keying
QSAFA	Quasi-Static Automatic Frequency Assignment
RA	Routing Area
RACH	Random Access Channel
RAI	Routing Area Identity
RCC	Reverse Control Channels
Rcom	Reliance Communications
RDF	Resource Description Framework
RECC	Reverse Control Channel
RF	Radio Frequency
RFCOMM	Radio Frequency Communication
RFID	Radio Frequency Identification
RITL	Radio on the Loop
RL	Reverse Link
RLC	Radio Link Control
RMD	RAM Mobile Data
RNC	Radio Network Controller
RPs	Radio Ports
RR	Radio Resource Management
RRME	Radio Resource Management Entity
RSAs	Rural Statistical Areas
R-SGW	Roaming Signalling Gateway
RSS	Received Signal Strength
RSSI	Received Signal Strength Indicator
RTS	Request To Send
RVC	Reverse Voice Channels
SA	Source Address
SACCH	Slow Associated Control Channels
SAN	Satellite Access Node
SAR	Specific Absorption Rate
SAT	Supervisory Audio Tone
SBC	System Broadcast Channel
SC	Slow Channel
SCC	Satellite Control Centre
SCCP	Signalling Connection Control Pan
SCFDMA	Single Carrier Frequency Division Multiple Access
SCH	Synchronizing Channel
SCO	Synchronous Connection Oriented
SCP	Service Control Point
SCPC	Single Channel per Carrier
SDCCH	Standalone Dedicated Control Channel
SDM	Spatial Division Multiplexing
SDSL	Symmetric Digital Subscriber Line
SF	Spreading Factor
SFD	Start of Frame Delimiter
SFIR	Spatial Filtering For Interference Reduction
SGSN	Serving GPRS Support Node
SGSN	Serving GSN
SHO	Soft Handoff
SIC	System Information Channel
SIFS	Short Inter Frame Spacing
SIG	Special Internet Group
SIM	Subscriber Identity Module
SIMO	Single Input Multiple Outputs
SINR	Signal to Interference plus Noise Ratio
SIR	Signal to Interference Ratio
SISO	Single Input Single Output
SLIP	Serial Line Internet Protocol
SM	Short Message
SME	Security Management Entity
SMS	Short Messaging Service
SMSC	SMS Centre
SMS-GMSC	SMS Gateway MSC
SMS-IWMSC	SMS Interworking MSC
SN	Service Node
SNDCP	Sub Network Dependent Convergence Protocol
SNR	Signal to Noise Ratio
SOFDMA	Scalable Orthogonal Frequency Division Multiple Access
SRES	Signed Response
SS	Spread Spectrum
SS7	Signalling System7

SSCS	Service Specific Convergence Sub Layer
SSLE	Symbol Spaced Linear Equalizer
SSMA	Spread Spectrum Multiple Access
SSNs	Sub System Numbers
SSP	Service Switching Points
STP	Signal Transfer Point
SVCs	Switched Virtual Circuits
TA	Transmitter Address
TAC	Type Approval Code
TACS	Total Access Communication System
TCAP	Transaction Capability Application Part
TCH	Traffic Channels
TCO	Total Cost of Ownership
TCS	Telephony Control Exchange Protocol
TD	Transmit Diversity
TD-CDMA	Time Division-Code Division Multiple Access
TDD	Time Division Duplex
TDL	Tapped Delay Line
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TD-SCDMA	Time Division Synchronous Code Division Multiple Access
TEI	Temporary Equipment Identifier
TH	Time Hopped
THSS	Time Hopping Spread Spectrum
TI	Transaction Identifier
TIA	Telecommunications Industry Association
TID	Tunnel Identifier
TL	Transfer Layer
TLLI	Temporary Logical Link Identity
TLS	Transport Layer Security
TMSI	Temporary Mobile Subscriber Identity
TMSI	Temporary Mobile Subscriber Identity
TR	Terminating Call
TRAU	Transcoder Rate Adaption Unit
T-SGW	Transport Media Gateway
TST	Time Slot Transfer
TUP	Telephone User Part
UART	Universal Asynchronous Receiver/Transmitter
UDP	User Datagram Protocol
UE	User Equipment
UF	Urban Factor
UHF	Ultra High Frequency
UMB	Ultra Mobile Broadband
UMC-136	Universal Wireless Communications-136
UML	User Mobile Link
UMTS	Universal Mobile Telecommunications System
UNII	Unlicensed National Infrastructure for Information
UPT	Universal Personal Telecommunication
URL	Uniform Resource Locators
USIM	UMTS Subscriber Identity Module
USSD	Unstructured Supplementary Services Data
UTRA	UMTS Terrestrial Radio Access
UWB	Ultra Wide Band
VAD	Voice Activity Detector
VCO	Voltage Controlled Oscillator
VCs	Virtual Circuits
VHF	Very High Frequency
VLR	Visitor Location Register
VoIP	Voice over Internet Protocol
VPN	Virtual Private Networking
VSAT	Very Small Aperture Terminal
VSF- CDMA	Variable Spreading Factor Code Division Multiple Access
VSF-OFCDM	Variable Spreading Factor Orthogonal Frequency and Code Division Multiplexing
W3C	World-Wide Web Consortium
WACS	Wireless Access Communication System
WAE	Wireless Application Environment
WAP	Wireless Access Protocol

WAP WTAI	WTA Application Programming Interface
WARC	World Administrative Radio Congress
WATM	Wireless Asynchronous Transfer Mode
WCDMA	Wideband Code Division Multiple Access
WDP	Wireless Datagram Protocol
WEI	Word Error Indicator
WEP	Wireless Equivalent Privacy
Wi-Fi	Wireless Fidelity
WiMAX	World Wide Interoperability for Microwave Access
WINCs	Wireless Network Interface Controllers
WLAN	Wireless Local Area Network
WLL	Wireless Local Loop
WMAN	Wireless Metropolitan Area Networks
WML	Wireless Markup Language
WPAN	Wireless Personal Area Network
WSP	Wireless Session Protocol
WTAs	Wireless Telephony Applications
WTP	Wireless Transaction Protocol
WUPE	Wireless User Premises Equipment
WWAN	Wireless Wide Area Network
WWW	World Wide Web
XCDR	Transcoder
XML	Extendable Markup Language
ZC	Zigbee Coordinator
ZED	Zigbee End Device
ZF	Zero Forcing
ZR	Zigbee Router

Glossary

1G	First generation cellular systems using analogue circuit-switched technology with FDMA (e.g. AMPS).
2G	Second generation systems are digital cellular including TDMA, CDMA, and GSM systems which support both data and voice services.
2.5G	It is extended to 2G which introduces new data-oriented services, packet switching capabilities, and applications. GPRS is the best-known 2.5G technology.
3G	Next generation system refers to a family of new air interfaces multiple access networks which provide global roaming with higher date rates (up to 384 kbps) to facilitate real-time applications like video and broadband Internet access to mobile phones.
4G	The next generation world wide area radio technology for high-speed wireless communications with new data services.
Access point (AP)	The AP is a wireless LAN transceiver or "base station" that can connect one or many wireless devices simultaneously to the Internet.
Adjacent channel interference (ACI)	The ACI is also known as inter channel interference, it is the interference caused by extraneous power from signal in an adjacent channel.
Ad-hoc network	An ad-hoc network topology is comprised of two or more computers communicating directly with each other using wireless network cards.
ALOHA	It is a simple protocol that allows multiple users to transmit data packets whenever they have data through the same communication channel.

Advanced mobile phone service (AMPS)	AMPS is also referred to as Analog Mobile phone Service. It was one of the leading analog cellular phone systems in the US, Mexico, Canada other countries. It uses the 800 MHz to 900 MHz frequency band.
Analogue modulation	Its aim is to transfer analogue signal, such as speech or TV signal, over band pass channel.
Antenna	A device used for transmitting and receiving electromagnetic waves.
Asynchronous transfer mode (ATM)	It is a standard for cell-based relay that carries voice, video, and data at speeds up to 2.488 Gbps.
Automatic gain control (AGC)	It is a system which holds the gain and, accordingly, the output of a receiver substantially remains constant in spite of input-signal amplitude fluctuations.
Bandwidth in bits per second	Speed of bit transmission (number of bits per second)/ transmission with modulation (channel/link/network can transmit).
Bandwidth in Hertz	It describes the transmission capacity of a communication channel in terms of a range of frequencies available to the system.
Base station controller	It handles radio channel setup, frequency hopping, and handovers of one or more BSs.
Base station	A low power transmission and reception station at a fixed location supporting radio access by mobile users within a given range, typically a cell site.
Base transceiver station	A radio transceiver that handles the radio link protocols with the MS.
Baud rate	It refers to the number of signal units per second that are required to represent these bits.
Beamwidth	It is the angle between the two directions in which the radiated power is half the maximum value of the beam.
Binary phase shift keying	It is an M-ary modulation scheme, with $M = 2$ in which phase of the sinusoidal carrier is changed according to the data bit to be transmitted. Number of bits sent in 1 s.
Bit rate	Baud rate × number of bits/signal element.
Bluetooth	A short-range radio technology for ad-hoc wireless communication of voice and data anywhere in the world.
Carrier sense multiple access (CSMA)	It is a method of listening to the shared communication channel before transmitting the data and while transmitting.

Carrier- to-interference ratio (CIR)	It is defined as the ratio of the desired average signal power at receiver to the total average interference power.
CDMA	Also known as IS-96, it is a multiple access method based on spread spectrum in which different users transmit on (approximately) the same carrier frequency, but use different spreading codes for different users.
CDMA2000	It is a radio access technique evolved as a popular cellular standard of 3G technology for its networks.
CDMAOne	IS-95 CDMA standard developed by Qualcomm is used to describe cellular networks based on IS-95A and IS-95B technology.
Cell	A city or other geographical area is divided into smaller areas called "cells", each having a low-powered radio transmitter/ receiver or base station to send or receive radio signals.
Cell sectoring	It is another way to increase the capacity by splitting each cell into multiple independent sectors.
Cell splitting	It is a process of subdividing a congested cell into smaller cells, thereby increasing capacity of cellular system.
Cellular digital packet data (CDPD)	It is a mobile data technology designed and deployed over analogue wide area networks, typically AMPS cellular networks.
Channel assignment	It does the allocation of specific channel to the cell sites and mobile units.
Control channel	A channel used for transmission of control information from base station to mobile station or mobile station to base station.
Decibel	Logarithmic way to express any value.
Diffraction	It is the ability of radio waves to turn sharp corners and bend around obstacles.
Digital (D)-AMPS	It is also known as IS-54 standard based on the AMPS system.
Digital cellular system (DCS)	One of the European 2G system of 1800 MHz GSM band which employs TDMA technique.
Digital enhanced cordless telephone (DECT)	It is a multicarrier/TDMA/TDD radio access system standard for cordless communications in residential, corporate, and public environments.
Digital modulation	A digital bit stream should be transferred over the band pass channel.
Digital subscriber line (DSL)	It is for a new technology that allows digital data to be sent over an ordinary copper telephone line at high speed.

Direct sequence (DS)	It is one form of spread spectrum technology of wireless communication in which the user's original signal is transformed into another form by multiplying with fast code sequence to increase the transmission bandwidth.
Directivity	This parameter tells us how well the antenna radiates in a given direction, compared to the total radiated power.
Distance to reuse ratio (D/R)	It defines the geographic distance required between cells that use same frequencies.
Diversity	It is a technique for improving the transmission of a signal, by receiving and processing multiple versions of the same transmitted signal, thereby reducing the effects of fading.
Doppler spread	It is measure of how rapidly the transmitted baseband signal changes as compared to the rate of change of the channel.
DQPSK	Differential quadrature phase shift keying (DQPSK) modulation uses differential encoding of the digital information stream.
Duplexer	It acts as signal separating filter which allows a transmitter operating on one frequency and a receiver operating on other frequency to share the same antenna.
Efficiency	It is defined as the total power radiated by the antenna divided by the input power to the antenna.
Electronic serial number (ESN)	It is a unique serial identification number assigned to each cellular phone which is automatically reached to cellular base station when a call is placed so that the wireless carrier processes the call.
Enhanced data rates for GSM evolution (EDGE)	This was designed specifically as an upgrade to GPRS for integration into GSM network. It supports data rates up to 384 kbps and is based on GSM technology that is being developed for TDMA networks.
Enhancer	It is used in an area that is a hole (weak spot) in the serving cell site.
Erlang	One Erlang is equivalent to an arrival rate of one call per minute multiplied by a holding time of one minute.
Ethernet	A group of standards for defining a LAN.
Fading	The performance of communication channel changes due to environmental changes causes time variations of the received signal strength over a radio link.
Forward error correction (FEC)	A technique used to combat the interference and fading effects of the cellular channels.

Frequency diversity	The multiple versions of signal received following different propagation paths being transmitted at different frequencies.
Frequency division multiple access(FDMA)	FDMA technique is the simplest scheme used to provide multiple access in analogue transmission in which different users transmit at different carrier frequencies.
Frequency hopping spread spectrum (FHSS)	It is a method of transmitting signals by rapidly switching channels, using a pseudo sequence known to both the transmitter and receiver.
Frequency management	It includes operations such as designation setup channels and the voice channels, numbering the channels, and grouping voice channels into subsets.
Frequency modulation	It is a method in which the instantaneous frequency of the radio-frequency wave (i.e. carrier) is varied in accordance with the modulating signal.
Frequency reuse	It refers to the usage of the same frequency carrier in different geographical locations that are distant enough by using basic design procedure.
Frequency shift keying	In this digital modulation scheme information data bits (logic 0 or logic 1) are represented by two different frequencies of the carrier signal.
Gaussian minimum shift keying	It is a binary digital modulation technique which uses Gaussian pulse for pulse shaping.
General packet radio service (GPRS)	It is an extension of the GSM system that uses packet-switching based data service which supports data rates up to 115 kbps.
Global positioning system (GPS)	A system of satellites computers use to measure receiver position using trilateration method.
Global system for mobile (GSM)	It was the first commercially operated digital cellular system and uses TDMA/FDD.
Half duplex	Two way communication, but uses the same radio channel for both transmission and reception.
Handoff	The process of transferring an active call from one radio frequency in one cell to another radio frequency in another successfully.
High-speed circuit switched data (HSCSD)	It is the easiest way to increase the GSM channel data rates up to 14.4 kbps by using multiple time slots.
IEEE 802.11	An international standard describing the characteristics of a wireless local area network. Its family of standards is often referred to as wireless fidelity (Wi-Fi).

IEEE 802.11g	The next extension to the 802.11 standard operates at the 2.4 GHz ISM band and allows for data rates ranging from 1 to 54 Mbps.
IEEE 802.16	Its family of standards is often referred to as Worldwide interoperability for microwave access (WiMAX).
IEEE802.11a	It is another PHY extension to the 802.11 standard that operates at the 5 GHz unlicensed band and allows for data rates of 6 to 54 Mbps.
IEEE802.11b	It is a PHY extension to the original 802.11 standard that operates at the 2.40 GHz band and allows for higher data rates of 5.5 and 11 Mbps.
Integrated services digital network (ISDN)	ISDN is one such unified service that can carry the voice, fax, and data simultaneously on a single line in a digital manner.
Internet	When two or more networks are connected they become internetwork. Individual networks are joined into internet using internetworking devices. They are routers and gateways.
IS-136	The latest generation of the digital standard time division multiple access (TDMA) technology.
IS-95	IS-95 is commonly referred to as CDMAone standard and is used in the United States and parts of Asia.
ISM	Industrial, scientific, and medical are a band of the radio spectrum that operate at 900 MHz to 2.4 GHz.
International Telecommunication Union (ITU)	ITU establishes regulations on international use of telegraph, telephone and radio and satellite communications services. ITU has its functions in managing Radio Transmission Technology (RTT) evaluation process and 3G spectrum allocations.
Local area network (LAN)	A network covering a limited area and allowing many users to share devices and data.
Macrocell	These are mainly used to cover large areas (radii between 1 and 10 km) with low traffic densities.
Medium access control (MAC)	The component of the data link layer that controls multi-access communications.
Microcell	Microcells are used in urban areas (radii between 200 m and 1 km) with high traffic density.
Minimum shift keying (MSK)	This is one such scheme of a continuous envelope digital modulation.
Mobile telephone switching office (MTSO)	MTSO coordinates the actions of the BSs, routing calls to them and controlling them as required.

Multimedia services (MMS)	It includes picture messaging, mobile internet browsing, and so on.
Multipath	It is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths.
Orthogonal frequency division multiplexing (OFDM)	It distributes the data over a large number of carriers that are spaced apart at precise frequencies.
Paging	A low-cost service used for sending information to a person on move.
Personal communication system (PCS)	It allows mobility of both users and services and operates in the 1900 MHz band.
Personal digital assistants (PDA)	A portable computing device used to transmit simple personal information functions.
Personal digital cellular (PDC)	One of the 2G system which operates in two frequency bands of 800 MHz and 1500 MHz.
Phase shift keying (PSK)	It comprises the manipulation of a carrier's phase in accordance with the transmitted bit stream.
Protocol	Set of rules that govern data communication.
Radiation pattern	A graphical description of the radiation of an antenna.
Repeaters	These are used in wireless communication for extending the range of the reception of the receiver.
Scattering	It occurs when the wave travels through or reflects from an object with dimensions smaller than the wavelength.
Short messaging service (SMS)	It is a method of exchanging alphanumeric messages of up to 160 characters within seconds of submission of the message.
Signal-to-noise ratio (SNR)	The ratio of received signal strength to the background noise.
Spread spectrum	It is used to allow multiple users to occupy the same transmission band for simultaneous transmission of signals without considerable interference.
Subscriber identity module (SIM) card	It is a "smart card" which plugs into the mobile equipment and contains information about the MS subscriber.
TCP/IP	A simplified network protocol used for movement of data/information across the multiple networks/Internet.
Throughput	It is a measure of how fast we can send the data through network.

Time division multiple access (TDMA)	This system divides the radio spectrum into time slots and each user is allowed to either transmit or receive in each time slots.
Ultra wideband (UWB)	It is a spread spectrum technique which provides high data rates and low power consumption capabilities for short range wireless networks.
UMTS	Universal mobile telecommunication system is an example of 3G systems.
Voice over IP (VoIP)	Voice over IP is the current standard for voice communication over data networks.
Wide area network (WAN)	It is a computer network that exchanges information across wide geographic areas.
Worldwide interoperability for microwave access (**WiMAX**)	It provides broadband wireless connectivity across a large geographical area such as a large metropolitan city.
Wireless application protocol (WAP)	A universal open standard which allows mobile users of wireless phones and other wireless terminals to access to telephony and information services, including the Internet and the Web.
Wireless CDMA (WCDMA)	A technology which provides higher capacity for voice and data at higher data rates of wideband digital radio communication applications.
Wireless fidelity (**Wi-Fi**)	It can be used to provide high-speed connections to wireless local area networks located within a radius of several dozen metres indoors (in general 20 m–50 m away) or within several hundred metres outdoors.
Wireless LAN (WLAN)	It transfers data through the air using radio waves.
Wireless Local Loop (**WLL**)	It is a system that connects subscribers to the local telephone station wirelessly.
Zigbee	A standard wireless networking protocol targeted towards automation and remote control applications.

List of Symbols

E_b	Energy per bit
η_b	Bandwidth efficiency factor
B_C	Coherence bandwidth/channel spacing
v_f	Voice activity factor
r_v	Voltage reflection coefficient
I_0	Interference density
σ_t	RMS delay spread
ξ	Trunking efficiency
μ	Bit efficiency
f_c	Carrier frequency
τ	Mean excess delay
R_L	Return loss
σ	Mean spacing factor
A	offered load
A^*	carried traffic load
A_c	Carrier Load
A_e	effective area of the antenna/effective aperture
A_{USER}	offered traffic intensity or offered load
B	channel bandwidth
B_D	Doppler spread
B_g	Guard band width
B_S	signal bandwidth
c	speed of light
C	Cellular system capacity/channel capacity
C/I	carrier-to-interference ratio
D	reuse distance
D/R ratio	frequency reuse factor/distance to reuse ratio
dB	Decibels
E	Electrical field
η	Efficiency
F/B ratio	Front-to-back ratio
f_m	Maximum Doppler spread
G	Antenna gain
G_p	Processing gain
H	Holding time
H	Magnetic field
k_f	Frequency sensitivity
L	Filter cut-off slope
N	Cluster Size
P_b	Probability of blocking
P_e	Probability of error
q	Frequency reuse factor/co-channel reuse ratio/co-channel interference reduction factor/co-channel reuse factor
R	Cell Radius
R_b	Bit rate
R_c	Chip rate
R_r	Radiation resistance
S	Normalized throughput
S/I	Signal to noise ratio
T_f	Frame duration
T_S	Symbol duration
v	path loss exponent
Z	Impedance
β	Modulation index of FM signal/degradation factor/positive step size
ε_c	Dielectric constant

ε_r	Relative permittivity
η_s	Spectrum efficiency
λ	Wavelength
σ	Standard deviation
α	Protection ratio
γ	Path loss exponential constant/path loss exponent/propagation path-loss slope

Index

A

B

D

E

F

G

H

I

N

O

X

Y

Z