WIRELESSCOMMUNICATIONS

EC412WIRELESS COMMUNICATIONS & CELLULAR NETWORKS

UNIT – I

INTRODUCTION TO MOBILE COMMUNICATION: Evolution of Mobile Radio Communication, Mobile Radio Telephony in US and around the world, Examples of Wireless Communication Systems: Paging system, Cordless telephones systems, Cellular telephone Systems, Trends in Cellular Radio and personal Communications. The Cellular concept: Frequency reuse, Channel Assignment strategies, Hand off Strategies, Interference and system capacity, improving coverage and capacity in cellular systems.

UNIT – II

MOBILE RADIO PROPAGATION: Large Scale Fading: Introduction, Free space propagation model, Relating power to electric field, The Three basic propagation mechanisms: Reflection, Ground reflection (Two-Ray) model, Diffraction, scattering, Practical Link budget design using path loss models.

Small Scale Fading: Small-scale Multipath Propagation, Impulse response model of a multipath channel, Parameters of mobile multipath channels, Types of small scale fading: Fading effects due to multipath time delay spread and Doppler spread Rayleigh and Ricean distributions.

UNIT – III

Equalization: Fundamentals of equalizers, Equalizers in a communication receiver, Linear equalizers, Nonlinear equalizers: Decision feedback equalizers, Maximum likelihood sequence Estimation (MLSE) equalizer. Diversity Techniques: Space diversity: Selection diversity, feedback, MRC, EGC diversity, Polarization diversity, Frequency diversity, Time diversity, Rake Receiver.

UNIT – IV

Multiple Access in Wireless communications: Principle and applications of Multiple Access Techniques FDMA, TDMA, CDMA, Spread Spectrum Multiple Access.

UNIT – V

Wireless Generations Technologies up to 3G:1G, TDMA-based 2G, IS-95, 2.5G, 3G development, Air interface technologies, Internet speeds of 2G, 2.5G, and 3G technologies, Limitations of 3G, Quality of services (QOS) in 3G. 4GTechnology:4G evolution,

TEXT / REFERENCE BOOKS:

1) WCY Lee, Mobile Cellular Telecommunications Systems, McGraw Hill, 1990.

2) WCY Lee, Mobile Communications Design Fundamentals, Prentice Hall, 1993.

3) Raymond Steele, Mobile Radio Communications, IEEE Press, New York, 1992.

4) AJ Viterbi, CDMA: Principles of Spread Spectrum Communications, Addison Wesley, 1995.

5) VK Garg &JE Wilkes, Wireless & Personal Communication Systems, Prentice Hall, 1996.

UNIT-1
IntroductiontoWirelessCommunicationSystems

The ability to communicate with people on the move has evolved remarkably since Guglielmo Marconi first demonstrated radio's ability to provide continuous contact with ships sailing the English Channel in 1897. Since then new wireless communications methods and services have been enthusiastically adopted by people throughout the world. Particularly during the past ten years, the mobile radio communications industry has grown by orders of magnitude, fueled by digital and RF circuit fabrication improvements, new large-scale circuit integration, and other miniaturization technologies which make portable radio equipment smaller, cheaper, and more reliable. Digital switching techniques have facilitated the large scale deployment of affordable, easy-to-use radio communication networks.

Evolution ofMobile RadioCommunications

The ability to provide wireless communications to an entire population was not even conceived until Bell Laboratories developed the cellular concept in the 1960s and 1970s. With the development of highly reliable, miniature, solid-state radio frequency hardware in the 1970s, the wireless communications era was born.

The following market penetration data show how wireless communications in the consumer sectorhasgrown in popularity.Figure 1.1 illustrates how mobile telephony has penetrated our daily lives compared with other popular inventions of the 20th century. Figure 1.1 shows that the first 35 years ofmobile telephony saw little market penetration due to high cost and the technological challenges involved, but however, in the past decade, wireless communications has been accepted by consumers at rates comparable to television and the video cassette recorder.

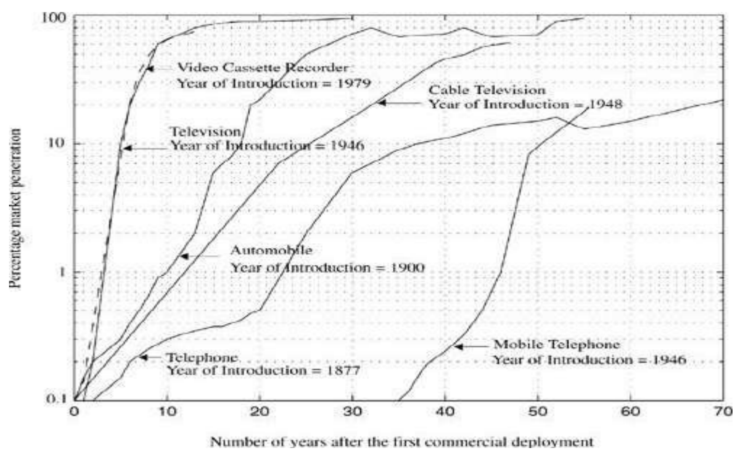
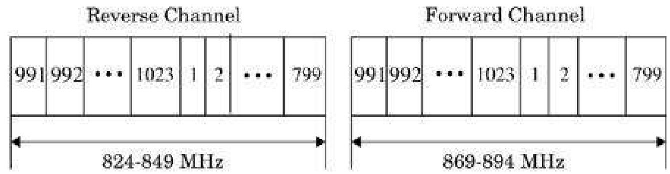


Figure 1.1. The growth of mobile telephony as compared with other popular inventions of the 20th century.

- By 1934, 194 municipal police radio systems and 58 state police stations had adopted amplitude modulation (AM) mobile communication systems for public safety in the U.S.
- In 1935, Edwin Armstrong demonstrated frequency modulation (FM) for the first time, and since the late 1930s, FM has been the primary modulation technique used for mobile communication systems throughout the world.
- With the boom in CB radio and cordless appliances suchas garage door openers and telephones, the number of users of mobile and portable radio in 1995 was about 100 million, or 37% of the U.S. population
- The number of worldwide cellular telephone users grew from 25,000 in 1984 to about 25 million in 1993, and since then subscription-based wireless services have been experiencing customer growth rates well in excess of 50% per year. At the beginning of the 21st century, over 1% of the worldwide wireless subscriber population had already abandoned wired telephone service for home use, and had begun to rely solely on their cellular service provider for telephone access.

MobileRadiotelephonyintheU.S.

In 1946, the first public mobile telephone service was introduced in twenty-five major American cities. Each systemuseda single, high-poweredtransmitterand largetowerinorder tocover distancesof over 50kmina particular market. During the 1950s and 1960s, AT&T Bell Laboratories and other telecommunications companies throughout the world developed the theory and techniques of cellular radiotelephony—the conceptof breaking acoveragezone(market)intosmallcells, eachof whichreuseportionsof thespectrumto increase spectrum usage at the expense of greater system infrastructure. AT&T proposed the concept of a cellular mobile system to the FCC in 1968, although technology was not available to implement cellular telephony until the late 1970s. In 1983, the FCCfinally allocated 666 duplex channels (40 MHz of spectrum in the 800 MHz band, each channel having a one-way bandwidth of 30 kHz for a total spectrum occupancy of 60 kHz for each duplex channel) for the U.S. Advanced Mobile Phone System (AMPS).



	Channel Number	Center Frequency (MHz)
Reverse Channel	$1 \le N \le 799$	$0.030N + 825.0$
	$991 \le N \le 1023$	$0.030(N - 1023) + 825.0$
Forward Channel	$1 \le N \le 799$	$0.030N + 870.0$
	$991 \le N \le 1023$	$0.030(N - 1023) + 870.0$

(Channels 800–990 are unused)

Figure1.2.FrequencyspectrumallocationfortheU.S.cellularradioservice

In late 1991, the first US Digital Cellular (USDC) system hardware was installed in major U.S. cities. The USDC standard(Electronic IndustryAssociationInterimStandardIS-54andlater IS-136)allowedcellular operators to replace gracefully some single-user analog channels with digital channels which support three users in the same 30 kHz bandwidth. In this way, U.S. carriers gradually phased out AMPS as more users accepted digital phones

A cellular system based on code division multiple access (CDMA) has been developed by Qualcomm, Inc. and standardized by the Telecommunications Industry Association (TIA) as an Interim Standard (IS-95). This system supports a variable number of users in 1.25 MHz wide channels using direct sequence spread spectrum. CDMA systems can operate at much larger interference levels because of theirinherent interference resistance properties. The ability of CDMA to operate with a much smaller signal-to-noise ratio (SNR) than conventional narrowbandFM techniques allowsCDMA systems to use the same set of frequencies in every cell, which provides a large improvement in capacity.

Personal Communication Service (PCS) licenses in the 1800/1900 MHz band were auctioned by the U.S. Government to wireless providers in early 1995, and these have spawned new wireless services that complement, as well as compete with, cellular and SMR.

MobileRadioSystemsAroundtheWorld

Many mobile radio standards have been developed for wireless systems throughout the world, and more standards are likely to emerge. Tables 1.1 through 1.3 list the most common paging, cordless, cellular, and personal communications standards used in North America, Europe, and Japan.

Table1.1.MajorMobileRadioStandardsinNorthAmerica

Standard	Type	Year Introduction	oMultiple Access	FrequencyBand	Modulation	Channel Bandwidth
AMPS	Cellular	1983	FDMA	824-894 MHz	FM	30kHz
NAMPS	Cellular	1992	FDMA	824-894 MHz	FM	10kHz
USDC	Cellular	1991	TDMA	824-894 MHz	π/4-DQPSK	30kHz
CDPD	Cellular	1993	FH/Packet	824-894 MHz	GMSK	30kHz
IS-95	Cellular/PCS	1993	CDMA	824-894MHz1.8 2.0GHz	QPSK/BPSK	1.25MHz
GSC	Paging	1970s	Simplex	Several	FSK	12.5kHz
POCSAG	Paging	1970s	Simplex	Several	FSK	12.5kHz
FLEX	Paging	1993	Simplex	Several	4-FSK	15kHz
DCS-1900 (GSM)	PCS	1994	TDMA	1.85-1.99GHz	GMSK	200kHz
PACS	Cordless/PCS	1994	TDMA/FDMA	1.85-1.99GHz	π/4-DQPSK	300kHz
MIRS	SMR/PCS	1994	TDMA	Several	16-QAM	25kHz
iDen	SMR/PCS	1995	TDMA	Several	16-QAM	25kHz

Table1.2.MajorMobileRadioStandardsinEurope

Standard	Type	Year Introduction	oMultiple Access	Frequency Band	Modulation	Channel Bandwidth
ETACS	Cellular	1985	FDMA	900MHz	FM	25kHz
NMT-450	Cellular	1981	FDMA	450-470 MHz	FM	25kHz
NMT-900	Cellular	1986	FDMA	890-960 MHz	FM	12.5kHz
GSM	Cellular/PCS	1990	TDMA	890-960 MHz	GMSK	200kHz
C-450	Cellular	1985	FDMA	450-465 MHz	FM	20kHz/10kHz

Standard	Type	Year Introduction	oMultiple Access	Frequency Band	Modulation	Channel Bandwidth
ERMES	Paging	1993	FDMA	Several	4-FSK	25kHz
CT2	Cordless	1989	FDMA	864-868 MHz	GFSK	100kHz
DECT	Cordless	1993	TDMA	1880-1900MHz	GFSK	1.728 MHz
DCS-1800	Cordless/PCS	1993	TDMA	1710-1880MHz	GMSK	200kHz

Table1.3.MajorMobileRadioStandardsinJapan

Standard	Type	YearofIntroduction	MultipleAccess	FrequencyBand	Modulation	ChannelBandwidth
JTACS	Cellular	1988	FDMA	860-925MHz	FM	25kHz
PDC	Cellular	1993	TDMA	810-1501 MHz	π/4-DQPSK	25kHz
NTT	Cellular	1979	FDMA	400/800 MHz	FM	25kHz
NTACS	Cellular	1993	FDMA	843-925MHz	FM	12.5kHz
NTT	Paging	1979	FDMA	280MHz	FSK	12.5kHz
NEC	Paging	1979	FDMA	Several	FSK	10kHz
PHS	Cordless	1993	TDMA	1895-1907MHz	π/4-DQPSK	300kHz

The world's first cellular system was implemented by the Nippon Telephone and Telegraph company (NTT)in Japan. The system, deployed in 1979, uses 600 FM duplex channels (25 kHz for each one-way link) in the 800 MHz band. In Europe, the Nordic Mobile Telephone system (NMT 450) was developed in 1981 forthe450 MHz band and uses 25 kHz channels. The European Total Access Cellular System (ETACS) was deployed in 1985 and is virtually identical to the U.S. AMPS system, except that the smaller bandwidth channelsresult in a slight degradation of signal-to-noise ratio (SNR) and coverage range. In Germany, a cellular standard called C-450 was introduced in 1985. The first generation European cellular systems are generally incompatiblewithoneanotherbecauseof thedifferentfrequenciesandcommunicationprotocolsused. These systems are now being replaced by the Pan European digital cellular standard GSM (GlobalSystem for Mobile) which was first deployed in 1990 in a new 900 MHz band which all of Europe dedicated for cellular telephone service. The GSM standard has gained worldwide acceptance as the first universal digital cellular systemwithmodernnetworkfeaturesextendedto eachmobileuser, andistheleading digitalairinterfacefor PCS services above 1800 MHz throughout the world. In Japan, the Pacific Digital Cellular (PDC) standard provides digital cellular coverage using a system similar to North America's USDC.

ExamplesofWirelessCommunicationSystems

Most people are familiar witha number of mobile radiocommunication systemsused in everyday life.Garage door openers, remote controllers for home entertainment equipment, cordless telephones, hand-held walkie-talkies, pagers (also called paging receivers or "beepers"), and cellular telephones are all examples of mobile radio communication systems. However, the cost, complexity, performance, and types of services offered by each of these mobile systems are vastly different.

Table1.4listsdefinitionsoftermsusedtodescribeelementsofwirelesscommunicationsystems.

Table1.4.WirelessCommunicationsSystemDefinitions

BaseStation	A fixed station in a mobile radio system used for radio communication with mobile stations Base stations are located at the center or on the edgeof acoverage region and consist o radio channels and transmitter and receiver antennas mounted on a tower.
ControlChannel	Radiochannelusedfor transmissionofcallsetup, callrequest, callinitiation, andother beacon or control purposes.
ForwardChannel	Radio channel used for transmission of information from the base station to the mobile.
FullDuplex Systems	Communication systems which allow simultaneous two-way communication. Transmission andreceptionistypicallyontwodifferentchannels(FDD)althoughnewcordless/PCS systems are using TDD.
HalfDuplexSystems	Communication systems which allow two-way communication by using the same radio channel for both transmission and reception. At any given time, the user can only either transmit or receive information.
Handoff	Theprocessoftransferringa mobilestationfromonechannelorbasestationtoanother.
MobileStation	Astationinthecellularradioserviceintendedforusewhileinmotionat unspecified locations.Mobilestationsmaybehand-heldpersonalunits(portables)orinstalledin vehicles(mobiles).
MobileSwitchingCenter	Switchingcenter which coordinates the routingofcalls in alarge service area.In a cellular radio system,the MSC connectsthe cellular base stationsand the mobilesto the PSTN. AnMSC is also called a mobile telephone switching office (MTSO).
Page	Abriefmessagewhichis broadcastovertheentireservicearea,usuallyina simulcast fashionby many base stations at the same time.

ReverseChannel	Radiochannelusedfortransmissionofinformationfromthemobiletobasestation.
Roamer	Amobilestationwhichoperatesina servicearea (market)other thanthatfromwhich service has been subscribed.
SimplexSystems	Communication systemswhichprovideonlyone-waycommunication.
Subscriber	Auserwhopayssubscriptionchargesforusinga mobilecommunicationssystem. A
Transceiver	device capable of simultaneously transmitting and receiving radio signals.

Mobile radio transmission systems may be classified as *simplex, half-duplex* or *full-duplex*. In simplex systems, communication is possible in only one direction. Paging systems, in which messages are received but not acknowledged, are simplex systems. Half-duplex radio systems allow two-way communication, but use the same radio channel for both transmission and reception. This means that at any given time, a user can only transmit or receive information. Constraints like "push-to-talk" and "release-to-listen" are fundamental features of half-duplex systems. Full duplex systems, on the other hand, allow simultaneous radio transmission and reception between a subscriber and a base station, by providing two simultaneous but separate channels (frequency division duplex, or FDD) or adjacent time slots on a single radio channel (time division duplex, or TDD) for communication to and from the user.

In FDD, a pair of simplex channels with a fixed and known frequency separation is used to define a specific radiochannelinthesystem. The channelused toconvey traffic tothemobileuser from a basestationiscalled the*forward channel*, whilethe channelusedtocarry traffic from themobileusertoa basestationis called the *reverse channel*. In the U.S. AMPS standard,the reverse channel has a frequency which is exactly 45 MHz lower than that of the forward channel. Full duplex mobile radio systems provide many of the capabilities of the standard telephone, with the added convenience of mobility. Full duplex and half-duplex systems use *transceivers* for radio communication. FDD is used exclusively in analog mobile radio systems.
Time division duplexing (TDD) uses the fact that it is possible to share a single radio channel in time, so that a portion of the time is used to transmit from the base station to the mobile, and the remaining time is used to transmit from the mobile to the base station. If the data transmission rate in the channel is muchgreater than the end-user's data rate, it is possible to store information bursts and provide the appearance of full duplex operation to a user, even though there are *not* two simultaneous radio transmissions at any instant. TDD is only possible with digital transmissionformatsand digital modulation, andis very sensitive to timing. It isfor this reason that TDD has only recently been used, and only for indoor or small area wireless applications where the physical coverage distances (and thus the radio propagation time delay)are much smaller than the many kilometers used in conventional cellular telephone systems.

PagingSystems
Paging systems are communication systems that send brief messages to a subscriber. Depending on the type of service, the message may be either a numeric message, an alphanumeric message, or a voice message. Paging systems are typically used to notify a subscriber of the need to call a particular telephone number or travel to a known location to receive further instructions. In modern paging systems, news headlines, stock quotations, and faxes may be sent. A message is sent to a paging subscriber via the paging system access number (usually a toll-free telephone number) with a telephone keypad or modem. The issued message is called a *page*. The paging system then transmits the page throughout the service area using base stations which broadcast the page on a radio carrier.

Paging systems vary widely in their complexity and coverage area. While simple paging systems may cover a limited range of 2 to 5 km, or may even be confined to within individual buildings, wide area paging systems can provide worldwide coverage. Though paging receivers are simple and inexpensive, the transmission system required is quite sophisticated. Wide area paging systems consist of a network of telephone lines, many base station transmitters, and large radio towers that simultaneously broadcast a page from each base station (this is called *simulcasting*). Simulcast transmitters may be located within the same service area or in different cities or countries. Paging systems are designed to provide reliable communication to subscribers wherever they are; whether inside a building, driving on a highway, or flying in an airplane. This necessitates large transmitter powers (on the order of kilowatts) and low data rates (a couple ofthousand bits persecond) for maximum coverage from each base station. Figure 1.3 shows a diagram of a wide area paging system.

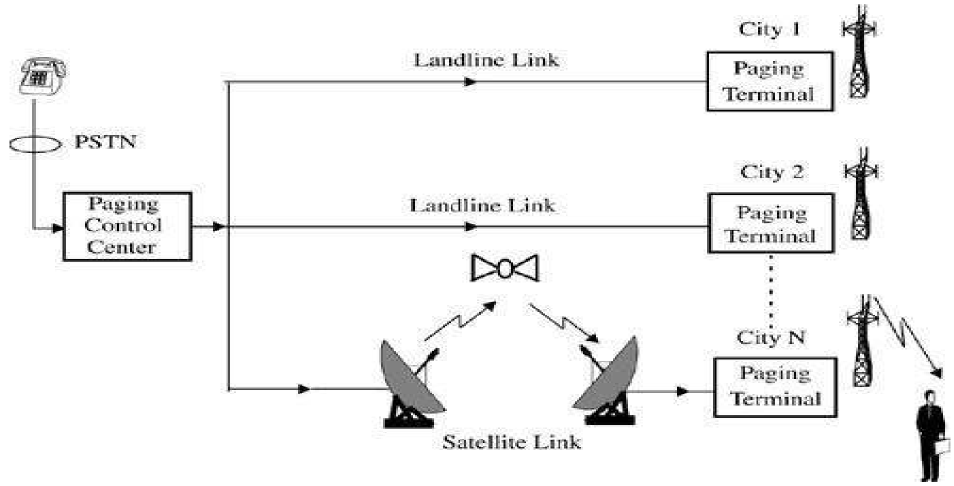


Figure1.3.Awideareapagingsystem.Thepagingcontrolcenterdispatchespagesreceivedfromthe PSTN throughout several cities at the same time.

Paging systems aredesignedtoprovideultra-reliablecoverage, even insidebuildings. Buildingscanattenuate radio signals by 20 or 30 dB, making the choice of base station locationsdifficult for the paging companies. For this reason, paging transmitters are usually located on tall buildings in the center of a city, and simulcasting is used in conjunction with additional base stations located on the perimeter of the city to flood the entire area. Small RF bandwidths are used to maximize the signal-to-noise ratio at each pagingreceiver, so low data rates (6400 bps or less) are used.

CordlessTelephoneSystems

Cordless telephone systems are full duplex communication systems that use radio to connect a portable handset to a dedicated base station, which is then connected to a dedicated telephone line with a specific telephone number on the public switched telephone network (PSTN). In first generation cordless telephone systems (manufactured in the 1980s), the portable unit communicates only to the dedicated base unit and only over distances of a few tens of meters. Early cordless telephones operate solely as extension telephones to a transceiver connected to a subscriber line on the PSTN and are primarily for in-home use.

Second generation cordless telephones have recently been introduced which allow subscribers to use their handsets at many outdoor locations within urban centers such as London or Hong Kong. Modern cordless telephones are sometimes combined with paging receivers so that a subscriber may first be paged and then respond to the page using the cordless telephone. Cordless telephone systems provide the user with limited range and mobility, as it is usually not possible to maintain a call if the user travels outside the range of the base station. Typical second generation base stations provide coverage ranges up to a few hundred meters. Figure 1.4 illustrates a cordless telephone system.

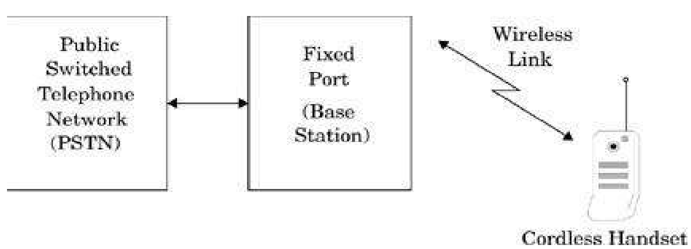


Figure1.4.Acordlesstelephonesystem.

CellularTelephoneSystems

A cellular telephone system provides a wireless connection to the PSTN for any user location within the radio range of the system. Cellular systems accommodate a large number of users over a large geographic area, within a limited frequency spectrum. Cellular radio systems provide high quality service that is often comparable to that of the landline telephone systems. High capacity is achieved by limiting the coverage of each base station transmitter to a small geographic area called a *cell* so that the same radio channels may be reused by another base station located some distance away. A sophisticated switching technique called a *handoff* enables a call to proceed uninterrupted when the user moves from one cell to another.

Figure1.5showsa basic cellularsystemwhichconsistsof *mobilestations,basestations* anda *mobileswitching center* (MSC). The mobile switching center is sometimes called a *mobile telephone switching office* (MTSO), since it is responsible for connecting all mobiles to the PSTN ina cellular system. Each mobile communicates via radio with one of the base stations and may be handed-off to any number of base stations throughout the duration of a call. The mobile station contains a transceiver, an antenna, and control circuitry, and may be mounted in a vehicle or used as a portable hand-held unit. The base stations consist of several transmitters and receivers which simultaneously handle full duplex communications and generally have towers which support several transmitting and receiving antennas. The base station serves as a bridge between all mobile users in the cell and connects the simultaneous mobile calls via telephone lines ormicrowave links to the MSC. The MSC coordinates the activities of all of the base stations and connects the entire cellular system to the PSTN. A typical MSChandles 100,000 cellular subscribers and 5,000 simultaneous conversations at atime, and accommodates all billing and system maintenance functions, as well.In large cities, several MSCsare used by a single carrier.

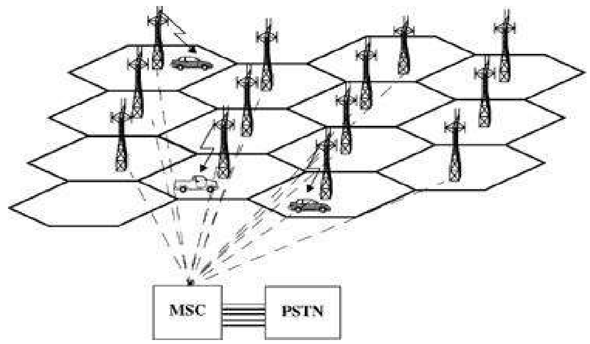


Figure1.5.Acellularsystem.Thetowersrepresent basestationswhichprovideradioaccessbetween mobile users and the mobile switching center (MSC).

Communicationbetweenthe basestationandthemobilesisdefinedby a standard *commonairinterface* (CAI) that specifies four different channels. The channels used for voice transmission from the base station to mobilesare called *forward voice channels*(FVC), and the channelsusedfor voice transmission from mobiles to the base station are called *reverse voice channels* (RVC). The two channels responsible for initiating mobile calls are the *forward control channels* (FCC) and *reverse control channels* (RCC). 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. Forward control channels also serve as beacons whichcontinually broadcast all of the traffic requests for all mobiles in the system.

Cellular systems rely on the frequency reuse concept,which requires that the forward control channels (FCCs) in neighboring cells be different. By defining a relatively small number of FCCs as part of the common air interface, cellular phones can be manufactured by many companies which can rapidly scan all of the possible FCCs to determine the strongest channel at any time. Once finding the strongest signal, the cellular phone receiver stays "camped" to the particular FCC. By broadcasting the same setup data on all FCCs at the same time, the MSC is able to signal all subscribers within the cellular system and can be certain that any mobile will be signaled when it receives a call via the PSTN.

HowaCellularTelephoneCallisMade

When a cellular phone is turned on, but is not yet engaged in a call, it first scans the group of forward control channels to determine the one with the strongest signal, and then monitors that control channel until the signal drops below a usable level. At this point, it again scans the control channels in search of the strongest base station signal. When a telephone call is placed to amobile user, the MSC dispatches the request to all base stations in the cellular system. The *mobile identification number* (MIN), which is the subscriber's telephone number, is then broadcast asa paging messageover all of the forward control channels throughout the cellular system. The mobile receives the paging message sent by the base station which it monitors, and responds by identifying itself over the reverse control channel. The base station relays the acknowledgment sent by the mobile and informs the MSC of the handshake. Then, the MSC instructs the base station to move the call to an unused voice channel within the cell (typically, between ten to sixty voice channels and just one control channel are used in each cell's base station). At this point, the base station signalsthe mobile to change frequencies to an unused forward and reverse voice channel pair, at which point anotherdata message (called an *alert*) is transmitted over the forward voice channel to instruct the mobile telephone to ring, thereby instructing the mobile user to answer the phone. Figure 1.6 shows the sequence of events involved with connecting a call to a mobile user in a cellular telephone system. All ofthese events occur within a few seconds and are not noticeable by the user.

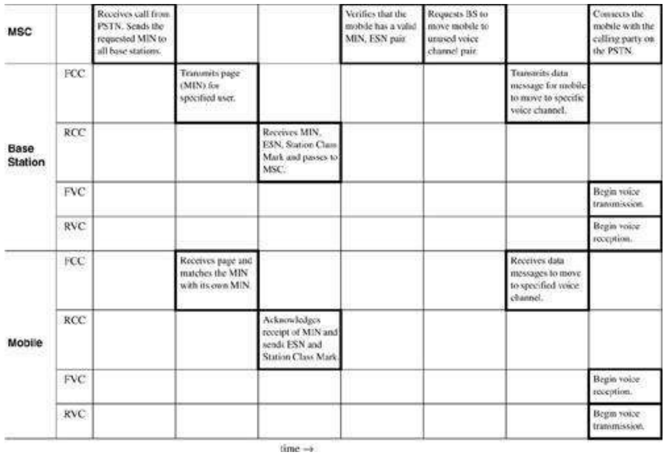


Figure1.6.Timing diagramillustratinghow acalltoamobileuserinitiatedbyalandline subscriberis established.

Oncea callisinprogress, the MSC adjuststhetransmittedpowerof themobileand changesthechannelof the mobile unit and base stations in order to maintain call quality as the subscriber moves in and out of range of each base station. This iscalled a *handoff*. Specialcontrol signaling isapplied to the voice channels so that the mobile unit may be controlled by the basestation and the MSC while a call is in progress.

When a mobile originates a call, a call initiation request is sent on the reverse control channel. With this request the mobile unit transmits its telephone number (MIN), *electronic serial number* (ESN), and the telephone number of the called party. The mobile also transmits a *station class mark* (SCM) which indicates what the maximum transmitter power level is for the particular user. The cell base station receives this data and sends it to the MSC. The MSC validates the request, makes connection to the called party through the PSTN, andinstructsthebasestationandmobileuserto movetoanunused forwardandreverse voicechannel pair to allow the conversation to begin. Figure 1.7 shows the sequence of events involved with connecting a call which is initiated by a mobile user in a cellular system.

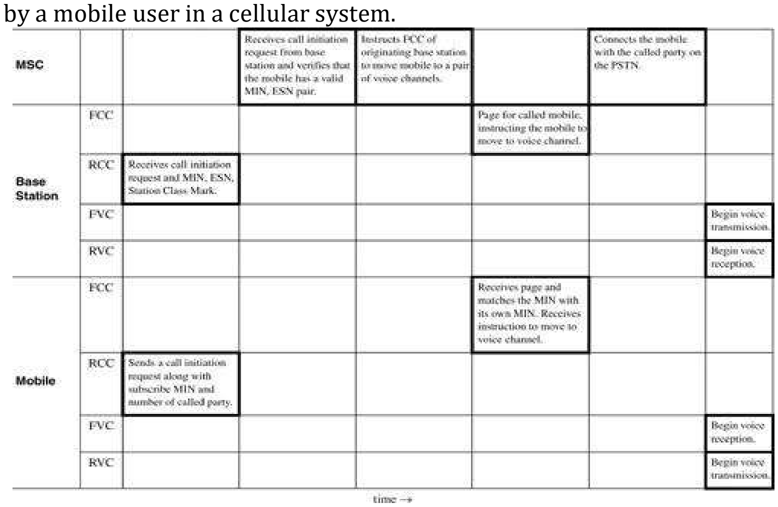


Figure1.7.Timingdiagramillustratinghowacallinitiatedbyamobileisestablished.

Allcellularsystems provide a servicecalled *roaming*. Thisallowssubscriberstooperateinserviceareasother than the one from which service is subscribed. When a mobile enters a city or geographicareathat is differentfrom itshomeservicearea, itisregisteredasa roamerinthenew servicearea. If a particularroamer has roaming authorization for billing purposes, the MSC registers the subscriber as a valid roamer. Once registered, roaming mobiles are allowed to receive and place calls from that area, and billing is routed automatically to the subscriber's home service provider.

ComparisonofCommonWirelessCommunicationSystems

Tables 1.5 and 1.6 illustrate the types of service, level of infrastructure, cost, and complexity required for the subscriber segment and base station segment of each of the five mobile or portable radio systems discussed earlier in this chapter. For comparison purposes, common household wireless remote devices are shown in the table. It is important to note that each of the five mobile radio systems given in Tables 1.5 and 1.6 use a fixed base station, and for good reason. Virtually all mobile radio communication systems strive to connect a moving terminal to a fixed distribution system of some sort and attempt to look invisible to the distribution system.

Table1.5.ComparisonofMobileCommunicationSystems—MobileStation

Service	Coverage Range	Required Infrastructure	Complexity	Hardware Cost	Carrier Frequency	Functionality
TV Remote Control	Low	Low	Low	Low	Infrared	Transmitter
Garage Doo Opener	Low	Low	Low	Low	<100MHz	Transmitter
PagingSystem	High	High	Low	Low	<1GHz	Receiver
CordlessPhone	Low	Low	Moderate	Low	<1GHz	Transceiver
CellularPhone	High	High	High	Moderate	<2GHz	Transceiver

Table1.6.ComparisonofMobileCommunicationSystems—BaseStation

Service	Coverage Range	Required Infrastructure	Complexity	Hardware Cost	Carrier Frequency	Functionality
TV Remote Control	Low	Low	Low	Low	Infrared	Receiver
Garage Doo Opener	Low	Low	Low	Low	<100MHz	Receiver
PagingSystem	High	High	High	High	<1GHz	Transmitter
CordlessPhone	Low	Low	Low	Moderate	<1GHz	Transceiver
CellularPhone	High	High	High	High	<2GHz	Transceiver

Notice that the expectations varywidely amongthe services,and the infrastructure costsare dependent upon the required coverage area. For the case of low power, hand-held cellular phones, a large number of base stations are required to insure that any phone is in close range to a base station within a city. If base stations were not within close range, a great deal of transmitter power would be required of the phone, thus limiting the battery life and rendering the service useless for hand-held users.

TrendsinCellular Radio and PersonalCommunications

Since 1989,there has been enormous activity throughout the world to develop personal wireless systemsthat combine the network intelligence of today's PSTN with modern digital signal processing and RF technology. The concept, called Personal Communication Services (PCS), originated in the United Kingdom when three companies were given spectrum in the 1800 MHz range to develop Personal Communication Networks (PCN) throughout Great Britain. PCNwas seenby theU.K.asa means of improving itsinternational competitiveness in the wireless field while developing new wireless systems and services for citizens.

Indoor wireless networking products are rapidly emerging and promise to become a major part of the telecommunications infrastructure within the next decade. An international standards body, IEEE 802.11, is developing standards for wireless access between computers inside buildings. The European TelecommunicationsStandardInstitute(ETSI) is alsodeveloping the20 Mbps HIPERLAN standard for indoor wireless networks. Products haveemerged thatallow users to link their phonewith their computer withinan office environment, as well as in a public setting, such asan airport or train station.

A worldwide standard, the Future Public Land Mobile Telephone System (FPLMTS)—renamed International Mobile Telecommunication 2000 (IMT-2000) in mid-1995—has been formulated by the International Telecommunications Union (ITU) which is the standards body for the United Nations, with headquarters in Geneva, Switzerland. FPLMTS (now IMT-2000) is a third generation universal, multi-function, globally compatible digital mobile radio system that will integrate paging, cordless, and cellular systems,as well as low earth orbit (LEO) satellites, into one universal mobile system.

In emerging nations, where existing telephone service is almost nonexistent, fixed cellular telephone systems are being installed at a rapid rate. This is due to the fact that developing nations are finding it is quicker and more affordable to install cellular telephone systems for fixed home use, rather than install wires in neighborhoods which have not yet received telephone connections to the PSTN.

ModernWirelessCommunicationSystems

Since the mid 1990s, the cellular communications industry has witnessed explosive growth. Wireless communications networks have become much more pervasive than anyone could have imagined when the cellular concept was first developed in the 1960s and 1970s. The widespread adoption of wireless communications was accelerated in the mid 1990s, when governments throughout the world provided increased competition and new radio spectrum licenses for personal communications services (PCS) in the 1800–2000 MHz frequency bands.

The rapid worldwide growth in cellular telephone subscribers has demonstrated conclusively that wireless communications is a robust, viable voice and data transport mechanism. New standards and technologies are being implemented to allow wireless networks to replace fiber optic or copper lines between fixed points several kilometers apart (*fixed wireless access*). Similarly, wireless networks have been increasingly used as a replacement for wires within homes, buildings, and office settings through the deployment of *wireless local area networks* (WLANs). Theevolving *Bluetooth* modem standard promises to replace troublesome appliance communication cords with invisible wireless connections within a person's personal workspace. Used primarilywithin buildings, WLANsand Bluetooth use low power levels and generally do not require a license for spectrum use.

SecondGeneration(2G)CellularNetworks
Most of today's ubiquitous cellular networks use what is commonly called *second generation* or *2G* technologies which conform to the second generation cellular standards. Unlike first generation cellular systems that relied exclusively on FDMA/FDD and analog FM, second generation standards use digital modulation formats and TDMA/FDD and CDMA/FDD multiple access techniques.

The most popular second generation standards include three TDMA standards and one CDMA standard: (a) *Global System Mobile (GSM)*, which supports eight time slotted users for each 200 kHz radio channel and has been deployed widely byservice providers in Europe, Asia, Australia, South America, and some parts of theUS (in the PCS spectrum band only); (b) *Interim Standard 136 (IS-136)*, also known as North American Digital Cellular (NADC), which supports three time slotted users for each 30 kHz radio channel and is a popular choice for carriers in North America, South America, and Australia (in both the cellular and PCS bands); (c) *Pacific Digital Cellular (PDC)*, a Japanese TDMA standard that is similar to IS-136 with more than 50 million users; and (d) the popular 2G CDMA standard *Interim Standard 95 Code Division Multiple Access (IS-95)*, also known as *cdmaOne*, which supports up to 64 users that are orthogonally coded and simultaneously transmitted on each 1.25 MHz channel. CDMA is widely deployed by carriers in North America (in both cellular and PCS bands), as well as in Korea, Japan, China, South America, and Australia.

In many countries, 2G wireless networks are designed and deployed for conventional mobile telephone service, as a high capacity replacement for, or in competition with, existing older first generation cellular telephone systems. Modern cellular systems are also being installed to provide fixed (non-mobile) telephone service to residences and businesses in developing nations—this is particularly cost effective for providing *plain old telephone service (POTS)* in countries that have poor telecommunications infrastructure and are unable to afford the installation of copper wire to all homes.

Table2.1.KeySpecifications ofLeading2G Technologies (adaptedfrom[Lib99])

	cdmaOne, IS-95,ANSI J STD-008	**GSM,DCS-1900,ANSIJ-STD 007**	**NADC,IS-54/IS-136,ANSIJ STD-011, PDC**
Uplink Frequencies	824-849MHz(USCellular) 1850-1910 MHz (US PCS)	890-915 MHz (Europe) 1850-1910MHz(USPCS)	800MHz,1500MHz(Japan) 1850-1910MHz(USPCS)
Downlink Frequencies	869-894MHz(USCellular) 1930-1990 MHz (US PCS)	935-960 MHz (Europe) 1930-1990MHz(USPCS)	869-894MHz(USCellular) 1930-1990 MHz (US PCS) 800MHz,1500MHz(Japan)
Duplexing	FDD	FDD	FDD
Multiple Acces Technology	CDMA	TDMA	TDMA
Modulation	BPSK with Quadrature Spreading	GMSKwith*BT*=0.3	π/4DQPSK
Carrier Separation	1.25MHz	200kHz	30kHz(IS-136) (25kHzfor PDC)
Channel Data Rate	1.2288 Mchips/sec	270.833 kbps	48.6kbps(IS-136) (42kbpsforPDC)
Voice channels percarrier	64	8	3
SpeechCoding	Code Excited Linea Prediction(CELP) @ 13 kbps, Enhanced VariableRateCodec (EVRC @ 8 kbps	Residual Pulse Excited Long TermPrediction(RPE-LTP)@ 13kbps	Vector Sum Excited Linea PredictiveCoder(VSELP)@ 7.95kbps

Evolutionto2.5GMobileRadioNetworks
Since the mid 1990s, the 2G digital standards have been widely deployed by wireless carriers for cellular and PCS, eventhoughthesestandards weredesignedbeforethewidespreaduseof theInternet. Consequently, 2G

technologies use circuit-switched data modems that limit data users to a single circuit-switchedvoicechannel.

In an effort to retrofit the 2G standards for compatibility with increased throughput data rates that are required to support modern Internet applications, new data-centric standards have been developed that can beoverlaid uponexisting 2G technologies. Thesenew standards represent *2.5G* technology and allow existing 2G equipment to be modified and supplemented with new base station add-ons and subscriber unit software upgrades to support higher data rate transmissions for web browsing, e-mail traffic, mobile commerce (m-commerce), and location-based mobile services. The 2.5G technologies also support a popular new web browsing format language,called Wireless Applications Protocol (WAP),that allows standard webpages to be viewed in a compressed formatspecifically designed for small, portable hand held wireless devices, a wide range of 2.5G standards havebeen developed to allow each of the major 2G technologies (GSM,CDMA,and IS-136) to be upgraded incrementally for faster Internet datarates.Figure 2.3 illustrates the various 2.5G and 3G upgrade paths for the major 2G technologies.

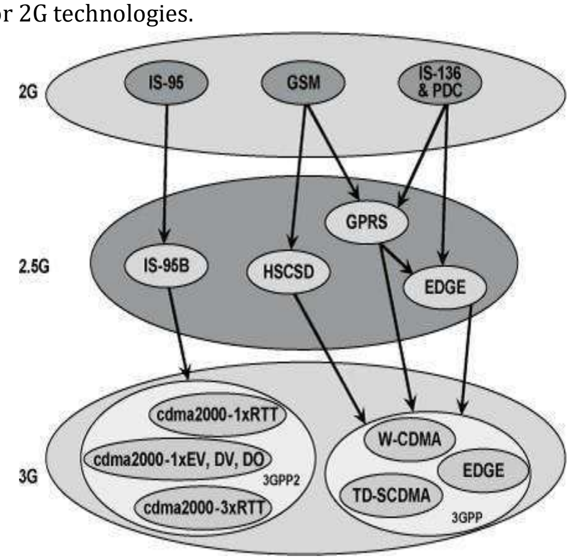


Figure2.3.Variousupgradepathsfor2Gtechnologies.

Table 2.2 describes the required changes to the network infrastructure (e.g., the base station and the switch) andthesubscriberterminals(e.g., thehandset)for the variousupgradeoptionsfor 2.5G and3G. Thetechnical features of each 2.5G upgrade path are described below.

Table2.2.CurrentandEmerging2.5Gand3GDataCommunicationStandards

WirelessData Technologies	Channel BW	Duplex	Infrastructurechange	Requires New Spectrum	RequiresNewHandsets
HSCSD	200KHz	FDD	Requiressoftwareupgradea basestation.	No	Yes New HSCSD handsets provide 57.6Kbps on HSCSD networks,and 9.6KbpsonGSMnetworkswithduamode phones. GSM-only phones willnotworkinHSCSDnetworks.
GPRS	200KHz	FDD	Requires new packet overla including routers an gateways.	No	Yes New GPRS handsets work on GPRSnetworksat171.2Kbps,9.6Kbpson GSMnetworks withdual modephones.GSM-onlyphoneswillnoworkin GPRS networks.
EDGE	200KHz	FDD	Requires new transceivera base station. Also, software upgradestothebasestation controller and base station.	No	Yes NewhandsetsworkonEDGEnetworksat 384 Kbps, GPRSnetworks at 144Kbps, and GSMnetworks at 9.6Kbpswithtri mode phones.GSM and GPRS-onlyphoneswillnotworkinEDGE networks.

WirelessData Technologies	Channel BW	Duplex	Infrastructurechange	Requires New Spectrum	RequiresNewHandsets
W-CDMA	5MHz	FDD	Requires completely newbasestations.	Yes	Yes New W-CDMA handsets will work on W-CDMA at 2 Mbps, EDGE networks at 384Kbps, GPRS networks at 144 Kbps, GSMnetworksat 9.6Kbps.Olderhandsetswillnot workinW CDMA.
IS-95B	1.25 MHz	FDD	Requiresnewsoftwareinbasestation controller.	No	Yes New handsets will work on IS-95B at 64Kbps andIS-95A at14.4Kbps.CdmaOnephonescanwork inIS-95Bat14.4Kbps.
cdma2000 1xRTT	1.25 MHz	FDD	Requiresnewsoftwareinbackbonean d new channecardsatbasestation.Alsoneed to build a newpacke service node.	No	Yes New handsets will work on1xRTT at 144Kbps, IS-95B at 64 Kbps, IS 95A at 14.4Kbps. Older handsets can work in 1xRTTbut at lower speeds.
cdma2000IxEV (DO and DV)	1.25 MHz	FDD	Requiressoftwareanddigitacarupgra de on 1xRTT networks.	No	Yes New handsets will workon1xEVat 2.4Mbps, 1xRTT at 144 Kbps IS-95B at64Kbps,IS-95Aat14.4 Kbps. Older handsets can work in 1xEVbut at lower speeds.
cdma2000 3xRTT	3.75 MHz	FDD	Requires backbone modificationsandnewchannelcardat base station.	Maybe	Yes Newhandsetswillworkon95Aa 14.4 Kbps, 95B at 64 Kbps, 1xRTT at 144Kbps, 3xRTT at 2 Mbps Older handsetscan work in 3X bu at lower speeds.

Evolutionfor2.5GTDMAStandards

Three different upgrade paths have been developed for GSM carriers, and two of these solutions also support IS-136. The three TDMA upgrade options include: (a) High Speed Circuit Switched Data (HSCSD); (b) General Packet Radio Service (GPRS); and (c) Enhanced Data Rates for GSM Evolution (EDGE). These options provide significant improvements in Internet access speed over today's GSM and IS-136 technology and support the creation of new Internet-ready cell phones.

HSCSDfor2.5GGSM

As the name implies, High Speed Circuit Switched Data is a circuit switched technique that allows a single mobile subscriber to use consecutive user time slots in the GSM standard.That is, instead of limitingeach user to only one specific time slot in the GSM TDMA standard, HSCSD allows individual data users to commandeer consecutive time slots in order to obtain higher speed data access on the GSM network. HSCSD relaxes the error control coding algorithms originally specified in the GSM standard for data transmissions and increases the available application data rate to 14,400 bps, as compared to the original 9,600 bps in the GSM specification. HSCSD is ideal for dedicated streaming Internet access or real-time interactive web sessions and simply requires the service provider to implement a software change at existing GSM base stations.

GPRSfor2.5GGSMandIS-136

General Packet Radio Service is a packet-based data network, which is well suited for non-real time Internet usage,including the retrieval of email, faxes, andasymmetric web browsing, where the user downloads much moredata thanituploadsontheInternet. UnlikeHSCSD,whichdedicatescircuitswitchedchannelstospecific users, GPRS supports multi-user network sharing of individual radio channels and time slots. Similar to the Cellular Digital Packet Data (CDPD) standard developed for the North American AMPS systems in the early 1990s, the GPRS standard provides a packet network on dedicated GSM or IS-136 radio channels. GPRSretains the original modulation formats specified in the original 2G TDMA standards, but uses a completely redefined air interface in order to better handle packet data access.

As is the case for any packet network, the data throughput experienced by an individual GPRS user decreases substantially as more users attempt to use the network or as propagation conditions become poor for particular users. It is worth noting that GPRS was originally designed to provide a packet data access overlay solely for GSM networks, but at the request ofNorth American IS-136 operators (see UWC-136 Air Interface in Table 2.3), GPRS was extended to include both TDMA standards.

Table2.3.LeadingIMT-2000CandidateStandardsasof1998(adaptedfrom[Lib99])

Air Interface	Modeof Operation	DuplexingMethod	KeyFeatures
cdma2000U STIATR45.5	Multi-Carrier and DirecSpreadingDS-CDMAa*N* =1.2288Mcpswith*N*=1, 3,6,9, 12	FDD andTDDModes	BackwardcompatibilitywithIS-95Aand IS-95B. Downlinkcan beimplemented usingeither Multi-Carrier or DirecSpreading. Uplink cansupport asimultaneousco mbinationof Multi-CarrieorDirectSpreadingAuxiliary carrierstohelp withdownlink channe estimation inforward link beamforming.
UTRA (UMTS Terrestria RadioAccess) ETSSMG2	DS_CDMA at Rates of *N* ×0.960Mcpswith*N*=4,8 16	FDD andTDDModes	WidebandDS_CDMASystem. Backwardcompatibilitywith GSM/DCS-1900. Upto 2.048 Mbps onDownlink in FDD Mode. Minimumforwardchannebandwidth of 5 GHz. Thecollectionof proposedstandardsrepresentedhere eachexhibituniquefeaturesbut support a common set ochip rates, 10 ms framestructure,with 16 slots pe frame. Connection-dedicated pilobitsassist in downlink beamforming.
W-CDMA/NA (Wideband CDMA NorthAmerica)USAT1P 1-ATIS			
W-CDMA/Japan (Wideband CDMA JapanARIB			
CDMAII SouthKorea TTA			
WIMS/W-CDMA USATIATR46.1			
CDMAI SouthKorea TTA	DS-CDMAat*N*×0.9216 Mcpswith *N*=1, 4, 16	FDD andTDDModes	Upto512kbps perspreadingcode,code aggregationupto 2.048Mbps.
UWC-136 (Universal Wireless CommunicationsConsortium) USA TIATR45.3	TDMA - Up to 722.2 kbp(Outdoor/Vehicular),Upto 5.2Mbps(Indoo Office)	FDD (Outdoor/Vehicular),TDD (Indoor Office)	Backwardcompatibilityandupgradepath for both IS-136 and GSM. FitsintoexistingIS-136andGSM.Explicit plans to suppor adaptiveantenna technology.
TD-SCDMA China Academy of TelecommunicationTechnology(CATT)	DS-CDMA1.1136Mcps	TDD	RFchannelbitrateupto 2.227Mbps. Useofsmartantennatechnologyisfundamenta (butnotstrictlyrequired)in TDSCMA.
DECT ETSI Project (EPDECT	1150-3456kbpsTDMA	TDD	Enhancedversionof2GDECTtechnology.

EDGEfor2.5GGSMandIS-136

EnhancedDataratesforGSM(or Global)Evolution isa more advancedupgrade tothe GSMstandard,and requiresthe additionofnewhardwareandsoftwareatexistingbasestations.EDGEintroducesanewdigital modulation format, 8-PSK (octal phase shift keying), which is used in addition to GSM's standard GMSK modulation. EDGE allows for nine different (autonomously and rapidly selectable) air interface formats, known as *multiple modulation and coding schemes* (MCS), with varying degrees of error control protection. Each MCS state may use either GMSK (low datarate) or 8-PSK (high datarate) modulation for network access, depending ontheinstantaneous demands of thenetwork and theoperating conditions. Becauseof the higher data rates and relaxed error control covering in many of the selectable air interface formats, the coverage rangeis smaller in EDGE than in HSDRC or GPRS. EDGE is sometimes referred to as EnhancedGPRS, orEGPRS.

IS-95Bfor2.5G CDMA

Unlike the several GSM and IS-136 evolutionary paths to high speed data access, CDMA (oftencalled *cdmaOne*) has a single upgrade path for eventual 3G operation. The interim data solution for CDMA is called IS-95B. Like GPRS, IS-95B is already being deployed worldwide, and provides high speed packet and circuit switched data access on a common CDMA radio channel by dedicating multiple orthogonal user channels (Walsh functions) for specific users and specific purposes. Each IS-95 CDMA radio channel supports up to 64 different user channels.

The 2.5G CDMA solution, IS-95B, supports *medium data rate* (MDR) service by allowing a dedicated user to command up to eight different user Walsh codes simultaneously and in parallel for an instantaneous throughput of 115.2 kbps per user (8 × 14.4 kbps). IS-95B also specifies hard handoff procedures that allow subscriber units to search different radio channels in the network without instruction from the switchso that subscriber units can rapidly tune to different base stations to maintain link quality.

ThirdGeneration(3G)WirelessNetworks

3G systemspromiseunparalleled wirelessaccessinwaysthathavenever beenpossiblebefore. Multi-megabit Internet access,communications using Voiceover Internet Protocol(VoIP), voice-activated calls, unparalleled network capacity, and ubiquitous "always-on" access are just some of the advantages being touted by 3G developers. Companies developing 3G equipment envision users having the ability to receive live music, conduct interactive web sessions, and have simultaneous voice and data access with multiple parties at the same time using a single mobile handset, whether driving, walking, or standing still in an office setting.
The eventual 3G evolution for CDMA systems leads to cdma2000. Several variants of CDMA 2000are currently being developed, but they all are based on the fundamentals of IS-95 and IS-95B technologies. The eventual 3G evolution for GSM, IS-136, and PDC systems leads to Wideband CDMA (W-CDMA), also called Universal Mobile Telecommunications Service (UMTS). W-CDMA is based on the network fundamentals of GSM, as well as the merged versions of GSM and IS-136 through EDGE.

3GW-CDMA(UMTS)

The Universal Mobile Telecommunications System (UMTS) is a visionary air interface standard that has evolved since late 1996 under the auspices of the European Telecommunications Standards Institute (ETSI). European carriers, manufacturers, and government regulators collectively developed the early versions of UMTS as a competitive open air-interface standard for third generation wireless telecommunications.

UMTS, orW-CDMA, assures backwardcompatibility withthe second generation GSM, IS-136, and PDC TDMA technologies, as well as all 2.5G TDMA technologies. Thenetwork structure and bit level packaging of GSM data is retained by W-CDMA, with additionalcapacity andbandwidth provided by a new CDMA air interface. The 3G W-CDMA air interface standard had been designed for "always-on" packet-based wireless service, so thatcomputers,entertainmentdevices,andtelephonesmayallsharethesamewirelessnetworkandbe connected to the Internet, anytime, anywhere. W-CDMA requires a minimum spectrum allocation of 5 MHz, which isanimportantdistinctionfrom theotherG standards. WithW-CDMAdata rates from as low as 8kbps to as high as 2 Mbps will be carried simultaneously on a single W-CDMA 5 MHz radio channel, and each channelwillbeableto support between100and350 simultaneous voicecalls atonce, depending onantenna sectoring, propagationconditions,uservelocity,andantennapolarizations.W-CDMA employsvariable/selectabledirect sequencespread spectrum chiprates thatcanexceed 16 Megachips persecondper user.

3Gcdma2000

The cdma2000 vision provides a seamless and evolutionary high data rate upgrade path for current users of 2G and 2.5GCDMA technology, usinga buildingblockapproach that centers on the original 2GCDMA channel bandwidth of 1.25 MHz per radio channel.

The first 3G CDMA air interface, cdma2000 1xRTT, implies that a single 1.25 MHz radio channel is used (the initials RTT stand for *Radio Transmission Technology).* The ultimate 3G solution for CDMA relies upon multicarrier techniques that gang adjacent cdmaOne radio channels together for increased bandwidth. The cdma2000 3xRTT standard uses three adjacent 1.25 MHz radio channels that are used together to provide packet data throughput speeds in excess of 2 Mbps per user, depending upon cell loading, vehicle speed, and propagationconditions.

3GTD-SCDMA
The China Academy ofTelecommunications Technology (CATT) and Siemens Corporation jointly submitted an IMT-2000 3G standard proposal in 1998, based on Time Division-Synchronous Code Division Multiple Access (TD-SCDMA). This proposal wasadopted by ITU as one of the 3G options in late 1999.

TD-SCDMA relies on the existing core GSM infrastructure and allows a 3G network to evolve through the addition of high data rate equipment at each GSM base station. TD-SCDMA combines TDMA and TDD techniques to provide a data-only overlay in an existing GSM network. By using TDD, different time slots within a single frame on a single carrier frequency are used to provide both forward channel and reverse channel transmissions. For the case of asynchronous traffic demand, such as when auser downloads afile, the forward link will require more bandwidth than the reverse link, and thus more time slots willbe dedicated to providing forward link traffic than for providing reverse link traffic.TD-SCDMA proponents claim that the TDD feature allows this 3G standard to be very easily and inexpensively added to existing GSM systems.

WirelessLocalLoop(WLL)andLMDS
Fixed wireless equipment is extremely well suited for rapidly deploying a broadband connection in many instances,andthisapproachissteadilybecoming morepopularforproviding “lastmile” broadbandlocalloop access, as well as for emergency or redundant point-to-point or point-to-multipoint private networks.
Modern fixed wireless systems are usually assigned microwave or millimeter radio frequencies in the 28 GHz band and higher, which is greater than ten times the carrier frequency of 3G terrestrial cellular telephone networks. At these higher frequencies, the wavelengths are extremely small, which in turn allows very high gain directional antennas to be fabricated in small physical form factors. Microwave wireless linkscan be used to create a wireless local loop (WLL) such as the one shown in Figure 2.4.

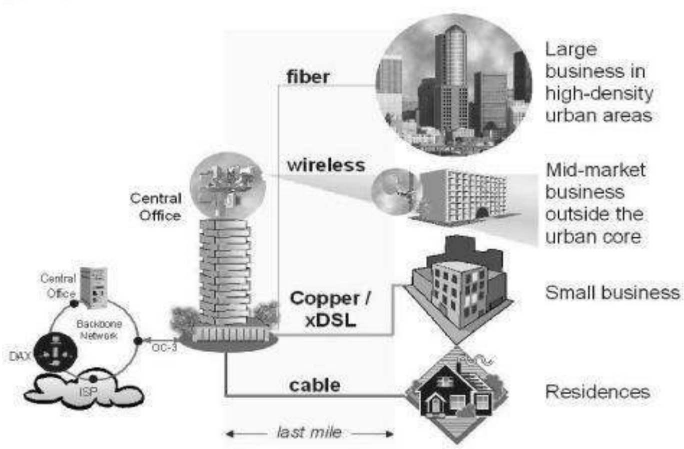


Figure2.4.Exampleoftheemergingapplicationsandmarkets for broadbandservices.

Thelocalloopcanbe thoughtofasthe“lastmile”ofthetelecommunicationnetworkthatresidesbetweenthe centraloffice(CO)andtheindividualhomesandbusinessesincloseproximitytotheCO.
Governments throughout the world have realized that WLL could greatly improve the efficiency of their citizens while stimulating competition that could lead to improved telecommunications services. A vast array of new services and applications have been proposed and are in the early stages of commercialization. These services include the concept of Local Multipoint Distribution Service (LMDS), which provides broadband telecommunications access in the local exchange.
Governments throughout the world have realized that WLL could greatly improve the efficiency of their citizens while stimulating competition that could lead to improved telecommunications services. A vast array of new services and applications have been proposed and are in the early stages of commercialization. These services include the concept of Local Multipoint Distribution Service (LMDS), which provides broadband telecommunications access in the local exchange.

In 1998, 1300 MHz of unused spectrum in the 27–31 GHz band was auctioned by the US government to support LMDS. Similar auctions have been held in other countries around the world.

One of the most promising applications for LMDS is in a local exchange carrier (LEC) network. Figure 2.7 shows a typical network configuration, where the LEC owns a very wide bandwidth asynchronous transfer mode (ATM) or Synchronous Optical Network (SONET) backbone switch, capable of connecting hundreds of megabitspersecondoftrafficwiththeInternet,thePSTN,ortoitsownprivatenetwork.AslongasaLOS pathexists,LMDSwillallowLECstoinstallwirelessequipmentonthepremisesofcustomersforrapid broadband connectivity without having to lease or install its own cables to the customers.

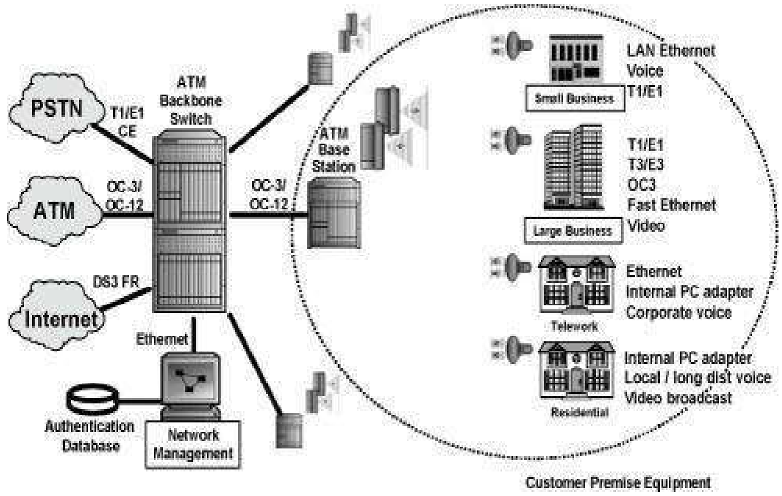


Figure2.7.AwirelessCompetitiveLocalExchangeCarrier(CLEC)usingAsynchronousTransferMode (ATM)distribution.

WirelessLocalArea Networks(WLANs)

In 1997 the FCC allocated 300 MHz of unlicensed spectrum in the Industrial Scientific and Medical (ISM) bands of 5.150–5.350 GHz and 5.725–5.825 GHz for the express purpose of supporting low-power license-free spread spectrum data communication. This allocation is called the *Unlicensed National Information Infrastructure (UNII)* band.

The IEEE 802.11 Wireless LAN working group was founded in 1987 to begin standardization of spread spectrum WLANs for use in the ISM bands. Figure 2.10 illustrates the evolution of IEEE 802.11 Wireless LAN standards, which also include infrared communications. Figure 2.10 shows how both frequency hopping and direct sequence approacheswere used in the original IEEE 802.11 standard (2 Mbps user throughput), butas of late 2001 only direct sequence spread spectrum (DS-SS) modems had thus far been standardized for high rate (11 Mbps) user data rates within IEEE 802.11.

The DS-SS IEEE 802.11b standard has been named *Wi-Fi* by the *Wireless Ethernet Compatibility Alliance (WECA)*, a group that promotes adoption of 802.11b DS-SS WLAN equipment and interoperability between vendors. IEEE 802.11g is developing *Complimentary Code Keying Orthogonal Frequency Division Multiplexing (CCK-OFDM)* standards in both the 2.4 GHz (802.11b) and 5 GHz (802.11a) bands, and will support roaming capabilities and dual-band use for public WLAN networks, while supporting backward compatibility with 802.11b technology.

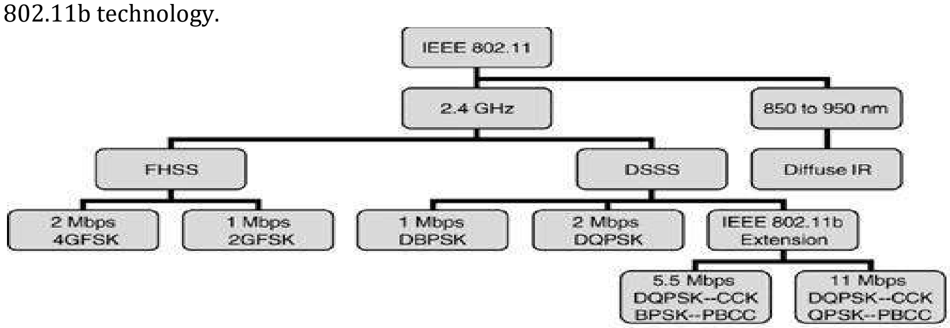


Figure2.10.OverviewoftheIEEE802.11WirelessLANstandard.

The frequency-hopping spread spectrum (FH-SS) proponents of IEEE 802.11 have formed the HomeRF standard that supports frequency hopping equipment. Itis worth noting thatboth DSand FH types of WLANs must operate in the same unlicensed bands that contain cordless phones, baby monitors, Bluetooth devices, and other WLAN users.

Table2.4.IEEE802.11bChannelsforBothDS-SSandFH-SSWLANStandards

Country	**FrequencyRangeAvailable**	**DSSSChannels Available**	**FHSSChannelsAvailable**
UnitedStates	2.4to2.4835 GHz	1through11	2through80
Canada	2.4to2.4835 GHz	1through11	2through80
Japan	2.4to2.497GHz	1through14	2through95
France	2.4465to2.4835 GHz	10through13	48through82
Spain	2.445to2.4835 GHz	10through11	47through73

Country	FrequencyRangeAvailable	DSSSChannels Available	FHSSChannelsAvailable
RemainderofEurope	2.4to2.4835	1through13	2through80

Figure 2.12 illustrates the unique WLAN channels that are specified in the IEEE 802.11b standard for the 2400–2483.5MHzband. AllWLANsaremanufacturedto operateonanyone of the specifiedchannelsandare assigned to a particular channel by the network operator when the WLAN system is first installed. The channelization scheme used by the network installer becomes very important for a high density WLAN installation, since neighboring access points must be separated from one another in frequency to avoid interference and significantly degraded performance.

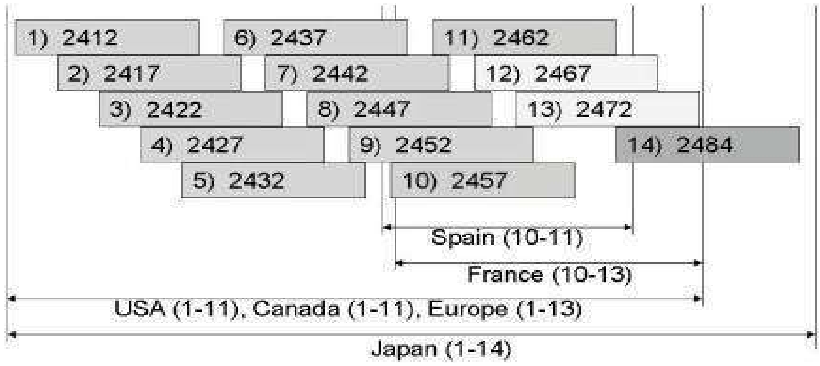


Figure 2.12.ChannelizationschemeforIEEE802.11b throughouttheworld.

In Europe in the mid 1990s, the *High Performance Radio Local Area Network* (HIPERLAN) standard was developed to provide a similar capability to IEEE 802.11. HIPERLAN was intended to provide individual wireless LANs for computer communications and used the 5.2 GHz and the 17.1 GHz frequency bands. HIPERLAN provides asynchronous user data rates of between 1 to 20 Mbps, as well as time bounded messaging at rates of 64 kbps to 2.048 Mbps. HIPERLAN was designed to operate up to vehicle speeds of 35 km/hr, and typically provided 20 Mbps throughput at 50 m range.

In 1997, Europe's ETSI established a standardization committee for Broadband Radio Access Networks (BRANs). The goal of BRAN is to develop a family of broadband WLAN-type protocols that allow user interoperability, covering both short range (e.g., WLAN) and long range (e.g., fixed wireless) networking. HIPERLAN/2 has emerged as the nextgeneration European WLAN standard and will provide up to 54 Mbpsof user data toa variety of networks, including theATM backbone, IP based networks, and theUMTS core.

BluetoothandPersonalAreaNetworks(PANs)

Bluetooth is an open standard that has been embraced by over 1,000 manufacturers of electronic appliances. It provides an ad-hoc approach for enabling various devices to communicate with one another within a nominal 10 meter range. Named after King Harald Bluetooth, the 10thcentury Viking who united Denmark and Norway, the Bluetooth standard aims to unify the connectivity chores of appliances within the personal workspace of an individual.

Bluetooth operates in the 2.4 GHz ISM Band (2400–2483.5 MHz) and uses a frequency hopping TDD scheme for each radio channel. Each Bluetooth radio channel has a 1 MHz bandwidth and hops at a rate of approximately 1600 hops per second. Transmissions are performed in 625 microsecond slots with a single packet transmitted over a single slot. For long data transmissions, particular users may occupy multiple slots using the same transmission frequency, thus slowing the instantaneous hopping rate to below 1600 hops/second. The frequency hopping scheme of each Bluetooth user is determined from a cyclic code of length $2^{27} - 1$, and each user has a channel symbol rate of 1 Mbps using GFSK modulation. The standard has been designed to support operation in very high interference levels and relies on a number of forward error control (FEC) coding and automatic repeat request (ARQ) schemes to support a raw channel bit error rate (BER) of about 10^{-3}.

Differentcountrieshaveallocatedvariouschannelsfor Bluetoothoperation. IntheUSandmostof Europe, the FHSS 2.4 GHz ISM band is available for Bluetooth use. A detailed list of states is defined in the Bluetooth standard to support a wide range of applications, appliances, and potential uses of the PersonalArea Network. Audio, text, data, and even video is contemplated in the Bluetooth standard [Tra01]. Figure 2.17 provides a depiction of the Bluetooth conceptwhere a gateway to the Internet via IEEE 802.11b isshownasa conceptualpossibility.

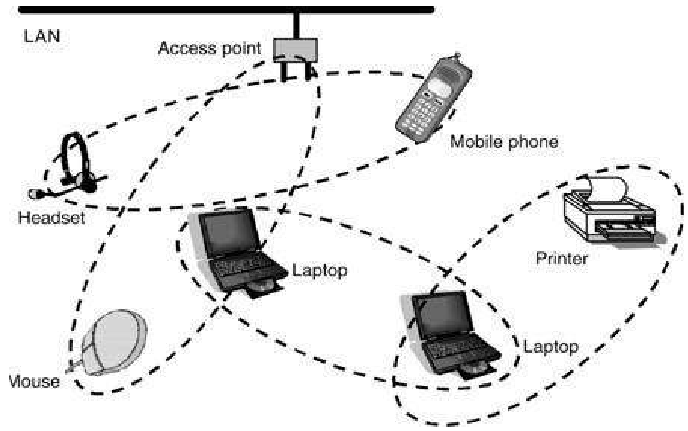


Figure2.17.ExampleofaPersonalAreaNetwork(PAN)asprovidedbytheBluetoothstandard.

The IEEE 802.15 standards committee has been formed to provide an international forum for developing Bluetooth and other PANs that interconnect pocket PCs, personal digital assistants (PDAs), cellphones, light projectors, and other appliances [Bra00]. With the rapid proliferation of *wearable computers*, such as PDAs, cellphones, smart cards, and position location devices, PANs may provide the connection to an entire new era of remote retrieval and monitoring of the world around us.

UNIT II
MobileRadioPropagation(Large-ScalePathLoss)

Introduction

Therearetwobasicwaysoftransmittinganelectro-magnetic(EM)signal,throughaguidedmedium or through an unguidedmedium.Guided mediums such as coaxial cables and fiber optic cables,are far less hostile toward the information carrying EM signal than the wireless or the unguided medium. It presents challenges and conditions which are unique for this kind of transmissions. A signal, as it travels through the wireless channel, undergoes many kinds of propagation effects such as reflection, diffraction and scattering, due tothe presence ofbuildings,mountains and other such obstructions. Reflection occurs when the EM waves impinge on objects which are much greater than the wavelength of the traveling wave. Diffraction is a phenomena occurring when the wave interacts with asurfacehaving sharp irregularities. Scatteringoccurs whenthemediumthroughthe wave is traveling contains objects which are much smaller than the wavelength of the EM wave. These varied phenomena's lead to large scale and small scale propagation losses. Due to the inherent randomness associated with such channels they are best described with the help of statistical models. Models which predict the mean signal strength for arbitrary transmitter receiver distances are termed as large scale propagation models. These are termed so because they predict the average signal strength for large Tx-Rx separations, typically for hundreds of kilometers.

FreeSpacePropagationModel:

The free space propagation model is used to predict received signal strength when the transmitter and receiver have a clear, unobstructed line-of-sight path between them. Satellite communication systems and microwave line-of-sight radio links typically undergo free space propagation. As with most large- scale radio wave propagation models, the free space model predicts that received power decays as a function of the T-R separation distance raised to some power (i.e. a power law function). The free space power received by a receiver antenna which is separated from a radiating transmitter antenna by a distance d, is given by the Friis free space equation

$$\mathbf{P_r(d)=P_tG_tG_r\lambda^2/(4\pi d)^2}$$

Where P_t is the transmitted power, P_r (d) is the received power which is a function of the T-R separation,G_t is the transmitter antenna gain, G_r is the receiver antenna gain, d is the T-R separation distance inmeters and λ is the wavelength in meters. The gain of an antenna is related to its effective aperture, A_e by,

$$\mathbf{G=4\pi A_e/\lambda^2}$$

TheeffectiveapertureA$_e$isrelatedtothephysicalsizeoftheantenna,andλisrelatedtothe carrierfrequency by,

$\lambda = c/f = 2\pi c/\omega_c$

where f is the carrier frequency in Hertz, ω_c is the carrier frequency in radians per second and c is the speed of light given in meters/s. An isotropic radiator is an ideal antenna which radiates power with unit gain uniformly in all directions, and is often used to reference antenna gains in wireless systems. The effective isotropic radiated power (EIRP) is defined as **$EIRP = P_tG_t$**

it represents the maximum radiated power available from a transmitter in the direction of maximum antenna gain, as compared to an isotropic radiator. In practice, effective radiated power (ERP) is used instead of EIRP to denote the maximum radiated power as compared to a half-wave dipole antenna (instead of an isotropic antenna).
The path loss, which represents signal attenuation as a positive quantity measured in dB, is defined as the difference (in dB) between the effective transmitted power and the received power, andmay or may not include the effect of the antenna gains. The path loss for the free space model when antenna gains are included is given by

$$PL(dB)=10\log(P_t/P_r)=-10\log\left[G_tG_r\lambda^2/(4\pi d)^2\right]$$

When antenna gains are excluded, the antennas are assumed to have unity gain, and path loss is given by **$PL\ (dB) = 10\log(P_t/P_r) = -10\log[\lambda^2/(4\pi d)^2]$**

The Friis free space model is only a valid predictor forP_r for valuesof d which are in the far-field of the transmitting antenna. The far-field or Fraunhofer region of a transmitting antenna is defined as the region beyond the far-field distance d_f, which is related to the largest linear dimension of the transmitter antenna aperture and the carrier wavelength. The Fraunhofer distance is given by

$$d_f = 2D^2/\lambda$$

WhereDisthelargestphysicallineardimensionoftheantenna.Thefar-fieldregion**d_f**mustsatisfy

$$d_f >> D$$

 # What will be the far-field distance for a Base station antenna with Largest dimensionD=0.5mFrequency of operation fc=900MHz

Sol:

λ=c/f=3*10^8/900*10^6)=0.33m

df=2D^2/λ=2(0.5)^2/0.33=1.5m

 # If a transmitter produces 50 watts of power, express the transmit power in units of (a) dBm, and(b) dBW If 50 watts is applied to a unity gain antenna with a 900 MHz carrier frequency, find the received power in dBm at a free space distance of 100 m fromthe antenna. What is Pr(10 km) ? Assume unity gain for the receiver antenna.

Given:
Transmitter power, $P_t = 50$ W.
Carrier frequency, $f_c = 900$ MHz

Using equation (3.9),
(a) Transmitter power,

$$P_t(\text{dBm}) = 10\log[P_t(\text{mW})/(1\text{ mW})]$$
$$= 10\log[50 \times 10^3] = 47.0\text{ dBm.}$$

(b) Transmitter power,

$$P_t(\text{dBW}) = 10\log[P_t(\text{W})/(1\text{ W})]$$
$$= 10\log[50] = 17.0\text{ dBW.}$$

The received power can be determined using equation (3.1).

$$P_r = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L} = \frac{50(1)(1)(1/3)^2}{(4\pi)^2(100)^2(1)} = 3.5 \times 10^{-6}\text{ W} = 3.5 \times 10^{-3}\text{ mW}$$

$$P_r(\text{dBm}) = 10\log P_r(\text{mW}) = 10\log\left(3.5 \times 10^{-3}\text{ mW}\right) = -24.5\text{ dBm.}$$

The received power at 10 km can be expressed in terms of dBm using equation (3.9), where $d_0 = 100$ m and $d = 10$ km

$$P_r(10\text{ km}) = P_r(100) + 20\log\left[\frac{100}{10000}\right] = -24.5\text{ dBm} - 40\text{ dB}$$
$$= -64.5\text{ dBm.}$$

ThreeBasicPropagationMechanisms

Reflection, Diffraction and Scattering is the three basic propagation mechanisms which impact propagation in a mobile communication system.

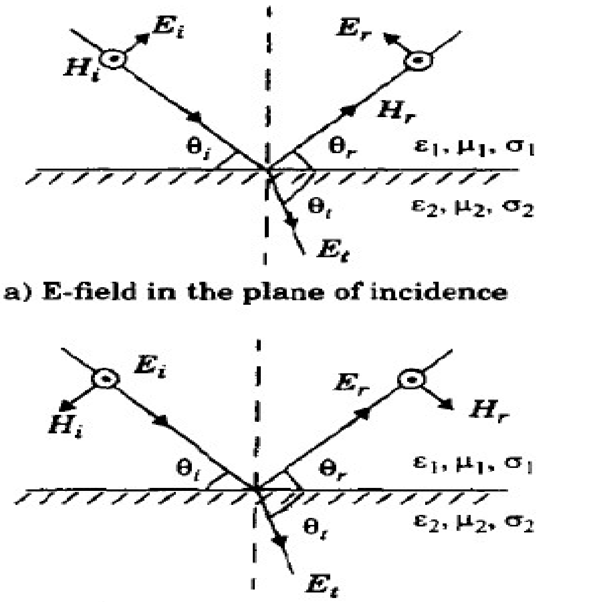
Reflection occurs when a propagating electromagnetic wave impinges upon an object which has very large dimensions when compared to the wavelength of the propagating wave. Reflections occur from the surface of the earth and from buildings andwalls.

Diffraction occurs when the radio path between the transmitter and receiver is obstructed by a surface that has sharp irregularities (edges). The secondary waves resulting from the obstructing surface are present throughout the space and even behind the obstacle, giving rise to a bending of waves around the obstacle, even when a line-of-sight path does not exist between transmitter and receiver. At high frequencies, diffraction, like reflection depends on the geometry of the object, as well as the amplitude, phase, and polarization of the incident wave at the point of diffraction.

Scattering occurs when the medium through which the wave travels consists of objects with dimensions that are small compared to the wavelength, and where the number of obstacles per unit volume is large. Scattered waves are produced by rough surfaces, small objects, or by other irregularities in the channel. In practice, foliage, street signs, and lamp posts induce scattering in a mobile communications system.

Reflection:

When a radio wave propagating in one medium impinges upon another medium having different electrical properties, the wave is partially reflected and partially transmitted. If the plane wave is incident on a perfect dielectric, part of the energy is transmitted into the second medium and part ofthe energy is reflected back into the first medium,and there is no loss ofenergy in absorption. If the second medium isa perfect conductor, then all incident energy is reflected back into the first medium without loss of energy. The electric field intensity of the reflected and transmitted waves may be related to the incident wave in the medium of origin through the Fresnel reflection coefficient (Γ). The reflection coefficient is a function of the material properties, and generally depends on the wave polarization, angle of incidence, and the frequency of the propagatingwave.



(b) E-field normal to the plane of incidence

Figure 3.4
Geometry for calculating the reflection coefficients between two dielectrics.

Reflectionfrom Dielectrics:

Figure 3.4 shows an electromagnetic wave incident at an angle θ_i with the plane of the boundary between two dielectric media. As shown in the figure, part of the energy is reflected back to the first mediaatanangleθ_r,andpartoftheenergyistransmitted(refracted)intothe secondmediaat an angle θ_t. The nature of reflection varies with the direction of polarization of the E-field. The behavior forarbitrary directionsof polarization can be studiedbyconsideringthe two distinct cases shown inFigure

The plane of incidence is defined as the plane containing the incident, reflected, and transmitted rays. In Figure 3.4a, the E—field polarization is parallel with the plane of incidence (that is, the E-field has a vertical polarization, or normal component, with respect to the reflecting surface) and in Figure 3.4b, the E-field polarization is perpendicular to the plane of incidence (that is, the incidentE-field is pointing out ofthe page towards the reader, and is perpendicular to the page and parallel tothereflectingsurface).

Because of superposition, only two orthogonal polarizations need be considered to solve general reflection problems. The reflection coefficients for the two cases of parallel and perpendicular E-field polarization at the boundary of two dielectrics are givenby

$$\Gamma_{\parallel} = \frac{E_r}{E_i} = \frac{\eta_2 \sin\theta_t - \eta_1 \sin\theta_i}{\eta_2 \sin\theta_t + \eta_1 \sin\theta_i} \qquad \text{(E-field in plane of incidence)}$$

$$\Gamma_{\perp} = \frac{E_r}{E_i} = \frac{\eta_2 \sin\theta_i - \eta_1 \sin\theta_t}{\eta_2 \sin\theta_i + \eta_1 \sin\theta_t} \qquad \text{(E-field not in plane of incidence)}$$

Whereɳistheintrinsicimpedanceoftherespectivemedium.

Where ε is the permittivity of the respective medium.

BrewsterAngle:

$$\Gamma_{\parallel} = \frac{-\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}}{\varepsilon_r \sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}}$$

$$\Gamma_{\perp} = \frac{\sin\theta_i - \sqrt{\varepsilon_r - \cos^2\theta_i}}{\sin\theta_i + \sqrt{\varepsilon_r - \cos^2\theta_i}}$$

The Brewster angle is the angle at which no reflection occurs in the medium of origin. It occurs when the incident angle BB is such that the reflection coefficient Γ|| is equal to zero. The Brewster angle is given by the value of θ_B which satisfies

$\sin(\theta_B) = \sqrt{\varepsilon_1}/\sqrt{\varepsilon_1+\varepsilon_2}$

Forthecasewhenthe firstmediumisfree spaceandthe secondmediumhasarelativepermittivity ε_r,aboveequationcanbeexpressedas

$\sin(\theta_B) = \sqrt{(\varepsilon_r-1)}/\sqrt{(\varepsilon_r^2-1)}$ Note that the Brewster angle occurs only for vertical (i.e. parallel) polarization.

ReflectionfromPerfectConductors:

Since electromagnetic energy cannot pass through a perfect conductor a plane wave incident on a conductor has all of its energy reflected. As the electric field at the surface of the conductor mustbe equal to zero at all times in order to obey Maxwel1's equations, the reflected wave must be equal in magnitude to the incident wave. For the case when E-field polarization is in the plane of incidence, the boundary conditions require that $\theta_i = \theta_r$ and $E_i = E_r$ (E-field in plane of incidence)

Similarly,forthecasewhentheE-fieldishorizontallypolarized,theboundaryconditionsrequire that **θ_i =θ_r and E_i = - E_r** (E-field not in plane of incidence)

GroundReflection(2-ray)Model:

In a mobile radio channel, a single direct path between the base station and a mobile is seldom the only physical means for propagation, and hence the free space propagation model is in most cases inaccurate when used alone. The 2-ray ground reflection model shown in Figure 3.7 is a useful propagation model that is based on geometric optics, and considers both the direct path and a groundreflectedpropagationpathbetweentransmitterandreceiver.Thismodelhasbeenfoundto be reasonably accurate for predicting the large-scale signal strength over distances of several kilometers for mobile radio systems that use tall towers (heights which exceed 50 m),as wellas for line of-sight, microcell channels in urban environments.

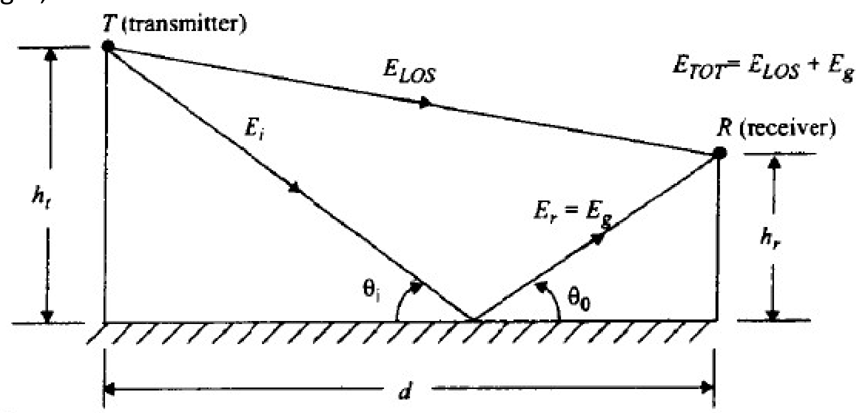


Figure 3.7
Two-ray ground reflection model.

Referringto Figure 3.7, $h_t$is the height of the transmitter and hr isthe height of the receiver. If $E_o$is the free space E-field (in units of V/m) at a reference distance do from the transmitter, then for d > do , the free space propagating E—field is given by

$$E(d,t) = \frac{E_0 d_0}{d}\cos\left(\omega_c\left(t-\frac{d}{c}\right)\right) \qquad (d > d_0)$$

Twopropagatingwavesarriveatthereceiver:thedirectwavethattravelsadistanced';andthe reflectedwavethattravelsadistanced".

TheelectricfieldE$_{TOT}$(d,t)canbeexpressedasthesumofequationsfordistancesd'andd"(i.e. direct waveandreflectedwave).

$$E_{TOT}(d,t) = \frac{E_0 d_0}{d'}\cos\left(\omega_c\left(t-\frac{d'}{c}\right)\right) + (-1)\frac{E_0 d_0}{d''}\cos\left(\omega_c\left(t-\frac{d''}{c}\right)\right)$$

Diffraction:

Diffraction allows radio signals to propagate around the curved surface of the earth, beyond the horizon, and to propagate behind obstructions. Although the received field strength decreases rapidly as a receiver moves deeper into the obstructed (shadowed) region, the diffraction field still exists and often has sufficient strength to produce a useful signal.

Thephenomenonofdiffractioncanbeexplainedby Huygens'sprinciple,which statesthatallpoints on a wave front can be considered as point sources for the production of secondary wavelets, and that these wavelets combine to produce a new wave front in the direction of` propagation. Diffraction is caused by the propagation of secondary wavelets into a shadowed region. The field strength of a diffracted wave in the shadowed region is the vector sum of the electric field components of all the secondary wavelets in the space around the obstacle.

FresnelZone Geometry:

Fresnel zones represent successive regions where secondary waves have a path length from the TX to the RX which arenλ/2 greater in path length than of the LOS path. The plane below illustrates successive Fresnelzones.

$$r_n = \sqrt{\frac{n\lambda d_1 d_2}{d_1 + d_2}}$$

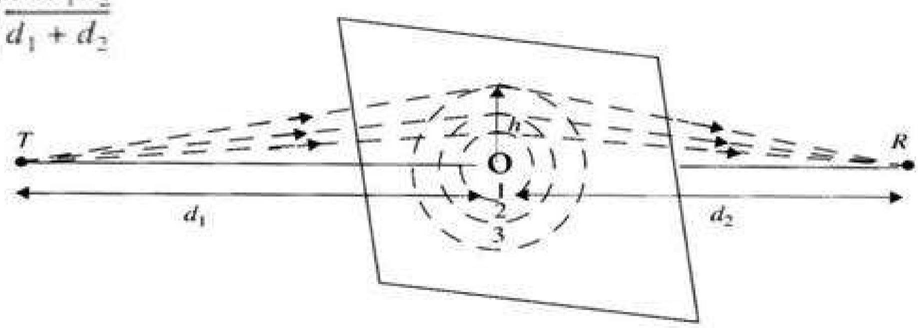


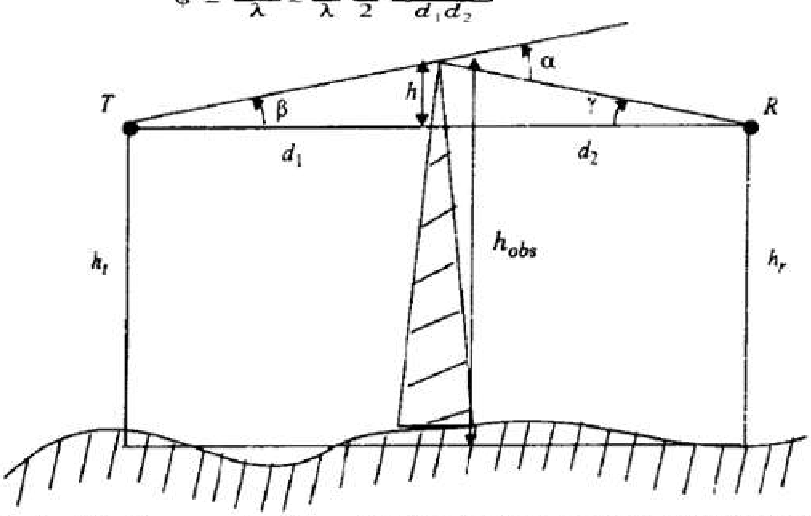
Figure 4.11 Concentric circles which define the boundaries of successive Fresnel zones.

Consider a transmitter and receiver separated in free space as shown in Figure 3.10a. Let an obstructingscreenofeffectiveheighthwithinfinitewidth(goingintoandoutofthepaper,)be

$$\Delta = \frac{h^2}{2}\frac{(d_1 + d_2)}{d_1 d_2}$$

The corresponding phase difference is given by

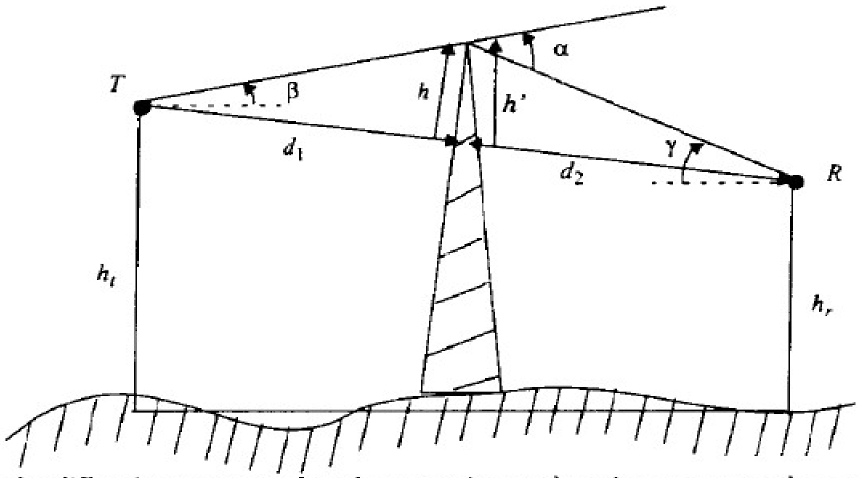
$$\phi = \frac{2\pi\Delta}{\lambda} = \frac{2\pi}{\lambda}\frac{h^2}{2}\frac{(d_1 + d_2)}{d_1 d_2}$$



(a) Knife-edge diffraction geometry. The point T denotes the transmitter and R denotes the receiver, with an infinite knife-edge obstruction blocking the line-of-sight path.

placed between them ata distanced1from thetransmitter and d2 from the receiver. It is apparent that the wave propagating from the transmitter to the receiver via the top of the screen travels a longerdistancethanifadirectline-of-sightpath(throughthescreen)existed.Assumingh<<d1,d2

andh>>λ,thenthedifferencebetweenthedirectpathandthediffractedpath,calledtheexcess pathlength(Δ),canbeobtainedfromthegeometryofFigureas



(b) Knife-edge diffraction geometry when the transmitter and receiver are not at the same height. Note that if α and β are small and $h << d_1$ and d_2,then h and h' are virtually identical and the geometry may be redrawn as shown in Figure 3.10c.

Knife-edgeDiffractionModel:

Estimating the signal attenuation caused by diffraction of radio waves over hills and buildings is essential in predicting the field strength in a given service area. Generally, it is impossible to make very precise estimates of the diffraction losses, and in practice prediction is a process of theoretical approximation modified by necessary empirical corrections. Though the calculation of diffraction losses over complex and irregular terrain is a mathematically difficult problem, expressions for diffraction losses for many simple cases have been derived. As a starting point, the limiting case of propagation over a knife-edge gives good insight into the order of magnitude of diffraction loss.

When shadowing is caused by a single object such as a hill or mountain, the attenuation caused by diffraction can be estimated by treating the obstruction as a diffracting knife edge. This is the simplest of diffraction models, and the diffraction loss in this case can be readily estimated using the classical Fresnel solution for the field behind a knife edge (also called a half-plane).

MultipleKnife-edgeDiffraction:

In many practical situations, especially in hilly terrain, the propagation path may consist of more than one obstruction, in which case the total diffraction loss due to all of the obstacles must be computed. Burlington suggested that the series of obstacles be replaced by a single equivalent obstacle so that the path loss can be obtained using single knife-edge diffraction models. This method, illustrated in Figure 3.15,oversimplifies the calculationsand often provides veryoptimistic estimatesofthereceivedsignalstrength.Inamorerigoroustreatment,Millingtonet.al.gavea wave-theory solution for the field behind two knife edges in series. This solution is very useful and can be applied easily for predicting diffraction losses due to two knife edges. However, extending this to morethantwo knife edges becomes a formidablemathematical problem.Many models that are mathematically less complicated have been developed to estimate the diffraction losses due to multiple obstructions.

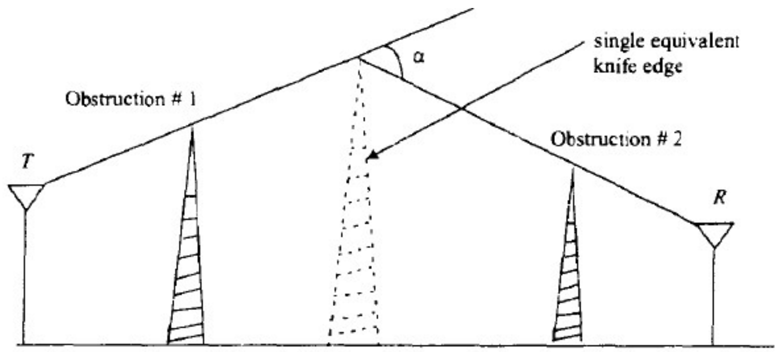


Figure 3.15
Bullington's construction of an equivalent knife edge {From [Bul47] © IEEE].

Scattering:

Theactualreceivedsignal inamobileradioenvironmentisoftenstrongerthanwhatispredictedby reflection and diffraction models alone. This is because when a radio wave impinges on a rough surface,thereflected energy is spread out(diffused) inalldirectionsdueto scattering.Objects such as lamp posts and trees tend to scatter energy in all directions, thereby providing additional radio energy at a receiver. Flat surfaces that have much larger dimension than a wavelength may be modeled as reflective surfaces.However, the roughness of such surfacesoften inducespropagation effects different from the specular reflection described earlier in this chapter.

Rayleigh criterion: used for testing surfaceroughness A surface is considered smooth if its min to maxprotuberance (bumps) h is less than critical heighth_c= $\lambda/8 \sin\Theta_i$

Scatteringpath lossfactor $\rho_S$isgivenby $\rho_S=\exp[-8[(\pi*\sigma_h*\sin\Theta_i)/\lambda]^2]$

Wherehissurfaceheightand$\sigma_h$isstandarddeviationofsurfaceheightaboutmeansurfaceheight. For rough surface, the flat surface reflection coefficient is multiplied by scattering loss factor ρ_S to account for diminished electricfield.

ReflectedE-fieldsfor$h>h_c$forroughsurfacecanbe calculatedas $\Gamma_{rough} = \rho_S\Gamma$

A surface is considered smooth if it’s minimum to maximum protuberance h is less than h_c, and is considered rough if the protuberance is greater than h_c. For rough surfaces, the flat surface reflection coefficient needs to be multiplied by a scattering loss factor, $\rho_s$to account for the diminished reflected field.

OUTDOORPROPAGATIONMODELS

Basedonthecoveragearea,theOutdoorpropagationenvironmentmaybedividedintothree categories
1. PropagationinMacrocells
2. PropagationinMicrocells
Outdoor radio transmission takes place over an irregular terrain. The terrain profile must be taken into consideration for estimating the path loss.
e.g.treesbuildingsandhillsmustbetakenintoconsideration
Longley-Ryce Model:

The Longley-Ryce model is applicable to point-to-point communication systems in the frequency range from 40 MHz to 100 GHz, over different kinds of terrain. The median transmission loss is predicted using the path geometry of the terrain profile and the refractivity of the troposphere. Geometric optics techniques (primarily the 2-ray ground reflection model) are used to predictsignal strengths within the radio horizon. Diffraction losses over isolated obstacles are estimated using the Fresnel-Kirchoff knife- edge models. Forward scatter theory is used to make troposcatter predictions over long distances.

The Longley-Ryce method operates in two modes. When a detailed terrain path profile is available, the path-specific parameters can be easily determined and the prediction is called a point-to-point mode prediction. On the other hand, if the terrain path profile is not available, the Longley-Ryce method provides techniques to estimate the path-specific parameters, and such a prediction is called an area mode prediction.

Okumura Model:

Okumura`s model is one of the most widely used models for signal prediction in urban areas. This model is applicable for frequencies in the range 150 MHz to 1920 MHz (although it is typically extrapolated up to 3000 MHz) and distances of 1 km to 100 km. It can be used for base station antenna heights ranging from 30 m to 1000 m. Okumura developed a set of curves giving the median attenuation relative to free space (Amu), in an urban area over a quasi-smooth terrain with a base station effective antenna height (hte) of 200 m and a mobile antenna height (hre) of 3 m. These curves were developed from extensive measurements using vertical omni-directional antennas at both the base and mobile, and are plotted as a function of frequency in the range 100 MHz to 1920 MHz and as a function of distance from the base station in the range 1 km to 100 km. To determine path loss using Okumura`s model, the free space path loss between the points of interest is first determined, and then the value of Amu(f, d) (as read from the curves) is added to it along with correction factors to account for the type of terrain. The model can be expressedas

$L_{50}(dB) = L_F + A_{mu}(f,d) - G(t_e) - G(r_e) - G_{AREA}$

where L50 is the 50th percentile (i.e., median) value of propagation path loss, L_F is the free space propagation loss, A_{mu} is the median attenuation relative to free space, $G(h_{te})$ is the base station antenna height gain factor, $G(h_{re})$ is the mobile antenna height gain factor, and G_{AREA} is thegain due to the type of environment. Note that the antenna height gains are strictly a function of height and have nothing to do with antenna patterns.

$$G(h_{te}) = 20\log\left(\frac{h_{te}}{200}\right) \qquad 1000\text{ m} > h_{te} > 30\text{ m}$$

$$G(h_{re}) = 10\log\left(\frac{h_{re}}{3}\right) \qquad h_{re} \le 3\text{ m}$$

$$G(h_{re}) = 20\log\left(\frac{h_{re}}{3}\right) \qquad 10\text{ m} > h_{re} > 3\text{ m}$$

HataModel:

The Hata model [Hat90] is an empirical formulation of the graphical path loss data provided by Okumura, and isvalid from 150 MHz to 1500 MHz. Hata presented the urban area propagation loss as a standard formula and supplied correction equations for application to other situations. The standard formula for median path loss in urban areas is given by

$L_{50}(urban)(dB)=69.55+26.16\log f_c-13.82\log h_{te}-a(h_{re})+(44.9-6.55\log h_{te})\log d$

where fc is the frequency (in MHz) from 150 MHz to 1500 MHz, h_{te} is the effective transmitter (base station) antenna height (in meters) ranging from 30 m to 200 m, h_{re} is the effective receiver (mobile) antenna height (in meters) ranging from 1 m to 10 m, d is the T-R separation distance (in km), and $a(h_{re})$ is the correction factor for effective mobile antenna height which is a function of the size of the coverage area. For a small to medium sized city, the mobile antenna correctionfactor is given by

$a(h_{re})=(1.11ogf_c-0.7)h_{re}-(1.56\log f_c-0.8)$dB for a
large city, it is given by

$a(h_{re})=8.29(\log 1.54h_{re})^2-1.1$dBfor$f_c$<=300MHz$a(h_{re})=3.2(\log 11.75h_{re})2—4.97$dB for f_c>=
300 MHz

ToobtainthepathlossinasuburbanareathestandardHataformulainequationsaremodifiedas $L_{50}(dB) =$
$L_{50}(urban)-2[\log(f_c/28)]^2-5.4$

andforpathlossinopenruralareas,theformulaismodifiedas $L_{50}(dB) =$

$L_{50}(urban) – 4.78(\log f_c)^2 + 18.33\log f_c - 40.94$

PCS Extension to Hata Model

The European Co-operative for Scientific and Technical research (EURO-COST) formed the COST-231 working committee to develop an extended version of the Hata model. COST-231 proposed the following formula to extend Hata's model to 2 GHz. The proposed model for path loss is [EUR91]

$$L_{50}(urban) = 46.3 + 33.9\log f_c - 13.82\log h_{te} - a(h_{re}) + (44.9 - 6.55\log h_{te})\log d + C_M \tag{3.87}$$

where $a(h_{re})$ is defined in equations (3.83), (3.84.a), and (3.84.b) and

$$C_M = \begin{array}{ll} 0\ \text{dB} & \text{for medium sized city and suburban areas} \\ 3\ \text{dB} & \text{for metropolitan centers} \end{array} \tag{3.88}$$

The COST-231 extension of the Hata model is restricted to the following range of parameters:

f : 1500 MHz to 2000 MHz
h_{te} : 30 m to 200 m
h_{re} : 1 m to 10 m
d : 1 km to 20 km

Walfisch and Bertoni Model

A model developed by Walfisch and Bertoni [Wal88] considers the impact of rooftops and building height by using diffraction to predict average signal strength at street level. The model considers the path loss, S, to be a product of three factors.

$$S = P_0 Q^2 P_1 \tag{3.89}$$

where P_0 represents free space path loss between isotropic antennas given by

$$P_0 = \left(\frac{\lambda}{4\pi R}\right)^2 \tag{3.90}$$

The factor Q^2 gives the reduction in the rooftop signal due to the row of buildings which immediately shadow the receiver at street level. The P_1 term is based upon diffraction and determines the signal loss from the rooftop to the street.

In dB, the path loss is given by

$$S\,(\mathrm{dB}) = L_0 + L_{rts} + L_{ms} \tag{3.91}$$

where L_0 represents free space loss, L_{rts} represents the "rooftop-to-street diffraction and scatter loss", and L_{ms} denotes multiscreen diffraction loss due to the rows of buildings [Xia92]. Figure 3.25 illustrates the geometry used in the Walfisch Bertoni model [Wal88], [Mac93]. This model is being considered for use by ITU-R in the IMT-2000 standards activities.

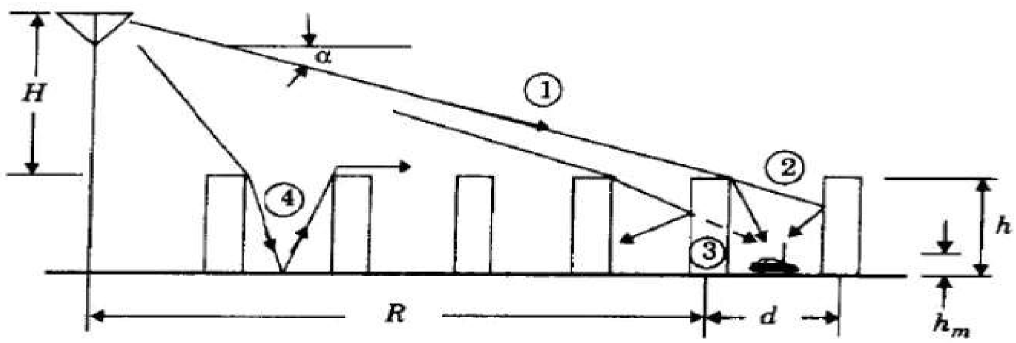


Figure 3.25
Propagation geometry for model proposed by Walfisch and Bertoni [From [Wal88] © IEEE].

Wideband PCS Microcell Model

Work by Feuerstein, et.al. in 1991 used a 20 MHz pulsed transmitter at 1900 MHz to measure path loss, outage, and delay spread in typical microcellular systems in San Francisco and Oakland. Using base station antenna heights of 3.7 m, 8.5 m, and 13.3 m, and a mobile receiver with an antenna height of 1.7 m above ground, statistics for path loss, multipath, and coverage area were developed from extensive measurements in line-of-sight (LOS) and obstructed (OBS) environments [Feu94]. This work revealed that a 2-ray ground reflection model (shown in Figure 3.7) is a good estimate for path loss in LOS microcells, and a simple log-distance path loss model holds well for OBS microcell environments.

For a flat earth ground reflection model, the distance d_f at which the first Fresnel zone just becomes obstructed by the ground (first Fresnel zone clearance) is given by

$$d_f = \frac{1}{\lambda}\sqrt{\left(\Sigma^2-\Delta^2\right)^2 - 2\left(\Sigma^2+\Delta^2\right)\left(\frac{\lambda}{2}\right)^2 + \left(\frac{\lambda}{2}\right)^4} \tag{3.92.a}$$

$$= \frac{1}{\lambda}\sqrt{16h_t^2h_r^2 - \lambda^2\left(h_t^2+h_r^2\right) + \frac{\lambda^4}{16}}$$

For LOS cases, a double regression path loss model that uses a regression breakpoint at the first Fresnel zone clearance was shown to fit well to measurements. The model assumes omnidirectional vertical antennas and predicts average path loss as

$$PL(d) = \begin{cases} 10n_1\log(d) + p_1 & \text{for } 1 < d < d_f \\ 10n_2\log(d/d_f) + 10n_1\log d_f + p_1 & \text{for } d > d_f \end{cases} \tag{3.92.b}$$

where p_1 is equal to $\overline{PL}(d_0)$ (the path loss in decibels at the reference distance of $d_0 = 1$ m), d is in meters and n_1, n_2 are path loss exponents which are a function of transmitter height, as given in Figure 3.26. It can easily be shown that at 1900 MHz, $p_1 = 38.0$ dB.

For the OBS case, the path loss was found to fit the standard log-distance path loss law of equation (3.69.a)

$$\overline{PL}(d)\,[dB] = 10n\log(d) + p_1 \tag{3.92.c}$$

INDOORPROPAGATIONMODELS

With the advent of Personal Communication Systems (PCS), there is a great deal of interest in characterizing radio propagation inside buildings. The indoor radio channel differs from the traditional mobile radio channel in two aspects - the distances covered are much smaller, and the variabilityoftheenvironmentismuchgreaterforamuchsmallerrangeofT-Rseparation distances. It has been observed that propagation within buildings is strongly influenced by specific features such as the layout of the building, the construction materials, and the building type. This section outlines models for path loss within buildings.

Indoor radio propagation is dominated by the same mechanisms as outdoor: reflection, diffraction, and scattering.However,conditionsare much more variable. For example,signallevels vary greatly depending on whether interior doors are open or closed inside a building. Where antennas are mounted also impacts large-scale propagation. Antennas mounted at desk level in a partitioned office receive vastly different signals than those mounted on the ceiling. Also, the smaller propagation distances make it more difficult to insure far-field radiation for all receiver locations and types of antennas.

PartitionLosses(samefloor):

Buildings have a wide variety of partitions and obstacles which form the internal and external structure.Housestypicallyuse awoodframepartitionwithplasterboardto form internalwalls and have wood or non-reinforced concrete between floors. Office buildings, on the other hand, often have large open areas (open plan) which are constructed by using moveable office partitions sothat the space may be reconfigured easily, and use metal reinforced concrete between floors. Partitions that are formed as part of the building structure are called hard partitions, and partitions that may 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.

PartitionLossesbetweenFloors:

The lossesbetweenfloors of abuilding are determined by theexternal dimensionsand materialsof 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 (which attenuates radio energy) can impact the loss between floors. It can be seen that for all three buildings, the attenuation between one floors of the building is greater than the incremental attenuation caused by each additional floor. After about five or six floor separations, very little additional path loss is experienced.

Log-distancePathLossModel:

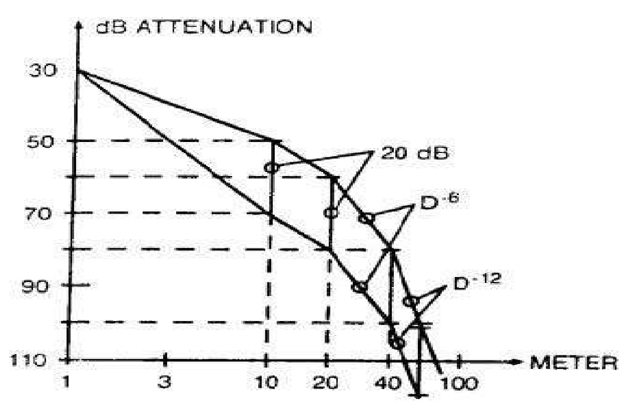
Accordingtothismodelthereceivedpoweratdistance*d*isgivenby,

$$PL\,(\mathrm{dB}) = PL\,(d_0) + 10n\log\left(\frac{d}{d_0}\right) + X_\sigma$$

Thevalueofnvarieswithpropagationenvironments.Thevalueofnis2forfreespace.Thevalueof n variesfrom4to6 for obstruction of building, and 3to5forurban scenarios. Theimportantfactor is to select the correct reference distance d_0. For large cell area it is 1 Km, while for micro-cell system it varies from 10m-1m.

Ericssonmultiplebreakpoint model:

It was obtained by measurements in a multiple floor office building. It has 4 breakpoints and considers both anupperandlowerboundonpath loss. Itassumesthatthere is30dBattenuationat d0 = 1m which is accurate for f = 900MHz & unity gain antennas.



AttenuationFactorModel:

The attenuation factor model incorporates a special path loss exponent and a floor attenuation factor to provide an estimate of indoor path loss. NSF is the path loss exponent for a same floor measurement and FAF is a floor attenuation factor based on the number of floors between transmitter and receiver.

RayTracingandSiteSpecificModeling

In physics, ray tracing is a method for calculating the path of waves or particles through a system with regions of varying propagation velocity, absorption characteristics, and reflecting surfaces. Under these circumstances, wave fronts may bend, change direction, or reflect off surfaces, complicating analysis. Ray tracing solves the problem by repeatedly advancing idealized narrow beams called rays through the medium by discrete amounts. Simple problems can be analyzed by propagatinga few rays using simple arithmetic.Moredetailedanalysis can beperformed by using a computer to propagate many rays.

When applied to problems of electromagnetic radiation, ray tracing often relies on approximate solutions to Maxwell's equations that are valid as long as the light waves propagate through and around objects whose dimensions are much greater than the light's wavelength. Ray theory does not describephenomenasuch as interference and diffraction, which require wavetheory(involving the phase of the wave).

UNIT-III

MobileRadioPropagation:Small–ScaleFadingandMultipath

Small Scale Multipath propagation-Factors influencing small scale fading, Doppler shift.Impulse Response Model of a multipath channel-Relationship between Bandwidth and Received power Small-Scale MultipathMeasurements-DirectRFPulseSystem,SpreadSpectrum SlidingCorrelator Channel Sounding, Frequency Domain Channels Sounding Parameters of MobileMultipathChannels-TimeDispersionParameters,CoherenceBandwidth,DopplerSpread and Coherence Time. Types of Small-Scale Fading-Fading effects Due to Multipath Time Delay Spread, Flat fading, Frequency selective fading.Fading effects Due to Doppler Spread-Fast fading, slow fading Statistical Models for multipath Fading Channels-Clarke‘s model for flat fading, spectral shape due to Doppler spread in Clarke‘s model, Simulation of Clarke and Gans Fading Model, Level crossing and fading statistics, Two-ray Rayleigh Fading Model.

SmallScale Multipath propagation:

Multipath in the radio channel creates small-scale fading effects. The three most important effects are:

- Rapid changes in signal strength over a small travel distance or time interval
- Random frequency modulation due to varying Doppler shifts on different multipath signals
- Time dispersion (echoes) caused by multipath propagation delays.

In built-up urban areas, fading occurs because the height of the mobile antennas are well below the height of surrounding structures, so there is no single line-of-sight path to the base station. Even when a line-of-sight exists, multipath still occurs due to reflections from the ground and surrounding structures. The incoming radio waves arrive from different directions with different propagation delays. The signal received by the mobile at any point in space may consist of a large number of plane waves having randomly distributed amplitudes, phases, and angles of arrival. These multipath components combine vectorially at the receiver antenna, and can cause the signal received by the mobile to distort or fade. Even when a mobile receiver is stationary, the received signal may fade due to movement of surrounding objects in the radio channel.

Factorsinfluencingsmallscalefading:

Many physical factors in the radio propagation channel influence small-scale fading. These include the following:

- **Multipath propagation** — The presence of reflecting objects and scatterers in the channel creates a constantly changing environment that dissipates the signal energy in amplitude, phase, and time. These effects result in multiple versions of the transmitted signal that arrive at the receiving antenna, displaced with respect to one another in time and spatial orientation. The random phase and amplitudes of the different multipath components cause fluctuations in signal strength, thereby inducing small-scale fading, signal distortion, or both. Multipath propagation often lengthens the time required for the baseband portion of the signal to reach the receiver which can cause signal smearing due to intersymbol interference.
- **Speed of the mobile** — The relative motion between the base station and the mobile results in random frequency modulation due to different Doppler shifts on each of the multipath components. Doppler shift will be positive or negative depending on whether the mobile receiver is moving toward or away from the base station.

- **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 the small-scale fading. Otherwise, motion of surrounding objects may be ignored, and only the speed of the mobile need be considered.
- **The transmission bandwidth of the signal** — If the transmitted radio signal bandwidth is greater than the "bandwidth" of the multipath channel, the received signal will be distorted, but the received signal strength will not fade much over a local area (i.e., the small-scale signal fading will not be significant). As will be shown, the bandwidth of the channel can be quantified by the *coherence bandwidth* which is related to the specific multipath structure of the channel. The coherence bandwidth is a measure of the maximum frequency difference for which signals are still strongly correlated in amplitude. If the transmitted signal has a narrow bandwidth as compared to the channel, the amplitude of the signal will change rapidly, but the signal will not be distorted in time. Thus, the statistics of small-scale signal strength and the likelihood of signal smearing appearing over small-scale distances are very much related to the specific amplitudes and delays of the multipath channel, as well as the bandwidth of the transmitted signal.

Dopplershift:

Consider a mobile moving at a constant velocity v, along a path segment having length d between points X and Y, while it receives signals from a remote source S, as illustrated in Figure 4.1. The difference in path lengths traveled by the wave from source S to the mobile at points X and Y is $\Delta l = d\cos\theta = v\Delta t\cos\theta$, where Δt is the time required for the mobile to travel from X to Y, and θ is assumed to be the same at points X and Y since the source is assumed to be very far away. The phase change in the received signal due to the difference in path lengths is therefore

$$\Delta\phi = \frac{2\pi\Delta l}{\lambda} = \frac{2\pi v\Delta t}{\lambda}\cos\theta \tag{4.1}$$

and hence the apparent change in frequency, or Doppler shift, is given by f_d, where

$$f_d = \frac{1}{2\pi}\cdot\frac{\Delta\phi}{\Delta t} = \frac{v}{\lambda}\cdot\cos\theta \tag{4.2}$$

Equation (4.2) relates the Doppler shift to the mobile velocity and the spatial angle between the direction of motion of the mobile and the direction of arrival of the wave. It can be seen from equation (4.2) that if the mobile is moving toward the direction of arrival of the wave, the Doppler shift is positive (i.e., the apparent received frequency is increased), and if the mobile is moving away from the direction of arrival of the wave, the Doppler shift is negative (i.e. the apparent received frequency is decreased). As shown in section 4.7.1, multipath components from a CW signal which arrive from different directions contribute to Doppler spreading of the received signal, thus increasing the signal bandwidth.

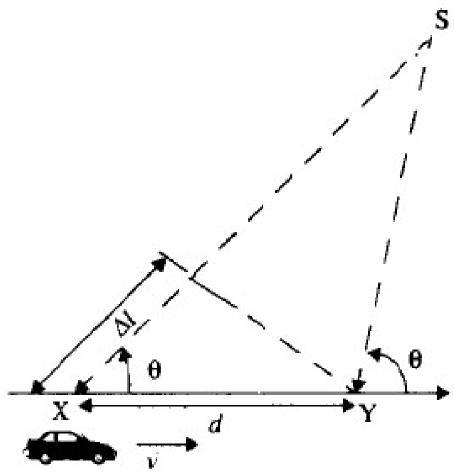


Figure 4.1
Illustration of Doppler effect.

Example 4.1

Consider a transmitter which radiates a sinusoidal carrier frequency of 1850 MHz. For a vehicle moving 60 mph, compute the received carrier frequency if the mobile is moving (a) directly towards the transmitter, (b) directly away from the transmitter, (c) in a direction which is perpendicular to the direction of arrival of the transmitted signal.

Solution to Example 4.1

Given:

Carrier frequency $f_c = 1850\,\text{MHz}$

Therefore, wavelength $\lambda = c/f_c = \dfrac{3\times10^8}{1850\times10^6} = 0.162\,\text{m}$

Vehicle speed $v = 60\text{ mph} = 26.82\text{ m/s}$

(a) The vehicle is moving directly towards the transmitter.
The Doppler shift in this case is positive and the received frequency is given by equation (4.2)

$$f = f_c + f_d = 1850\times10^6 + \frac{26.82}{0.162} = 1850.00016\text{ MHz}$$

(b) The vehicle is moving directly away from the transmitter.
The Doppler shift in this case is negative and hence the received frequency is given by

$$f = f_c - f_d = 1850 \times 10^6 - \frac{26.82}{0.162} = 1849.999834 \text{ MHz}$$

(c) The vehicle is moving perpendicular to the angle of arrival of the transmitted signal.
In this case, $\theta = 90°$, $\cos\theta = 0$, and there is no Doppler shift.
The received signal frequency is the same as the transmitted frequency of 1850 MHz.

Impulse Response Model of a Multipath Channel

The small-scale variations of a mobile radio signal can be directly related to the impulse response of the mobile radio channel. The impulse response is a wideband channel characterization and contains all information necessary to simulate or analyze any type of radio transmission through the channel. This stems from the fact that a mobile radio channel may be modeled as a linear 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. The impulse response is a useful characterization of the channel, since it may be used to predict and compare the performance of many different mobile communication systems and transmission bandwidths for a particular mobile channel condition.

To show that a mobile radio channel may be modeled as a linear filter with a time varying impulse response, consider the case where time variation is due strictly to receiver motion in space. This is shown in Figure 4.2.

Figure 4.2
The mobile radio channel as a function of time and space.

In Figure 4.2, the receiver moves along the ground at some constant velocity v. For a fixed position d, the channel between the transmitter and the receiver can be modeled as a linear time invariant system. However, due to the different multipath waves which have propagation delays which vary over different spatial locations of the receiver, the impulse response of the linear time invariant channel should be a function of the position of the receiver. That is, the channel impulse response can be expressed as $h(d,t)$. Let $x(t)$ represent the transmitted signal, then the received signal $y(d,t)$ at position d can be expressed as a convolution of $x(t)$ with $h(d,t)$.

$$y(d,t) = x(t) \otimes h(d,t) = \int_{-\infty}^{\infty} x(\tau)h(d,t-\tau)d\tau \tag{4.3}$$

For a causal system, $h(d,t) = 0$ for $t<0$, thus equation (4.3) reduces to

$$y(d,t) = \int_{-\infty}^{t} x(\tau)h(d,t-\tau)d\tau \tag{4.4}$$

Since the receiver moves along the ground at a constant velocity v, the position of the receiver can by expressed as

$$d = vt \tag{4.5}$$

Substituting (4.5) in (4.4), we obtain

$$y(vt,t) = \int_{-\infty}^{t} x(\tau)h(vt,t-\tau)d\tau \tag{4.6}$$

Since v is a constant, $y(vt,t)$ is just a function of t. Therefore, equation (4.6) can be expressed as

$$y(t) = \int_{-\infty}^{t} x(\tau)h(vt,t-\tau)d\tau = x(t) \otimes h(vt,t) = x(t) \otimes h(d,t) \tag{4.7}$$

From equation (4.7) it is clear that the mobile radio channel can be modeled as a linear time varying channel, where the channel changes with time and distance.

Since v may be assumed constant over a short time (or distance) interval, we may let $x(t)$ represent the transmitted bandpass waveform, $y(t)$ the received waveform, and $h(t,\tau)$ the impulse response of the time varying multipath radio channel. The impulse response $h(t,\tau)$ completely characterizes the channel and is a function of both t and τ. The variable t represents the time variations due to motion, whereas τ represents the channel multipath delay for a fixed value of t. One may think of τ as being a vernier adjustment of time. The received signal $y(t)$ can be expressed as a convolution of the transmitted signal $x(t)$ with the channel impulse response (see **Figure 4.3a**).

$$y(t) = \int_{-\infty}^{\infty} x(\tau)h(t,\tau)d\tau = x(t) \otimes h(t,\tau) \tag{4.8}$$

If the multipath channel is assumed to be a bandlimited bandpass channel, which is reasonable, then $h(t,\tau)$ may be equivalently described by a complex baseband impulse response $h_b(t,\tau)$ with the input and output being the complex envelope representations of the transmitted and received signals, respectively (see Figure 4.3b). That is,

$$r(t) = c(t) \otimes \frac{1}{2}h_b(t,\tau) \tag{4.9}$$

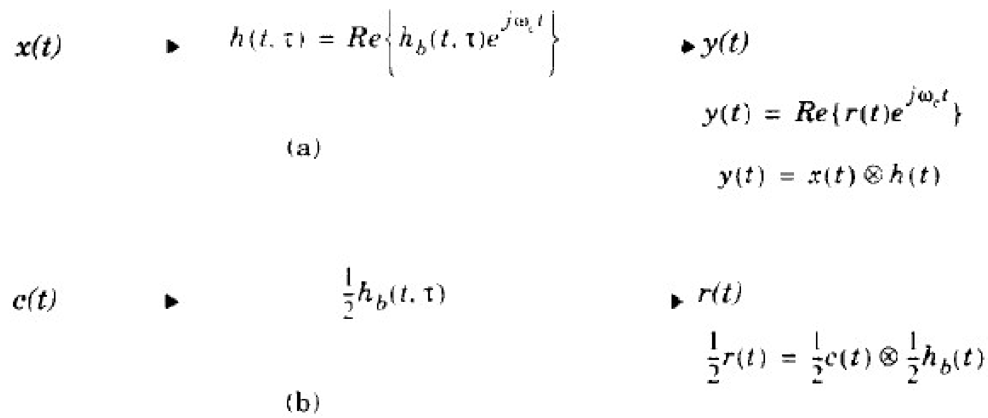


Figure 4.3
(a) Bandpass channel impulse response model.
(b) Baseband equivalent channel impulse response model.

where $c(t)$ and $r(t)$ are the complex envelopes of $x(t)$ and $y(t)$, defined as

$$x(t) = Re\{c(t)\exp(j2\pi f_c t)\} \tag{4.10}$$

$$y(t) = Re\{r(t)\exp(j2\pi f_c t)\} \tag{4.11}$$

The factor of $1/2$ in equation (4.9) is due to the properties of the complex envelope, in order to represent the passband radio system at baseband. The low-pass characterization removes the high frequency variations caused by the carrier, making the signal analytically easier to handle. It is shown by Couch [Cou93] that the average power of a bandpass signal $\overline{x^2(t)}$ is equal to $\frac{1}{2}\overline{|c(t)|^2}$, where the overbar denotes ensemble average for a stochastic signal, or time average for a deterministic or ergodic stochastic signal.

It is useful to discretize the multipath delay axis τ of the impulse response into equal time delay segments called *excess delay bins*, where each bin has a time delay width equal to $\tau_{i+1} - \tau_i$, where τ_0 is equal to 0, and represents the first arriving signal at the receiver. Letting i = 0, it is seen that $\tau_1 - \tau_0$ is equal to the time delay bin width given by $\Delta\tau$. For convention, $\tau_0 = 0$, $\tau_1 = \Delta\tau$, and $\tau_i = i\Delta\tau$, for $i = 0$ to $N-1$, where N represents the total number of possible equally-spaced multipath components, including the first arriving component. Any number of multipath signals received within the i th bin are represented by a single resolvable multipath component having delay τ_i. This technique of quantizing the delay bins determines the time delay resolution of the channel model, and the useful frequency span of the model can be shown to be $1/(2\Delta\tau)$. That is, the model may be used to analyze transmitted signals having bandwidths which are less than $1/(2\Delta\tau)$. Note that $\tau_0 = 0$ is the excess time delay of the first arriving multipath component, and neglects the propagation delay between the transmitter and receiver. *Excess delay* is the relative delay of the i th multipath component as compared to the first arriving component and is given by τ_i. The *maximum excess delay* of the channel is given by $N\Delta\tau$.

Since the received signal in a multipath channel consists of a series of attenuated, time-delayed, phase shifted replicas of the transmitted signal, the baseband impulse response of a multipath channel can be expressed as

$$h_b(t, \tau) = \sum_{i=0}^{N-1} a_i(t, \tau)\exp[j(2\pi f_c\tau_i(t) + \phi_i(t, \tau))]\delta(\tau - \tau_i(t)) \tag{4.12}$$

where $a_i(t, \tau)$ and $\tau_i(t)$ are the real amplitudes and excess delays, respectively, of i th multipath component at time t [Tur72]. The phase term $2\pi f_c\tau_i(t) + \phi_i(t, \tau)$ in (4.12) 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. In general, the phase term is simply represented by a single variable $\theta_i(t, \tau)$ which lumps together all the mechanisms for phase shifts of a single multipath component within the ith excess delay bin. Note that some excess delay bins may have no multipath at some time t and delay τ_i, since $a_i(t, \tau)$ may be zero. In equation (4.12), N is the total possible number of multipath components (bins), and $\delta(\cdot)$ is the unit impulse function which determines the specific multipath bins that have components at time t and excess delays τ_i. Figure 4.4 illustrates an example of different snapshots of $h_b(t, \tau)$, where t varies into the page, and the time delay bins are quantized to widths of $\Delta\tau$.

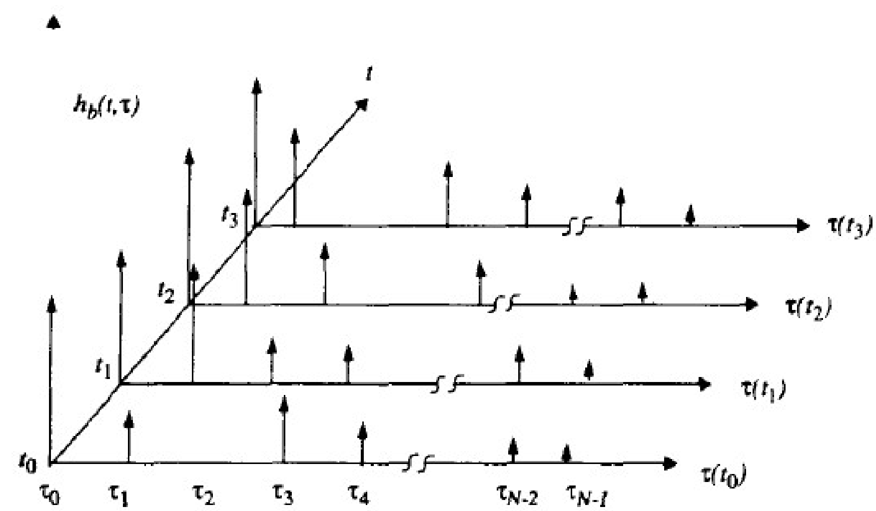


Figure 4.4
An example of the time varying discrete-time impulse response model for a multipath radio channel.

If the channel impulse response is assumed to be time invariant, or is at least wide sense stationary over a small-scale time or distance interval, then the channel impulse response may be simplified as

$$h_b(\tau) = \sum_{i=0}^{N-1} a_i \exp(-j\theta_i)\,\delta(\tau - \tau_i) \tag{4.13}$$

When measuring or predicting $h_b(\tau)$, a probing pulse $p(t)$ which approximates a delta function is used at the transmitter. That is

$$p(t) \approx \delta(t - \tau) \tag{4.14}$$

is used to sound the channel to determine $h_b(\tau)$.

For small-scale channel modeling, the *power delay profile* of the channel is found by taking the spatial average of $|h_b(t;\tau)|^2$ over a local area. By making several local area measurements of $|h_b(t;\tau)|^2$ in different locations, it is possible to build an ensemble of power delay profiles, each one representing a possible small-scale multipath channel state [Rap91a].

Based on work by Cox [Cox72], [Cox75], if $p(t)$ has a time duration much smaller than the impulse response of the multipath channel, $p(t)$ does not need to be deconvolved from the received signal $r(t)$ in order to determine relative multipath signal strengths. The received power delay profile in a local area is given by

$$P(t;\tau) \approx k|h_b(t;\tau)|^2 \tag{4.15}$$

and many snapshots of $|h_b(t;\tau)|^2$ are typically averaged over a local (small-scale) area to provide a single time-invariant multipath power delay profile $P(\tau)$. The gain k in equation (4.15) relates the transmitted power in the probing pulse $p(t)$ to the total power received in a multipath delay profile.

Relationship Between Bandwidth and Received Power

In actual wireless communication systems, the impulse response of a multipath channel is measured in the field using channel sounding techniques. We now consider two extreme channel sounding cases as a means of demonstrating how the small-scale fading behaves quite differently for two signals with different bandwidths in the identical multipath channel.

Consider a pulsed, transmitted RF signal of the form

$$x(t) = \mathrm{Re}\{p(t)\exp(j2\pi f_c t)\}$$

where $p(t)$ is a repetitive baseband pulse train with very narrow pulse width T_{bb} and repetition period T_{REP} which is much greater than the maximum measured excess delay τ_{max} in the channel. Now let

$$p(t) = 2\sqrt{\tau_{max}/T_{bb}} \text{ for } 0 \le t \le T_{bb}$$

and let $p(t)$ be zero elsewhere for all excess delays of interest. The low pass channel output $r(t)$ closely approximates the impulse response $h_b(t)$ and is given by

$$\begin{aligned} r(t) &= \frac{1}{2}\sum_{i=0}^{N-1} a_i(\exp(-j\theta_i)) \cdot p(t-\tau_i) \\ &= \sum_{i=0}^{N-1} a_i \exp(-j\theta_i) \cdot \sqrt{\frac{\tau_{max}}{T_{bb}}} \operatorname{rect}\left[t - \frac{T_{bb}}{2} - \tau_i\right] \end{aligned} \qquad (4.16)$$

To determine the received power at some time t_0, the power $|r(t_0)|^2$ is measured. The quantity $|r(t_0)|^2$ is called the *instantaneous multipath power delay profile* of the channel, and is equal to the energy received over the time duration of the multipath delay divided by τ_{max}. That is, using equation (4.16)

$$|r(t_0)|^2 = \frac{1}{\tau_{max}} \int_0^{\tau_{max}} r(t) \times r^*(t)\, dt \tag{4.17}$$

$$= \frac{1}{\tau_{max}} \int_0^{\tau_{max}} \frac{1}{4} \mathrm{Re} \left\{ \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} a_j(t_0) a_i(t_0) p(t-\tau_j) p(t-\tau_i) \exp(-j(\theta_j - \theta_i)) \right\} dt$$

Note that if all the multipath components are resolved by the probe *p(t)*, then $|\tau_j - \tau_i| > T_{bb}$ for all $j \neq i$, and

$$|r(t_0)|^2 = \frac{1}{\tau_{max}} \int_0^{\tau_{max}} \frac{1}{4} \left(\sum_{k=0}^{N-1} a_k^2(t_0) p^2(t-\tau_k) \right) dt \tag{4.18}$$

$$= \frac{1}{\tau_{max}} \sum_{k=0}^{N-1} a_k^2(t_0) \int_0^{\tau_{max}} \left\{ \sqrt{\frac{\tau_{max}}{T_{bb}}} \mathrm{rect}\left[t - \frac{T_{bb}}{2} - \tau_k \right] \right\}^2 dt$$

$$= \sum_{k=0}^{N-1} a_k^2(t_0)$$

For a wideband probing signal *p(t)*, T_{bb} is smaller than the delays between multipath components in the channel, and equation (4.18) shows that the total received power is simply related to the sum of the powers in the individual multipath components, and is scaled by the ratio of the probing pulse's width and amplitude, and the maximum observed excess delay of the channel. Assuming that the received power from the multipath components forms a random process where each component has a random amplitude and phase at any time t, the average small-scale received power for the wideband probe is found from equation (4.17) as

$$E_{a,\theta}[P_{WB}] = E_{a,\theta}\left[\sum_{i=0}^{N-1}\left|a_i \exp(j\theta_i)\right|^2\right] \approx \sum_{i=0}^{N-1}\overline{a_i^2} \tag{4.19}$$

In equation (4.19), $E_{a,\theta}[\cdot]$ denotes the ensemble average over all possible values of a_i and θ_i in a local area, and the overbar denotes sample average over a local measurement area which is generally measured using multipath measurement equipment. The striking result of equations (4.18) and (4.19) is that if a transmitted signal is able to resolve the multipaths, then the *small-scale received power is simply the sum of the powers received in each multipath component*. In practice, the amplitudes of individual multipath components do not fluctuate widely in a local area. Thus, the received power of a wideband signal such as $p(t)$ does not fluctuate significantly when a receiver is moved about a local area [Rap89].

Now, instead of a pulse, consider a CW signal which is transmitted into the exact same channel, and let the complex envelope be given by $c(t) = 2$. Then, the instantaneous complex envelope of the received signal is given by the phasor sum

$$r(t) = \sum_{i=0}^{N-1} a_i \exp(j\theta_i(t,\tau)) \tag{4.20}$$

and the instantaneous power is given by

$$|r(t)|^2 = \left|\sum_{i=0}^{N-1} a_i \exp(j\theta_i(t,\tau))\right|^2 \tag{4.21}$$

As the receiver is moved over a local area, the channel changes, and the received signal strength will vary at a rate governed by the fluctuations of a_i and θ_i. As mentioned earlier, a_i varies little over local areas, but θ_i will vary greatly due to changes in propagation distance over space, resulting in large fluctuations of $r(t)$ as the receiver is moved over small distances (on the order of a wavelength). That is, since $r(t)$ is the phasor sum of the individual multipath components, the instantaneous phases of the multipath components cause the large fluctuations which typifies small-scale fading for CW signals. The average received power over a local area is then given by

$$E_{a,\theta}[P_{CW}] = E_{a,\theta}\left[\left|\sum_{i=0}^{N-1} a_i \exp(j\theta_i)\right|^2\right] \tag{4.22}$$

$$E_{a,\theta}[P_{CW}] \approx \overline{\left[\left(a_0 e^{j\theta_0} + a_1 e^{j\theta_1} + \dots + a_{N-1} e^{j\theta_{N-1}}\right) \times \left(a_0 e^{-j\theta_0} + a_1 e^{-j\theta_1} + \dots + a_{N-1} e^{-j\theta_{N-1}}\right)\right]} \tag{4.23}$$

$$E_{a,\theta}[P_{CW}] \approx \sum_{i=0}^{N-1} \overline{a_i^2} + 2\sum_{i=0}^{N-1} \sum_{i,j \neq i}^{N} r_{ij} \overline{\cos(\theta_i - \theta_j)} \tag{4.24}$$

where r_{ij} is the path amplitude correlation coefficient defined to be

$$r_{ij} = E_a[a_i a_j] \tag{4.25}$$

and the overbar denotes time average for CW measurements made by a mobile receiver over the local measurement area [Rap89]. Note that when $\overline{\cos(\theta_i - \theta_j)} = 0$ and/or $r_{ij} = 0$, then the average power for a CW signal is equivalent to the average received power for a wideband signal in a small-scale region. This is seen by comparing equation (4.19) and equation (4.24). This can occur when either the multipath phases are identically and independently distributed (i.i.d uniform) over $[0, 2\pi]$ or when the path amplitudes are uncorrelated. The i.i.d uniform distribution of θ is a valid assumption since multipath components traverse differential path lengths that measure hundreds of wavelengths and are likely to arrive with random phases. If for some reason it is believed that the phases are not independent, the average wideband power and average CW power will still be equal if the paths have uncorrelated amplitudes. However, if the phases of the paths are dependent upon each other, then the amplitudes are likely to be correlated, since the same mechanism which affects the path phases is likely to also affect the amplitudes. This situation is highly unlikely at transmission frequencies used in wireless mobile systems.

Thus it is seen that the *received local ensemble average power of wideband and narrowband signals are equivalent*. When the transmitted signal has a bandwidth much greater than the bandwidth of the channel, then the multipath structure is completely resolved by the received signal at any time, and the received power varies very little since the individual multipath amplitudes do not change rapidly over a local area. However, if the transmitted signal has a very narrow bandwidth (e.g., the baseband signal has a duration greater than the excess delay of the channel), then multipath is not resolved by the received signal, and large signal fluctuations (fading) occur at the receiver due to the phase shifts of the many unresolved multipath components.

Figure 4.5 illustrates actual indoor radio channel measurements made simultaneously with a wideband probing pulse having $T_{bb} = 10$ ns, and a CW transmitter. The carrier frequency was 4 GHz. It can be seen that the CW signal undergoes rapid fades, whereas the wideband measurements change little over the 5λ measurement track. However, the local average received powers of both signals were measured to be virtually identical [Haw91].

Small-Scale Multipath Measurements

Because of the importance of the multipath structure in determining the small-scale fading effects, a number of wideband channel sounding techniques have been developed. These techniques may be classified as *direct pulse measurements*, *spread spectrum sliding correlator measurements*, and *swept frequency measurements*.

1. Direct RF Pulse System

A simple channel sounding approach is the direct RF pulse system (see Figure 4.6). This technique allows engineers to determine rapidly the power delay profile of any channel, as demonstrated by Rappaport and Seidel [Rap89], [Rap90]. Essentially a wide band pulsed bistatic radar, this system transmits a repetitive pulse of width τ_{bb} s, and uses a receiver with a wide bandpass filter ($BW = 2/\tau_{bb}$Hz). The signal is then amplified, detected with an envelope detector, and displayed and stored on a high speed oscilloscope. This gives an immediate measurement of the square of the channel impulse response convolved with the probing pulse (see equation (4.17)). If the oscilloscope is set on averaging mode, then this system can provide a local average power delay profile. Another attractive aspect of this system is the lack of complexity, since off-the-shelf equipment may be used.

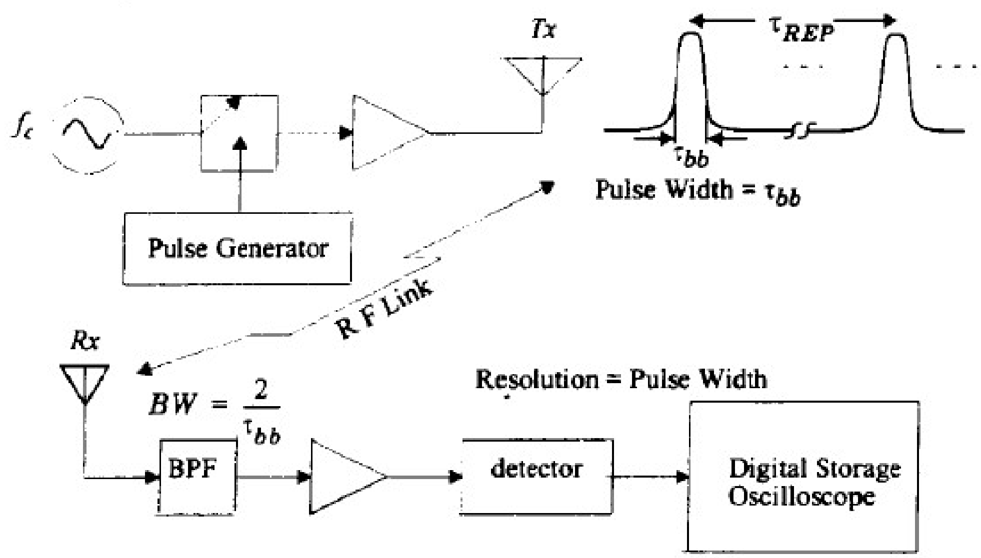


Figure 4.6
Direct RF channel impulse response measurement system.

The minimum resolvable delay between multipath components is equal to the probing pulse width τ_{bb}. The main problem with this system is that it is subject to interference and noise, due to the wide passband filter required for multipath time resolution. Also, the pulse system relies on the ability to trigger the oscilloscope on the first arriving signal. If the first arriving signal is blocked or fades, severe fading occurs, and it is possible the system may not trigger properly. Another disadvantage is that the phases of the individual multipath components are not received, due to the use of an envelope detector. However, use of a coherent detector permits measurement of the multipath phase using this technique.

2. Spread Spectrum Sliding Correlator Channel Sounding

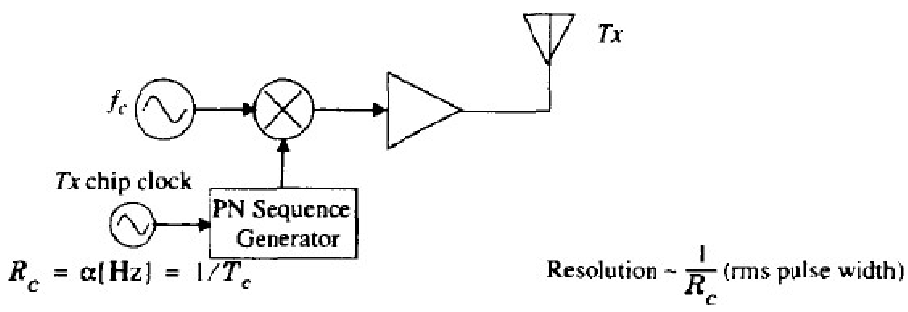
The basic block diagram of a spread spectrum channel sounding system is shown in Figure 4.7. The advantage of a spread spectrum system is that, while the probing signal may be wideband, it is possible to detect the transmitted signal using a narrowband receiver preceded by a wideband mixer, thus improving the dynamic range of the system as compared to the direct RF pulse system.

In a spread spectrum channel sounder, a carrier signal is "spread" over a large bandwidth by mixing it with a binary pseudo-noise (PN) sequence having a chip duration T_c and a chip rate R_c equal to $1/T_c$ Hz. The power spectrum envelope of the transmitted spread spectrum signal is given by [Dix84] as

$$S(f) = \left[\frac{\sin \pi(f - f_c)T_c}{\pi(f - f_c)T_c}\right]^2 \qquad (4.26)$$

and the null-to-null bandwidth is

$$BW = 2R_c \qquad (4.27)$$



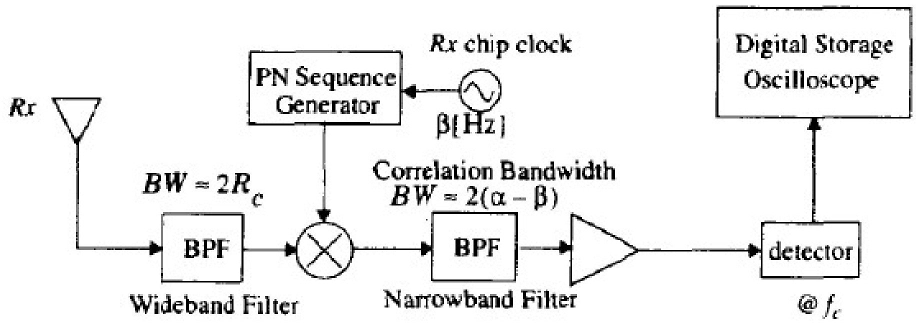


Figure 4.7
Spread spectrum channel impulse response measurement system.

The spread spectrum signal is then received, filtered, and *despread* using a PN sequence generator identical to that used at the transmitter. Although the two PN sequences are identical, the transmitter chip clock is run at a slightly faster rate than the receiver chip clock. Mixing the chip sequences in this fashion implements a *sliding correlator* [Dix84]. When the PN code of the faster chip clock catches up with the PN code of the slower chip clock, the two chip sequences will be virtually identically aligned, giving maximal correlation. When the two sequences are not maximally correlated, mixing the incoming spread spectrum signal with the unsynchronized receiver chip sequence will spread this signal into a bandwidth at least as large as the receiver's reference PN sequence. In this way, the narrowband filter that follows the correlator can reject almost all of the incoming signal power. This is how *processing gain* is realized in a spread spectrum receiver and how it can reject passband interference, unlike the direct RF pulse sounding system.

Processing gain (PG) is given as

$$PG = \frac{2R_c}{R_{bb}} = \frac{2\tau_{bb}}{T_c} = \frac{(S/N)_{out}}{(S/N)_{in}} \tag{4.28}$$

where $\tau_{bb} = 1/R_{bb}$, is the period of the baseband information. For the case of a sliding correlator channel sounder, the baseband information rate is equal to the frequency offset of the PN sequence clocks at the transmitter and receiver.

When the incoming signal is correlated with the receiver sequence, the signal is collapsed back to the original bandwidth (i.e., "despread"), envelope detected, and displayed on an oscilloscope. Since different incoming multipaths will have different time delays, they will maximally correlate with the receiver PN sequence at different times. The energy of these individual paths will pass through the correlator depending on the time delay. Therefore, after envelope detection, the channel impulse response convolved with the pulse shape of a single chip is displayed on the oscilloscope. Cox [Cox72] first used this method to measure channel impulse responses in outdoor suburban environments at 910 MHz. Devasirvatham [Dev86], [Dev90a] successfully used a direct sequence spread spectrum channel sounder to measure time delay spread of multipath components and signal level measurements in office and residential buildings at 850 MHz. Bultitude [Bul89] used this technique for indoor and microcellular channel sounding work, as did Landron [Lan92].

The time resolution $(\Delta\tau)$ of multipath components using a spread spectrum system with sliding correlation is

$$\Delta\tau = 2T_c = \frac{2}{R_c} \tag{4.29}$$

In other words, the system can resolve two multipath components as long as they are equal to or greater than $2T_c$ seconds apart. In actuality, multipath components with interarrival times smaller than $2T_c$ can be resolved since the rms pulse width is smaller than the absolute width of the triangular correlation pulse, and is on the order of T_c.

The sliding correlation process gives *equivalent time* measurements that are updated every time the two sequences are maximally correlated. The time between maximal correlations (T) can be calculated from equation (4.30)

$$\Delta T = T_c \gamma l = \frac{\gamma l}{R_c} \tag{4.30}$$

where T_c = chip period (s)
R_c = chip rate (Hz)
γ = slide factor (dimensionless)
l = sequence length (chips)

The slide factor is defined as the ratio between the transmitter chip clock rate and the difference between the transmitter and receiver chip clock rates [Dev86]. Mathematically, this is expressed as

$$\gamma = \frac{\alpha}{\alpha - \beta} \tag{4.31}$$

where α = transmitter chip clock rate (Hz)
β = receiver chip clock rate (Hz)

For a maximal length PN sequence, the sequence length is

$$l = 2^n - 1 \tag{4.32}$$

where n is the number of shift registers in the sequence generator [Dix84].

Since the incoming spread spectrum signal is mixed with a receiver PN sequence that is slower than the transmitter sequence, the signal is essentially down-converted ("collapsed") to a low-frequency narrowband signal. In other words, the relative rate of the two codes slipping past each other is the rate of information transferred to the oscilloscope. This narrowband signal allows narrowband processing, eliminating much of the passband noise and interference. The processing gain of equation (4.28) is then realized using a narrowband filter ($BW = 2(\alpha - \beta)$).

The equivalent time measurements refer to the relative times of multipath components as they are displayed on the oscilloscope. The observed time scale on the oscilloscope using a sliding correlator is related to the actual propagation time scale by

$$\text{Actual Propagation Time} = \frac{\text{Observed Time}}{\gamma} \tag{4.33}$$

This effect is due to the relative rate of information transfer in the sliding correlator. For example, T_c of equation (4.30) is an observed time measured on an oscilloscope and not actual propagation time. This effect, known as *time dilation*, occurs in the sliding correlator system because the propagation delays are actually expanded in time by the sliding correlator.

Caution must be taken to ensure that the sequence length has a period which is greater than the longest multipath propagation delay. The PN sequence period is

$$\tau_{PNseq} = T_c l \tag{4.34}$$

The sequence period gives an estimate of the maximum unambiguous range of incoming multipath signal components. This range is found by multiplying the speed of light with τ_{PNseq} in equation (4.34).

There are several advantages to the spread spectrum channel sounding system. One of the key spread spectrum modulation characteristics is the ability to reject passband noise, thus improving the coverage range for a given transmitter power. Transmitter and receiver PN sequence synchronization is eliminated by the sliding correlator. Sensitivity is adjustable by changing the sliding factor and the post-correlator filter bandwidth. Also, required transmitter powers can be considerably lower than comparable direct pulse systems due to the inherent "processing gain" of spread spectrum systems.

A disadvantage of the spread spectrum system, as compared to the direct pulse system, is that measurements are not made in real time, but they are compiled as the PN codes slide past one another. Depending on system parameters and measurement objectives, the time required to make power delay profile measurements may be excessive. Another disadvantage of the system described here is that a noncoherent detector is used, so that phases of individual multipath components can not be measured. Even if coherent detection is used, the sweep time of a spread spectrum signal induces delay such that the phases of individual multipath components with different time delays would be measured at substantially different times, during which the channel might change.

3. Frequency Domain Channel Sounding

Because of the dual relationship between time domain and frequency domain techniques, it is possible to measure the channel impulse response in the frequency domain. Figure 4.8 shows a frequency domain channel sounder used for measuring channel impulse responses. A vector network analyzer controls a synthesized frequency sweeper, and an S-parameter test set is used to monitor the frequency response of the channel. The sweeper scans a particular frequency band (centered on the carrier) by stepping through discrete frequencies. The number and spacings of these frequency steps impact 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 level at port 2. These signal levels allow the analyzer to determine the complex response (i.e., transmissivity $S_{21}(\omega)$) of the channel over the measured frequency range. The transmissivity response is a frequency domain representation of the channel impulse response. This response is then converted to the time domain using inverse discrete Fourier transform (IDFT) processing, giving a band-limited version of the impulse response. In theory, this technique works well and indirectly provides amplitude and phase information in the time domain. However, the system requires careful calibration and hardwired synchronization between the transmitter and receiver, making it useful only for very close measurements (e.g., indoor channel sounding). Another limitation with this system is the non-real-time nature of the measurement. For time varying channels, the channel frequency response can change rapidly, giving an erroneous impulse response measurement. To mitigate this effect, fast sweep times are necessary to keep the total swept frequency response measurement interval as short as possible. A faster sweep time can be accomplished by reducing the number of frequency steps, but this sacrifices time resolution and excess delay range in the time domain. The swept frequency system has been used successfully for indoor propagation studies by Pahlavan [Pah95] and Zaghloul, et.al. [Zag91a], [Zag91b].

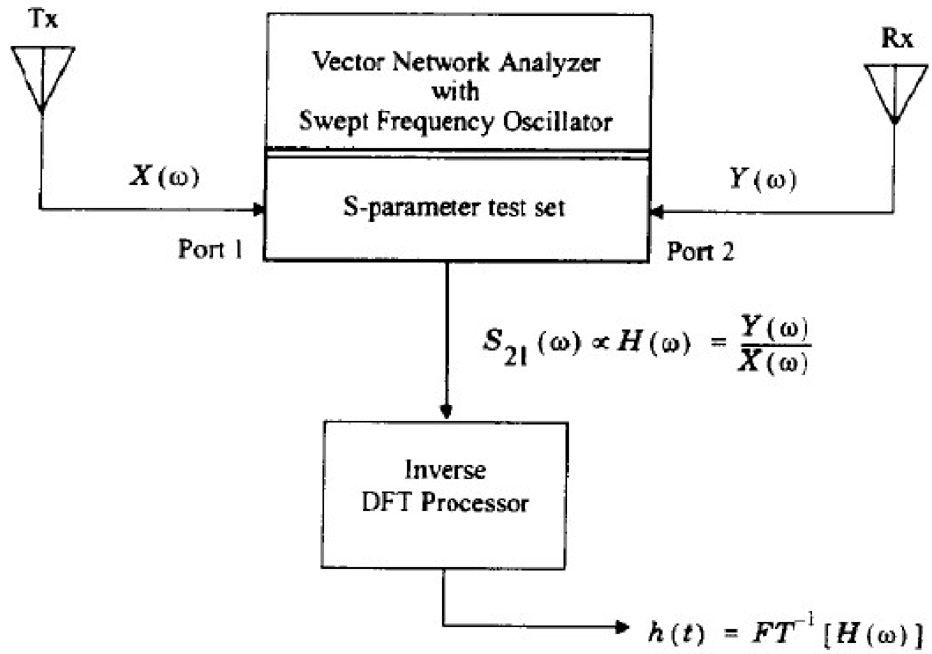
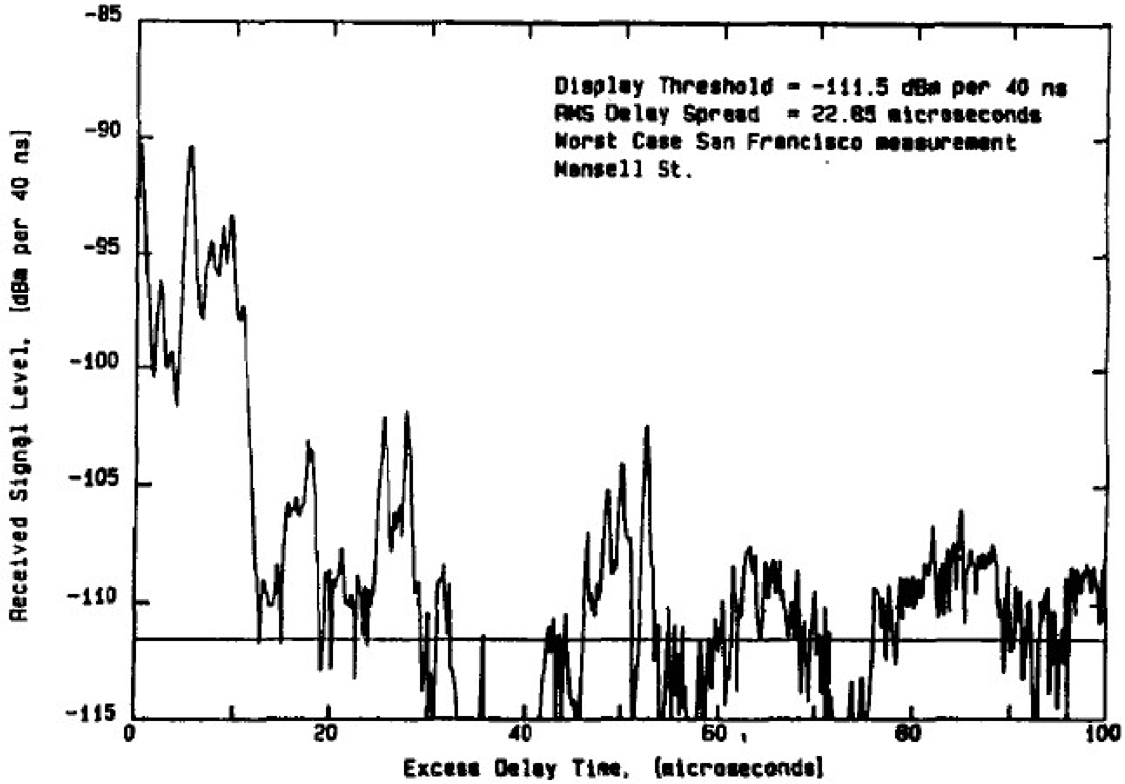


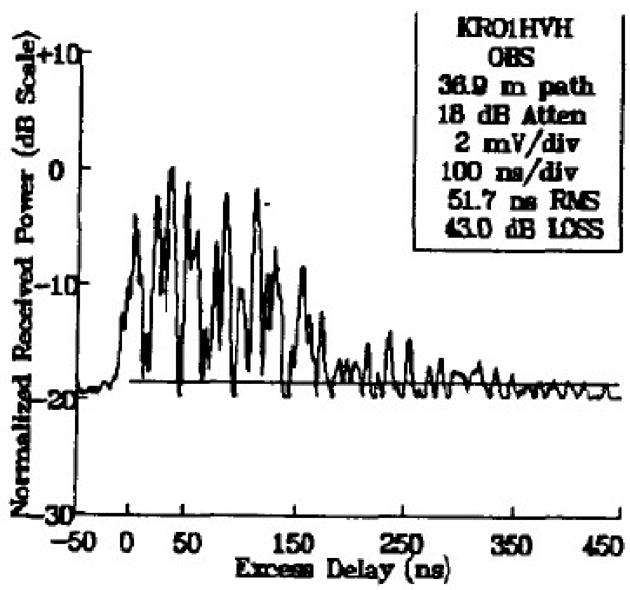
Figure 4.8
Frequency domain channel impulse response measurement system.

Parameters of Mobile Multipath Channels

Many multipath channel parameters are derived from the power delay profile, given by equation (4.18). Power delay profiles are measured using the techniques discussed in Section 4.4 and are generally represented as plots of relative received power as a function of excess delay with respect to a fixed time delay reference. Power delay profiles are found by averaging instantaneous power delay profile measurements over a local area in order to determine an average small-scale power delay profile. Depending on the time resolution of the probing pulse and the type of multipath channels studied, researchers often choose to sample at spatial separations of a quarter of a wavelength and over receiver movements no greater than 6 m in outdoor channels and no greater than 2 m in indoor channels in the 450 MHz - 6 GHz range. This small-scale sampling avoids large-scale averaging bias in the resulting small-scale statistics. Figure 4.9 shows typical power delay profile plots from outdoor and indoor channels, determined from a large number of closely sampled instantaneous profiles.



(a)



(b)

Figure 4.9
Measured multipath power delay profiles
a) From a 900 MHz cellular system in San Francisco [From [Rap90] © IEEE].
b) Inside a grocerv store at 4 GHz [From [Haw91] © IEEE].

Time Dispersion Parameters

In order to compare different multipath channels and to develop some general design guidelines for wireless systems, parameters which grossly quantify the multipath channel are used. The *mean excess delay, rms delay spread,* and *excess delay spread (X* dB*)* are multipath channel parameters that can be determined from a power delay profile. The time dispersive properties of wide band multipath channels are most commonly quantified by their mean excess delay ($\bar{\tau}$) and rms delay spread (σ_τ). The mean excess delay is the first moment of the power delay profile and is defined to be

$$\bar{\tau} = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)\tau_k}{\sum_k P(\tau_k)} \tag{4.35}$$

The rms delay spread is the square root of the second central moment of the power delay profile and is defined to be

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\bar{\tau})^2} \tag{4.36}$$

where

$$\overline{\tau^2} = \frac{\sum_k a_k^2 \tau_k^2}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)\tau_k^2}{\sum_k P(\tau_k)} \tag{4.37}$$

These delays are measured relative to the first detectable signal arriving at the receiver at $\tau_0 = 0$. Equations (4.35) - (4.37) do not rely on the absolute power level of $P(\tau)$, but only the relative amplitudes of the multipath components within $P(\tau)$. Typical values of rms delay spread are on the order of microseconds in outdoor mobile radio channels and on the order of nanoseconds in indoor radio channels. Table 4.1 shows the typical measured values of rms delay spread.

It is important to note that the rms delay spread and mean excess delay are defined from a single power delay profile which is the temporal or spatial average of consecutive impulse response measurements collected and averaged over a local area. Typically, many measurements are made at many local areas in order to determine a statistical range of multipath channel parameters for a mobile communication system over a large-scale area [Rap90].

The *maximum excess delay (X* dB*)* of the power delay profile is defined to be the time delay during which multipath energy falls to X dB below the maxi-

Table 4.1 Typical Measured Values of RMS Delay Spread

Environment	Frequency (MHz)	RMS Delay Spread (σ_τ)	Notes	Reference
Urban	910	1300 ns avg. 600 ns st. dev. 3500 ns max.	New York City	[Cox75]
Urban	892	10-25 μs	Worst case San Francisco	[Rap90]
Suburban	910	200-310 ns	Averaged typical case	[Cox72]
Suburban	910	1960-2110 ns	Averaged extreme case	[Cox72]
Indoor	1500	10-50 ns 25 ns median	Office building	[Sal87]
Indoor	850	270 ns max.	Office building	[Dev90a]
Indoor	1900	70-94 ns avg. 1470 ns max.	Three San Francisco buildings	[Sei92a]

mum. In other words, the maximum excess delay is defined as $\tau_X - \tau_0$, where τ_0 is the first arriving signal and τ_X is the maximum delay at which a multipath component is within X dB of the strongest arriving multipath signal (which does not necessarily arrive at τ_0). Figure 4.10 illustrates the computation of the maximum excess delay for multipath components within 10 dB of the maximum. The maximum excess delay (X dB) defines the temporal extent of the multipath that is above a particular threshold. The value of τ_X is sometimes called the *excess delay spread* of a power delay profile, but in all cases must be specified with a threshold that relates the multipath noise floor to the maximum received multipath component.

In practice, values for $\bar{\tau}$, $\overline{\tau^2}$, and σ_τ depend on the choice of noise threshold used to process $P(\tau)$. The noise threshold is used to differentiate between received multipath components and thermal noise. If the noise threshold is set too low, then noise will be processed as multipath, thus giving rise to values of $\bar{\tau}$, $\overline{\tau^2}$, and σ_τ that are artificially high.

It should be noted that the power delay profile and the magnitude frequency response (the spectral response) of a mobile radio channel are related through the Fourier transform. It is therefore possible to obtain an equivalent description of the channel in the frequency domain using its frequency response characteristics. Analogous to the delay spread parameters in the time domain, *coherence bandwidth* is used to characterize the channel in the frequency domain. The rms delay spread and coherence bandwidth are inversely proportional to one another, although their exact relationship is a function of the exact multipath structure.

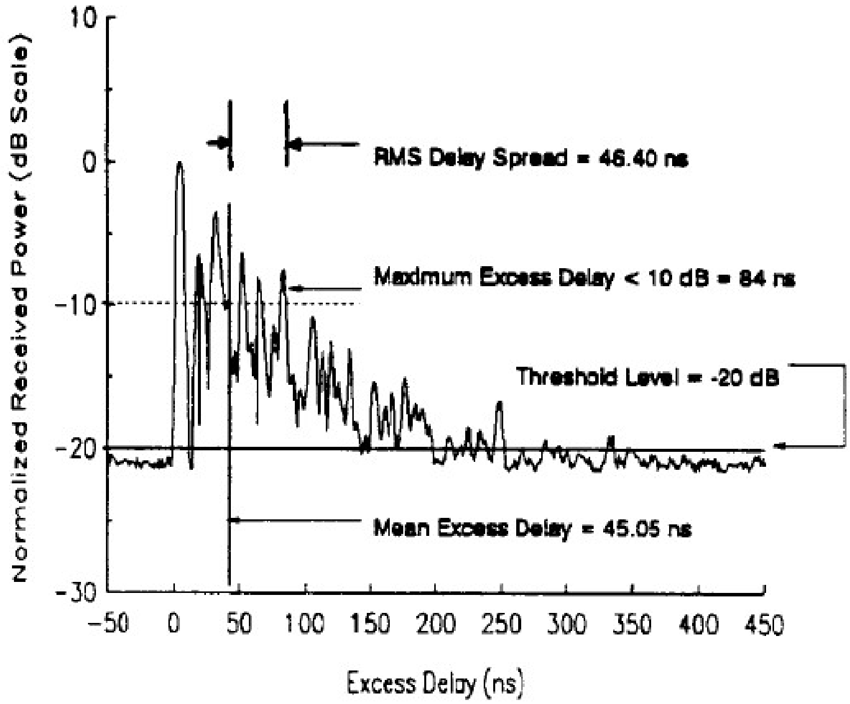


Figure 4.10
Example of an indoor power delay profile; rms delay spread, mean excess delay, maximum excess delay (10 dB), and threshold level are shown.

Coherence Bandwidth

While the delay spread is a natural phenomenon caused by reflected and scattered propagation paths in the radio channel, the coherence bandwidth, B_c, is a defined relation derived from the rms delay spread. Coherence bandwidth is a statistical measure of the range of frequencies over which the channel can be considered "flat" (i.e., a channel which passes all spectral components with approximately equal gain and linear phase). In other words, coherence bandwidth is the range of frequencies over which two frequency components have a strong potential for amplitude correlation. Two sinusoids with frequency separation greater than B_c are affected quite differently by the channel. If the coherence bandwidth is defined as the bandwidth over which the frequency correlation function is above 0.9, then the coherence bandwidth is approximately [Lee89b]

$$B_c \approx \frac{1}{50\sigma_\tau} \tag{4.38}$$

If the definition is relaxed so that the frequency correlation function is above 0.5, then the coherence bandwidth is approximately

$$B_c \approx \frac{1}{5\sigma_\tau} \tag{4.39}$$

It is important to note that an exact relationship between coherence bandwidth and rms delay spread does not exist, and equations (4.38) and (4.39) are "ball park estimates". In general, spectral analysis techniques and simulation are required to determine the exact impact that time varying multipath has on a particular transmitted signal [Chu87], [Fun93], [Ste94]. For this reason, accurate multipath channel models must be used in the design of specific modems for wireless applications [Rap91a], [Woe94].

Doppler Spread and Coherence Time

Delay spread and coherence bandwidth are parameters which describe the time dispersive nature of the channel in a local area. However, they do not offer information about the time varying nature of the channel caused by either relative motion between the mobile and base station, or by movement of objects in the channel. *Doppler spread* and *coherence time* are parameters which describe the time varying nature of the channel in a small-scale region.

Doppler spread B_D is a measure of the spectral broadening caused by the time rate of change of the mobile radio channel and is defined as the range of frequencies over which the received Doppler spectrum is essentially non-zero. When a pure sinusoidal tone of frequency f_c is transmitted, the received signal spectrum, called the Doppler spectrum, will have components in the range $f_c - f_d$ to $f_c + f_d$, where f_d is the Doppler shift. The amount of spectral broadening depends on f_d which is a function of the relative velocity of the mobile, and the angle θ between the direction of motion of the mobile and direction of arrival of the scattered waves. *If the baseband signal bandwidth is much greater than B_D, the effects of Doppler spread are negligible at the receiver.* This is a *slow* fading channel.

Coherence time T_c is the time domain dual of Doppler spread and is used to characterize the time varying nature of the frequency dispersiveness of the channel in the time domain. The Doppler spread and coherence time are inversely proportional to one another. That is,

$$T_C \approx \frac{1}{f_m} \tag{4.40.a}$$

Coherence time is actually a statistical measure of the time duration over which the channel impulse response is essentially invariant, and quantifies the similarity of the channel response at different times. In other words, coherence time is the time duration over which two received signals have a strong potential for amplitude correlation. If the reciprocal bandwidth of the baseband signal is greater than the coherence time of the channel, then the channel will change during the transmission of the baseband message, thus causing distortion at the receiver. If the coherence time is defined as the time over which the time correlation function is above 0.5, then the coherence time is approximately [Ste94]

$$T_C \approx \frac{9}{16\pi f_m} \tag{4.40.b}$$

where f_m is the maximum Doppler shift given by $f_m = \upsilon / \lambda$. In practice, (4.40.a) suggests a time duration during which a Rayleigh fading signal may fluctuate wildly, and (4.40.b) is often too restrictive. A popular rule of thumb for modern digital communications is to define the coherence time as the geometric mean of equations (4.40.a) and (4.40.b). That is,

$$T_C = \sqrt{\frac{9}{16\pi f_m^2}} = \frac{0.423}{f_m} \qquad (4.40.c)$$

The definition of coherence time implies that two signals arriving with a time separation greater than T_C are affected differently by the channel. For example, for a vehicle traveling 60 mph using a 900 MHz carrier, a conservative value of T_C can be shown to be 2.22 ms from (4.40.b). If a digital transmission system is used, then as long as the symbol rate is greater than $1/T_C = 454$ bps, the channel will not cause distortion due to motion (however, distortion could result from multipath time delay spread, depending on the channel impulse response). Using the practical formula of (4.40.c), $T_C = 6.77$ ms and the symbol rate must exceed 150 bits/s in order to avoid distortion due to frequency dispersion.

Example 4.5

Determine the proper spatial sampling interval required to make small-scale propagation measurements which assume that consecutive samples are highly correlated in time. How many samples will be required over 10 m travel distance if $f_c = 1900$ MHz and $v = 50$ m/s. How long would it take to make these measurements, assuming they could be made in real time from a moving vehicle? What is the Doppler spread B_D for the channel?

Solution to Example 4.5

For correlation, ensure that the time between samples is equal to $T_C/2$, and use the smallest value of T_C for conservative design.
Using equation (4.40.b)

$$T_C \approx \frac{9}{16\pi f_m} = \frac{9\lambda}{16\pi v} = \frac{9c}{16\pi v f_c} = \frac{9 \times 3 \times 10^8}{16 \times 3.14 \times 50 \times 1900 \times 10^6}$$

$$T_C = 565\mu s$$

Taking time samples at less than half T_C, at $282.5\mu s$ corresponds to a spatial sampling interval of

$$\Delta x = \frac{vT_C}{2} = \frac{50 \times 565\mu s}{2} = 0.014125 \text{ m} = 1.41 \text{ cm}$$

Therefore, the number of samples required over a 10 m travel distance is

$$N_x = \frac{10}{\Delta x} = \frac{10}{0.014125} = 708 \text{ samples}$$

The time taken to make this measurement is equal to $\frac{10\text{m}}{50 \text{ m/s}} = 0.2$ s
The Doppler spread is

$$B_D = f_m = \frac{vf_c}{c} = \frac{50 \times 1900 \times 10^6}{3 \times 10^8} = 316.66 \text{ Hz}$$

Types of Small-Scale Fading

Small-Scale Fading
(Based on multipath time delay spread)

Flat Fading
1. BW of signal $<$ BW of channel
2. Delay spread $<$ Symbol period

Frequency Selective Fading
1. BW of signal $>$ BW of channel
2. Delay spread $>$ Symbol period

Small-Scale Fading
(Based on Doppler spread)

Fast Fading
1. High Doppler spread
2. Coherence time $<$ Symbol period
3. Channel variations faster than baseband signal variations

Slow Fading
1. Low Doppler spread
2. Coherence time $>$ Symbol period
3. Channel variations slower than baseband signal variations

Figure 4.11
Types of small-scale fading.

Fading Effects Due to Multipath Time Delay Spread

Time dispersion due to multipath causes the transmitted signal to undergo either flat or frequency selective fading.

1. Flat fading

If the mobile radio channel has a constant gain and linear phase response over a bandwidth which is greater than the bandwidth of the transmitted signal, then the received signal will undergo *flat fading*. This type of fading is historically the most common type of fading described in the technical literature. In flat fading, the multipath structure of the channel is such that the spectral characteristics of the transmitted signal are preserved at the receiver. However the strength of the received signal changes with time, due to fluctuations in the gain of the channel caused by multipath. The characteristics of a flat fading channel are illustrated in Figure 4.12.

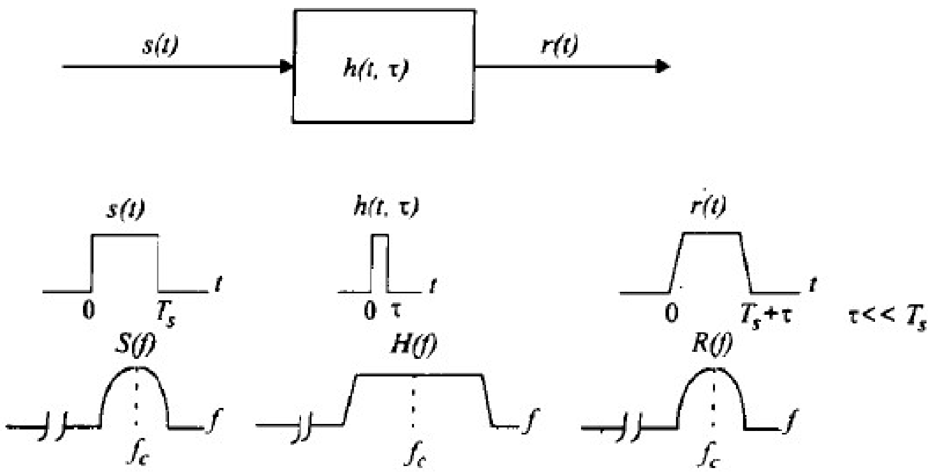


Figure 4.12
Flat fading channel characteristics.

It can be seen from Figure 4.12 that if the channel gain changes over time, a change of amplitude occurs in the received signal. Over time, the received signal $r(t)$ varies in gain, but the spectrum of the transmission is preserved. In a flat fading channel, the reciprocal bandwidth of the transmitted signal is much larger than the multipath time delay spread of the channel, and $h_b(t, \tau)$ can be approximated as having no excess delay (i.e., a single delta function with $\tau = 0$). Flat fading channels are also known as *amplitude varying channels* and are sometimes referred to as *narrowband channels*, since the bandwidth of the applied signal is *narrow* as compared to the channel flat fading bandwidth. Typical flat fading channels cause deep fades, and thus may require 20 or 30 dB more transmitter power to achieve low bit error rates during times of deep fades as compared to systems operating over non-fading channels. The distribution of the instantaneous gain of flat fading channels is important for designing radio links, and the most common amplitude distribution is the Rayleigh distribution. The Rayleigh flat fading channel model assumes that the channel induces an amplitude which varies in time according to the Rayleigh distribution.

To summarize, a signal undergoes flat fading if

$$B_S \ll B_C \tag{4.41}$$

and

$$T_S \gg \sigma_\tau \tag{4.42}$$

where T_S is the reciprocal bandwidth (e.g., symbol period) and B_S is the bandwidth, respectively, of the transmitted modulation, and σ_τ and B_C are the rms delay spread and coherence bandwidth, respectively, of the channel.

2. **Frequency Selective Fading**

If the channel possesses a constant-gain and linear phase response over a bandwidth that is smaller than the bandwidth of transmitted signal, then the channel creates *frequency selective fading* on the received signal. Under such conditions the channel impulse response has a multipath delay spread which is greater than the reciprocal bandwidth of the transmitted message waveform. When this occurs, the received signal includes multiple versions of the transmitted waveform which are attenuated (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. Thus the channel induces *intersymbol interference* (ISI). Viewed in the frequency domain, certain frequency components in the received signal spectrum have greater gains than others.

Frequency selective fading channels are much more difficult to model than flat fading channels since each multipath signal must be modeled and the channel must be considered to be a linear filter. It is for this reason that wideband multipath measurements are made, and models are developed from these measurements. When analyzing mobile communication systems, statistical impulse response models such as the 2-ray Rayleigh fading model (which considers the impulse response to be made up of two delta functions which independently fade and have sufficient time delay between them to induce frequency selective fading upon the applied signal), or computer generated or measured impulse responses, are generally used for analyzing frequency selective small-scale fading. Figure 4.13 illustrates the characteristics of a frequency selective fading channel.

For frequency selective fading, the spectrum $S(f)$ of the transmitted signal has a bandwidth which is greater than the coherence bandwidth B_C of the channel. Viewed in the frequency domain, the channel becomes frequency selective, where the gain is different for different frequency components. Frequency selec-

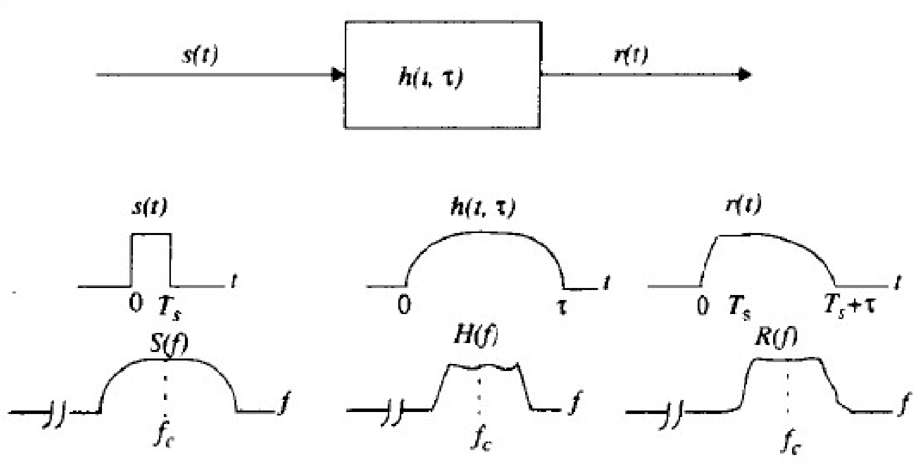


Figure 4.13
Frequency selective fading channel characteristics.

tive fading is caused by multipath delays which approach or exceed the symbol period of the transmitted symbol. Frequency selective fading channels are also known as *wideband channels* since the bandwidth of the signal $s(t)$ is wider than the bandwidth of the channel impulse response. As time varies, the channel varies in gain and phase across the spectrum of $s(t)$, resulting in time varying distortion in the received signal $r(t)$. To summarize, a signal undergoes frequency selective fading if

$$B_S > B_C \tag{4.43}$$

and

$$T_S < \sigma_\tau \tag{4.44}$$

A common rule of thumb is that a channel is frequency selective if $T_s \leq 10\sigma_\tau$, although this is dependent on the specific type of modulation used. Chapter 5 presents simulation results which illustrate the impact of time delay spread on bit error rate (BER).

Fading Effects Due to Doppler Spread

1. Fast Fading

Depending on how rapidly the transmitted baseband signal changes as compared to the rate of change of the channel, a channel may be classified either as a *fast fading* or *slow fading* channel. In a *fast fading channel*, 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, which leads to signal distortion. Viewed in the frequency domain, signal distortion due to fast fading increases with increasing Doppler spread relative to the bandwidth of the transmitted signal. Therefore, a signal undergoes fast fading if

$$T_S > T_C \tag{4.45}$$

and

$$B_S < B_D \tag{4.46}$$

It should be noted that when a channel is specified as a fast or slow fading channel, it does not specify whether the channel is flat fading or frequency selective in nature. Fast fading only deals with the rate of change of the channel due to motion. In the case of the flat fading channel, we can approximate the impulse response to be simply a delta function (no time delay). Hence, a *flat fading, fast fading* channel is a channel in which the amplitude of the delta function varies faster than the rate of change of the transmitted baseband signal. In the case of a *frequency selective, fast fading* channel, the amplitudes, phases, and time delays of any one of the multipath components vary faster than the rate of change of the transmitted signal. In practice, fast fading only occurs for very low data rates.

2. **Slow Fading**

In a *slow fading channel*, the channel impulse response changes at a rate much slower than the transmitted baseband signal *s(t)*. In this case, the channel may be assumed to be static over one or several reciprocal bandwidth intervals. In the frequency domain, this implies that the Doppler spread of the channel is much less than the bandwidth of the baseband signal. Therefore, a signal undergoes slow fading if

$$T_S \ll T_C \tag{4.47}$$

and

$$B_S \gg B_D \tag{4.48}$$

It should be clear that the velocity of the mobile (or velocity of objects in the channel) and the baseband signaling determines whether a signal undergoes fast fading or slow fading.

The relation between the various multipath parameters and the type of fading experienced by the signal are summarized in Figure 4.14. Over the years, some authors have confused the terms fast and slow fading with the terms large-scale and small-scale fading. It should be emphasized that fast and slow fading deal with the relationship between the time rate of change in the channel and the transmitted signal, and not with propagation path loss models.

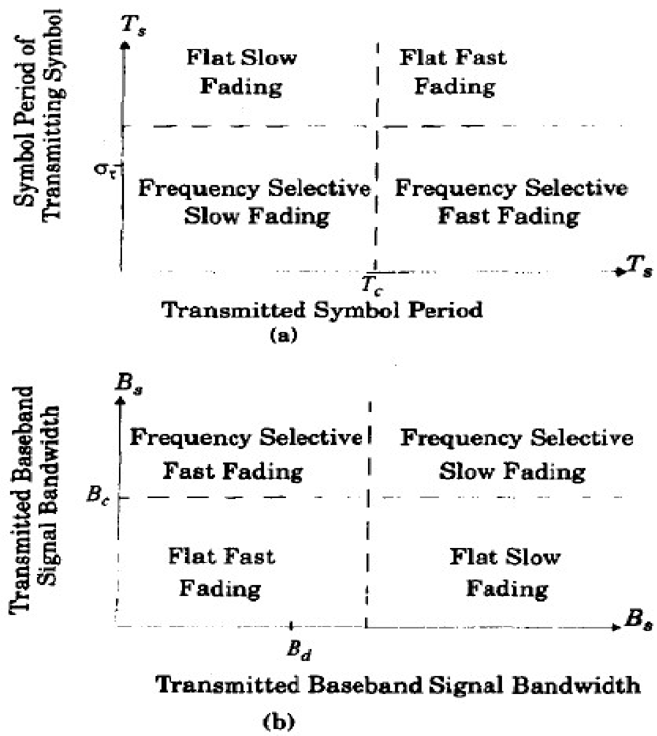


Figure 4.14
Matrix illustrating type of fading experienced by a signal as a function of
(a) symbol period
(b) baseband signal bandwidth.

StatisticalModelsformultipathfadingChannels:

Several multipath models have been suggested to explain the observed statistical nature of a mobile channel. The first model presented by Ossana [Oss64] was based on interference of waves incident and reflected from the flat sides of randomly located buildings. Although Ossana's model [Oss64] predicts flat fading power spectra that were in agreement with measurements in suburban areas, it assumes the existence of a direct path between the transmitter and receiver, and is limited to a restricted range of reflection angles. Ossana's model is therefore rather inflexible and inappropriate for urban areas where the direct path is almost always blocked by buildings or other obstacles. Clarke's model [Cla68] is based on scattering and is widely used.

.1 Clarke's Model for Flat Fading

Clarke [Cla68] developed a model where the statistical characteristics of the electromagnetic fields of the received signal at the mobile are deduced from scattering. The model assumes a fixed transmitter with a vertically polarized antenna. The field incident on the mobile antenna is assumed to be comprised of N azimuthal plane waves with arbitrary carrier phases, arbitrary azimuthal angles of arrival, and each wave having equal average amplitude. It should be noted that the equal average amplitude assumption is based on the fact that in the absence of a direct line-of-sight path, the scattered components arriving at a receiver will experience similar attenuation over small-scale distances.

Figure 4.19 shows a diagram of plane waves incident on a mobile traveling at a velocity v, in the x-direction. The angle of arrival is measured in the x-y plane with respect to the direction of motion. Every wave that is incident on the mobile undergoes a Doppler shift due to the motion of the receiver and arrives at the receiver at the same time. That is, no excess delay due to multipath is assumed for any of the waves (flat fading assumption). For the n th wave arriving at an angle α_n to the x-axis, the Doppler shift in Hertz is given by

$$f_n = \frac{v}{\lambda}\cos\alpha_n \tag{4.57}$$

where λ is the wavelength of the incident wave.

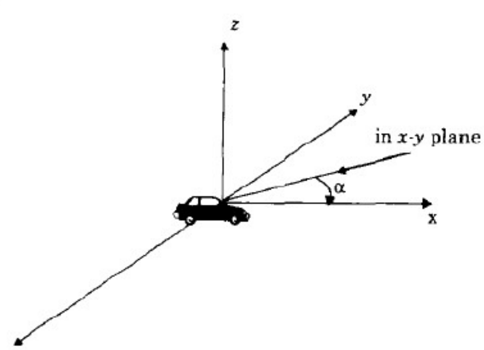


Figure 4.19
Illustrating plane waves arriving at random angles.

The vertically polarized plane waves arriving at the mobile have E and H field components given by

$$E_z = E_o \sum_{n=1}^{N} C_n \cos(2\pi f_c t + \theta_n) \tag{4.58}$$

$$H_x = -\frac{E_o}{\eta} \sum_{n=1}^{N} C_n \sin\alpha_n \cos(2\pi f_c t + \theta_n) \tag{4.59}$$

$$H_y = -\frac{E_o}{\eta} \sum_{n=1}^{N} C_n \cos\alpha_n \cos(2\pi f_c t + \theta_n) \tag{4.60}$$

where E_0 is the real amplitude of local average E-field (assumed constant), C_n is a real random variable representing the amplitude of individual waves, η is the intrinsic impedance of free space (377Ω), and f_c is the carrier frequency. The random phase of the n th arriving component θ_n is given by

$$\theta_n = 2\pi f_n t + \phi_n \tag{4.61}$$

The amplitudes of the E-and H-field are normalized such that the ensemble average of the C_n's is given by

$$\sum_{n=1}^{N} \overline{C_n^2} = 1 \tag{4.62}$$

Since the Doppler shift is very small when compared to the carrier frequency, the three field components may be modeled as narrow band random processes. The three components E_z, H_x, and H_y can be approximated as Gaussian random variables if N is sufficiently large. The phase angles are assumed to have a uniform probability density function (pdf) on the interval $(0,2\pi]$. Based on the analysis by Rice [Ric48] the E-field can be expressed in an in-phase and quadrature form

$$E_z = T_c(t)\cos(2\pi f_c t) - T_s(t)\sin(2\pi f_c t) \tag{4.63}$$

where

$$T_c(t) = E_0 \sum_{n=1}^{N} C_n \cos(2\pi f_n t + \phi_n) \tag{4.64}$$

and

$$T_s(t) = E_0 \sum_{n=1}^{N} C_n \sin(2\pi f_n t + \phi_n) \tag{4.65}$$

Both $T_c(t)$ and $T_s(t)$ are Gaussian random processes which are denoted as T_c and T_s, respectively, at any time t. T_c and T_s are uncorrelated zero-mean Gaussian random variables with an equal variance given by

$$\overline{T_c^2} = \overline{T_s^2} = \overline{|E_z|^2} = E_0^2/2 \tag{4.66}$$

where the overbar denotes the ensemble average.

The envelope of the received E-field, $E_z(t)$, is given by

$$|E_z(t)| = \sqrt{T_c^2(t) + T_s^2(t)} = r(t) \tag{4.67}$$

Since T_c and T_s are Gaussian random variables, it can be shown through a Jacobean transformation [Pap91] that the random received signal envelope r has a Rayleigh distribution given by

$$p(r) = \begin{cases} \dfrac{r}{\sigma^2}\exp\left(-\dfrac{r^2}{2\sigma^2}\right) & 0 \le r \le \infty \\ 0 & r < 0 \end{cases} \tag{4.68}$$

where $\sigma^2 = E_0^2/2$

1.1 Spectral Shape Due to Doppler Spread in Clarke's Model

Gans [Gan72] developed a spectrum analysis for Clarke's model. Let $p(\alpha)d\alpha$ denote the fraction of the total incoming power within $d\alpha$ of the angle α, and let A denote the average received power with respect to an isotropic antenna. As $N \to \infty$, $p(\alpha)d\alpha$ approaches a continuous, rather than a discrete, distribution. If $G(\alpha)$ is the azimuthal gain pattern of the mobile antenna as a function of the angle of arrival, the total received power can be expressed as

$$P_r = \int_0^{2\pi} AG(\alpha)p(\alpha)\,d\alpha \tag{4.69}$$

where $AG(\alpha)p(\alpha)d\alpha$ is the differential variation of received power with angle. If the scattered signal is a CW signal of frequency f_c, then the instantaneous frequency of the received signal component arriving at an angle α is obtained using equation (4.57)

$$f(\alpha) = f = \frac{v}{\lambda}\cos(\alpha) + f_c = f_m\cos\alpha + f_c \tag{4.70}$$

where f_m is the maximum Doppler shift. It should be noted that $f(\alpha)$ is an even function of α, (i.e., $f(\alpha) = f(-\alpha)$).

If $S(f)$ is the power spectrum of the received signal, the differential variation of received power with frequency is given by

$$S(f)\,|df| \tag{4.71}$$

Equating the differential variation of received power with frequency to the differential variation in received power with angle, we have

$$S(f)\,|df| = A\,[p(\alpha)\,G(\alpha) + p(-\alpha)\,G(-\alpha)]\,|d\alpha| \tag{4.72}$$

Differentiating equation (4.70), and rearranging the terms, we have

$$|df| = |d\alpha|\,|-\sin\alpha|\,f_m \tag{4.73}$$

Using equation (4.70), α can be expressed as a function of f as

$$\alpha = \cos^{-1}\left[\frac{f-f_c}{f_m}\right] \tag{4.74}$$

This implies that

$$\sin\alpha = \sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2} \tag{4.75}$$

Substituting equation (4.73) and (4.75) into both sides of (4.72), the power spectral density $S(f)$ can be expressed as

$$S(f) = \frac{A\,[p(\alpha)\,G(\alpha) + p(-\alpha)\,G(-\alpha)]}{f_m\sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2}} \tag{4.76}$$

where

$$S(f) = 0, \qquad |f-f_c| > f_m \tag{4.77}$$

The spectrum is centered on the carrier frequency and is zero outside the limits of $f_c \pm f_m$. Each of the arriving waves has its own carrier frequency (due to its direction of arrival) which is slightly offset from the center frequency. For the case of a vertical $\lambda/4$ antenna ($G(\alpha) = 1.5$), and a uniform distribution $p(\alpha) = 1/2\pi$ over 0 to 2π, the output spectrum is given by (4.76) as

$$S_{E_z}(f) = \frac{1.5}{\pi f_m\sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2}} \tag{4.78}$$

In equation (4.78) the power spectral density at $f = f_c \pm f_m$ is infinite, i.e., Doppler components arriving at exactly $0°$ and $180°$ have an infinite power spectral density. This is not a problem since α is continuously distributed and the probability of components arriving at exactly these angles is zero.

Figure 4.20 shows the power spectral density of the resulting RF signal due to Doppler fading. Smith [Smi75] demonstrated an easy way to simulate Clarke's model using a computer simulation as described Section 4.7.2.

After envelope detection of the Doppler-shifted signal, the resulting baseband spectrum has a maximum frequency of $2f_m$. It can be shown [Jak74] that the electric field produces a baseband power spectral density given by

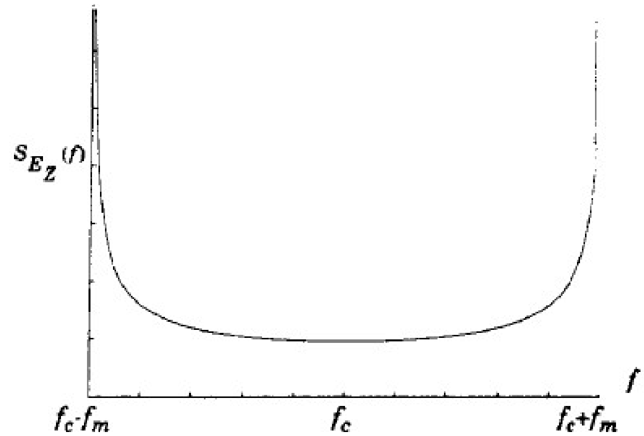


Figure 4.20
Doppler power spectrum for an unmodulated CW carrier [From [Gan72]] © IEEE].

$$S_{bbE_z}(f) = \frac{1}{8\pi f_m} K\left[\sqrt{1-\left(\frac{f}{2f_m}\right)^2}\right] \tag{4.79}$$

where $K[\bullet]$ is the complete elliptical integral of the first kind. Equation (4.79) is not intuitive and is a result of the temporal correlation of the received signal when passed through a nonlinear envelope detector. Figure 4.21 illustrates the baseband spectrum of the received signal after envelope detection.

The spectral shape of the Doppler spread determines the time domain fading waveform and dictates the temporal correlation and fade slope behaviors. Rayleigh fading simulators must use a fading spectrum such as equation (4.78) in order to produce realistic fading waveforms that have proper time correlation.

1.2 Simulation of Clarke and Gans Fading Model

It is often useful to simulate multipath fading channels in hardware or software. A popular simulation method uses the concept of in-phase and quadrature modulation paths to produce a simulated signal with spectral and temporal characteristics very close to measured data.

As shown in Figure 4.22, two independent Gaussian low pass noise sources are used to produce in-phase and quadrature fading branches. Each Gaussian source may be formed by summing two independent Gaussian random variables which are orthogonal (i.e., $g = a + jb$, where a and b are real Gaussian random variables and g is complex Gaussian). By using the spectral filter defined by equation (4.78) to shape the random signals in the frequency domain, accurate time domain waveforms of Doppler fading can be produced by using an inverse fast Fourier transform (IFFT) at the last stage of the simulator.

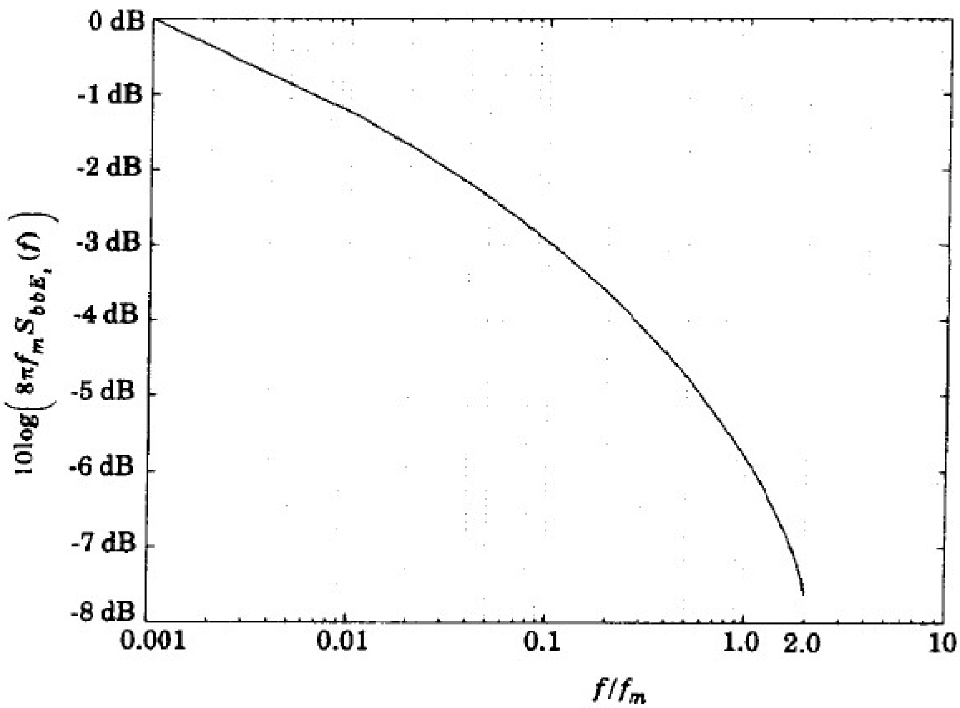


Figure 4.21
Baseband power spectral density of a CW Doppler signal after envelope detection.

Smith [Smi75] demonstrated a simple computer program that implements Figure 4.22(b). His method uses a complex Gaussian random number generator (noise source) to produce a baseband line spectrum with complex weights in the positive frequency band. The maximum frequency component of the line spectrum is f_m. Using the property of real signals, the negative frequency components are constructed by simply conjugating the complex Gaussian values obtained for the positive frequencies. Note that the IFFT of this signal is a purely real Gaussian random process in the time domain which is used in one of the quadrature arms shown in Figure 4.22. The random valued line spectrum is then multiplied with a discrete frequency representation of $\sqrt{S_{E_z}(f)}$ having the same number of points as the noise source. To handle the case where equation (4.78) approaches infinity at the passband edge, Smith truncated the value of $S_{E_z}(f_m)$ by computing the slope of the function at the sample frequency just prior to the passband edge and extended the slope to the passband edge. Simulations using the architecture in Figure 4.22 are usually implemented in the frequency domain using complex Gaussian line spectra to take advantage of easy implementation of equation (4.78). This, in turn, implies that the low pass Gaus-

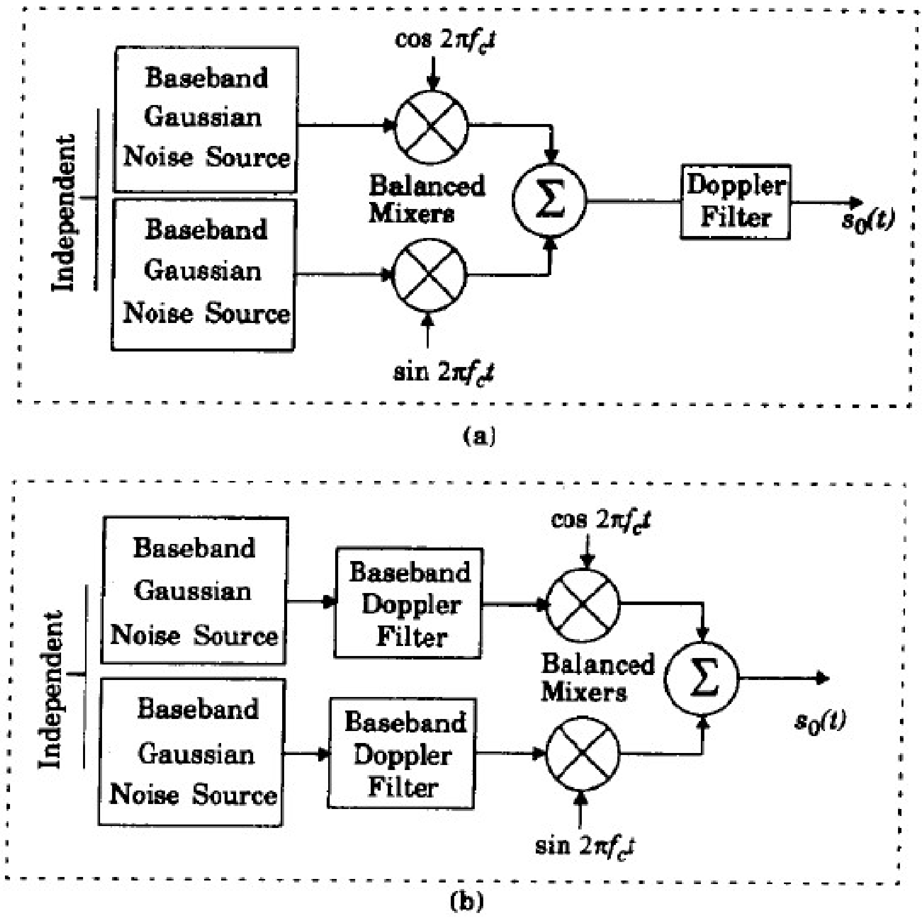


Figure 4.22
Simulator using quadrature amplitude modulation with (a) RF Doppler filter and (b) baseband Doppler filter.

sian noise components are actually a series of frequency components (line spectrum from $-f_m$ to f_m), which are equally spaced and each have a complex Gaussian weight. Smith's simulation methodology is shown in Figure 4.23.

To implement the simulator shown in Figure 4.23, the following steps are used:

(1) Specify the number of frequency domain points (N) used to represent $\sqrt{S_{E_z}(f)}$ and the maximum Doppler frequency shift (f_m). The value used for N is usually a power of 2.

(2) Compute the frequency spacing between adjacent spectral lines as $\Delta f = 2f_m/(N-1)$. This defines the time duration of a fading waveform, $T = 1/\Delta f$.

(3) Generate complex Gaussian random variables for each of the $N/2$ positive frequency components of the noise source.

(4) Construct the negative frequency components of the noise source by conju-

gating positive frequency values and assigning these at negative frequency values.

(5) Multiply the in-phase and quadrature noise sources by the fading spectrum $\sqrt{S_{E_z}(f)}$.

(6) Perform an IFFT on the resulting frequency domain signals from the in-phase and quadrature arms to get two N-length time series, and add the squares of each signal point in time to create an N-point time series like under the radical of equation (4.67).

(7) Take the square root of the sum obtained in step 6 to obtain an N point time series of a simulated Rayleigh fading signal with the proper Doppler spread and time correlation.

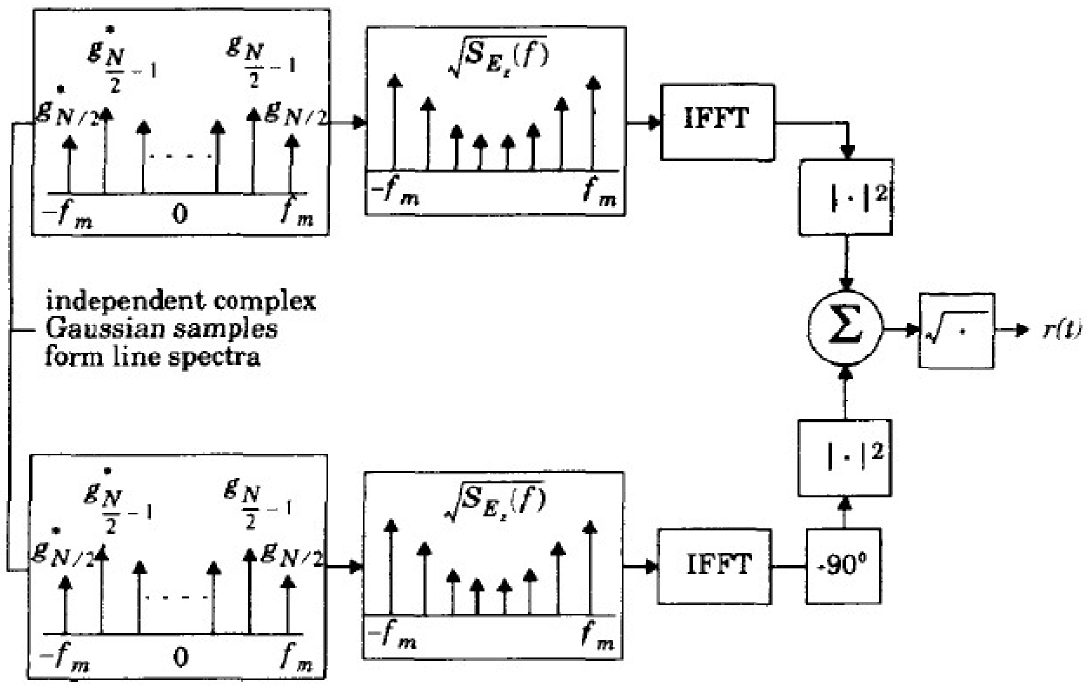


Figure 4.23
Frequency domain implementation of a Rayleigh fading simulator at baseband

Several Rayleigh fading simulators may be used in conjunction with variable gains and time delays to produce frequency selective fading effects. This is shown in Figure 4.24.

By making a single frequency component dominant in amplitude within $\sqrt{S_{E_z}(f)}$, the fading is changed from Rayleigh to Ricean. This can be used to alter the probability distributions of the individual multipath components in the simulator of Figure 4.24.

To determine the impact of flat fading on an applied signal $s(t)$, one merely needs to multiply the applied signal by $r(t)$, the output of the fading simulator. To determine the impact of more than one multipath component, a convolution must be performed as shown in Figure 4.24.

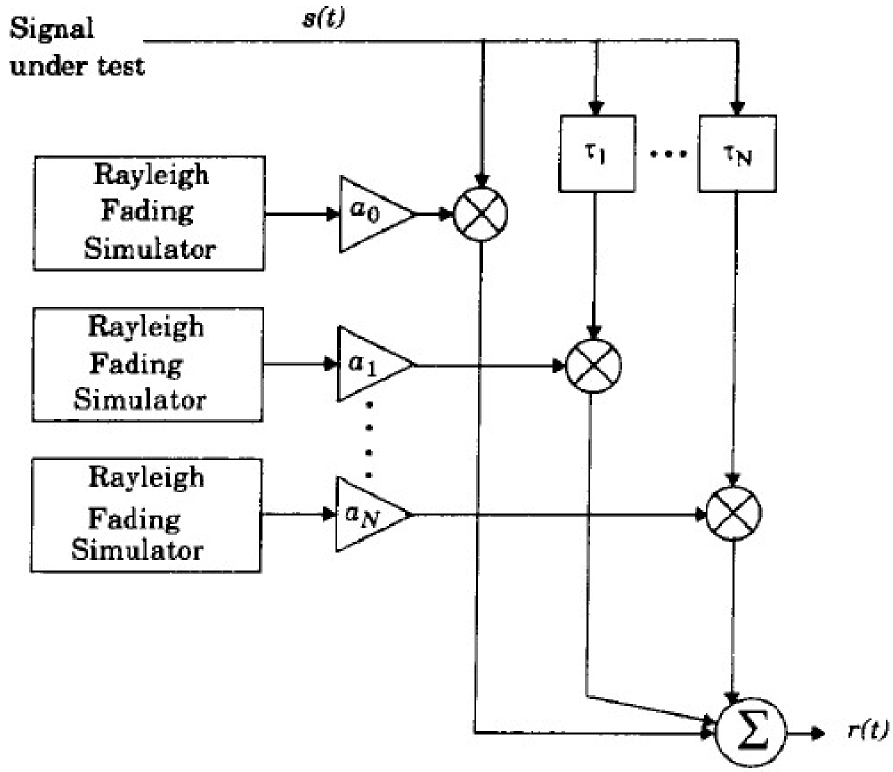


Figure 4.24
A signal may be applied to a Rayleigh fading simulator to determine performance in a wide range of channel conditions. Both flat and frequency selective fading conditions may be simulated, depending on gain and time delay settings.

1.3 Level Crossing and Fading Statistics

Rice computed joint statistics for a mathematical problem which is similar to Clarke's fading model [Cla68], and thereby provided simple expressions for computing the average number of level crossing and the duration of fades. The *level crossing rate* (LCR) and *average fade duration* of a Rayleigh fading signal are two important statistics which are useful for designing error control codes and diversity schemes to be used in mobile communication systems, since it becomes possible to relate the time rate of change of the received signal to the signal level and velocity of the mobile.

The *level crossing rate* (LCR) is defined as the expected rate at which the Rayleigh fading envelope, normalized to the local rms signal level, crosses a specified level in a positive-going direction. The number of level crossings per second is given by

$$N_R = \int_0^\infty \dot{r}\, p(R, \dot{r})\, d\dot{r} = \sqrt{2\pi} f_m \rho e^{-\rho^2} \tag{4.80}$$

where $\dot{r}$ is the time derivative of $r(t)$ (i.e., the slope), $p(R, \dot{r})$ is the joint density function of r and $\dot{r}$ at $r = R$, f_m is the maximum Doppler frequency and $\rho = R/R_{rms}$ is the value of the specified level R, normalized to the local rms amplitude of the fading envelope [Jak74]. Equation (4.80) gives the value of N_R, the average number of level crossings per second at specified R. The level crossing rate is a function of the mobile speed as is apparent from the presence of f_m in equation (4.80). There are few crossings at both high and low levels, with the maximum rate occurring at $\rho = 1/\sqrt{2}$, (i.e., at a level 3 dB below the rms level). The signal envelope experiences very deep fades only occasionally, but shallow fades are frequent.

Example 4.6

For a Rayleigh fading signal, compute the positive-going level crossing rate for $\rho = 1$, when the maximum Doppler frequency (f_m) is 20 Hz. What is the maximum velocity of the mobile for this Doppler frequency if the carrier frequency is 900 MHz?

Solution to Example 4.6

Using equation (4.80), the number of zero level crossings is

$$N_R = \sqrt{2\pi}(20)(1)e^{-1} = 18.44 \text{ crossings per second}$$

The maximum velocity of the mobile can be obtained using the Doppler relation, $f_{d,max} = v/\lambda$.

Therefore velocity of the mobile at $f_m = 20$ Hz is

$$v = f_d\lambda = 20\text{ Hz}(1/3\text{ m}) = 6.66\text{ m/s} = 24\text{ km/hr}$$

The *average fade duration* is defined as the average period of time for which the received signal is below a specified level R. For a Rayleigh fading signal, this is given by

$$\bar{\tau} = \frac{1}{N_R}Pr[r \le R] \tag{4.81}$$

where $Pr[r \le R]$ is the probability that the received signal r is less than R and is given by

$$Pr[r \le R] = \frac{1}{T}\sum_i \tau_i \tag{4.82}$$

where τ_i is the duration of the fade and T is the observation interval of the fading signal. The probability that the received signal r is less than the threshold R is found from the Rayleigh distribution as

$$P_r[r \le R] = \int_0^R p(r)\,dr = 1 - \exp(-\rho^2) \tag{4.83}$$

where $p(r)$ is the pdf of a Rayleigh distribution. Thus, using equations (4.80), (4.81), and (4.83), the average fade duration as a function of ρ and f_m can be expressed as

$$\bar{\tau} = \frac{e^{\rho^2} - 1}{\rho f_m \sqrt{2\pi}} \tag{4.84}$$

1.4 Two-ray Rayleigh Fading Model

Clarke's model and the statistics for Rayleigh fading are for flat fading conditions and do not consider multipath time delay. In modern mobile communication systems with high data rates, it has become necessary to model the effects of multipath delay spread as well as fading. A commonly used multipath model is an independent Rayleigh fading 2-ray model (which is a specific implementation of the generic fading simulator shown in Figure 4.24). Figure 4.25 shows a block diagram of the 2-ray independent Rayleigh fading channel model. The impulse response of the model is represented as

$$h_b(t) = \alpha_1 \exp(j\phi_1)\,\delta(t) + \alpha_2 \exp(j\phi_2)\,\delta(t-\tau) \tag{4.85}$$

where α_1 and α_2 are independent and Rayleigh distributed, ϕ_1 and ϕ_2 are independent and uniformly distributed over $[0, 2\pi]$, and τ is the time delay between the two rays. By setting α_2 equal to zero, the special case of a flat Rayleigh fading channel is obtained as

$$h_b(t) = \alpha_1 \exp(j\phi_1)\,\delta(t) \tag{4.86}$$

By varying τ, it is possible to create a wide range of frequency selective fading effects. The proper time correlation properties of the Rayleigh random variables α_1 and α_2 are guaranteed by generating two independent waveforms, each produced from the inverse Fourier transform of the spectrum described in Section 4.7.2.

CHAPTER6

MultipleAccessTechniques

6.1 Introduction

In Chapter 5 we discussed that cellular systems divide a geographic region into cellswheremobileunitsineachcellcommunicatewiththecell'sbasestation.The goal in the design of a cellular system is to be able to handle as many calls as pos- sible in a given bandwidth with the specified blocking probability (reliability).

Multiplexingdealswiththedivisionoftheresourcestocreatemultiplechan- nels. Multiplexing can create channels in frequency, time, etc., and the corre- sponding terms are then frequency division multiplexing (FDM), time division multiplexing (TDM), etc. [1,3]. Since the amount of spectrum available is limited, we need to find ways to allow multiple users to share the available spectrum simultaneously.Sharedaccessisusedtoimplementamultipleaccessschemewhen access by many users to a channel is required [13,14,15]. For example, one can create multiple channels using TDM, but each of these channels can be accessed by a group of users using the ALOHA multiple access scheme [8,9]. The multiple access schemes can be either reservation-based or random.

Multiple access schemes allow many users to share the radio spectrum. Sharing the bandwidth efficiently among users is one of the main objectives of multiple access schemes [16,17]. The variability of wireless channels presents both challenges and opportunities in designing multiple access communications systems. Multiple access strategy has an impact on robustness and interference levelsgeneratedinothercells.Therefore,multipleaccessschemesaredesignedto maintain orthogonality and reduce interference effects [10].

Multiple access schemes can be classified as *reservation-based* multiple access(e.g.,FDMA,TDMA,CDMA)[4,5]and*random*multipleaccess(e.g., ALOHA, CSMA) (see Figure 6.1) [9,23]. If data traffic is continuous and a small transmissiondelayisrequired(forexampleinvoicecommunication)reservation-based multiple access is used. The family of reservation-based multiple access includesfrequencydivisionmultipleaccess(FDMA),timedivisionmultipleaccess (TDMA), and code division multiple access (CDMA) [6,7,12,21,22]. In many wirelesssystemsforvoicecommunication,thecontrolchannelisbasedonrandom multiple access and the communication channel is based on FDMA, TDMA, or CDMA. The reservation-based multiple access technique has a disadvantage in thatoncethechannelisassigned,itremainsidleiftheuserhasnothingtotransmit,

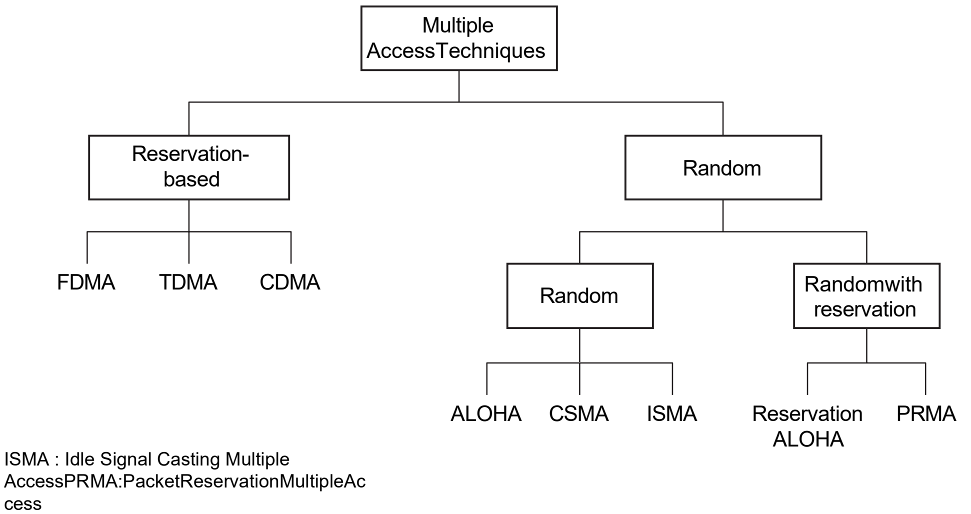


Figure6.1Multipleaccessschemes.

whileotherusersmayhavedatawaitingtobetransmitted.Thisproblemiscritical when data generation is random and has a high peak-rate to average-rate ratio.In this situation, random multiple access is more efficient, because a communica-tionchannelissharedbymanyusersanduserstransmittheirdatainarandomor partially coordinated fashion. ALOHA and carrier sense multiple access (CSMA) are examples of random multiple access [8]. If the data arrives in a random man-ner,andthedatalengthislarge,thenrandommultipleaccesscombinedwith a reservation protocol will perform better than both random- and reservation-based schemes.

We first focus on the reservation-based multiple access schemes including narrowband channelized and wideband nonchannelized systems for wireless communications.Wediscussaccesstechnologies—FDMA,TDMA,andCDMA. We examine FDMA and TDMA from a capacity, performance, and spectral effi- ciency viewpoint. As networks have evolved, the demand for higher capacitieshasencouragedresearchersandsystemdesignerstoexamineaccessschem esthat are even more spectrally efficient than TDMA. Therefore, we also examine the CDMA system. Work in standards bodies around the world indicates that the 3G/4G wireless systems are evolving to wideband CDMA (WCDMA) to achieve highefficienciesandhighaccessdatarates.Thelaterpartofthechapterisdevoted to the discussion of random multiple access schemes.

6.2 NarrowbandChannelizedSystems

Traditional architectures for analog and digital wireless systems are channelized [6,11].In achannelized system, thetotal spectrumis divided intoa large numberof relativelynarrowradiochannelsthataredefinedbycarrierfrequency.Eachradio channel consists of a pair of frequencies. The frequency used for transmission from the base station to the mobile station is called the *forward channel* (downlink channel) and the frequency used for transmission from the mobile station to the base station is called the *reverse channel* (uplink channel). A user is assigned both frequencies for the duration of the call. The forward and reverse channels areassignedwidelyseparatedfrequenciestokeeptheinterferencebetweentrans- mission and reception to a minimum.

A narrowband channelized system demands precise control of output frequencies for an individual transmitter. In this case, the transmission by a given mobilestationoccurswithinthespecifiednarrowbandwidthtoavoidinterference with adjacent channels. The tightness of bandwidth limitations plays a dominant role in the evaluation and selection of modulation technique. It also influencesthe design of transmitter and receiver elements, particularly the filters which can greatly affect the cost of a mobile station.

A critical issue with regulators and operators around the world is how efficiently the radio spectrum is being used. Regulatory bodies want to encourage competition for cellular services. Thus for a given availability of bandwidth, more operators can be licensed. For a particular operator, a more efficient technology can support more users within the assigned spectrum and thus increase profits.

Whenweexamineefficienciesofvarioustechnologies,wefindthateachsys- temhasmadedifferenttrade-offsindeterminingtheoptimummethodforaccess. Some of the parameters that are used in the trade-off are bandwidth per user, guard bands between channels, frequency reuse among different cells in the sys- tem, the signal-to-noise and signal-to-interference ratio, the methods of channel and speech coding, and the complexity of the system.

The first-generation analog cellular systems showed signs of capacity saturation in major urban areas, even with a modest total user population. A major capacity increase was needed to meet future demand. Several digital techniques were deployed to solve the capacity problem of analog cellular systems. Thereare two basic types of systems whereby a fixed spectrum resource is partitioned and shared among different users [13,16]. The channels are created by dividing the total system bandwidth into frequency channels through the use of FDM and then further dividing each frequency channel into time channels through the use ofTDM.MostsystemsuseacombinationofFDMAandTDMA.

6.2.1 Frequency Division Duplex (FDD) and Time Division Duplex (TDD) System

Many cellular systems (such as AMP, GSM, DAMP, etc.) use *frequency division duplex* (*FDD*) in which the transmitter and receiver operate simultaneously on different frequencies. Separation is provided between the downlink and uplink channelstoavoidthetransmittercausingselfinterferencetoitsreceiver.Other precautions are also needed to prevent self interference, such as the use of two antennas, or alternatively one antenna with a duplexer (a special design of RF filters protecting the receiver from the transmit frequency). A duplexer adds weight, size, and cost to a radio transceiver and can limit the minimum size of a subscriber unit.

A cellular system can be designed to use one frequency band by using *time division duplex* (*TDD*). In TDD a bidirectional flow of information is achievedusing the simplex-type scheme by automatically alternating in time the direction of transmission on a single frequency. At best TDD can only provide a quasi-simultaneousbidirectionalflow,sinceonedirectionmustbeoffwhiletheotheris using the frequency. However, with a high enough transmission rate on the chan- nel, the off time is not noticeable during conversations, and with a digital speech system, the only effect is a very short delay.

The amount of spectrum required for both FDD and TDD is the same. The differenceliesintheuseoftwobandsofspectrumseparatedbytherequiredband-width for FDD, whereas TDD requires only one band of frequencies but twice the bandwidth. It may be easier to find a single band of unassigned frequencies than finding two bands separated by the required bandwidth.

With TDD systems, the transmit time slot and the receiver time slot of the subscriber unit occur at different times. With the use of a simple RF switch in the subscriber unit, the antenna can be connected to the transmitter when a transmit burst is required (thus disconnecting the receiver from the antenna) and to the receiver for the incoming signal. The RF switch thus performs the function of the duplexer, but is less complex, smaller in size, and less costly. TDD uses a burst mode scheme like TDMA and therefore also does not require a duplexer. Sincethe bandwidth of a TDD channel is twice that of a transmitter and receiver in an FDD system, RF filters in all the transmitters and receivers for TDD systems must be designed to cover twice the bandwidth of FDD system filters.

6.2.2 FrequencyDivisionMultipleAccess

The FDMA is the simplest scheme used to provide multiple access. It separates differentusersbyassigningadifferentcarrierfrequency(seeFigure6.2).Multiple users are isolated using bandpass filters. In FDMA, signals from various users are assigneddifferentfrequencies,justasinananalogsystem.Frequencyguardbands areprovidedbetweenadjacentsignalspectratominimizecrosstalkbetweenadja- cent channels.

The advantages and disadvantages of FDMA with respect to TDMAor CDMA are:

Advantages

1. Capacity can be increased by reducing the information bit rate and using an efficient digital speech coding scheme (See Chapter 8) [20].

2. Technological advances required for implementation are simple. A system canbeconfiguredsothatimprovementsintermsofalowerbitratespeech coding could be easily incorporated.
3. Hardware simplicity, because multiple users are isolated by employing simple bandpass filters.

Disadvantages

1. The system architecture based on FDMA was implemented in first-generationanalogsystemssuchasadvancedmobilephonesystem(AMPS) or total access communication system (TACS). The improvement in capac- ity depends on operation at a reduced signal-to-interference (S/I) ratio. But the narrowband digital approach gives only limited advantages in this regard so that modest capacity improvements could be expected from the allocated spectrum.
2. The maximum bit-rate per channel is fixed and small, inhibiting the flexibilityinbit-ratecapabilitythatmaybearequirementforcomputerfile transfer in some applications in the future.

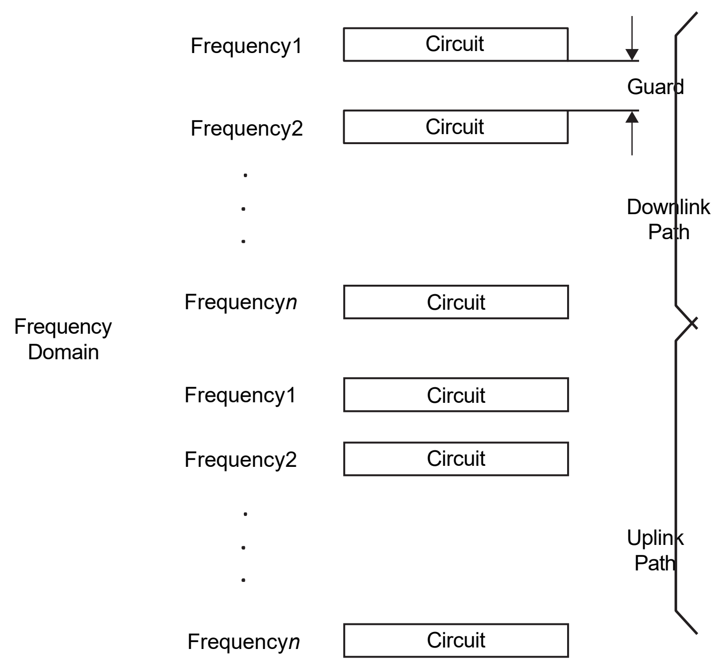


Figure6.2FDMA/FDDchannelarchitecture.

3. Inefficient use of spectrum, in FDMA if a channel is not in use, it remainsidle and cannot be used to enhance the system capacity.
4. Crosstalk arising from adjacent channel interference is produced by non-linear effects.

6.2.3 TimeDivisionMultipleAccess

In a TDMA system, each user uses the whole channel bandwidth for a fraction of time (see Figure 6.3) compared to an FDMA system where a single user occupies the channel bandwidth for the entire duration (see Figure 6.2) [2]. 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 eachuser'stransmissiontominimizecrosstalkbetweenchannels(seeFigure6.4).
Each user is assigned a frequency and a time slot to transmit data. The data is transmittedviaaradio-carrierfromabasestationtoseveralactivemobilesinthe 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 mobile stations use for identification.

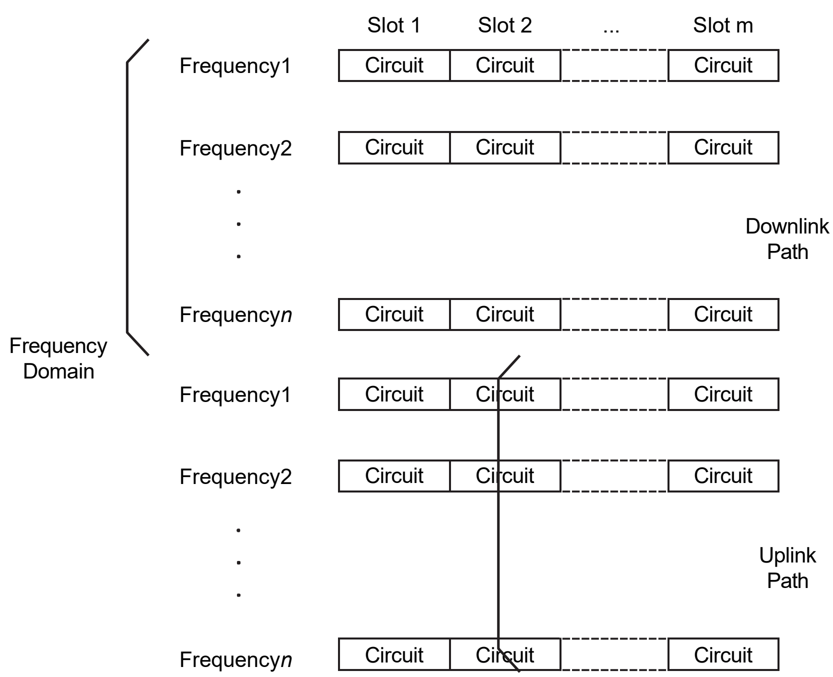


Figure6.3TDMA/FDDchannelarchitecture.

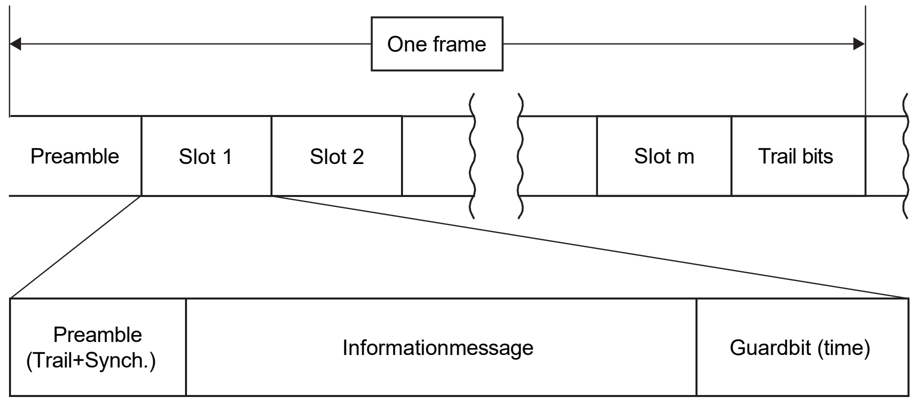


Figure6.4TDMAframe.

In a TDMA system, the user can use multiple slots to support a wide rangeofbitratesbyselectingthelowestmultiplexingrateormultipleofit.Thisenables supporting a variety of voice coding techniques at different bit rates with differ- ent voice qualities. Similarly, data communications customers could use the same kinds of schemes, choosing and paying for the digital data rate as required. This would allow customers to request and pay for a bandwidth on demand.

Depending on the data rate used and the number of slots per frame, aTDMA system can use the entire bandwidth of the system or can employ an FDD scheme. The resultant multiplexing is a mixture of frequency division and time division. The entire frequency band is divided into a number of duplex channels (about 350 to 400kHz). These channels are deployed in a frequency-reuse pat- tern, in which radio-port frequencies are assigned using an autonomous adaptive frequency assignment algorithm. Each channel is configured in a TDM mode for thedownlinkdirectionandaTDMAmodefortheuplinkdirection.

TheadvantagesanddisadvantagesofTDMAare:

Advantages

1. TDMApermitsaflexiblebitrate,notonlyformultiplesofthebasic single channel rate but also submultiples for low bit rate broadcast-type traffic.
2. TDMA offers the opportunity for frame-by-frame monitoring of signal strength/biterrorratestoenableeithermobilesorbasestationstoinitiate and execute handoffs.
3. TDMA, when used exclusively and not with FDMA, utilizes bandwidthmore efficiently because no frequency guard band is required between channels.

4. TDMA transmits each signal with sufficient guard time between time slots to accommodate time inaccuracies because of clock instability, delay spread, transmission delay because of propagation distance, and the tails of signal pulse because of transient responses.

Disadvantages

1. For mobiles and particularly for hand-sets, TDMA on the uplink demands high peak power in transmit mode, that shortens battery life.
2. TDMA requires a substantial amount of signal processing for matched filtering and correlation detection for synchronizing with a time slot.
3. TDMA requires synchronization. If the time slot synchronization is lost,the channels may collide with each other.
4. One complicating feature in a TDMA system is that the propagation time for a signal from a mobile station to a base station varies with its distance to the base station.

6.3 Spectral Efficiency

An efficient use of the spectrum is the most desirable feature of a mobile communications system. To realize this, a number of techniques have been proposed or already implemented. Some of these techniques used to improve spectral efficiency are reducing the channel bandwidth, information compression (low-rate speech coding), variable bit rate codec (see Chapter 8), improved channel assignment algorithms (dynamic channel assignment), and so on [11,17,19]. 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. The measure of spectral efficiency enables one to estimate the capacity of a mobile communications system.

The overall spectral efficiency of a mobile communications system can be estimated by knowing the *modulation* and the *multiple access* spectral efficiencies separately [16].

6.3.1 Spectral Efficiency of Modulation

The spectral efficiency with respect to modulation is defined as [16]:

$$\eta_m = \frac{(\text{Total Number of Channels Available in the System})}{(\text{Bandwidth})(\text{Total Coverage Area})} \tag{6.1a}$$

$$\eta_m = \frac{\frac{B_w}{B_c} \times \frac{N_c}{N}}{B_w \times N_c \times A} \tag{6.1b}$$

$$\eta_m = \frac{1}{B_c \times N \times A_c} \text{ Channels/MHz/km}^2 \tag{6.1c}$$

where:

η_m=modulationefficiency(channels/MHz/km^2)
B_w=bandwidthofthesystem(MHz)
B_c=channelspacing(MHz)
N_c=totalnumberofcellsinthecoveredarea
N=frequencyreusefactorofthesystem(orclustersize)
A_c=areacoveredbyacell(km^2)

Equation 6.1c indicates that the spectral efficiency of modulation does not depend on the bandwidth of the system. It only depends on the channel spacing, the cell area, and the frequency reuse factor, N. By reducing the channel spacing, thespectralefficiencyofmodulationforthesystemisincreased,providedthecell 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.

Anotherdefinitionofspectralefficiencyofmodulationis*Erlangs/MHz/km*^{2}

$$\eta_m = \frac{\text{MaximumTotalTrafficCarriedbySystem}}{\text{(SystemBandwidth)(TotalCoverageArea)}} \tag{6.2a}$$

$$\eta_m = \frac{\text{TotalTrafficCarriedby} \left(\frac{B_w/B_c}{N}\right) \text{Channels}}{B_w A_c} \tag{6.2b}$$

Byintroducingthetrunkingefficiencyfactor,$\eta_t$inEquation6.2b(<1,itis a function of the blocking probability and number of available channels per cell), the total traffic carried by the system is given as:

$$\eta_m = \frac{\eta_t \left(\frac{B_w/B_c}{N}\right)}{B_w A_c} \tag{6.2c}$$

$$\eta_m = \frac{\eta_t}{B_c N A_c} \tag{6.2d}$$

where:

$\eta_t$isafunctionoftheblockingprobabilityandthetotalnumberofavailable channelspercell $\left[\frac{B_w/B_c}{N}\right]$

BasedonEquation6.2dwecanconclude:

1. 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).
2. The relationship between system bandwidth, B_w, and the amount of traffic carried by the system is nonlinear, i.e., for a given percentage increase in B_w,theincreaseinthetrafficcarriedbythesystemismorethantheincrease in B_w.
3. 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.
4. Thespectralefficiencyofmodulationdependsontheblockingprobability.

Example 6.1

In the GSM800 digital channelized cellular system, the one-way bandwidth of the systemis12.5MHz.TheRFchannelspacingis200kHz.Eightusersshareeach RF channel and three channels per cell are used for control channels. Calculatethe spectral efficiency of modulation (for a dense metropolitan area with small cells) using the following parameters:

- Areaofacell=8km^2
- Totalcoveragearea=4000km^2
- Averagenumberofcallsperuserduringthebusyhour=1.2
- Averageholdingtimeofacall=100seconds
- Callblockingprobability=2%
- Frequencyreusefactor=4

Solution

Numberof200kHzRFchannels$=\frac{12.5\times1000}{200}=62$

Number of traffic channels $=62\times8=496$

Number of signaling channels per cell $=3$

Numberoftrafficchannelspercell$=\frac{496}{4}-3=121$

Numberofcells$=\frac{4000}{8}=500$

With2%blockingforanomnidirectionalcase,thetotaltrafficcarriedby 121channels(usingErlang-Btables)=108.4(1.0—0.02)=106.2Erlangs/cell or13.28Erlangs/km^2

Numberofcallsperhourpercell$=\frac{106.2\times3600}{100}=3823$,calls/hour/km$^2=\frac{3823}{8}=477.9$calls/hour/km^2

Max. numberofusers/cell/hour$=\frac{3823}{1.2}=3186$,users/hour/channel$=\frac{3186}{121}=26.33$

$$\eta_m=\frac{(\text{Erlangspercell})\times\text{no.ofcells}}{B_w\times\text{CoverageArea}}=\frac{106.2\times500}{12.5\times4000}=1.06\text{Erlangs/MHz/km}^2$$

6.3.2 MultipleAccessSpectralEfficiency

Multiple access spectral efficiency is defined as the ratio of the total time or frequencydedicatedfortraffictransmissiontothetotaltimeorfrequencyavailable to the system. Thus, the multiple access spectral efficiency is a dimensionless number with an upper limit of unity.

InFDMA,userssharetheradiospectruminthefrequencydomain.InFDMA, the multiple access efficiency is reduced because of guard bands between chan- nels and also because of signaling channels. In TDMA, the efficiency is reduced because of guard time and synchronization sequence.

FDMASpectralEfficiency

ForFDMA,multipleaccessspectralefficiencyisgivenas:

$$\eta_a=\frac{B_cN_T}{B_w}\leq1 \tag{6.3}$$

where:

η_a=multipleaccessspectralefficiency

N_T=totalnumberoftrafficchannelsinthecoveredarea

B_c=channelspacing

B_w=systembandwidth

Example 6.2

Inafirst-generationAMPsystemwherethereare395channelsof30kHzeach in a bandwidth of 12.5MHz, what is the multiple access spectral efficiency for FDMA?

Solution

$$\eta_a=\frac{30\times395}{12.5\times1000}=0.948$$

TDMASpectral Efficiency

TDMA can operate as wideband or narrowband. In the wideband TDMA, the entirespectrumisusedbyeachindividualuser.ForthewidebandTDMA,multiple access spectral efficiency is given as:

$$h_a = \frac{\tau M_t}{T_f} \tag{6.4}$$

where:

τ=durationofatimeslotthatcarriesdata

T_f=frameduration

M_t=numberoftimeslotsperframe

InEquation6.4itisassumedthatthetotalavailablebandwidthisshared byallusers.ForthenarrowbandTDMAschemes,thetotalbandisdividedinto 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 mul- tiple access spectral efficiency of the narrowband TDMA system is given as:

$$h_a = \left(\frac{(\tau M_t)}{T_f}\right)\left(\frac{(B_u N_u)}{B_w}\right) \tag{6.5}$$

where:

B_u=bandwidthofanindividualuserduringhisorhertimeslot

N_u=number of users sharing the same time slot in the system, but having access to different frequency sub-bands

6.3.3 OverallSpectralEfficiencyofFDMAandTDMASystems

The overall spectral efficiency, h, of a mobile communications system is obtained by considering both the modulation and multiple access spectral efficiencies

$$h = h_m h_a \tag{6.6}$$

Example 6.3

In the North American Narrowband TDMA cellular system, the one-way band- width of the system is 12.5MHz. The channel spacing is 30kHz and the total numberof voice channels in the system is 395. The frame duration is 40ms, with sixtime slots per frame. The system has an individual user data rate of 16.2kbps in which the speech with error protection has a rate of 13kbps. Calculate the multiple access spectral efficiency of the TDMA system.

Solution

Thetimeslotdurationthatcarriesdata: $\tau = \left(\frac{13}{16.2}\right)\left(\frac{40}{6}\right) = 5.35\,\text{ms}$

$T_f = 40\,\text{ms}, M_t = 6, N_u = 395, B_u = 30\,\text{kHz}$, and $B_w = 12.5\,\text{MHz}$

$$\eta_a = \frac{5.35 \times 6}{40} \times \frac{30 \times 395}{12500} = 0.76$$

Theoverheadportionoftheframe $= 1.0 - 0.76 = 24\%$

CapacityandFrameEfficiencyofaTDMASystem

CellCapacity

Thecellcapacityisdefinedasthemaximumnumberofusersthatcanbesup- ported simultaneously in each cell.

ThecapacityofaTDMAsystemisgivenby[16]:

$$N_u = \frac{\eta_b \mu}{\nu_f} \times \frac{B_w}{RN} \tag{6.7}$$

where:

N_u = numberofchannels(mobileusers)percell
η_b = bandwidthefficiencyfactor(<1.0)
μ = bitefficiency (=2 bit/symbol for QPSK, =1 bit/symbol for GMSK as used in GSM)
ν_f = voiceactivityfactor(equaltooneforTDMA)
B_w = one-waybandwidthofthesystem
R = information(bitrateplusoverhead)peruser
N = frequencyreusefactor

$$\text{Spectralefficiency}\ \eta = \frac{N_u \times R}{B_w}\ \text{bit/sec/Hz} \tag{6.8}$$

Example6.4

Calculate the capacity and spectral efficiency of a TDMA system using the following parameters: bandwidthefficiencyfactor $\eta_b = 0.9$, bitefficiency(withQPSK) $\mu = 2$, voice activity factor $\nu_f = 1.0$, one-way system bandwidth $B_w = 12.5\,\text{MHz}$, informationbitrate $R = 16.2\,\text{kbps}$, andfrequencyreusefactor $N = 19$.

Solution

$$N_u = \frac{0.9 \times 2}{1.0} \times \frac{12.5 \times 10^6}{16.2 \times 10^3 \times 19}$$

$N = 73.1$ (say73mobileuserspercell)

$$\text{Spectral efficiency } \eta = \frac{73 \times 16.2}{12.5 \times 1000} = 0.094 \text{ bit/sec/Hz}$$

Efficiency of a TDMA Frame

We refer to Figure 6.4 that shows a TDMA frame. The number of overhead bits per frame is:

$$b_0 = N_r b_r + N_t b_p + (N_t + N_r) b_g \tag{6.9}$$

where:

N_r = number of reference bursts per frame
N_t = number of traffic bursts (slots) per frame
b_r = number of overhead bits per reference burst
b_p = number of overhead bits per preamble per slot
b_g = number of equivalent bits in each guard time interval

The total number of bits per frame is:

$$b_T = T_f \times R_{rf} \tag{6.10a}$$

where:

T_f = frame duration
R_{rf} = bit rate of the RF channel

$$\text{Frame efficiency } \eta = (1 - b_0/b_T) \times 100\% \tag{6.10b}$$

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 = bit rate of each channel (user).

$$\text{No. of channels/frame} \quad N_{CF} = \frac{(\text{Total Data Bits})/(\text{frame})}{(\text{Bits per Channel})/(\text{frame})}$$

$$N_{CF} = \frac{\eta R_{rf} T_f}{R T_f} \tag{6.11a}$$

$$N_{CF} = \frac{\eta R_{rf}}{R} \tag{6.11b}$$

Equation 6.11b indicates the number of time slots per frame.

Example 6.5

ConsidertheGSMTDMAsystemwiththefollowingparameters:

$N_r = 2$

N_t=24framesof120mseachwitheighttimeslotsperframe

b_r=148bits in each of 8 time slots

b_p= 34bitsineachof8timeslots

b_g=8.25bits in each of 8 time slots

T_f= 120 ms

R_{rf}=270.8333333kbps

R=22.8kbps

Calculatetheframeefficiencyandthenumberofchannelsperframe.

Solution

$$b_0 = 2\times(8\times148)+24\times(8\times34)+8\times8.25 = 10{,}612 \text{ bits per frame}$$

$$b_T = 120\times10^{-3}\times270.8333333\times10^{3} = 32{,}500 \text{ bits per frame}$$

$$h = \left(1-\frac{10612}{32500}\right)\times100 = 67.35\%$$

$$\text{Number of channels/frame} = \frac{0.6735\times270.8333333}{22.8} = 8$$

The last calculation, with an answer of 8 channels, confirms that our calculation of efficiency is correct.

6.4 WidebandSystems

Inwidebandsystems,theentiresystembandwidthismadeavailabletoeachuser, and is many times larger than the bandwidth required to transmit information. Such systems are known as *spread spectrum* (SS) systems. There are two funda- mental types of spread spectrum systems: (1) direct sequence spread spectrum (DSSS) and (2) frequency hopping spread spectrum (FHSS) [3,26].

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 (see Chapter 11 for details). 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 (see Figure 6.5).

OneoftheadvantagesoftheDSSSsystemisthatthetransmissionbandwidth exceedsthecoherencebandwidth(seeChapter3).Thereceivedsignal,afterdespreading (see Chapter 11 for details), resolves into multiple signals with different time delays.ARakereceiver(seeChapter11)canbeusedtorecoverthemultipletime

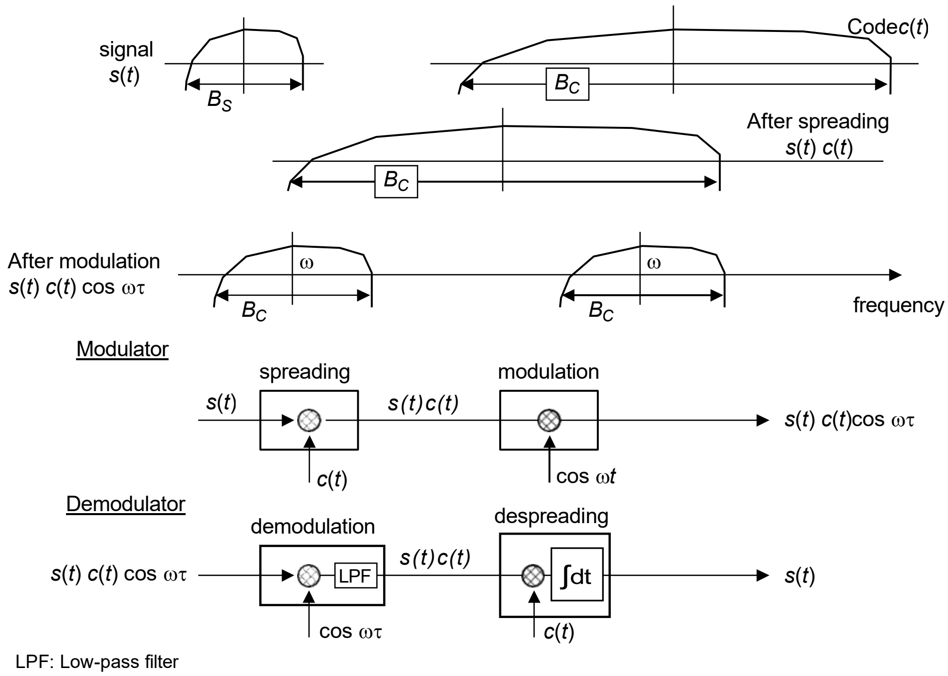


Figure6.5Directsequencespreadspectrum.

delayedsignalsandcombinethemintoonesignaltoprovideatimediversitywitha lowerfrequencyofdeepfades.Thus,theDSSSsystemprovidesaninherentrobustness againstmobile-channeldegradations.AnotherpotentialbenefitofaDSSSsystemis thegreaterresistancetointerferenceeffectsinafrequencyreusesituation.Also,there may be no hard limit on the number of mobile users who can simultaneously gain access.ThecapacityofaDSSSsystemdependsuponthedesiredvalueofE_b/I_0instead of resources (frequencies or time slots) as in FDMA or TDMA systems.

Frequency hopping (FH) is the periodic changing of the frequency or the frequency set associated with transmission (see Figure 6.6). If the modulation is M-*ary* frequency-shift-keying (MFSK) (see Chapter 9 for details), two or more frequenciesareinthesetthatchangeateachhop.Forothermodulations,asingle center or carrier frequency is changed at each hop.

AnFHsignalmaybeconsideredasequenceofmodulatedpulseswithpseudo-randomcarrierfrequencies.Thesetofpossiblecarrierfrequenciesiscalledthehop set. Hopping occurs over a frequency band that includes a number of frequency channels. The bandwidth of a frequency channel is called the *instantaneous bandwidth* (B_I). The bandwidth of the frequency band over which the hopping occursis called the total *hopping bandwidth* (B_H). The time duration between hops is called the *hop duration* or hopping period (T_H).

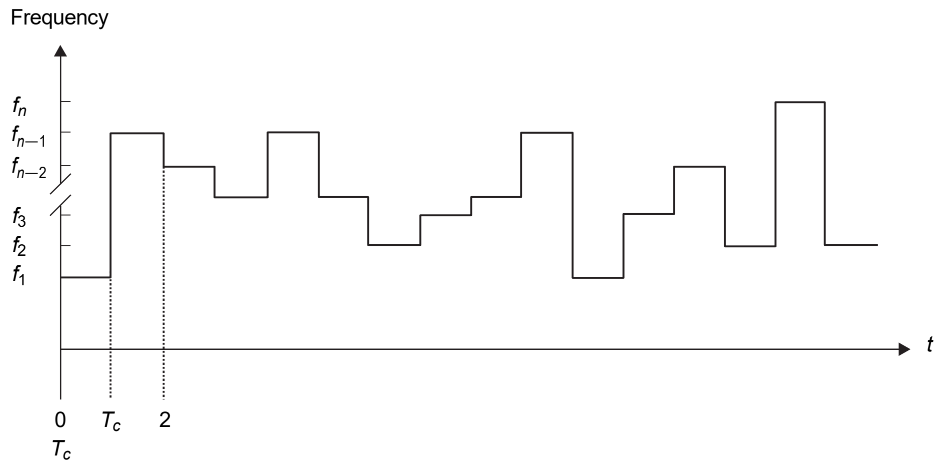


Figure 6.6 Frequency hopping spread spectrum system.

Frequency hopping can be classified as fast or slow. *Fast frequency hopping* occurs if there is frequency hop for each transmitted symbol. Thus, fast frequency hopping implies that the hopping rate equals or exceeds the information symbol rate. *Slow frequency hopping* occurs if two or more symbols are transmitted in the time interval between frequency hops.

Frequency hopping allows communicators to hop out of frequency channels with interference or to hop out of fades. To exploit this capability, error-correcting codes, appropriate interleaving, and disjoint frequency channels are nearly always used. A frequency synthesizer is required for frequency hopping systems to convert a stable reference frequency into the various frequency of hop set.

Frequency hopping communicators do not often operate in isolation. Instead, they are usually elements of a network of frequency hopping systems that create mutual multiple-access interference. This network is called a frequency-hopping multiple-access (FHMA) network.

If the hoppers of an FHMA network all use the same M frequency channels, but coordinate their frequency transitions and their hopping sequence, then the multiple-access interference for a lightly loaded system can be greatly reduced compared to a non-hopped system. For the number of hopped signals (M_h) less than the number of channels (N_c), a coordinated hopping pattern can eliminate interference. As the number of hopped signals increases beyond N_c, then the interference will increase in proportion to the ratio of the number of signals to the number of channels. In the absence of fading or multipath interference, since there is no interference suppression system in frequency hopping, for a high channel loading the performance of a frequency hopping system is no better than a non-hopped system. Frequency hopping systems are best for light channel loadings in the presence of conventional non-hopped systems.

When fading or multipath interference is present, the frequency hopping system has better error performance than a non-hopped system. If the transmitter hops to a channel in a fade, the errors are limited in duration since the system will shortly hop to a new frequency where the fade may not be as deep.

Network coordination for frequency hopping systems are simpler to implement than that for DSSS systems because the timing alignments must be within a fraction of a hop duration, rather than a fraction of a sequence chip (narrow pulse). In general, frequency hopping systems reject interference by trying to avoid it, whereas DSSS systems reject interference by spreading it. The interleaving and error-correcting codes that are effective with frequency hopping systems are also effective with DSSS systems.

The major problems with frequency hopping systems with increasing hopping rates are the cost of the frequency synthesizer increases and its reliability decreases, and synchronization becomes more difficult.

In theory, a wideband system can be overlaid on existing, fully loaded, narrowband channelized systems (as an example, the IS-95 CDMA system overlays on existing AMPS [FDMA]). Thus, it may be possible to create a wideband network right on top of the narrowband cellular system using the same spectrum.

6.5 Comparisons of FDMA, TDMA, and DS-CDMA

The DSSS approach is the basis to implementation of the direct sequence code division multiple access (DS-CDMA) technique introduced by Qualcom. The DS-CDMA has been used in commercial applications of mobile communications. The primary advantage of DS-CDMA is its ability to tolerate a fair amount of interfering signals compared to FDMA and TDMA that typically cannot tolerate any such interference(Figure 6.7). As a result of the interference tolerance of CDMA, the problems of frequency band assignment and adjacent cell interference are greatly simplified. Also, flexibility in system design and deployment are significantly improved since interference to others is not a problem. On the other hand, FDMA and TDMA radios must be carefully assigned a frequency or time slot to assure that there is no interference with other similar radios. Therefore, sophisticated filtering and guard band protection is needed with FDMA and TDMA technologies. With DS-CDMA, adjacent microcells share the same frequencies whereas with FDMA/TDMA it is not feasible for adjacent microcells to share the same frequencies because of interference. In both FDMA and TDMA systems, a time-consuming frequency planning task is required whenever a network changes, whereas no such frequency planning is needed for a CDMA network since each cell uses the same frequencies.

Capacity improvements with DS-CDMA also result from voice activity patterns during two-way conversations, (i.e., times when a party is not talking) that cannot be cost-effectively exploited in FDMA or TDMA systems. DS-CDMA radios can, therefore, accommodate more mobile users than FDMA/TDMA radios

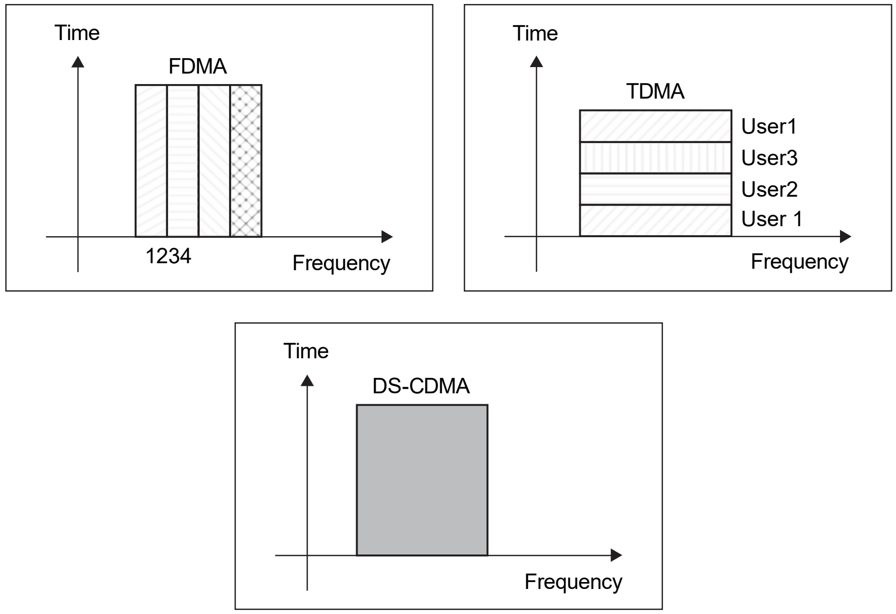


Figure6.7Comparisonofmultipleaccessmethods.

onthesamebandwidth.FurthercapacitygainsforFDMA,TDMA,andCDMA can also result from antenna technology advancement by using directional anten- nas that allow the microcell area to be divided into sectors. Table 6.1 provides a summary of access technologies used for various wireless systems.

Table6.1Accesstechnologiesforwirelesssystem.

System	Accesstechnology	Mode of operation	Framerate(kbps)
NorthAmerican IS-54(DualMode)	TDMA/FDD FDMA/FDD	Digital/ AnalogFM	48.6 —
NorthAmerican IS-95(DualMode)	DS-CDMA/FDD FDMA/FDD	Digital/ AnalogFM	1228.8 —
NorthAmerican IS-136	TDMA/FDD	Digital	48.6
GSM(usedallover world)	TDMA/FDD	Digital	270.833
EuropeanCT-2 Cordless	FDMA/TDD	Digital	72.0
DECTCordless	TDMA/TDD	Digital	1152.0

6.6 Capacity of a DS-CDMA System

The capacity of a DS-CDMA system depends on the processing gain, G_p (a ratio of spreading bandwidth, B_w, and information rate, R), the bit energy-to-interference ratio, E_b/I_0, the voice duty cycle, v_f, the DS-CDMA omnidirectional frequency reuse efficiency, h_f, and the 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 signal-to-interference ratio is

$$\frac{S}{I} = \frac{R}{B_w} \times \frac{E_b}{I_0} \tag{6.12}$$

where:

E_b = energy per bit

I_0 = 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 center 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):

$$\frac{S}{I} = \frac{1}{(N_u - 1) \times [1 + 6 \times k_1 + 12 \times k_2 + 18 \times k_3 + \ldots]} \tag{6.13}$$

where:

N_u = number of mobile users in the band, B_w

$k_i, i = 1, 2, 3, \ldots$ = the interference contribution from all terminals in individual cells in tiers 1, 2, 3, etc., relative to the interference from the center cell. This loss contribution is a function of both the path loss to the center cell and the power reduction because of power control to an interfering mobile's own cell center.

If we define a frequency reuse efficiency, h_f, as in Equation 6.14a, then E_b/I_0 is given by Equation 6.15.

$$h_f = \frac{1}{[1 + 6 \times k_1 + 12 \times k_2 + 18 \times k_3 + \ldots]} \tag{6.14a}$$

$$\frac{S}{I} = \frac{h_f}{(N_u - 1)} \tag{6.14b}$$

$$\frac{E_b}{I_0} = \frac{B_w}{R} \times \frac{h_f}{(N_u - 1)} \tag{6.15}$$

This equation does not include the effect of background thermal and spurious noise (i.e., p) 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:

$$\frac{E_b}{I_0} = \frac{B_w}{R} \times \frac{h_f}{(N_u - 1) + p/S} \tag{6.16}$$

Note that from Equation 6.16 the capacity of the DS-CDMA system isreduced by p/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 codingscheme.Byusingapowerfulcodingscheme,theE_b/I_0ratiocanbereduced, but this increases system complexity. Also, 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 behavior ofhumanspeechandtheotherisbasedontheapplicationofthesectorizedantennas.Fromexperimentalstudiesithasbeenfoundthattypicallyinafullduplex 2-way voice conversation, the duty cycle of each voice is, on the average, lessthan 40%. 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 inter- fering channels. It is not cost-effective to exploit the voice activity in the FDMA or TDMAsystembecauseofthetimedelayassociatedwithreassigningthechannel resourceduringthespeechpauses.Ifwedefine$v_f$asthevoiceactivityfactor(<1), then Equation 6.16 can be written as:

$$\frac{E_b}{I_0} = \frac{h_f}{v_f} \times \frac{B_w}{R} \times \frac{1}{(N_u - 1) + p/S} \tag{6.17a}$$

$$(N_u - 1) + \frac{p}{S} = \left[\frac{h_f}{v_f}\right] \times \left[\frac{B_w}{R}\right] \times \left[\frac{I_0}{E_b}\right] \tag{6.17b}$$

TheequationtodeterminethecapacityofaDS-CDMAsystemshould 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 6.17b is augmented by these additional factorstoprovidethefollowingequationforDS-CDMAcapacityatonecell:

$$N_u=\frac{h_f h_b c_d \lambda}{v_f}\times\frac{B_w}{R\times(E_b/I_0)}+1-\frac{p}{S} \tag{6.18a}$$

Equation6.18acanberewrittenasEquation6.18bbyneglectingthelast two terms.

$$N_u=\frac{h_f h_b c_d \lambda}{v_f}\times\frac{B_w}{R\times(E_b/I_0)} \tag{6.18b}$$

where:

h_f = frequency reuse efficiency <1
h_b =bandwidth efficiency factor <1
c_d =capacitydegradationfactortoaccountforimperfectAPC <1
v_f =voiceactivityfactor <1
B_w=one-way bandwidth of the systemR=information bit rate plus overhead
E_b=energyperbitofthedesiredsignal
E_b/I_0=desiredenergy-to-interferenceratio(dependentonqualityofservice)
λ=efficiency of sector-antenna in cell (<G, number of sectors in thecell-site antenna)

For digital voice transmission, E_b/I_0 is the required value for a bit error rate (BER) of about 10^{-3}or better, and $h_f$depends on the quality of the diversity. Under the most optimistic assumption, $h_f<0.5$. The voice activity factor, $v_f$is usuallyassumedtobelessthanorequalto0.6.E_b/I_0foraBERof10^{-3}canbeas highas63(18dB)ifnocodingisusedandaslowas5(7dB)forasystemusing a powerful coding scheme. The capacity degradation factor, $c_d$will depend on the implementation but will always be less than 1.

Example 6.6

CalculatethecapacityandspectralefficiencyoftheDS-CDMAsystemwithan omnidirectional cell using the following data:

- bandwidthefficiency$h_b=0.9$
- frequencyreuse efficiency $h_f=0.45$

- capacity degradation factor $c_d = 0.8$
- voice activity factor $v_f = 0.4$
- information bit rate $R = 16.2$ kbps
- $E_b/I_0 = 7$ dB
- one-way system bandwidth $B_w = 12.5$ MHz

Neglect other sources of interference.

Solution

$E_b/I_0 = 5.02$ (7 dB)

$$N_u = 0.45 \times 0.9 \times 0.8 \times 1 \times \frac{1}{0.4} \times \frac{12.5 \times 10^6}{16.2 \times 10^3 \times 5.02}$$

$N_u = 124.5$ (say 125)

The spectral efficiency, $\eta = \frac{125 \times 16.2}{12.5 \times 10^3} = 0.162$ bits/sec/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/sec/Hz.

6.7 Comparison of DS-CDMA vs. TDMA System Capacity

Using Equations 6.7 and 6.18b with $v_f = 1$ (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 systems is given as:

$$\frac{N_{\text{CDMA}}}{N_{\text{TDMA}}} = \frac{c_d N \eta_f}{E_b/I_0} \times \frac{1}{v_{f_{\text{cdma}}}} \times \frac{1}{\mu} \times \frac{R_{\text{TDMA}}}{R_{\text{CDMA}}} \tag{6.19}$$

Example 6.7

Using the data given in Examples 6.4 and 6.6, compare the capacity of the DS-CDMA and TDMA omnidirectional cell.

Solution

$$\frac{N_{\text{CDMA}}}{N_{\text{TDMA}}} = \frac{0.8 \times 19 \times 0.45}{5.02} \times \frac{1}{0.4} \times \frac{1}{2} \times \frac{16.2}{16.2} = 1.703$$

6.8 Frequency Hopping Spread Spectrum with M-*ary* Frequency Shift Keying

The FHSS system uses M-*ary* frequency shift keying modulation (MFSK) and involves the hopping of the carrier frequency in a random manner. It uses MFSK, in which $b = \log_2 M$ information bits determine which one of M frequencies is to beused[19].TheportionoftheM-*ary*signalsetisshiftedpseudo-randomlyby the frequency synthesizer over a hopping bandwidth, B_{ss}. A typical block diagram is shown in Figure 6.8.

InaconventionalMFSKsystem,thedatasymbolismodulatedona carrier whose frequency is pseudo-randomly determined. The frequency synthesizerproducesatransmissiontonebasedonsimultaneousdictatesofthepseudo-noise(PN)code(seeChapter11)andthedata.Ateachfrequencyhoptimea PNgeneratorfeedsthefrequencysynthesizerafrequencyword(a sequenceof Lchips),whichdictatesoneof2^Lsymbol-setpositions.TheFHbandwidth, B_{ss}, and the minimum frequency spacing between consecutive hop positions, Δf, dictate the minimum number of chips required in the frequency word.

Example6.8

Ahoppingbandwidth,B_{ss},of600MHzandafrequencystepsize,Δf,of400Hz are used. What is the minimum number of PN chips that are required for each frequency word?

Solution

$$\text{Numberoftonescontainedin}B_{ss} = \frac{B_{ss}}{\Delta f} = \frac{600\times10^6}{400} = 1.5\times10^6$$

$$\text{Minimumnumberofchipsrequired} = \lceil \log_2(1.5\times10^6) \rceil = 20\text{chips}$$

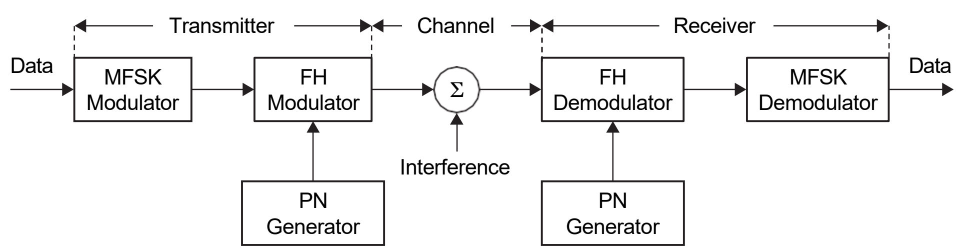


Figure6.8FrequencyhoppingusingMFSK.

6.9 OrthogonalFrequencyDivisionMultiplexing(OFDM)

In this section we briefly introduce OFDM. For more details readers should refer to [19]. OFDM uses three transmission principles, multirate, multisymbol, and multicarrier. OFDM is similar to frequency division multiplexing (FDM). OFDM distributes the data over a large number of carriers that are spaced apart at precise frequencies. The spacing provides the orthogonality in this technique, which prevents the demodulator from seeing frequencies other than their own.

Multiple Input, Multiple Output-OFDM (MIMO-OFDM) uses multiple antennas to transmit and receive radio signals. MIMO-OFDM allows service providerstodeployabroadbandwirelessaccesssystemthathasnon-line-of-sight (NLOS)functionality.MIMO-OFDMtakesadvantageofthemultipathproper- ties of the environment using base station antennas that do not have LOS. The MIMO-OFDMsystemusesmultipleantennastosimultaneouslytransmitdata in small pieces to the receiver, which can process the data flow and put it back together. This process, called *spatial multiplexing*, proportionally boosts the data transmission speed by a factor equal to the number of transmitting antennas. In addition, since all data is transmitted both in the same frequency band and with separate spatial signatures, this technique utilizes spectrum efficiently. VOFDM (vectorOFDM)usestheconceptofMIMOtechnology.

We consider a data stream operating at R bps and an available bandwidth of $N\Delta f$ centered at 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 substreams using a serial-to-parallel converter, each substream has a data rate of R/N and is transmitted on a separate subcarrier, with spacing between adjacent subcarriers of Δf **(**see Figure 6.9**)**. The bit duration is N/R. The advantage ofOFDM is that on a multiple channel the multipath is reduced relative to the symbolintervalbyaratioof$1/N$andthusimposeslessdistortionineachmodulated symbol. OFDM overcomes inter-symbol interference (ISI) in a multipath environment. ISI has a greater impact at higher data rates because the distance between bitsorsymbolsissmaller.WithOFDM,thedatarateisreducedbyafactorof$N$, whichincreasesthesymboldurationbyafactorof$N$.Thus,ifthesymboldura- tion is $T_s$for the source stream, the duration of OFDM signals isNT_s. This significantlyreducestheeffectofISI.Asadesigncriterion,NisselectedsothatNT_sissign ificantlygreaterthanτ_{rms}(rmsdelayspread)ofthechannel.Withtheuse of OFDM,itmaynotbenecessarytodeployanequalizer.OFDMisanidealsolution for broadband communications, because increasing the data rate is simply a matterofincreasingthenumberofsubcarriers.Toavoidoverlapbetweenconsecutive symbols, a time guard is enforced between the transmissions of two OFDM pulses that will reduce the effective data rate. Also, some subcarriers are devoted to synchronization of signal, and some are reserved for redundancy.

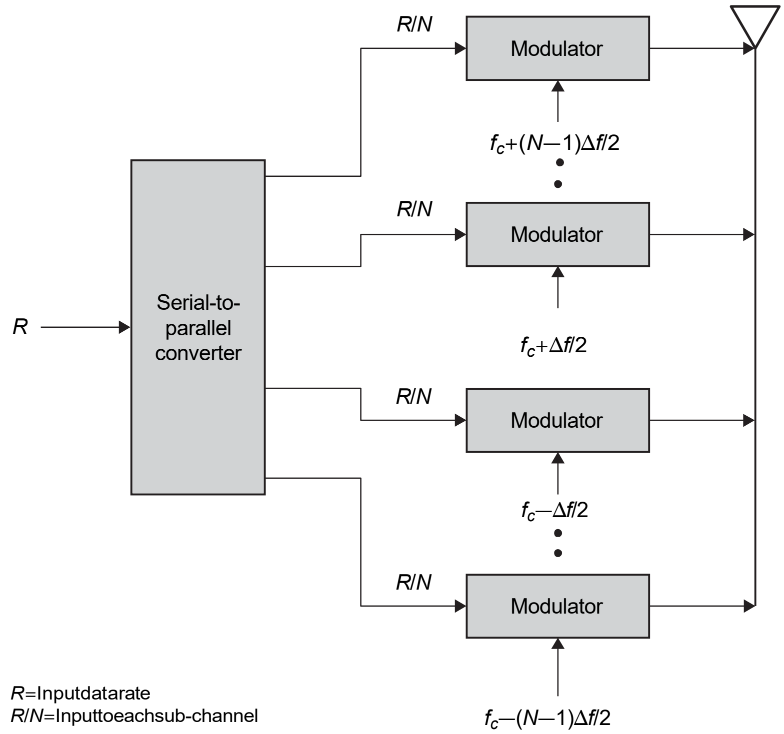


Figure6.9Orthogonalfrequencydivisionmultiplexing(OFDM).

ThemostimportantfeatureofOFDMistheorthogonalrelationshipbetween the subcarrier signals. Orthogonality allows the OFDM subcarriers to overlapeach other without interference. OFDM uses FH to create a spread spectrum system. FH has several advantages over DSSS, for example, no near-far problem, easier synchronization, less complex receivers, and so on.

In the OFDM the input information sequence is first converted into parallel datasequencesandeachserial/parallelconverteroutputismultipliedwithspreadingc ode.Datafromallsubcarriersismodulatedinbasebandbyinverse fastFouriertransform(IFFT)andconvertedbackintoserialdata.Theguard intervalisinsertedbetweensymbolstoavoidISIcausedbymultipathfading and finally the signal is transmitted after RF up-conversion. At the receiver, after down-conversion, the m-subcarrier component corresponding to the received data isfirstcoherentlydetectedwithFFTandthenmultipliedwithgaintocombinethe energyofthereceivedsignalscatteredinthefrequencydomain(seeFigure6.10). Wireless Local Area Networks (WLAN) development is ongoing for wireless point-to-pointandpoint-to-multipointconfigurationsusingOFDMtechnology.

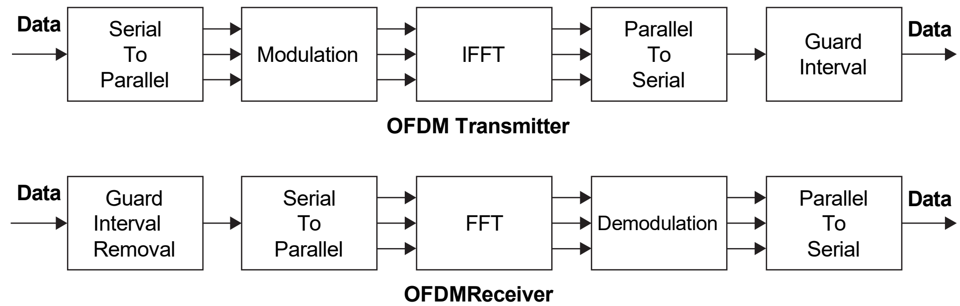


Figure6.10IEEE802.11aTransmitandReceiveOFDM.

In a supplement to the IEEE 802.11 standard, the IEEE 802.11 working group publishedIEEE802.11a,whichoutlinestheuseofOFDMinthe5.8GHzband.

Thebasicprincipalofoperationistodivideahigh-speedbinarysignalto be transmitted into a number of lower data rate subcarriers. There are 48 data subcarriers and 4 pilot subcarriers for a total of 52 subcarriers. Each lower data rate bit stream is used to modulate a separate subcarrier from one of the channels in the 5GHz band. Prior to transmission the data is encoded using convolutionalcode(seeChapter8)ofrate,$R=1/2$andbitinterleavedforthe desireddatarate.Eachbitisthenmappedintoacomplexnumberaccording to the modulation type and subdivided in 48 data subcarriers and 4 pilot subcarriers. The subcarriers are combined using an IFFT and transmitted. At the receiver,thecarrierisconvertedbacktoamulticarrierlowerdatarateformusing FFT.Thelowerdatasubcarriersarecombinedtoformahighratedataunit.

6.10 MulticarrierDS-CDMA(MC-DS-CDMA)

Future wireless systems such as a fourth-generation (4G) system will need flexibilitytoprovidesubscriberswithavarietyofservicessuchasvoice,data,images, and video. Because these services have widely differing data rates and traffic profiles, future generation systems will have to accommodate a wide variety of data rates. DS-CDMA has proven very successful for large-scale cellular voice systems, but there are concerns whether DS-CDMA will be well-suited to non-voice traf-fic. The DS-CDMA system suffers inter-symbol interference (ISI) and multi-user interference (MUI) caused by multipath propagation, leading to a high loss of performance.

With OFDM, the time dispersive channel is seen in the frequency domain asa set of parallel independent flat subchannels and can be equalized at a low complexity. There are potential benefits to combining OFDM and DS-CDMA. Basically thefrequency-selective channel is first equalized inthe frequency domain using the

OFDM modulation technique. DS-CDMA is applied on top of theequalized channel, keeping the orthogonality properties of spreading codes. The combination ofOFDMandDS-CDMAisusedinMC-DS-CDMA.MC-DS-CDMA[4,5,12,25] marries the best of the OFDM and DS-CDMA world and, consequently, it can ensure good performance in severe multipath conditions. MC-DS-CDMA can achieve very large average throughput. To further enhance the spectral efficiency of the system, some form of adaptive modulation can be used.

Basically,threemaindesignsexistintheliterature,namely,MC-CDMA, MC-DS-CDMA,andmultitone(MT)-CDMA.InMC-CDMA,thespreadingcodeisappliedacrossanumberoforthogonalsubcarriersinthefrequencydomain. In MC-DS-CDMA, the data stream is first divided into a number of substreams. Each substream is spread in time through a spreading code and then transmitted over one of a set of orthogonal subcarriers. In MT-CDMA the system undergoes similar operations as MC-DS-CDMA except that the different subcarriers are not orthogonal after spreading. This allows higher spectral efficiencies and longer spreadingcodes;however,differentsubstreamsinterferewithoneother.TheMC- DS-CDMA transmitter spreads the original data stream over different orthogonal subcarriers using a given spreading code in the frequency domain.

6.11 RandomAccessMethods

So far we have discussed the reservation-based schemes, now we focus on random-access schemes [8]. When each user has a steady flow of information to transmit(forexample,adatafiletransferorafacsimiletransmission),reservation-based access methods are useful as they make an efficient use of communication resources. However, when the information to be transmitted is bursty in nature, thereservation-basedaccessmethodsresultinwastingcommunicationresources. Furthermore,inacellularsystemwheresubscribersarechargedbasedonachan- nel connection time, the reservation-based access methods may be too expensive totransmitshortmessages.Random-accessprotocolsprovideflexibleandefficient methods for managing a channel access to transmit short messages. The random-accessmethodsgivefreedomforeachusertogainaccesstothenetworkwhenever the user has information to send. Because of this freedom, these schemes resultin contention among users accessing the network. Contention may cause colli- sions and may require retransmission of the information. The commonly used random-accessprotocolsarepureALOHA,slotted-ALOHA,andCSMA/CD.In the following section we briefly describe details of each of these protocols and provide the necessary throughput expressions.

6.11.1 Pure ALOHA

In the pure ALOHA [18,23] scheme, each user transmits information whenever theuserhasinformationtosend.Ausersendsinformationinpackets.After sending a packet, the user waits a length of time equal to the round-trip delay for an acknowledgment (ACK) of the packet from the receiver. If no ACK is received, the packet is assumed to be lost in a collision and it is retransmitted with a randomly selected delay to avoid repeated collisions.* The normalized throughput S (average new packet arrival rate divided by the maximum packet throughput) of the pure ALOHA protocol is given as:

$$S = Ge^{-2G} \tag{6.20}$$

where$G$=normalizedofferedtrafficload

FromEquation6.20itshouldbenotedthatthemaximumthroughputoccurs attrafficload$G$=50%andis$S$=1/2e.Thisisabout0.184.Thus,thebestchannel utilization with the pure ALOHA protocol is only 18.4%.

6.11.2 SlottedALOHA

In the slotted-ALOHA [23] system, the transmission time is divided into timeslots.Eachtimeslotismadeexactlyequaltopackettransmissiontime.Users are synchronized to the time slots, so that whenever a user has a packet to send, the packet is held and transmitted in the next time slot. With the synchronized time slots scheme, the interval of a possible collision for any packet is reduced to one packet time from two packet times, as in the pure ALOHA scheme. The normalizedthroughput$S$fortheslotted-ALOHAprotocolisgivenas:

$$S = Ge^{-G} \tag{6.21}$$

where$G$=normalizedofferedtrafficload

The maximum throughput for the slotted ALOHA occurs at $G = 1.0$ (Equation6.21)anditisequalto1/$e$orabout0.368.Thisimpliesthatatthemaximum throughput, 36.8% of the time slots carry successfully transmitted packets. The best channel utilization with the slotted ALOHA protocol is 36.8%—twice the pure ALOHA protocol.

---

*It should be noted that the protocol on CDMA access channels as implemented in TIA IS-95-A is baseduponthepureALOHAapproach.Themobilestationrandomizesitsattemptforsendingamessageonthe access channel and may retry if an acknowledgment is not received from the base station. For furtherdetails, one should reference Section 6.6.3.1.1.1 of TIA IS-95-A.

6.11.3 CarrierSenseMultipleAccess(CSMA)

The carrier sense multiple access (CSMA) [8,18] protocols have been widely used inbothwiredandwirelessLANs.Theseprotocolsprovideenhancementsoverthe pure and slotted ALOHA protocols. The enhancements are achieved through the use of the additional capability at each user station to sense the transmissions of other user stations. The carrier sense information is used to minimize the length of collision intervals. For carrier sensing to be effective, propagation delays must be less than packet transmission times. Two general classes of CSMA protocolsare nonpersistent and p-persistent.

- **Nonpersistent CSMA:** A user station does not sense the channel continuously while it is busy. Instead, after sensing the busy condition, it waits fora randomly selected interval of time before sensing again. The algorithm worksasfollows:ifthechannelisfoundtobeidle,thepacketistransmitted; or if the channel is sensed busy, the user station backs off to reschedule the packettoalatertime.Afterbackingoff,thechannelissensedagain,andthe algorithm is repeated again.
- **p-persistentCSMA:**Theslotlengthistypicallyselectedtobethemaximum propagation delay. When a station has information to transmit, it sensesthe channel. If the channel is found to be idle, it transmits with probability p. With probability $q = 1 - p$, the user station postpones its action to the next slot, where it senses the channel again. If that slot is idle, the station transmitswithprobabilityporpostponesagainwithprobabilityq.The procedure is repeated until either the frame has been transmitted or the channel is found to be busy. If the station initially senses the channel to be busy, it simply waits one slot and applies the above procedure.
- **1-persistentCSMA:**1-persistentCSMAisthesimplestformofthep-persistent CSMA.Itsignifiesthetransmissionstrategy,whichistotransmitwithprobability1assoonasthechannelbecomesidle.Aftersendingthepacket,theuser stationwaitsforanACK,andifitisnotreceivedwithinaspecifiedamountof time, the user station waits for a random amount of time, and then resumes listeningtothechannel.Whenthechannelisagainfoundtobeidle,thepacket is retransmitted immediately.

Formoredetails,thereadershouldreferto[18].

ThethroughputexpressionsfortheCSMAprotocolsare:

- **UnslottednonpersistentCSMA**

$$S = \frac{Ge^{-aG}}{G(1+2a)+e^{-aG}} \tag{6.22}$$

- **SlottednonpersistentCSMA**

$$S=\frac{aGe^{-aG}}{1-e^{-aG}+a} \tag{6.23}$$

- **Unslotted1-persistentCSMA**

$$S=\frac{G[1+G+aG(1+G+(aG)/2)]e^{-G(1+2a)}}{G(1+2a)-(1-e^{-aG})+(1+aG)e^{-G(1+a)}} \tag{6.24}$$

- **Slotted1-persistentCSMA**

$$S=\frac{Ge^{-G(1+a)}[1+a-e^{-aG}]}{(1+a)(1-e^{-aG})+ae^{-G(1+a)}} \tag{6.25}$$

where:

S=normalizedthroughput
G=normalizedofferedtrafficload
$a=\tau/T_p$
τ=maximumpropagationdelay
T_p=packettransmissiontime

Example6.9

WeconsideraWLANinstallationinwhichthemaximumpropagationdelayis 0.4sec.The WLAN operates at a data rate of 10Mbps, and packets have 400 bits. Calculate the normalized throughput with: (1) an unslotted nonpersistent, (2) a slotted persistent, and (3) a slotted 1-persistent CSMA protocol.

Solution

$$T_p=\frac{400}{10}=40\mu s$$

$$a=\frac{\tau}{T_p}=\frac{0.4}{40}=0.01$$

$$G=\frac{40\times10^{-6}\times10\times10^{6}}{400}=1$$

- Slotted nonpersistent:

$$S=\frac{0.01\times 1\times e^{-0.01}}{1-e^{-0.01}+0.01}=0.495$$

- Unslotted nonpersistent:

$$S=\frac{1\times e^{-0.01}}{(1+0.02)+e^{-0.01}}=0.493$$

- Slotted 1-persistent:

$$S=\frac{e^{-1.01}(1+0.01-e^{-0.01})}{(1+0.01)(1-e^{-0.01})+0.01e^{-1.01}}=0.531$$

6.11.4 Carrier Sense Multiple Access with Collision Detection

A considerable performance improvement in the basic CSMA protocols can be achieved by means of the *carrier sense multiple access with collision detection* (CSMA/CD) technique. The CSMA/CD protocols are essentially the same as those for CSMA with addition of the collision-detection feature. Similar to CSMA protocols, there are nonpersistent, 1-persistent, and p-persistent CSMA/CD protocols. More details about CSMA/CD protocols can be found in [27].

When a CSMA/CD station senses that a collision has occurred, it immediately stops transmitting its packets and sends a brief jamming signal to notify all stations of this collision. Collisions are detected by monitoring the analog waveform directly from the channel. When signals from two or more stations are present simultaneously, the composite waveform is distorted from that of a single station. This is manifested in the form of larger than normal voltage amplitude on the cable. In the Ethernet the collision is recognized by the transmitting station, which goes into a retransmission phase based on an exponential random backoff algorithm.

The normalized throughput for unslotted nonpersistent and slotted nonpersistent CSMA/CD is given as:

Unslotted nonpersistent CSMA/CD

$$S=\frac{Ge^{-aG}}{Ge^{-aG}+bG(1-e^{-aG})+2aG(1-e^{-aG})+(2-e^{-aG})} \tag{6.26}$$

where b = jamming signal length

SlottednonpersistentCSMA/CD

$$S=\frac{aGe^{-aG}}{aGe^{-aG}+b(1-e^{-aG}-aGe^{-aG})+a(2-e^{-aG}-aGe^{-aG})} \tag{6.27}$$

Whilethesecollisiondetectionmechanismsareagoodideaonawired local area network (LAN), they cannot be used on a wireless local area network (WLAN) environment for two main reasons:

- Implementingacollisiondetectionmechanismwouldrequiretheimplementation of a full duplex radio capable of transmitting and receiving at the same time—an approach that would increase the cost significantly.
- In a wireless environment we cannot assume that all stations hear each other(whichisthebasicassumptionofthecollisiondetectionscheme),and thefactthatastationwantstotransmitandsensesthemediumasfreedoes not necessarily mean that the medium is free around the receiver area.

6.11.5 CarrierSenseMultipleAccesswithCollisionAvoidance (CSMA/CA)

IEEE 802.11 uses a protocol known as *carrier sense multiple access with collision avoidance* (CSMA/CA) or distributed coordination function (DCF). CSMA/CA attempts to avoid collisions by using *explicit packet acknowledgment* (ACK),which means an ACK packet is sent by the receiving station to confirm that the data packet arrived intact.

The CSMA/CA protocol works as follows. A station wishing to transmit sensesthemedium,ifthemediumisbusy(i.e.,someotherstationistransmitting) thenthestationdefersitstransmissiontoalatertime.Ifnoactivityisdetected,the stationwaitsanadditional,randomlyselectedperiodoftimeandthentransmitsif themediumisstillfree.Ifthepacketisreceivedintact,thereceivingstationissues an ACK frame that, once successfully received by the sender, completes the pro- cess. If the ACK frame is not detected by the sending station, either because the originaldatapacketwasnotreceivedintactortheACKwasnotreceivedintact, a collision is assumed to have occurred and the data packet is transmitted again after waiting another random amount of time. The CSMA/CA provides a way to share access over the medium. This explicit ACK mechanism also handles interference and other radio-related problems very effectively. However, it does add some overhead to 802.11 that 802.3 does not have, so that an 802.11 WLAN will always have slower performance than the equivalent Ethernet LAN (802.3).

TheCSMA/CAprotocolisveryeffectivewhenthemediumisnotheavilyloadedsinceitallowsstationstotransmitwithminimumdelay.Butthere isalwaysachanceofstationssimultaneouslysensingthemediumasbeingfree and transmitting at the same time, causing a collision. These collisions must be identified, so that the media access control (MAC) layer can retransmit the packet by itself and not by the upper layers, which would cause significant delay. In particular, the hidden node and exposed node problems should be addressed by MAC. Both of them give rise to many performance problems including throughput degradtion, unfair throughput distribution, and throughput instability (see Chapter 18 for details).

The IEEE 802.11 uses a collision avoidance (CA) mechanism together with a positive ACK. The MAC layer of a station wishing to transmit senses the medium. If the medium is free for a specified time (called *distributed inter-frame space* (DIFS)), then the station is able to transmit the packet; if the medium is busy (or becomes busy during the DIFS interval) the station defers using the *exponential backoff algorithm*.

This scheme implies that, except in cases of very high network congestion, no packets will be lost, because retransmission occurs each time a packet is not acknowledged. This entails that all packets sent will reach their destination in sequence.

The IEEE 802.11 MAC layer provides cyclic redundancy check (*CRC*) *checksum* and *packet fragmentation*. Each packet has a CRC checksum calcu-lated and attached to ensure that the data was not corrupted in transmit. Packet fragmentation is used to segment large packets into smaller units when sent over the medium. This is useful in very congested environments or when interference is a factor, since large packets have a better chance of being corrupted. This technique reduces the need for retransmission in many cases and improves overall wireless network performance. The MAC layer is responsible for reassembling fragments received, rendering the process transparent to higher-level protocols. The following are some of the reasons it is preferable to use smaller packets in a WLAN environment.

- DuetohigherBERofaradiolink,theprobabilityofapacketgettingcor- rupted increases with packet size.
- Incaseofpacketcorruption(eitherduetocollisionorinterference),the smaller the packet, the less overhead it needs to retransmit.

A simple *stop-and-wait* algorithm is used at the MAC sublayer. In this mechanismthetransmittingstationisnotallowedtotransmitanewfragmentuntilone of the following happens:

- ReceivesanACKforthefragment,or
- Decides that the fragment was retransmitted too many times and drops the whole frame.

Exponential backoff scheme is used to resolve contention problems among different stations wishing to transmit data at the same time. When a station goes into the backoff state, it waits an additional, randomly selected number of time slots known as a contention window (in 802.11b a slot has a 20μs duration and therandomnumbermustbegreaterthan0andsmallerthanamaximumvalue referred to as a contention window (CW)). During the wait, the station continues sensing the medium to check whether it remains free or if another transmission begins. At the end of its window, if the medium is still free the station can send its frame. If during the window another station begins transmitting data, the backoff counter is frozen and counting down starts again as the channel returns to theidle state.

There is a problem related to the CW dimension. With a small CW, if many stationsattempttotransmitdataatthesametimeitispossiblethatsomeofthem may have the same backoff interval. This means that there will continuously be collisions, with serious effects on network performance. On the other hand, witha large CW, if a few stations wish to transmit data they will likely have long back-offdelaysresultingindegradationofnetworkperformance.Thesolutionisto use an exponentially growing CW size. It starts from a small value (CW_{min}=31) and doubles after each collision, until it reaches the maximum value CW_{max}(CW_{max}= 1023). In 802.11 the backoff algorithm is executed in three cases:

1. Whenthestationsensesthemediumbusybeforethefirsttransmissionofa packet
2. Beforeeachretransmission
3. Afterasuccessfultransmission

This is necessary to avoid a single host wanting to transmit a large quantity of data and occupying the channel for too long, denying access to all other stations.Thebackoffmechanismisnotusedwhenthestationdecidestotransmit

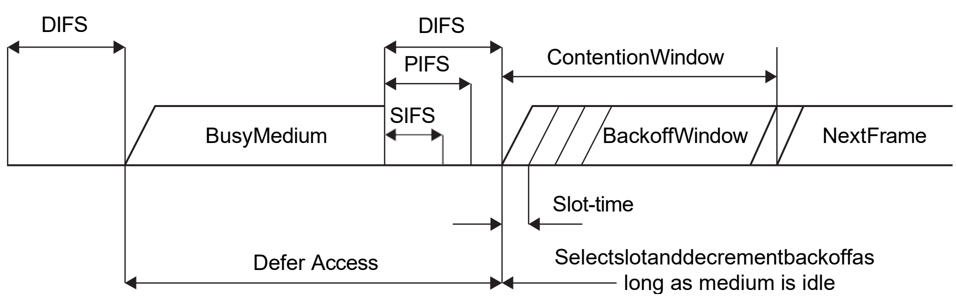


Figure6.11CSMA/CAinIEEE802.11b.

a new packet after an idle period and the medium has been free for more than a distributed inter-frame space (DIFS) (see Figure 6.11).

To support time-bounded services, the IEEE 802.11 standard defines the point coordinate function (PCF) to let stations have priority access to the wireless medium, coordinated by a station called point coordinate (PC). The PCF has higher priority than DIFS, because it may start transmissions after a shorter duration than DIFS; this time space is called PCF interframe space (PIFS), which is 25 μs for IEEE 802.11 and larger than SIFS.

$$\text{The transmission time for a data frame} = \left(\text{PLCP} + \frac{D}{R}\right)\mu\text{s}$$

where:

PLCP = the time required to transmit the physical layer convergence protocol (PLCP)

D = the frame size

R = the channel bit rate

$$\text{CSMA/CA packet transmission time} = \text{BO} + \text{DIFS} + 2\text{PLCP} + \frac{D}{R} + \text{SIFS} + \frac{A}{R}\mu\text{s}$$

where:

A = the ACK frame size

BO = the backoff time

DIFS = the distributed inter-frame space

SIFS = the short inter-frame space

PIFS = the point coordinate interframe space

6.12 Idle Signal Casting Multiple Access

In the CSMA scheme, each terminal must be able to detect the transmissions of all other terminals. However, not all packets transmitted from different terminals can be sensed, or terminals may be hidden from each other by buildings or some other obstacles. This is known as the *hidden terminal problem*, which severely degrades the throughput of the CSMA. The idle signal casting multiple access (ISMA) system transmits an idle/busy signal from the base station to indicate the presence or absence of another terminal's transmission. The ISMA and CSMA are basically the same. In the CSMA, each terminal must listen to all other terminals, whereas in the ISMA, each terminal is informed from the base station of the other terminals' transmission. Similar to CSMAs, there are nonpersistent ISMAs and 1-persistent ISMAs.

6.13 Packet Reservation Multiple Access

Packet reservation multiple access (PRMA) allows a variety of information sources to share the same communication channel and obtains a statistical multiplexing effect. In PRMA, time is divided into frames, each of which consists of a fixed number of time slots. For voice terminals, voice activity detection is adopted. The voice signal comprises a sequence of talk spurts. At the beginning of a talk spurt, the terminal transmits the first packet based on slotted ALOHA. Once the packetis transmitted successfully, that terminal is allowed to use the same time slots in thesucceedingframes(reservationismade).Thereservationiskeptuntiltheend of the talk spurt. The status “reserved” or “unreserved” of each slot is broadcast from the base station.

6.14 ErrorControlSchemesforLinkLayer

Errorcontrolschemesforthelinklayerareusedtoimprovetheperformance of mobile communication systems [8]. Several automatic repeat request (ARQ) schemes are used. At the physical layer of wireless mobile communication sys- tems, error detection and correction techniques such as forward error correction (FEC) schemes are used. For some of the data services, higher layer protocols use ARQ schemes to enable retransmission of any data frames in which an error is detected. The ARQ schemes are classified as follows [23]:

- **Stop and Wait:** The sender transmits the first packet numbered 0 after storing a copy of that packet. The sender then waits for an ACK numbered0, (ACK0) of that packet. If the ACK0 does not arrive before a time-out, the sendertransmitsanothercopyofthefirstpacket.IftheACK0arrivesbefore a time-out, the sender discards the copy of the first packet and is ready to transmit the next packet, which it numbers 1. The sender repeats the previ- ous steps, with numbers 0 and 1 interchanged. The advantages of the Stop and Wait protocol are its simplicity and its small buffer requirements. The sender needs to keep only the copy of the packet that it last transmitted, and the receiver does not need to buffer packets at the data link layer. The main disadvantage of the Stop and Wait protocol is that it does not use the communication link very efficiently.

Thetotaltimetakentotransmitapacketandtopreparefortransmittingthe next one is

$$T = T_p + 2T_{\text{prop}} + 2T_{\text{proc}} + T_a \tag{6.28}$$

where:

T=totaltimefortransmissiontime
T_p=transmissiontimeforapacket
T_{prop}=propagation time of a packet or an ACK
T_{proc}=processing time for a packet or an ACK
T_a=transmission time for an ACK

The protocol efficiency without any error is:

$$\eta(0) = \frac{T_p}{T} \quad (6.29)$$

If p is the probability that a packet or its ACK is corrupted by transmission errors, and a successful transmission of a packet and its ACK takes T seconds and occurs with probability $1 - p$, the protocol efficiency for full duplex (FD) is given as:

$$\eta_{FD} = \frac{(1-p)T_p}{(1-p)T + pT_p} \quad (6.30)$$

- **Selective Repeat Protocol (SRP):** In case of the SRP, only the selected packets are retransmitted. The data link layer in the receiver delivers exactly one copy of every packet in the correct order. The data link layer in the receiver may get the packets in the wrong order from the physical layer. This occurs, for example, when transmission errors corrupt the first packet and not the second one. The second packet arrives correctly at the receiver before the first. The data link layer in the receiver uses a buffer to store the packets that arrive out of order. Once the data link layer in the receiver has a consecutive group of packets in its buffer, it can deliver them to the network layer. The sender also uses a buffer to store copies of the unacknowledged packets. The number of the packets which can be held in the sender/receiver buffer is a design parameter.

Let W = the number of packets which the sender and receiver buffers can each hold and SRP = number of packets in modulo 2W. The protocol efficiency without any error and with a packet error probability of p is given as:

$$\eta(0) = \min\left\{\frac{WT_p}{T}, 1\right\} \quad (6.31)$$

For very large W, the protocol efficiency is

$$\eta(p) = 1 - p \quad (6.32)$$

where:

WT_p = time-out

$$\eta(p) = \frac{2 + p(W-1)}{2 + p(3W-1)} \quad (6.33)$$

SRPisveryefficient,butitrequiresbufferingpacketsatboththesenderand the receiver.

- **Go-Back-N (GBN):** The Go-Back-N protocol allows the sender to have multipleunacknowledgedpacketswithoutthereceiverhavingtostorepackets. This is done by not allowing the receiver to accept packets that are out of order. When a time-out timer expires for a packet, the transmitter resends that packet and all subsequent packets. The Go-Back-N protocol improves ontheefficiencyoftheStopandWaitprotocol,butislessefficientthanSRP. The protocol efficiency for full duplex is given as:

$$\eta_{FD} = \frac{1}{1+\left(\frac{p}{1-p}\right)W} \tag{6.34}$$

- **Window-controlOperationBasedonReceptionMemory(WORM)ARQ:** In digital cellular systems, bursty errors occur by multipath fading, shadowing,andhandoffs.Thetypicalbit-errorratefluctuatesfrom10^{-1}to10^{-6}.Therefore,theconventionalARQschemesdonotoperatewellin a digital cellular system. WORM ARQ has been suggested for control of dynamicerrorcharacteristics.ItisahybridschemethatcombinesSRPGBN protocol.GBNprotocolischoseninthesevereerrorconditionwhereasSRP is selected in the normal error condition.
- **VariableWindowandFrameSizeGBNandSRP[24]:**Sincewirelesssystems have bursty error characteristics, the error control schemes should have a dynamic adaptation to a bursty channel environment. The SRP and GBN with variable window and frame size have been proposed to improve error control in wireless systems. Table 6.2 provides the window and frame size for different BER. If the error rate increases, the window and frame sizeare decreased. In the case of the error rate being small, the window and frame size are increased. The optimum threshold values of BER, window and frame size were obtained through computer simulation.

Table6.2Bit-errorrateversuswindowandsize.

Bit-errorrate(BER)	Windowsize(W)	Framesize(bits)
BER$\leq 10^{-4}$	32	172
$10^{-4}<$BER$<10^{-3}$	8	80
$10^{-3}<$BER$<10^{-2}$	4	40
$10^{-2}<$BER	2	16

Example6.10

We consider a WLAN in which the maximum propagation delay is 4 sec. The WLANoperatesatadatarateof10Mbps.ThedataandACKpacketsareof 400and20bits,respectively.TheprocessingtimeforadataorACKpacketis 1 sec. If the probability p that a data packet or its ACK can be corrupted during transmission is 0.01, find the data link protocol efficiency with (1) Stop and Wait protocol—fullduplex,(2)SRPwithwindowsizeW=8,and(3)Go-Back-N protocolwithwindowsizeW=8.

Solution

$$T_p = \frac{400}{10} = 40\,\mu\text{s}$$

$$T_a = \frac{20}{10} = 2\,\mu\text{s}$$

$$T_{\text{prop}} = 4\,\mu\text{s}$$

$$T_{\text{proc}} = 1\,\mu\text{s}$$

$$T = 40 + 2\times 4 + 2\times 1 + 2 = 52\,\mu\text{s}$$

StopandWait:

$$h = \frac{(1-0.01)\times 40}{(1-0.01)\times 52 + 001\times 40} = 0.763$$

SRP:

$$h = \frac{2+0.01(8-1)}{2+0.01(24-1)} = 0.954$$

GBN:

$$h = \frac{1}{1+8\left(\frac{0.01}{1-0.01}\right)} = 0.925$$

6.15 Summary

Thechapterdescribedtheaccesstechnologiesusedinwirelesscommunications includingreservation-basedmultipleaccessandrandommultipleaccess.FDMA, added CDMA technology to become 3G systems. Cordless telephone systems started with CT0 and CT1, became digital with CT2, and ended in Europe in the fully digital standard**DECT**. This standard has even been chosen as one of the candidates for a 3G system (IMT-FT).

While the number of different systems might be confusing, there are some “natural” development paths. Most network providers offering GSM service todaywilldeployUMTS,whilecdmaOneuserswillchoosecdma2000forsimpler migration. The reasons for this are quite simple. With the introduction of GPRS in GSM networks, the core of the network was already enhanced in a way that it canbedirectlyusedforUMTSwiththeradiotechnologies**UTRAFDD**and**UTRA TDD**. A similar path for evolution exists for**TD-SCDMA**, the Chinese proposal for a 3G system (which has been integrated into UTRA TDD). With some simpli- fication it can be said that UMTS mainly adds a new radio interface but relies in its initial phase on the same core network as GSM/GPRS. Also for cdmaOne the evolution to cdma2000 technologies is quite natural, as the new standardis backwardcompatibleandcanreusefrequencies.Cdma20001xstillusesthesame 1.25 MHz channels as cdmaOne does, but offers data rates of up to 153 kbit/s. The**cdma2000 3x**standard uses three 1.25 MHz channels to fit into ITU’s fre- quency scheme for 3G. However, this standard is not pushed as much as the following enhancements of cdma2000 1x. These enhancements are:

- **cdma2000 1x EV-DO**(evolution-data optimized, also known as high data rate (HDR), some call it data only) promising peak data rates of 2.4 Mbit/s using a second 1.25 MHz channel; and
- **cdma2000 1x EV-DV**(evolution-data and voice) aiming at 1.2 Mbit/s for mobile and 5.2 Mbit/s for stationary users.

Cdma2000 1x EV-DO was the first version of cdma2000 accepted by the ITU as 3G system. More information about the technologies and acronyms used in the diagram is provided in the following sections.

4.1 GSM

GSM is the most successful digital mobile telecommunication system in the world today. It is used by over 800 million people in more than 190 countries.In the early 1980s, Europe had numerous coexisting analog mobile phone sys- tems, which were often based on similar standards (e.g., NMT 450), but ran on slightlydifferentcarrierfrequencies.Toavoidthissituationforasecondgenera- tion fully digital system, the**groupe spéciale mobile (GSM)**was founded in 1982.Thissystemwassoonnamedthe**globalsystemformobilecommunica- tions (GSM)**, with the specification process lying in the hands of ETSI (ETSI, 2002), (GSM Association, 2002). In the context of UMTS and the creation of 3GPP (Third generation partnership project, 3GPP, 2002a) the wholedevelop- ment process of GSM was transferred to 3GPP and further development is combined with 3G development. 3GPP assigned new numbers to all GSM standards. However, to remain consistent with most of the GSM literature, this GSM section stays with the original numbering (see 3GPP, 2002a, for conversion). Section 4.4 will present the ongoing joint specification process in more detail.

The primary goal of GSM was to provide a mobile phone system that allows userstoroamthroughoutEuropeandprovidesvoiceservicescompatibleto ISDN and other PSTN systems. The specification for the initial system alreadycovers more than 5,000 pages; new services, in particular data services, now add even more specification details. Readers familiar with the ISDN reference modelwill recognize many similar acronyms, reference points, and interfaces. GSM standardization aims at adopting as much as possible.

GSM is a typical second generation system, replacing the first generationanalog systems, but not offering the high worldwide data rates that the third gen- eration systems, such as UMTS, are promising. GSM has initially been deployed in Europe using 890–915 MHz for uplinks and 935–960 MHz for downlinks – this system is now also called **GSM 900** to distinguish it from the later versions. These versions comprise GSM at 1800 MHz (1710–1785 MHz uplink, 1805–1880 MHz downlink), also called**DCS (digital cellular system) 1800**, and the GSM system mainly used in the US at 1900 MHz (1850–1910 MHz uplink, 1930–1990 MHz downlink),alsocalled**PCS(personalcommunicationsservice)1900**.Two more versions of GSM exist.**GSM 400**is a proposal to deploy GSM at 450.4–457.6/478.8–486 MHz for uplinks and 460.4–467.6/488.8–496 MHz for downlinks. This system could replace analog systems in sparsely populated areas.

A GSM system that has been introduced in several European countries for railroadsystemsis**GSM-Rail**(GSM-R,2002),(ETSI,2002).Thissystemdoesnot onlyuseseparatefrequenciesbutoffersmanyadditionalserviceswhichare unavailable using the public GSM system. GSM-R offers 19 exclusive channelsforrailroadoperatorsforvoiceanddatatraffic(seesection4.1.3formoreinf or- mationaboutchannels).Specialfeaturesofthissystemare,e.g.,emergencycalls with acknowledgements, voice group call service (VGCS), voice broadcast ser- vice(VBS).Theseso-calledadvancedspeechcallitems(ASCI)resemblefeatures typically available in trunked radio systems only (see section 4.3). Calls are pri- oritized:highprioritycallspre-emptlowprioritycalls.Callshaveveryshort set-up times: emergency calls less than 2 s, group calls less than 5 s. Calls can be directed for example, to all users at a certain location, all users with a certainfunction, or all users within a certain number space. However, the most sophis- ticated use of GSM-R is the control of trains, switches, gates, and signals. Trains goingnotfasterthan160km/hcancontrolallgates,switches,andsignals themselves.Ifthetraingoesfasterthan160km/h(manytrainsarealreadycap- able of going faster than 300 km/h) GSM-R can still be used to maintain control. Thefollowingsectiondescribesthearchitecture,services,andprotocolsof GSM that are common to all three major solutions,**GSM 900**,**GSM 1800**, and**GSM 1900**. GSM has mainly been designed for this and voice services and this stillconstitutesthemainuseofGSMsystems.However,onecanforeseethat manyfutureapplicationsformobilecommunicationswillbedatadriven.The relationshipofdatatovoicetrafficwillshiftmoreandmoretowardsdata.

4.1.1 Mobileservices

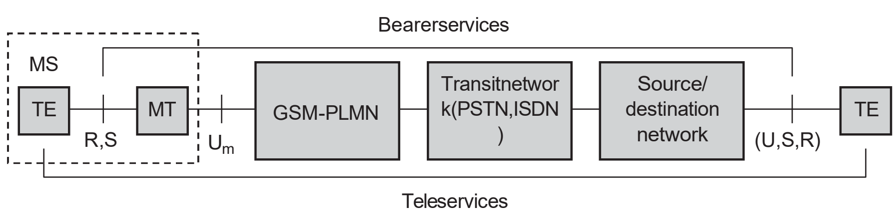
GSM permits the integration of different voice and data services and the inter-working with existing networks. Services make a network interesting forcustomers.GSMhasdefinedthreedifferentcategoriesofservices:bearer,tele, and supplementary services. These are described in the following subsections. Figure 4.3 shows a reference model for GSM services. A**mobile station MS**is connectedtothe**GSMpubliclandmobilenetwork(PLMN)**viatheU$_m$inter- face.(GSM-PLMNistheinfrastructureneededfortheGSMnetwork.)This network is connected to transit networks, e.g.,**integrated services digital net- work (ISDN)**or traditional**public switched telephone network (PSTN)**. There might be an additional network, the source/destination network, before another **terminal TE**is connected.**Bearer services**now comprise all services that enable the transparent transmission of data between the interfaces to the network, i.e., Sincaseofthemobilestation,andasimilarinterfacefortheotherterminal(e.g., S_0 for ISDN terminals). Interfaces like U, S, and R in case of ISDN have not been defined for all networks, so it depends on the specific network which interface is used as a reference for the transparent transmission of data. In the classical GSM model, bearer services are connection-oriented and circuit- or packet-switched. These services only need the lower three layers of the ISO/OSI reference model.

Within the mobile station MS, the**mobile termination (MT)**performs all networkspecifictasks(TDMA,FDMA,codingetc.)andoffersaninterfacefor datatransmission(S)totheterminalTEwhichcanthenbenetworkindepen-dent.DependingonthecapabilitiesofTE,furtherinterfacesmaybeneeded, suchasR,accordingtotheISDNreferencemodel(Halsall,1996).**Teleservices** are application specific and may thus need all seven layers of the ISO/OSI refer-encemodel.Theseservicesarespecifiedend-to-end,i.e.,fromoneterminalTE to another.

4.1.1.1 Bearer services

GSM specifies different mechanisms for data transmission, the original GSM allowingfordataratesofupto9600bit/sfornon-voiceservices.Bearerservices permit transparent and non-transparent, synchronous or asynchronous data transmission.**Transparentbearerservices**onlyusethefunctionsofthephysi- cal layer (layer 1) to transmit data. Data transmission has a constant delay and throughputifnotransmissionerrorsoccur.Theonlymechanismtoincrease

Figure4.3
Bearer and tele servicesreference model



transmission quality is the use of**forward error correction (FEC)**, which codes redundancyintothedatastreamandhelpstoreconstructtheoriginaldatain caseoftransmissionerrors.DependingontheFEC,dataratesof2.4,4.8,or 9.6 kbit/s are possible. Transparent bearer services do not try to recover lost dataincaseof,forexample,shadowingorinterruptionsduetohandover.

Non-transparent bearer servicesuse protocols of layers two and three to implement error correction and flow control. These services use the transparent bearer services, adding a**radio link protocol (RLP)**. This protocol comprises mechanismsof**high-leveldatalinkcontrol(HDLC)**,(Halsall,1996)andspe- cial selective-reject mechanisms to trigger retransmission of erroneous data. The achieved bit error rate is less than 10^{-7}, but now throughput and delay may vary depending on transmission quality.

Usingtransparentandnon-transparentservices,GSMspecifiesseveral bearerservicesforinterworkingwithPSTN,ISDN,andpacketswitchedpublic data networks (PSPDN) like X.25, which is available worldwide. Data transmis-sioncanbefull-duplex,synchronouswithdataratesof1.2,2.4,4.8,and 9.6 kbit/s or full-duplex, asynchronous from 300 to 9,600 bit/s (ETSI, 1991a). Clearly, these relatively low data rates reflect the assumption that data services willonlyconstitutesomesmallpercentageoftheoveralltraffic.Whilethis is still true of GSM networks today, the relation of data and voice services is changing,withdatabecomingmoreandmoreimportant.Thisdevelopment is also reflected in the new data services (see section 4.1.8).

4.1.1.2 Teleservices

GSM mainly focuses on voice-oriented tele services. These comprise encrypted voicetransmission,messageservices,andbasicdatacommunicationwithtermi-nalsasknownfromthePSTNorISDN(e.g.,fax).However,asthemainserviceis **telephony**, the primary goal of GSM was the provision of high-quality digital voice transmission, offering at least the typical bandwidth of 3.1 kHz of analog phonesystems.Specialcodecs(coder/decoder)areusedforvoicetransmission, while other codecs are used for the transmission of analog data for communica-tion with traditional computer modems used in, e.g., fax machines.

Another service offered by GSM is the**emergency number**. The same number can be used throughout Europe. This service is mandatory for all providers and free of charge. This connection also has the highest priority, possibly pre-empting other connections, and will automatically be set up with the closest emergency center.

Ausefulserviceforverysimplemessagetransferisthe**shortmessageser- vice (SMS)**, which offers transmission of messages of up to 160 characters. SMS messages do not use the standard data channels of GSM but exploit unused capacity in the signalling channels (see section 4.1.3.1). Sending and receivingof SMS is possible during data or voice transmission. SMS was in the GSM stan-dardfromthebeginning;however,almostnooneuseditunntilmillionsof youngpeoplediscoveredthisserviceinthemid-ninetiesasafunservice.SMS can be used for “serious” applications such as displaying road conditions, e-mail headersorstockquotes,butitcanalsotransferrlogos,ringtones,horoscopes and love letters. Today more than 30 billion short messages are transferred worldwide per month! SMS is big business today, not only for the network operators,butalsoformanycontentproviders.ItshouldbenotedthatSMSis typicallytheonlywaytoreachamobilephonefromwithinthenetwork.Thus, SMS is used for updating mobile phone software or for implementing so-calledpush services (see chapter 10).

The successor of SMS, the**enhanced message service (EMS)**, offers a larger messagesize(e.g.,760characters,concatenatingseveralSMs),formattedtext, and the transmission of animated pictures, small images and ring tones in a standardizedway(somevendorsofferedsimilarproprietaryfeaturesbefore). EMS never really took off as the **multimedia message service (MMS)** was available.(NokianeverlikedEMSbutpushedSmartMessaging,aproprietary system.) MMS offers the transmission of larger pictures (GIF, JPG, WBMP), short videoclipsetc.andcomeswithmobilephonesthatintegratesmallcameras. MMSisfurtherdiscussedinthecontextofWAPinchapter10.

Anothernon-voiceteleserviceis**group3fax**,whichisavailableworldwide. In this service, fax data is transmitted as digital data over the analog telephone networkaccordingtotheITU-TstandardsT.4andT.30usingmodems.Typically, atransparentfaxserviceisused,i.e.,faxdataandfaxsignalingistransmitted using a transparent bearer service. Lower transmission quality causes an automaticadaptationofthebearerservicetolowerdataratesandhigher redundancy for better FEC.

4.1.1.3 Supplementaryservices

Inadditiontoteleandbearerservices,GSMproviderscanoffer**supplementary services.** Similar to ISDN networks, these services offer various enhancements for the standard telephony service, and may vary from provider to provider. Typical services are user**identification**, call**redirection**, or**forwarding**of ongoing calls. Standard ISDN features such as**closed user groups**and**multiparty**communication may be available. Closed user groups are of special interest to companies because they allow, for example, a company-specific GSM sub-network, to which only members of the group have access.

4.1.2 Systemarchitecture

As with all systems in the telecommunication area, GSM comes with a hierarchical, complex system architecture comprising many entities, interfaces,and acronyms. Figure 4.4 gives a simplified overview of the GSM system as specified in ETSI (1991b). A GSM system consists of three subsystems, the**radio sub system(RSS)**,the**networkandswitchingsubsystem(NSS)**,andthe**operation subsystem(OSS)**.Eachsubsystemwillbediscussedinmoredetailinthefollowingsections.Generally,aGSMcustomeronlynoticesaverysmallfractionof the whole network – the mobile stations (MS) and some antenna masts of the base transceiver stations (BTS).

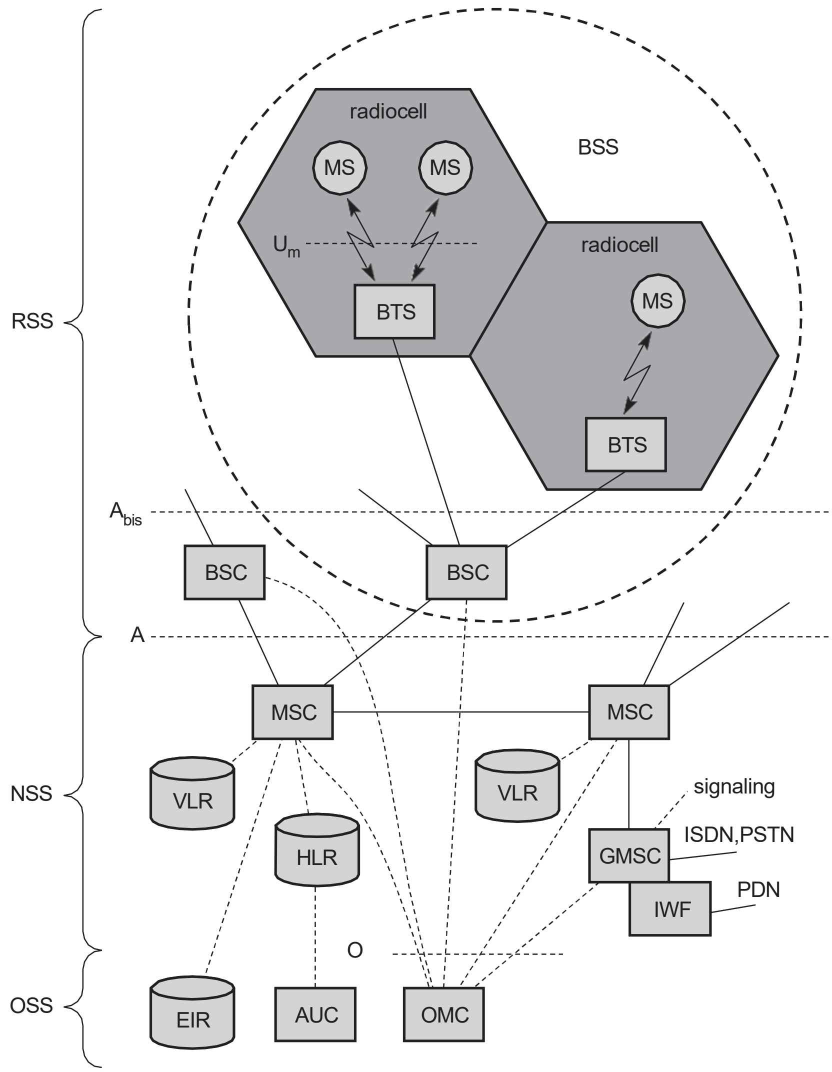


Figure4.4
Functional architecture of a GSM system

4.1.2.1 Radiosubsystem

As the name implies, the**radio subsystem (RSS)**comprises all radio specific entities, i.e., the**mobile stations (MS)**and the**base station subsystem (BSS)**. Figure 4.4 shows the connection between the RSS and the NSS via the**A interface**(solid lines) and the connection to the OSS via the**O interface**(dashed lines). The A interface is typically based on circuit-switched PCM-30 systems (2.048Mbit/s),carryingupto3064kbit/sconnections,whereastheOinterface usestheSignallingSystemNo.7(SS7)basedonX.25carryingmanagementdata to/from the RSS.

- **Basestationsubsystem(BSS)**:AGSMnetworkcomprisesmanyBSSs,each controlledbyabasestationcontroller(BSC).TheBSSperformsallfunctions necessary to maintain radio connections to an MS, coding/decoding ofvoice,andrateadaptationto/fromthewirelessnetworkpart.BesidesaBSC, the BSS contains several BTSs.
- **Base transceiver station (BTS)**: A BTS comprises all radio equipment, i.e., antennas, signal processing, amplifiers necessary for radio transmission. A BTS can form a radio cell or, using sectorized antennas, several cells (see section 2.8), and is connected to MS via the **$U_m$interface** (ISDN U interface for mobileuse),andtotheBSCviathe**A_{bis}interface**.The$U_m$interfacecontains allthemechanismsnecessaryforwirelesstransmission(TDMA,FDMAetc.) andwillbediscussedinmoredetailbelow.TheA_{bis}interfaceconsistsof16or 64 kbit/s connections. A GSM cell can measure between some 100 m and35 km depending on the environment (buildings, open space, mountains etc.) but also expected traffic.
- **Base station controller (BSC)**: The BSC basically manages the BTSs. It reserves radio frequencies, handles the handover from one BTS to another withintheBSS,andperformspagingoftheMS.TheBSCalsomultiplexes theradiochannelsontothefixednetworkconnectionsattheAinterface.

Table4.1givesanoverviewofthetasksassignedtotheBSCandBTSoroftasks in which these entities support other entities in the network.

- **Mobile station (MS)**: The MS comprises all user equipment and software needed for communication with a GSM network. An MS consists of user independent hard- and software and of the**subscriber identity module (SIM)**, which stores all user-specific data that is relevant to GSM.[3]While an MS can be identified via the**international mobile equipment identity (IMEI)**,ausercanpersonalizeanyMSusinghisorherSIM,i.e.,user-specific mechanisms like charging and authentication are based on the SIM, not on the device itself. Device-specific mechanisms, e.g., theft protection, use the devicespecificIMEI.WithouttheSIM,onlyemergencycallsarepossible.The SIM card contains many identifiers and tables, such as card-type, serial number, a list of subscribed services, a**personal identity number (PIN)**, a **PIN unblocking key (PUK)**, an**authentication key K_i**, and the**inter- national mobile subscriber identity (IMSI)** (ETSI, 1991c). The PIN is used to unlock the MS. Using the wrong PIN three times will lock the SIM. In such cases,thePUKisneededtounlocktheSIM.TheMSstoresdynamicinforma- tion while logged onto the GSM system, such as, e.g., the **cipher key K_c**and thelocationinformationconsistingofa**temporarymobilesubscriberiden- tity(TMSI)**andthe**locationareaidentification(LAI)**.TypicalMSsforGSM 900 have a transmit power of up to 2 W, whereas for GSM 1800 1 W is enoughduetothesmallercellsize.Apartfromthetelephoneinterface,an

3Manyadditionalitemscanbestoredonthemobiledevice.However,thisisirrelevanttoGSM.

Function	BTS	BSC
Managementofradiochannels		X
Frequencyhopping	X	X
Managementofterrestrialchannels		X
Mappingofterrestrialontoradiochannels		X
Channelcodinganddecoding	X	
Rateadaptation	X	
Encryptionanddecryption	X	X
Paging	X	X
Uplink signalmeasurement	X	
Trafficmeasurement		X
Authentication		X
Locationregistry,locationupdate		X
Handovermanagement		X

Table4.1 Tasksof the BTS and BSC within a BSS

MS can also offer other types of interfaces to users with display, loudspeaker, microphone, and programmable soft keys. Further interfaces comprise computermodems,IrDA,orBluetooth.TypicalMSs,e.g.,mobilephones, comprise many more vendor-specific functions and components, such as cameras, fingerprint sensors, calendars, address books, games, and Internet browsers. Personal digital assistants (PDA) with mobile phone functions are also available. The reader should be aware that an MS could also be inte-grated into a car or be used for location tracking of a container.

4.1.2.2 Networkandswitchingsubsystem

The "heart" of the GSM system is formed by the**network and switching subsystem (NSS)**. The NSS connects the wireless network with standard public networks, performs handovers between different BSSs, comprises functions for worldwidelocalizationofusersandsupportscharging,accounting,androaming of users between different providers in different countries. The NSS consistsof the following switches and databases:

- **Mobile services switching center (MSC)**: MSCs are high-performance digitalISDNswitches.TheysetupconnectionstootherMSCsandtotheBSCs via the A interface, and form the fixed backbone network of a GSMsystem. Typically, an MSC manages several BSCs in a geographical region. A **gateway MSC(GMSC)**has additional connections to other fixed networks, such as**PSTN**and**ISDN**.Usingadditional**interworkingfunctions(IWF)**,anMSC can also connect to**public data networks (PDN)**such as X.25. An MSChandles all signaling needed for connection setup, connection release and handoverofconnectionstootherMSCs.The**standardsignalingsystem No. 7 (SS7)**is used for this purpose. SS7 covers all aspects of control signal-ing for digital networks (reliable routing and delivery of control messages, establishing and monitoring of calls). Features of SS7 are number portability, free phone/toll/collect/credit calls, call forwarding, three-way calling etc. An MSC also performs all functions needed for supplementary services such ascall forwarding, multi-party calls, reverse charging etc.

- **Home location register (HLR)**: The HLR is the most important database in a GSM system as it stores all user-relevant information. This comprises static information, such as the**mobile subscriber ISDN number (MSISDN)**, sub-scribed services (e.g., call forwarding, roaming restrictions, GPRS), and the **international mobile subscriber identity (IMSI)**. Dynamic information isalso needed, e.g., the current**location area (LA)**of the MS, the**mobile sub- scriber roaming number (MSRN)**, the current VLR and MSC. As soon as anMS leaves its current LA, the information in the HLR is updated. This informa- tion is necessary to localize a user in the worldwide GSM network. All these user-specific information elements only exist once for each user in a singleHLR, which also supports charging and accounting. The parameters will be explained in more detail in section 4.1.5. HLRs can manage data for several million customers and contain highly specialized data bases which must fulfill certainreal-timerequirementstoanswerrequestswithincertaintime-bounds.
- **Visitor location register (VLR)**: The VLR associated to each MSC is a dynamic database which stores all important information needed for theMS users currently in the LA that is associated to the MSC (e.g., IMSI, MSISDN, HLR address). If a new MS comes into an LA the VLR is respons- ible for, it copies all relevant information for this user from the HLR. This hierarchy of VLR and HLR avoids frequent HLR updates and long-distance signaling of user information. The typical use of HLR and VLR for user localization will be described in section 4.1.5. Some VLRs in existence, are capable of managing up to one million customers.

4.1.2.3 Operationsubsystem

The third part of a GSM system, the**operation subsystem (OSS)**, contains the necessaryfunctionsfornetworkoperationandmaintenance.TheOSSpossesses network entities of its own and accesses other entities via SS7 signaling (see Figure 4.4). The following entities have been defined:

- **Operation and maintenance center (OMC)**: The OMC monitors and con-trols all other network entities via the O interface (SS7 with X.25). Typical OMC management functions are traffic monitoring, status reports of net-work entities, subscriber and security management, or accounting and billing. OMCs use the concept of**telecommunication management net-work (TMN)** as standardized by the ITU-T.

- **Authentication centre (AuC)**: As the radio interface and mobile stations areparticularlyvulnerable,aseparateAuChasbeendefinedtoprotectuser identity and data transmission. The AuC contains the algorithms for authentication as well as the keys for encryption and generates the values needed for user authentication in the HLR. The AuC may, in fact, be situated in a special protected part of the HLR.
- **Equipment identity register (EIR)**:TheEIRisadatabaseforallIMEIs,i.e., it stores all device identifications registered for this network. As MSs are mobile, they can be easily stolen. With a valid SIM, anyone could use the stolen MS. The EIR has a blacklist of stolen (or locked) devices. In theory an MS is useless as soon as the owner has reported a theft. Unfortunately, the blacklistsofdifferentprovidersarenotusuallysynchronizedandtheillegal use of a device in another operator's network is possible (the reader may speculate as to why this is the case). The EIR also contains a list of valid IMEIs (white list), and a list of malfunctioning devices (gray list).

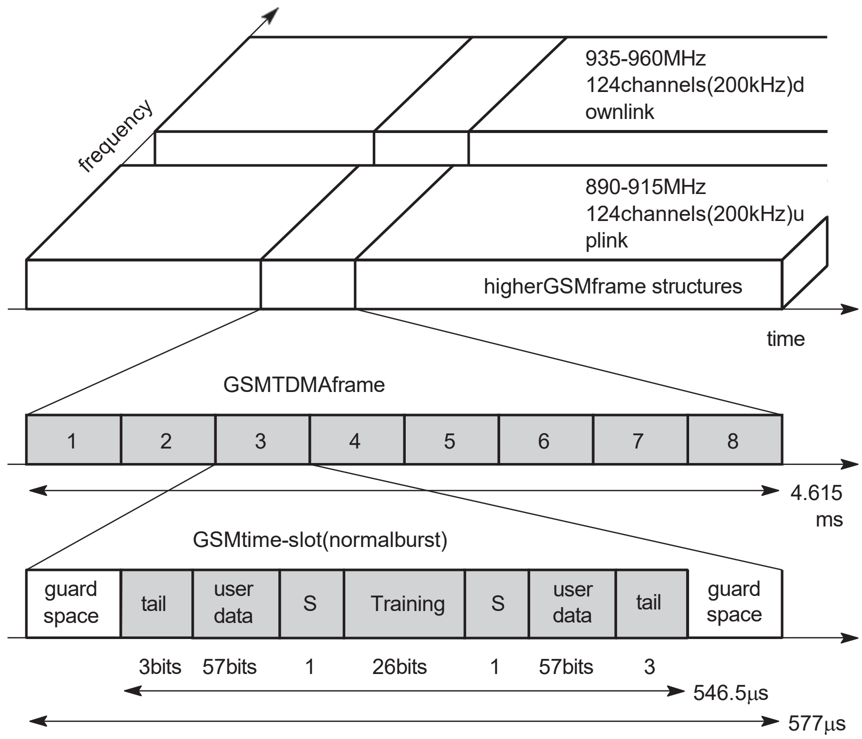
4.1.3 Radiointerface

The most interesting interface in a GSM system is U_m, the radio interface, as it comprises many mechanisms presented in chapters 2 and 3 for multiplexingand media access. GSM implements SDMA using cells with BTS and assigns an MS to a BTS. Furthermore, FDD is used to separate downlink and uplink as shown in Figures 3.3 and 4.5. Media access combines TDMA and FDMA. In GSM 900, 124 channels, each 200 kHz wide, are used for FDMA, whereas GSM 1800 uses, 374 channels. Due to technical reasons, channels 1 and 124 are not used for transmission in GSM 900. Typically, 32 channels are reserved for organizational data; the remaining 90 are used for customers. Each BTS then manages a single channel for organizational data and, e.g., up to 10 channels for user data. The following example is based on the GSM 900 system, but GSM works in a similar way at 1800 and 1900 MHz.

WhileFigure3.3inchapter3hasalreadyshowntheFDMinGSM,Figure 4.5alsoshowstheTDMused.Eachofthe248channelsisadditionallyseparated in time via a**GSM TDMA frame**, i.e., each 200 kHz carrier is subdivided into frames that are repeated continuously. The duration of a frame is 4.615 ms. A frame is again subdivided into 8**GSM time slots**, where each slot represents a physicalTDMchannelandlastsfor577μs.EachTDMchanneloccupiesthe 200 kHz carrier for 577 μs every 4.615 ms.

Data is transmitted in small portions, called**bursts**. Figure 4.5 shows a so-called**normal burst**as used for data transmission inside a time slot (user and signaling data). In the diagram, the burst is only 546.5 μs long and contains 148bits. The remaining 30.5μs are used as**guard space**to avoid overlapping with otherburstsduetodifferentpathdelaysandtogivethetransmittertimetoturn onandoff.Fillingthewholeslotwithdataallowsforthetransmissionof

Figure4.5
GSMTDMAframe, slots, and bursts



156.25 bit within 577μs. Each physical TDM channel has a raw data rate of about33.8kbit/s,eachradiocarriertransmitsapproximately270kbit/soverthe $U_m$interface.

The first and last three bits of a normal burst (**tail**) are all set to 0 and can be usedtoenhancethereceiverperformance.The**training**sequenceinthemiddle of a slot is used to adapt the parameters of the receiver to the current path propa- gation characteristics and to select the strongest signal in case ofmulti-path propagation. A flag**S**indicates whether the**data**field contains user or network control data. Apart from the normal burst, ETSI (1993a) defines four more bursts for data transmission: a **frequency correction** burst allows the MS to correct the local oscillator to avoid interference with neighboring channels, a**synchroniza- tionburst**withanextendedtrainingsequencesynchronizestheMSwiththe BTS in time, an **access burst** is used for the initial connection setup between MSand BTS, and finally a **dummy burst** is used if no data is available for a slot.

Twofactorsallowfortheuseofsimpletransmitterhardware:ontheone hand,theslotsforuplinkanddownlinkofaphysicalTDMchannelaresepa- ratedinfrequency(45MHzforGSM900,95MHzforGSM1800usingFDD). Ontheotherhand,theTDMAframesareshiftedintimeforthreeslots,i.e.,if theBTSsendsdataattime$t_0$inslotoneonthedownlink,theMSaccessesslot one on the uplink at time $t_0+3\cdot577$ μs. An MS does not need a full-duplex transmitter, a simpler half-duplex transmitter switching between receiving and sending is enough.

To avoid frequency selective fading, GSM specifies an optional**slow frequency hopping**mechanism. MS and BTS may change the carrier frequency after each frame based on a common hopping sequence. An MS changes its frequency between up and downlink slots respectively.

4.1.3.1 Logicalchannelsandframehierarchy

While the previous section showed the physical separation of the medium into 8*124duplexchannels,thissectionpresentslogicalchannelsandahierarchyof frames based on the combination of these physical channels. A physical channelconsistsofaslot,repeatedevery4.615ms.ThinkofalogicalchannelC$_1$ that only takes up every fourth slot and another logical channel C_2 that uses every other slot. Both logical channels could use the same physical channelwith the pattern $C_1C_2xC_2C_1C_2xC_2C_1$etc. (The x indicates that the physicalchannel still has some capacity left.)

GSM specifies two basic groups of logical channels, i.e., traffic channels and control channels:[4]

- **Traffic channels (TCH)**: GSM uses a TCH to transmit user data (e.g., voice, fax). Two basic categories of TCHs have been defined, i.e.,**full-rate TCH (TCH/F)**and**half-rateTCH(TCH/H)**.ATCH/Fhasadatarateof22.8kbit/s, whereas TCH/H only has 11.4 kbit/s. With the voice codecs available at the beginningoftheGSMstandardization,13kbit/swererequired,whereasthe remainingcapacityoftheTCH/F(22.8kbit/s)wasusedforerrorcorrection (**TCH/FS**). Improved codes allow for better voice coding and can use a TCH/H. Using these TCH/HSs doubles the capacity of the GSM system for voice transmission. However, speech quality decreases with the use of TCH/HS and many providers try to avoid using them. The standard codecs forvoicearecalled**fullrate**(FR,13kbit/s)and**halfrate**(HR,5.6kbit/s).A newercodec,**enhancedfullrate**(EFR),providesbettervoicequalitythanFR aslongasthetransmissionerrorrateislow.Thegenerateddatarateisonly 12.2kbit/s.Newcodecs,whichautomaticallychoosethebestmodeofopera- tion depending on the error rate (AMR, adaptive multi-rate), will be used togetherwith3Gsystems.Anadditionalincreaseinvoicequalityisprovided bytheso-called**tandemfreeoperation(TFO)**.Thismodecanbeusediftwo MSs exchange voice data. In this case, coding to and from PCM encoded voice (standard in ISDN) can be skipped and the GSM encoded voice data is directly exchanged. Data transmission in GSM is possible at many different data rates, e.g.,**TCH/F4.8**for 4.8 kbit/s,**TCH/F9.6**for 9.6 kbit/s, and, as a newerspecification,**TCH/F14.4**for14.4kbit/s.Theselogicalchannelsdiffer intermsoftheircodingschemesanderrorcorrectioncapabilities.

4Moreinformationabout channelscanbefoundin Goodman(1997)andETSI(1993a).

- **Controlchannels(CCH)**:ManydifferentCCHsareusedinaGSMsystem to control medium access, allocation of traffic channels or mobility management. Three groups of control channels have been defined, each again with subchannels (maybe you can imagine why the initial specification already needed over 5,000 pages):
  - **Broadcast control channel (BCCH)**: A BTS uses this channel to signal information to all MSs within a cell. Information transmitted in thischannel is, e.g., the cell identifier, options available within this cell (fre- quency hopping), and frequencies available inside the cell and in neighboringcells.TheBTSsendsinformationforfrequencycorrection via the**frequency correction channel (FCCH)**and information abouttime synchronization via the**synchronization channel (SCH)**, wherebothchannelsaresubchannelsoftheBCCH.
  - **Common control channel (CCCH)**: All information regarding connection setup between MS and BS is exchanged via the CCCH. Forcalls toward an MS, the BTS uses the**paging channel (PCH)**for paging the appropriate MS. If an MS wants to set up a call, it uses the**random accesschannel(RACH)**tosenddatatotheBTS.TheRACHimplements multipleaccess(allMSswithinacellmayaccessthischannel)usingslot- ted Aloha. This is where a collision may occur with other MSs in a GSM system.TheBTSusesthe**accessgrantchannel(AGCH)**tosignalanMS that it can use a TCH or SDCCH for further connection setup.
  - **Dedicated control channel (DCCH)**: While the previous channelshave all been unidirectional, the following channels are bidirectional.As long as an MS has not established a TCH with the BTS, it uses the **stand-alonededicatedcontrolchannel(SDCCH)**withalowdatarate (782bit/s)forsignaling.Thiscancompriseauthentication,registration or other data needed for setting up a TCH. Each TCH and SDCCH has a **slow associated dedicated control channel (SACCH)**associated with it, which is used to exchange system information, such as the channel quality and signal power level. Finally, if more signaling information needstobetransmittedandaTCHalreadyexists,GSMusesa**fastasso-ciated dedicated control channel (FACCH).** The FACCH uses the time slotswhichareotherwiseusedbytheTCH.Thisisnecessaryinthecase of handovers where BTS and MS have to exchange larger amounts of data in less time.

However, these channels cannot use time slots arbitrarily – GSM specifies a veryelaboratemultiplexingschemethatintegratesseveralhierarchiesofframes. If we take a simple TCH/F for user data transmission, each TCH/F will have an associated SACCH for slow signaling. If fast signaling is required, the FACCHuses the time slots for the TCH/F. A typical usage pattern of a physical channel for data transmission now looks like this (with T indicating the user traffic inthe TCH/F and S indicating the signalling traffic in the SACCH):

TTTTTTTTTTTTSTTTTTTTTTTTxTTTTTTTTTTTSTTTTTTTTTTT

x

Twelve slots with user data are followed by a signalling slot. Again 12 slots with user data follow, then an unused slot. This pattern of 26 slots is repeated over and over again. In this case, only 24 out of 26 physical slots are used for the TCH/F.Now recall that each normal burst used for data transmission carries 114 bit user data and is repeated every 4.615 ms. This results in a data rate of 24.7 kbit/s. Asthe TCH/F only uses 24/26 of the slots, the final data rate is 22.8 kbit/s as speci- fied for the TCH/F. The SACCH thus has a capacity of 950 bit/s.

This periodic pattern of 26 slots occurs in all TDMA frames with a TCH. The combinationofthese framesiscalled**trafficmultiframe**.Figure4.6showsthelogi- cal combination of 26 frames (TDMA frames with a duration of 4.615 ms) to a multiframe with a duration of 120 ms. This type of multiframe is used for TCHs, SACCHs for TCHs, and FACCHs. As these logical channels are all associated withusertraffic,themultiframeiscalledtrafficmultiframe.TDMAframescontaining(sig- naling) data for the other logical channels are combined to a **control multiframe**. Control multiframes consist of 51 TDMA frames and have a duration of 235.4 ms.

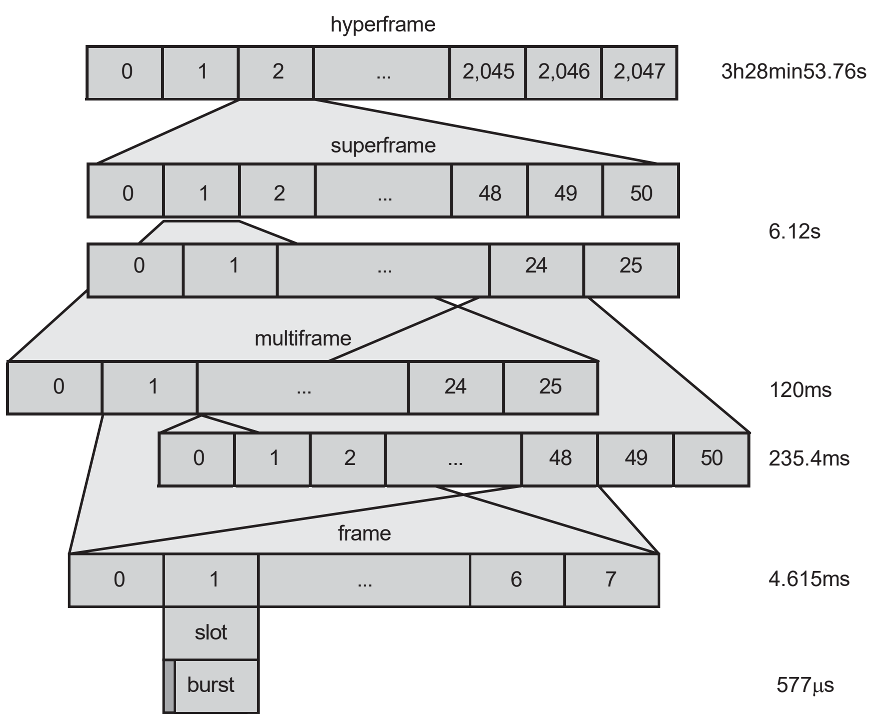


Figure4.6 GSMstructuringof timeusingaframe hierarchy

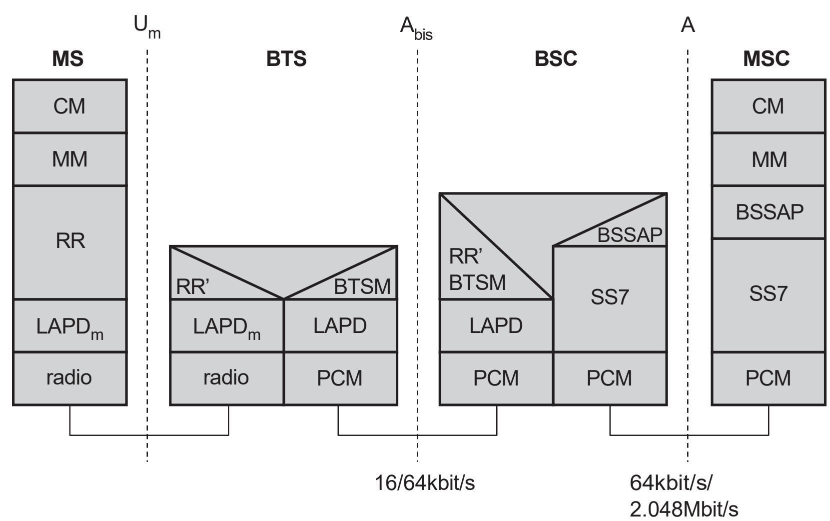
This logical frame hierarchy continues, combining 26 multiframes with 51 frames or 51 multiframes with 26 frames to form a**superframe**. 2,048 superframes build a**hyperframe**with a duration of almost 3.5 hours. Altogether, 2,715,648 TDMA frames form a hyperframe. This large logical structure is needed for encryption – GSM counts each TDMA frame, with the frame number forming input for the encryption algorithm. The frame number plus the slot number uniquely identify each time slot in GSM.

4.1.4 Protocols

Figure 4.7 shows the protocol architecture of GSM with signaling protocols, interfaces,aswellastheentitiesalreadyshowninFigure4.4.Themaininterest lies in the $U_m$interface, as the other interfaces occur between entities in a fixed network.**Layer 1**, the physical layer, handles all**radio**-specific functions. This includes the creation of bursts according to the five different formats,**multiplexing**ofburstsintoaTDMAframe,**synchronization**withtheBTS,detection ofidlechannels,andmeasurementofthe**channelqualit**yonthedownlink. The physical layer at $U_m$uses GMSK for digital**modulation**and performs **encryption/decryption**ofdata,i.e.,encryptionisnotperformedend-to-end, but only between MS and BSS over the air interface.

Synchronization also includes the correction of the individual path delay betweenanMSandtheBTS.AllMSswithinacellusethesameBTSandthusmust besynchronizedtothisBTS.TheBTSgeneratesthetime-structureofframes,slots etc.Aproblematicaspectinthiscontextarethedifferentroundtriptimes(RTT). AnMSclosetotheBTShasaveryshortRTT,whereasanMS35kmawayalready exhibitsanRTTofaround0.23ms.IftheMSfarawayusedtheslotstructurewith-

Figure4.7
Protocolarchitecture forsignaling



out correction, large guard spaces would be required, as 0.23 ms are already 40 per centofthe0.577msavailableforeachslot.Therefore,theBTSsendsthecurrent RTTtotheMS,whichthenadjustsitsaccesstimesothatallburstsreachtheBTS withintheirlimits.Thismechanismreducestheguardspacetoonly30.5μsorfive percent(seeFigure4.5).Adjustingtheaccessiscontrolledviathevariable**timing advance**, where a burst can be shifted up to 63 bit times earlier, with each bit havingadurationof3.69μs(whichresultsinthe0.23msneeded).Asthevariable timing advance cannot be extended a burst cannot be shifted earlier than 63 bit times. This results in the 35 km maximum distance between an MS and a BTS. It mightbepossibletoreceivethesignalsoverlongerdistances;toavoidcollisionsat the BTS, access cannot be allowed.[5]

The main tasks of the physical layer comprise**channel coding**and**error detection/correction**, which is directly combined with the coding mechanisms. Channelcodingmakesextensiveuseofdifferent**forwarderrorcorrection (FEC)**schemes.FECaddsredundancytouserdata,allowingforthedetection and correction of selected errors. The power of an FEC scheme depends on the amount of redundancy, coding algorithm and further interleaving of data to minimize the effects of burst errors. The FEC is also the reason why error detec-tion and correction occurs in layer one and not in layer two as in the ISO/OSI reference model. The GSM physical layer tries to correct errors, but it does not deliver erroneous data to the higher layer.

DifferentlogicalchannelsofGSMusedifferentcodingschemeswithdiffer-entcorrectioncapabilities.Speechchannelsneedadditionalcodingofvoice dataafteranalogtodigitalconversion,toachieveadatarateof22.8kbit/s (using the 13 kbit/s from the voice codec plus redundancy, CRC bits, and inter-leaving(Goodman,1997).AsvoicewasassumedtobethemainserviceinGSM, the physical layer also contains special functions, such as**voice activity detec-tion (VAD)**, which transmits voice data only when there is a voice signal. This mechanism helps to decrease interference as a channel might be silent approxi-mately60percentofthetime(undertheassumptionthatonlyoneperson speaks at the same time and some extra time is needed to switch between the speakers). During periods of silence (e.g., if a user needs time to think before talking), the physical layer generates a**comfort noise**to fake a connection (completesilencewouldprobablyconfuseauser),butnoactualtransmission takes place. The noise is even adapted to the current background noise at the communication partner's location.

All this interleaving of data for a channel to minimize interference due to bursterrorsandtherecurrencepatternofalogicalchannelgeneratesa**delay**for transmission.Thedelayisabout60msforaTCH/FSand100msforaTCH/F9.6

5 Aspecialtrickallowsforlargercells.IfthetimingadvanceforMSsthatarefurtherawaythan35kmissettozero,theb urstsarrivingfromtheseMSswillfallintothefollowingtimeslot.Receptionofdataissimplyshiftedonetimeslotan dagainthetimingadvancemaybeuseduptoadistanceof70 km (under simplified assumptions). Using this special trick, the capacity of a cell is decreased (nearand far MSs cannot be mixed arbitrarily), but coverage of GSM is extended. Network operators maychoose this approach, e.g., in coastal regions.

(within 100 ms signals in fixed networks easily travel around the globe). These times have to be added to the transmission delay if communicating with an MS insteadofastandardfixedstation(telephone,computeretc.)andmayinfluence the performance of any higher layer protocols, e.g., for computer data transmission (see chapter 9).

Signaling between entities in a GSM network requires higher layers(see Figure4.7).Forthispurpose,the**LAPD$_m$**protocolhasbeendefinedattheU$_m$ interfacefor**layertwo**.LAPD$_m$,asthenamealreadyimplies,hasbeenderived fromlinkaccessprocedurefortheD-channel(**LAPD**)inISDNsystems,whichisa versionofHDLC(Goodman,1997),(Halsall,1996).LAPD$_m$isalightweightLAPD becauseitdoesnotneedsynchronizationflagsorchecksummingforerrordetection. (The GSM physical layer already performs these tasks.) LAPD$_m$offers reliable data transfer over connections, re-sequencing of data frames, and flow control (ETSI, 1993b), (ETSI, 1993c). As there is no buffering between layer one and two, LAPD$_m$has to obey the frame structures, recurrence patterns etc. defined for theU$_m$interface. Further services provided by LAPD$_m$include segmentation and reassembly of data and acknowledged/unacknowledged data transfer.

ThenetworklayerinGSM,**layerthree**,comprisesseveralsublayersas Figure 4.7 shows. The lowest sublayer is the**radio resource management (RR)**. Onlyapartofthislayer,**RR**',isimplementedintheBTS,theremainderissitu- ated in the BSC. The functions of RR' are supported by the BSC via the**BTS management(BTSM)**.ThemaintasksofRRaresetup,maintenance,and release of radio channels. RR also directly accesses the physical layer for radio informationandoffersareliableconnectiontothenexthigherlayer.

Mobility management (MM) contains functions for registration, authentication,identification,locationupdating,andtheprovisionofa**temporary mobile subscriber identity (TMSI)** that replaces the **international mobile subscriber identity(IMSI)**andwhichhidestherealidentityofanMSuserovertheairinter- face. While the IMSI identifies a user, the TMSI is valid only in the current location area of a VLR. MM offers a reliable connection to the next higher layer.

Finally,the**callmanagement(CM)**layercontainsthreeentities:**callcon- trol (CC)**,**short message service (SMS)**, and**supplementary service (SS)**. SMS allowsformessagetransferusingthecontrolchannelsSDCCHandSACCH(if nosignalingdataissent),whileSSoffersthe servicesdescribedinsection 4.1.1.3. CC provides a point-to-point connection between two terminals and is used by higher layers for call establishment, call clearing and change of call parameters.Thislayeralsoprovidesfunctionstosendin-bandtones,called**dual tonemultiplefrequency(DTMF)**,overtheGSMnetwork.Thesetonesareused, e.g., for the remote control of answering machines or the entry of PINs in elec- tronic banking and are, also used for dialing in traditional analog telephone systems. These tones cannot be sent directly over the voice codec of a GSM MS, as the codec would distort the tones. They are transferred as signals and then converted into tones in the fixed network part of the GSM system.

AdditionalprotocolsareusedattheA$_{bis}$andAinterfaces(theinternalinter- faces of a GSM system not presented here). Data transmission at the physical layertypicallyuses**pulsecodemodulation(PCM)**systems.WhilePCMsystems offertransparent64kbit/schannels,GSMalsoallowsforthesubmultiplexingof four16kbit/schannelsintoasingle64kbit/schannel(16kbit/sareenoughfor user data from an MS). The physical layer at the A interface typically includes leasedlineswith2.048Mbit/scapacity.LAPDisusedforlayertwoatA$_{bis}$,BTSM for BTS management.

Signaling system No. 7 (SS7)is used for signaling between an MSC and a BSC. This protocol also transfers all management information betweenMSCs, HLR, VLRs, AuC, EIR, and OMC. An MSC can also control a BSS via a **BSS application part (BSSAP)**.

4.1.5 Localizationandcalling

One fundamental feature of the GSM system is the automatic, worldwide localizationofusers.Thesystemalwaysknowswhereausercurrentlyis,andthesame phone number is valid worldwide. To provide this service, GSM performsperi- odic location updates even if a user does not use the mobile station (provided that the MS is still logged into the GSM network and is not completely switched off). The HLR always contains information about the current location (only the location area, not the precise geographical location), and the VLR currently responsible for the MS informs the HLR about location changes. As soon as an MS moves into the range of a new VLR (a new location area), the HLR sends all userdataneededtothenewVLR.ChangingVLRswithuninterruptedavailability ofallservicesisalsocalled**roaming.**Roamingcantakeplacewithinthenetwork of one provider, between two providers in one country (national roaming is, often not supported due to competition between operators), but also between different providers in different countries (international roaming). Typically, peopleassociateinternationalroamingwiththetermroamingasitisthistypeof roaming that makes GSM very attractive: one device, over 190 countries!

TolocateanMSandtoaddresstheMS,severalnumbersareneeded:

- **MobilestationinternationalISDNnumber(MSISDN):**[6]The only important numberforauserofGSMisthephonenumber.Rememberthatthe phonenumberisnotassociatedwithacertaindevicebutwiththeSIM, whichispersonalizedforauser.TheMSISDNfollowstheITU-Tstandard E.164 for addresses as it is also used in fixed ISDN networks. This number consists of the**country code (CC)**(e.g., +49 179 1234567 with 49 for Germany), the**national destination code (NDC)**(i.e., the address of the network provider, e.g., 179), and the **subscriber number (SN)**.

---

6 In other types of documentation, this number is also called ‘Mobile Subscriber ISDN Number’ or‘Mobile Station ISDN Number’. Even the original ETSI standards use different wordings for the sameacronym.However,theterm‘subscriber’ismuchbettersuitedasitexpressestheindependenceoftheuser related number from the device (station).

- **International mobile subscriber identity (IMSI):**GSM uses the IMSI for internal unique identification of a subscriber. IMSI consists of a**mobilecountry code (MCC)**(e.g., 240 for Sweden, 208 for France), the**mobile networkcode(MNC)**(i.e.,thecodeofthenetworkprovider),andfinally the **mobile subscriber identification number (MSIN)**.
- **Temporary mobile subscriber identity (TMSI):**To hide the IMSI, which would give away the exact identity of the user signaling over the air interface, GSM uses the 4 byte TMSI for local subscriber identification. TMSI is selected by the current VLR and is only valid temporarily and within the location area of the VLR (for an ongoing communication TMSI and LAI are sufficient to identify a user; the IMSI is not needed). Additionally, a VLRmay change the TMSI periodically.
- **Mobile station**[7]**roaming number (MSRN):**Another temporary address thathidestheidentityandlocationofasubscriberisMSRN.TheVLRgener- ates this address on request from the MSC, and the address is also stored in the HLR. MSRN contains the current**visitor country code (VCC)**, the**visi- tor national destination code (VNDC)**, the identification of the current MSC together with the subscriber number. The MSRN helps the HLR to find a subscriber for an incoming call.

All these numbers are needed to find a subscriber and to maintain the connection with a mobile station. The interesting case is the**mobile terminated call (MTC)**, i.e., a situation in which a station calls a mobile station (the calling stationcouldbeoutsidetheGSMnetworkoranothermobilestation).Figure4.8 shows the basic steps needed to connect the calling station with the mobileuser. In step 1, a user dials the phone number of a GSM subscriber. The fixed network (PSTN) notices (looking at the destination code) that the number belongs to a user in the GSM network and forwards the call setup to theGateway MSC (2). The GMSC identifies the HLR for the subscriber (which is coded in the phone number) and signals the call setup to the HLR (3). The HLR now checks whether the number exists and whether the user has subscribed to the requested services, and requests an MSRN from the current VLR (4). After receiving the MSRN (5), the HLR can determine the MSC responsible for the MS andforwardsthisinformationtotheGMSC(6).TheGMSCcannowforward the call setup request to the MSC indicated (7).

From this point on, the MSC is responsible for all further steps. First, it requeststhecurrentstatusoftheMSfromtheVLR(8).IftheMSisavailable,the MSC initiates paging in all cells it is responsible for (i.e. the location area, LA,10), as searching for the right cell would be too time consuming (but this approach puts some load on the signaling channels so optimizations exist). The

---

7 Here,adiscrepancyexistsbetweenITU-TstandardsandETSI'sGSM.MScandenotemobilestationor mobile subscriber. Typically, almost all MS in GSM refer to subscribers, as identifiers are not dependentonthestation,butonthesubscriberidentity(storedintheSIM).

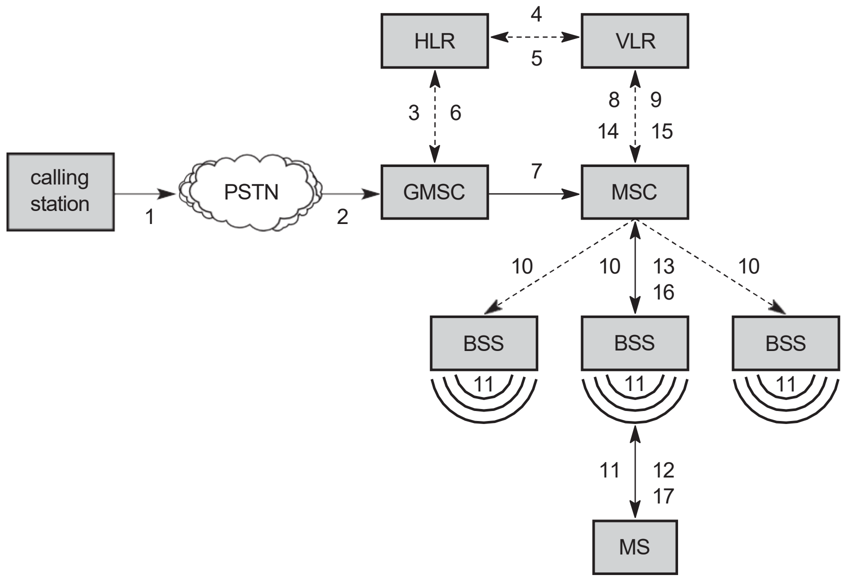


Figure4.8 Mobileterminated call (MTC)

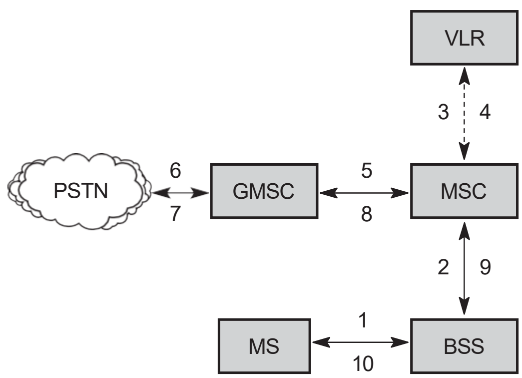
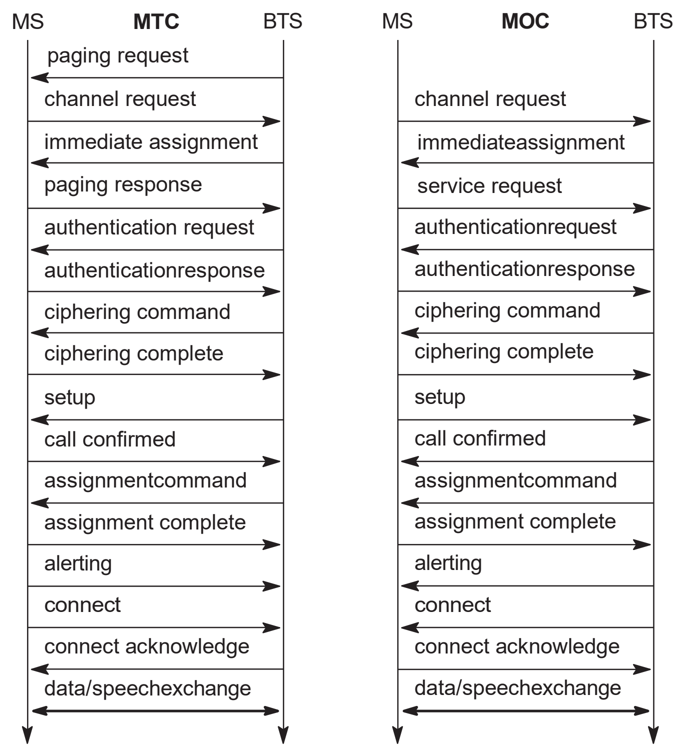


Figure 4.9 Mobileoriginated call (MOC)

BTSsofallBSSstransmitthispagingsignaltotheMS(11).IftheMSanswers(12 and 13), the VLR has to perform security checks (set up encryption etc.). The VLR then signals to the MSC to set up a connection to the MS (steps 15 to 17).

Itismuchsimplertoperforma**mobileoriginatedcall(MOC)**comparedtoa MTC(seeFigure4.9).TheMStransmitsarequestforanewconnection(1),theBSS forwardsthisrequesttotheMSC(2).TheMSCthenchecksifthisuserisallowed tosetupacallwiththerequestedservice(3and4)andcheckstheavailabilityof resourcesthroughtheGSMnetworkandintothePSTN.Ifallresourcesareavail- able, the MSC sets up a connection between the MS and the fixed network.

Figure4.10
Messageflowfor MTC and MOC



In addition to the steps mentioned above, other messages are exchanged betweenanMSandBTSduringconnectionsetup(ineitherdirection).Thesemess-agescanbequiteoftenheardinradiosorbadlyshieldedloudspeakersascrackling noise before the phone rings. Figure 4.10 shows the messages for an MTC and MOC. Paging is only necessary for an MTC, then similar message exchanges follow. The first step in this context is the channel access via the random access channel (RACH) with consecutive channel assignment; the channel assigned could be a traffic channel (TCH) or a slower signalling channel SDCCH.

The next steps, which are needed for communication security, comprisethe authentication of the MS and the switching to encrypted communication. The system now assigns a TCH (if this has not been done). This has the advantage of only having to use an SDCCH during the first setup steps. If the setup fails, no TCH has been blocked. However, using a TCH from the begin-ning has a speed advantage.

The following steps depend on the use of MTC or MOC. If someone is call-ing the MS, it answers now with ‘alerting’ that the MS is ringing and with ‘connect’thattheuserhaspressedtheconnectbutton.Thesameactions happen the other way round if the MS has initiated the call. After connection acknowledgement, both parties can exchange data.

Closingtheconnectioncomprisesauser-initiateddisconnectmessage(both sides can do this), followed by releasing the connection and the radio channel.

4.1.6 Handover

Cellular systems require**handover**procedures, as single cells do not cover the whole service area, but, e.g., only up to 35 km around each antenna on the countryside and some hundred meters in cities (Tripathi, 1998). The smaller thecellsizeandthefasterthemovementofamobilestationthroughthecells(up to250km/hforGSM),themorehandoversofongoingcallsarerequired. However,ahandovershouldnotcauseacut-off,alsocalled**calldrop**.GSM aims at maximum handover duration of 60 ms.

Therearetwobasicreasonsforahandover(about40havebeenidentified in the standard):

- Themobilestation**movesoutoftherange**ofaBTSoracertainantennaof a BTS respectively. The received**signal level**decreases continuously until it fallsbelowtheminimalrequirementsforcommunication.The**errorrate** maygrowduetointerference,thedistancetotheBTSmaybetoohigh (max.35km)etc.–alltheseeffectsmaydiminishthe**qualityoftheradio link** and make radio transmission impossible in the near future.
- Thewiredinfrastructure(MSC,BSC)maydecidethatthe**trafficinonecell is too high**and shift some MS to other cells with a lower load (if possible). Handover may be due to **load balancing**.

Figure4.11showsfourpossiblehandoverscenariosinGSM:

- **Intra-cell handover:**Within a cell, narrow-band interference couldmake transmissionatacertainfrequencyimpossible.TheBSCcouldthendecide to change the carrier frequency (scenario 1).
- **Inter-cell, intra-BSC handover:**This is a typical handover scenario. The mobilestationmovesfromonecelltoanother,butstayswithinthecontrol of the same BSC. The BSC then performs a handover, assigns a new radio channel in the new cell and releases the old one (scenario 2).
- I**nter-BSC,intra-MSChandover:**AsaBSConlycontrolsalimitednumberof cells;GSMalsohastoperformhandoversbetweencellscontrolledbydiffer- ent BSCs. This handover then has to be controlled by the MSC (scenario 3). This situation is also shown in Figure 4.13.
- **Inter MSC handover:**A handover could be required between two cells belongingtodifferentMSCs.NowbothMSCsperformthehandover together (scenario 4).

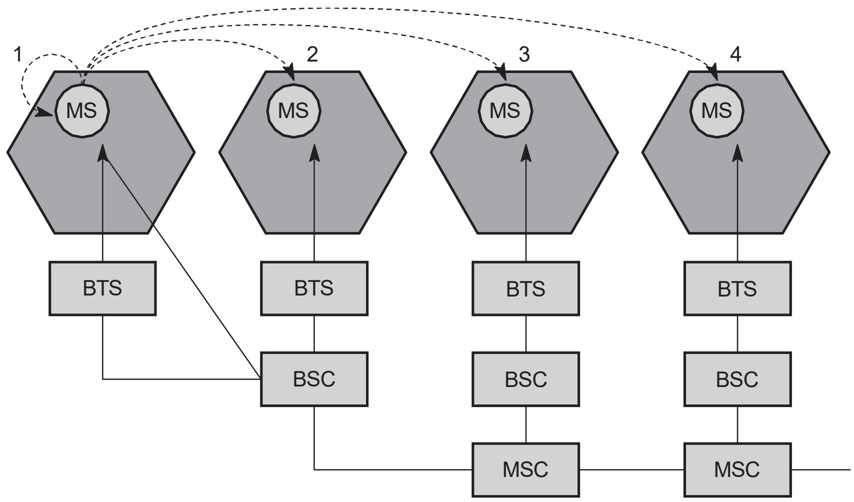


Figure4.11 Typesofhandover inGSM

Toprovideallthenecessaryinformationforahandoverduetoaweaklink, MS and BTS both perform periodic measurements of the downlink and uplink quality respectively. (Link quality comprises signal level and bit error rate.) MeasurementreportsaresentbytheMSabouteveryhalf-secondandcontain the quality of the current link used for transmission as well as the quality of certainchannelsinneighboringcells(theBCCHs).

Figure4.12showsthetypicalbehaviorofthereceivedsignallevelwhilean MSmovesawayfromoneBTS(BTS$_{old}$)closertoanotherone(BTS$_{new}$).Inthis case,thehandoverdecisiondoesnotdependontheactualvalueofthereceived signal level, but on the average value. Therefore, the BSC collects all values (bit error rate and signal levels from uplink and downlink) from BTS and MS and calculatesaveragevalues.Thesevaluesarethencomparedtothresholds,i.e.,the handover margin (HO_MARGIN), which includes some hysteresis to avoid a ping-pong effect (Wong, 1997). (Without hysteresis, even short-term interference,e.g.,shadowingduetoabuilding,couldcauseahandover.)Still,even withtheHO_MARGIN,theping-pongeffectmayoccurinGSM–avaluewhich is too high could cause a cut-off, and a value which is too low could cause too many handovers.

Figure 4.13 shows the typical signal flow during an inter-BSC,intra-MSC handover.TheMSsendsitsperiodicmeasurementsreports,theBTS$_{old}$forwards thesereportstotheBSC$_{old}$togetherwithitsownmeasurements.Basedonthese valuesand,e.g.,oncurrenttrafficconditions,theBSC$_{old}$maydecidetoperform a handover and sends the message HO_required to the MSC. The task of theMSC then comprises the request of the resources needed for the handover from the new BSC, BSC$_{new}$. This BSC checksif enough resources (typically frequencies ortimeslots)areavailableandactivatesaphysicalchannelattheBTS$_{new}$topre- pareforthearrivaloftheMS.

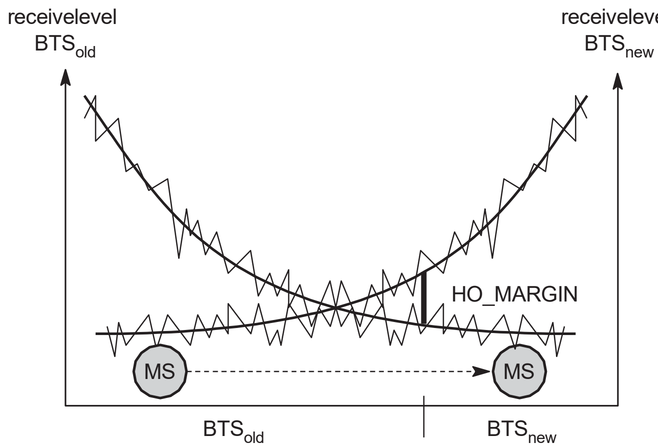


Figure 4.12 Handoverdecision depending on receive level

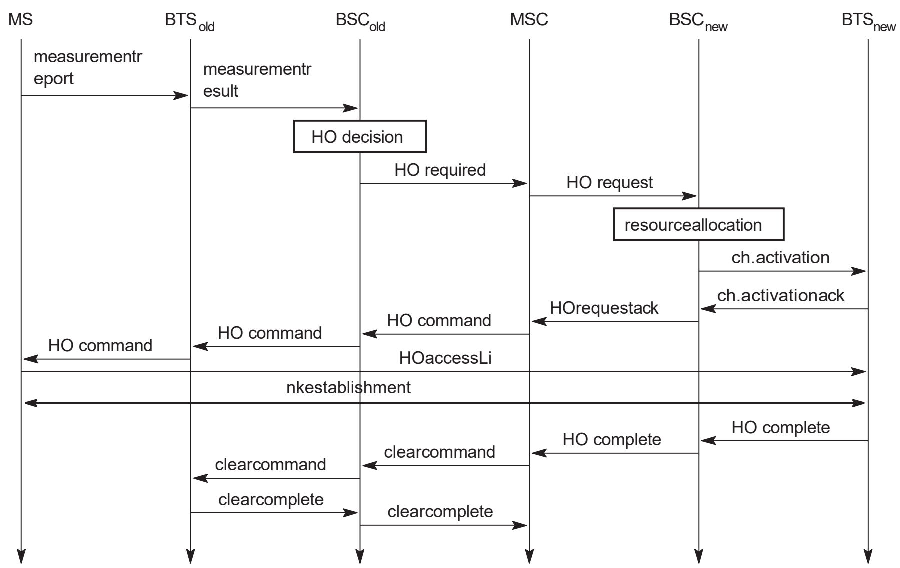


Figure4.13 Intra-MSChandover

The BTS_{new}acknowledges the successful channel activation, BSC_{new}acknowledgesthehandoverrequest.TheMSCthenissuesahandovercomma nd that is forwarded to the MS. The MS now breaks its old radio link and accesses the new BTS. The next steps include the establishment of the link (this includes layer two link establishment and handover complete messages from the MS). Basically, the MS has then finished the handover, but it is important to release the resources at the old BSC and BTS and to signal the successful handoverusing the handover and clear complete messages as shown.

Moresophisticatedhandovermechanismsareneededforseamless handoversbetweendifferentsystems.Forexample,future3Gnetworkswillnotcoverwh olecountriesbutfocusoncitiesandhighways.Handoverfrom, e.g., UMTS to GSM without service interruption must be possible. Even more challenging is the seamless handover between wireless LANs (see chapter 7) and 2G/3G networks. This can be done using multimode mobile stations and a more sophisticated roaming infrastructure. However, it is still not obvious how these systems may scale for a large number of users and many handovers, and what handover quality guarantees they can give.

4.1.7 Security

GSMoffersseveralsecurityservicesusingconfidentialinformationstoredinthe AuC and in the individual SIM (which is plugged into an arbitrary MS). TheSIM storespersonal,secretdataandisprotectedwithaPINagainstunauthorizeduse. (Forexample,thesecretkey$K_i$usedforauthenticationandencryptionprocedures isstoredintheSIM.)ThesecurityservicesofferedbyGSMareexplainedbelow:

- **Access control and authentication:**Thefirststepincludestheauthentica-tionofavaliduserfortheSIM.TheuserneedsasecretPINtoaccesstheSIM. Thenextstepisthesubscriberauthentication(seeFigure4.10).Thisstepis basedonachallenge-responseschemeaspresentedinsection4.1.7.1.
- **Confidentiality:** All user-related data is encrypted. After authentication, BTS andMSapplyencryptiontovoice,data,andsignalingasshowninsection 4.1.7.2. This confidentiality exists only between MS and BTS, but it does not exist end-to-end or within the whole fixed GSM/telephone network.
- **Anonymity:**To provide user anonymity, all data is encrypted beforetrans-mission, and user identifiers (which would reveal an identity) are not used over the air. Instead, GSM transmits a temporary identifier (TMSI), which is newlyassignedbytheVLRaftereachlocationupdate.Additionally,theVLR can change the TMSI at any time.

Three algorithms have been specified to provide security services in GSM. **Algorithm A3**is used for**authentication**,**A5**for**encryption**, and**A8**for the **generation of a cipher key**. In the GSM standard only algorithm A5 was publicly available, whereas A3 and A8 were secret, but standardized with open interfaces.BothA3andA8arenolongersecret,butwerepublishedontheinter- net in 1998. This demonstrates that security by obscurity does not really work. Asitturnedout,thealgorithmsarenotverystrong.However,networkproviders can use stronger algorithms for authentication – or users can apply stronger end-to-end encryption. Algorithms A3 and A8 (or their replacements) are located on the SIM and in the AuC and can be proprietary. Only A5 which is implemented in the devices has to be identical for all providers.

4.1.7.1 Authentication

BeforeasubscribercanuseanyservicefromtheGSMnetwork,heorshemustbe authenticated. Authentication is based on the SIM, which stores the**individual authenticationkeyK_i**,the**useridentificationIMSI**,andthealgorithmusedfor authentication**A3.**Authenticationuses achallenge-responsemethod:theaccess

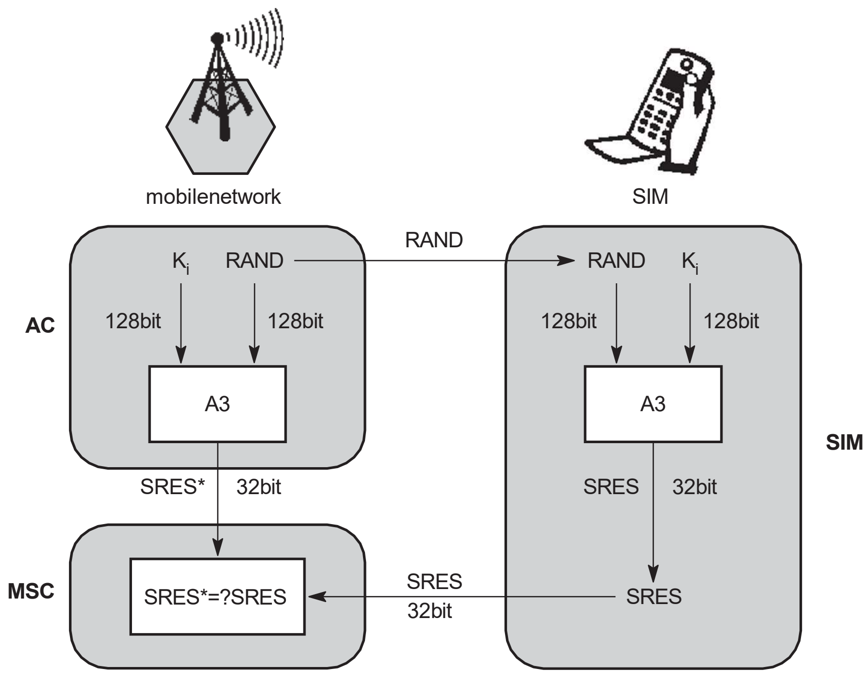


Figure 4.14 Subscriber authentication

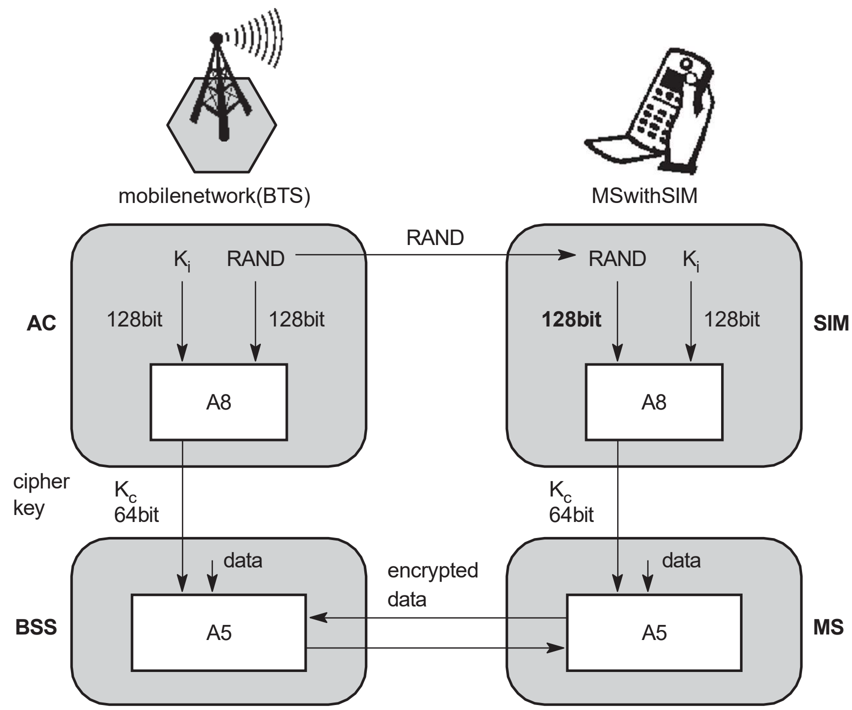
controlACgeneratesarandomnumber**RAND**aschallenge,andtheSIMwithin theMSanswerswith**SRES**(signedresponse)asresponse(seeFigure4.14).TheAuC performs the basic generation of random values RAND, signed responses SRES, and cipher keys $K_c$for each IMSI, and then forwards this information to the HLR. The current VLR requests the appropriate values for RAND, SRES, and $K_c$from the HLR.

Forauthentication,theVLRsendstherandomvalueRANDtotheSIM. Both sides, network and subscriber module, perform the same operation with RANDandthekeyK_i,calledA3.TheMSsendsbacktheSRESgeneratedbythe SIM;theVLRcannowcomparebothvalues.Iftheyarethesame,theVLR acceptsthe subscriber,otherwise the subscriberis rejected.

4.1.7.2 Encryption

To ensure privacy, all messages containing user-related information are encrypted in GSM over the air interface. After authentication, MS and BSS can start using encryption by applying the cipher key K_c(the precise location of security functions for encryption, BTS and/or BSC are vendor dependent). $K_c$is generated using the individual key $K_i$and a random value by applying the algorithm A8. Note that the SIM in the MS and the network both calculate the same $K_c$based on the random value RAND. The key $K_c$itself is not transmitted over the air interface.

Figure4.15
Dataencryption



MS and BTS can now encrypt and decrypt data using the algorithm A5 and the cipher key K_c. As Figure 4.15 shows, $K_c$should be a 64 bit key – which is not very strong, but is at least a good protection against simple eavesdropping. However, the publication of A3 and A8 on the internet showed that in certain implementations 10 of the 64 bits are always set to 0, so that the real length of the key is thus only 54 consequently, the encryption is much weaker.

4.1.8 Newdataservices

Asmentionedabove,thestandardbandwidthof9.6kbit/s(14.4kbit/swithsome providers)availablefordatatransmissionisnotsufficientfortherequirementsof today's computers. When GSM was developed, not many people anticipated the tremendous growth of data communication compared to voice communication. At that time, 9.6 kbit/s was a lot, or at least enough for standard group 3 fax machines. But with the requirements of, e.g., web browsing, file download, or even intensive e-mail exchange with attachments, this is not enough.

To enhance the data transmission capabilities of GSM, two basic approachesarepossible.AsthebasicGSMisbasedonconnection-orientedtrafficchannels, e.g., with 9.6 kbit/s each, several channels could be combined to increase band-width.ThissystemiscalledHSCSDandispresentedinthefollowingsection.A more progressive step is the introduction of packet-oriented traffic in GSM, i.e., shifting the paradigm from connections/telephone thinking to packets/internet thinking. The system, called GPRS, is presented in section 4.1.8.2.

4.1.8.1 HSCSD

A straightforward improvement of GSM's data transmission capabilities is**high speed circuit switched data (HSCSD)**, which is available with some providers. In this system, higher data rates are achieved by bundling several TCHs. An MS requests one or more TCHs from the GSM network, i.e., it allocates severalTDMA slots within a TDMA frame. This allocation can be asymmetrical, i.e.,more slots can be allocated on the downlink than on the uplink, which fits the typical user behavior of downloading more data compared to uploading. Basically, HSCSD only requires software upgrades in an MS and MSC (both have to be able to split a traffic stream into several streams, using a separate TCH each, and to combine these streams again).

In theory, an MS could use all eight slots within a TDMA frame to achieveanairinterfaceuserrate(AIUR)of,e.g.,8TCH/F14.4channelsor115.2kbit/s (ETSI, 1998e). One problem of this configuration is that the MS is required to send and receive at the same time. Standard GSM does not require this capability – uplink and downlink slots are always shifted for three slots. ETSI (1997a) specifiestheAIURavailableat57.6kbit/s(duplex)usingfourslotsintheuplink and downlink (Table 4.2 shows the permitted combinations of traffic channels and allocated slots for non-transparent services).

Although it appears attractive at first glance, HSCSD exhibits some major disadvantages.Itstillusestheconnection-orientedmechanismsofGSM.These are not at all efficient for computer data traffic, which is typically bursty and asymmetrical.Whiledownloadingalargerfilemayrequireallchannels reserved, typical web browsing would leave the channels idle most of the time. Allocatingchannelsisreflecteddirectlyintheservicecosts,as,oncethechan- nels have been reserved, other users cannot use them.

AIUR	TCH/F4.8	TCH/F9.6	TCH/F14.4
4.8kbit/s	1	–	–
9.6kbit/s	2	1	–
14.4kbit/s	3	–	1
19.2kbit/s	4	2	–
28.8kbit/s	–	3	2
38.4kbit/s	–	4	–
43.2kbit/s	–	–	3
57.6kbit/s	–	–	4

Table 4.2 Available dataratesforHSCSD in GSM

For*n*channels, HSCSD requires*n*times signaling during handover, connection setup and release. Each channel is treated separately. The probability of blockingorservicedegradationincreasesduringhandover,asinthiscaseaBSC has to check resources for*n*channels, not just one. All in all, HSCSD may be an attractiveinterimsolutionforhigherbandwidthandratherconstanttraffic(e.g., filedownload).However,itdoesnotmakemuchsenseforburstyinternettraffic as long as a user is charged for each channel allocated for communication.

4.1.8.2 GPRS

The next step toward more flexible and powerful data transmission avoids the problemsofHSCSDbybeingfullypacket-oriented.The**generalpacketradioser- vice (GPRS)**provides packet mode transfer for applications that exhibit traffic patterns such as frequent transmission of small volumes (e.g., typical web requests) or infrequent transmissions of small or medium volumes (e.g., typical web responses) according to the requirement specification (ETSI, 1998a). Compared to existing data transfer services, GPRS should use the existing network resources more efficiently for packet mode applications, and should provide a selection of QoS parameters for the service requesters. GPRS should also allow for broadcast, multicast, and unicast service. The overall goal in this context is the provision of a more efficient and, thus, cheaper packet transfer service for typicalinternetapplicationsthatusuallyrelysolelyonpackettransfer.Network providers typically support this model by charging on volume and not on connection time as is usual for traditional GSM data services and for HSCSD. The main benefit for users of GPRS is the 'always on' characteristic – no connection has to be set up prior to data transfer. Clearly, GPRS was driven by the tremendous success of the packet-oriented internet, and by the new traffic models and applications. However, GPRS, as shown in the following sections, needs additionalnetworkelements,i.e.,softwareandhardware.UnlikeHSCSD,GPRSdoes notonlyrepresentasoftwareupdatetoallowforthebundlingofchannels,italso representsabigsteptowardsUMTSasthemaininternalinfrastructureneededfor UMTS(initsinitialrelease)isexactlywhatGPRSuses(seesection4.4).

ThemainconceptsofGPRSareasfollows(ETSI,1998b).ForthenewGPRS radio channels, the GSM system can allocate between one and eight time slotswithin a TDMA frame. Time slots are not allocated in a fixed,pre-determined manner but on demand. All time slots can be shared by the active users; up- and downlinkareallocatedseparately.Allocationoftheslotsisbasedoncurrent load and operator preferences. Depending on the coding, a transfer rate of up to170kbit/sispossible.ForGPRS,operatorsoftenreserveatleastatimeslotper cell to guarantee a minimum data rate. The GPRS concept is independent ofchannel characteristics and of the type of channel (traditional GSM traffic orcontrol channel), and does not limit the maximum data rate (only the GSMtransport system limits the rate). All GPRS services can be used in parallel to conventionalservices.Table4.3showsthetypicaldataratesavailablewithGPRS ifitisusedtogetherwithGSM(GPRScanalsobeusedforotherTDMAsystems).

Coding scheme	1slot	2slots	3slots	4slots	5slots	6slots	7slots	8slots
CS-1	9.05	18.2	27.15	36.2	45.25	54.3	63.35	72.4
CS-2	13.4	26.8	40.2	53.6	67	80.4	93.8	107.2
CS-3	15.6	31.2	46.8	62.4	78	93.6	109.2	124.8
CS-4	21.4	42.8	64.2	85.6	107	128.4	149.8	171.2

Table4.3GPRSdata rates in kbit/s

In the beginning, only coding schemes CS-1 and CS-2 are available. The system chooses a coding scheme depending on the current error rate (CS-4 provides no error correction capabilities).

It should be noted that the real available data rate heavily depends on the currentloadofthecellasGPRStypicallyonlyusesidletimeslots.Thetransfer rate depends on the capabilities of the MS as not all devices are able to send and receiveatthesametime.Table4.4givesexamplesfordeviceclassestogether withtheirabilitytousetimeslotsforsendingandreceivingdata.Themaxi- mum possible number of slots limits the transfer rate even more. For example, a class12devicemayreceivedatausing4slotswithinaGSMtimeframeorit maysenddatausing4slots.However,amaximumnumberof5slotsmaybe usedaltogether.Usingall8slotsfordataencodedusingCS-4yieldsthemaxi- mumrateof171.2kbit/s.Today,atypicalMSisaclass10deviceusingCS-2, which results in a receiving rate of 53.6 kbit/s and a sending rate of 26.8 kbit/s.

Inphase1,GPRSoffersa**point-to-point (PTP)**packettransferservice(ETSI, 1998c). One of the PTP versions offered is the**PTP connection oriented network service (PTP-CONS)**, which includes the ability of GPRS to maintain a virtualcircuituponchangeofthecellwithintheGSMnetwork.Thistypeof

Class	Receivingslots	Sendingslots	Maximumnumberofslots
1	1	1	2
2	2	1	3
3	2	2	3
5	2	2	4
8	4	1	5
10	4	2	5
12	4	4	5

Table4.4 Examples for GPRS device classes

Table4.5 Reliability classes in GPRS accordingtoETSI (1998c)

Reliability class	Lost SDU probability	DuplicateSDU probability	Outofsequence SDUprobability	CorruptSDU probability
1	10^{-9}	10^{-9}	10^{-9}	10^{-9}
2	10^{-4}	10^{-5}	10^{-5}	10^{-6}
3	10^{-2}	10^{-5}	10^{-5}	10^{-2}

service corresponds to **X.25**, the typical circuit-switched packet-oriented transfer protocol available worldwide. The other PTP version offered is the **PTP connectionless network service (PTP-CLNS)**, which supports applications that arebasedontheInternetProtocol**IP**.Multicasting,called**point-to-multipoint (PTM)** service, is left for GPRS phase 2.

UsersofGPRScanspecifya**QoS-profile**.Thisdeterminesthe**serviceprecedence**(high,normal,low),**reliabilityclass**and**delayclass**ofthetransmission, and**user data throughput**. GPRS should adaptively allocate radio resources to fulfill these user specifications. Table 4.5 shows the three reliability classes together with the maximum probabilities for a lost service data unit (SDU), a duplicated SDU, an SDU out of the original sequence, and the probability of deliveringacorruptSDUtothehigherlayer.Reliabilityclass1couldbeusedfor very error-sensitive applications that cannot perform error corrections them- selves. If applications exhibit greater error tolerance, class 2 could be appropriate. Finally, class 3 is the choice for error-insensitive applications or applications that can handle error corrections themselves.

DelaywithinaGPRSnetworkisincurredbychannelaccessdelay,coding forerrorcorrection,andtransferdelaysinthefixedandwirelesspartofthe GPRS network. The delay introduced by external fixed networks is out of scope. However, GPRS does not produce additional delay by buffering packets as store- and-forward networks do. If possible, GPRS tries to forward packets as fast as possible. Table 4.6 shows the specified maximum mean and 95 percentile delay values for packet sizes of 128 and 1,024 byte. As we can clearly see, no matterwhichclass,alldelaysareordersofmagnitudehigherthanfixednetwork delays. This is a very important characteristic that has to be taken into account whenimplementinghigherlayerprotocolssuchasTCPontopofGPRSnet- works(seechapter9).Typicalroundtriptimes(RTT)infixednetworksarein theorderof10to100ms.UsingrealunloadedGPRSnetworksroundtriptimes ofwellabove1sforevensmallpackets(128–512byte)arecommon. Additionally,GPRSexhibitsalargejittercomparedtofixednetworks(several 100 ms are not uncommon). This characteristic has a strong impact on user experiencewhen,e.g.,interactiveInternetapplicationsareusedontopofGPRS.

Table 4.6 Delay classes in GRPS according to ETSI (1998c)

	SDU size 128 byte		SDU size 1,024 byte	
Delay Class	**Mean**	**95 percentile**	**Mean**	**95 percentile**
1	<0.5s	<1.5s	<2s	<7s
2	<5s	<25s	<15s	<75s
3	<50s	<250s	<75s	<375s
4		Unspecified		

Finally, GPRS includes several **security services** such as authentication, access control, user identity confidentiality, and user information confidentiality. Even a completely **anonymous service** is possible, as, e.g., applied for road toll systems that only charge a user via the MS independent of the user's identity.

The **GPRS architecture** introduces two new network elements, which are called **GPRS support nodes (GSN)** and are in fact routers. All GSNs are integrated into the standard GSM architecture, and many new interfaces have been defined (see Figure 4.16). The **gateway GPRS support node (GGSN)** is the inter-working unit between the GPRS network and external **packet data networks (PDN)**. This node contains routing information for GPRS users, performs address conversion, and tunnels data to a user via encapsulation. The GGSN is connected to external networks (e.g., IP or X.25) via the G_i interface and transfers packets to the SGSN via an IP-based GPRS backbone network (G_n interface).

The other new element is the **serving GPRS support node (SGSN)** which supports the MS via the G_b interface. The SGSN, for example, requests user addresses from the **GPRS register (GR)**, keeps track of the individual MSs' location, is responsible for collecting billing information (e.g., counting bytes), and performs several security functions such as access control. The SGSN is connected to a BSC via frame relay and is basically on the same hierarchy level as an MSC. The GR, which is typically a part of the HLR, stores all GPRS-relevant data. GGSNs and SGSNs can be compared with home and foreign agents, respectively, in a mobile IP network (see chapter 8).

As shown in Figure 4.16, packet data is transmitted from a PDN, via the GGSN and SGSN directly to the BSS and finally to the MS. The MSC, which is responsible for data transport in the traditional circuit-switched GSM, is only used for signaling in the GPRS scenario. Additional interfaces to further network elements and other PLMNs can be found in ETSI (1998b).

Before sending any data over the GPRS network, an MS must attach to it, following the procedures of the **mobility management**. The attachment procedure includes assigning a temporal identifier, called a **temporary logical link identity (TLLI)**, and a **ciphering key sequence number (CKSN)** for data encryption. For each MS, a **GPRS context** is set up and stored in the MS and in

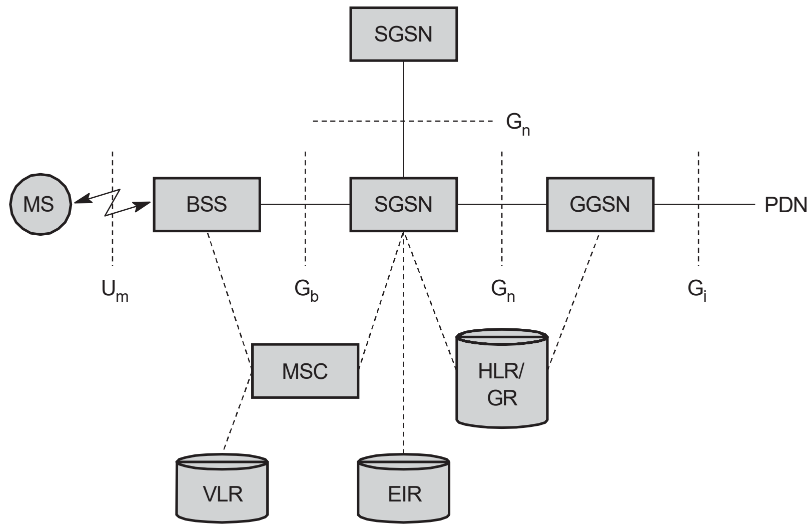


Figure4.16 GPRSarchitecture reference model

the corresponding SGSN. This context comprises the status of the MS (whichcan be ready, idle, or standby; ETSI, 1998b), the CKSN, a flag indicating if compression is used, and routing data (TLLI, the routing area RA, a cell identifier, and a packet data channel, PDCH, identifier). Besides attaching and detaching, mobility management also comprises functions for authentication, location management, and ciphering (here, the scope of ciphering lies between MS and SGSN, which is more than in standard GSM). In**idle**mode an MS is not reachable and all context is deleted. In the**standby**state only movement across routing areas is updated to the SGSN but not changes of the cell. Permanent updating would waste battery power, no updating would require system-wide paging. The update procedure in standby mode is a compromise. Only in the **ready** state every movement of the MS is indicated to the SGSN.

Figure 4.17 shows the protocol architecture of the transmission plane forGPRS.ArchitecturesforthesignalingplanescanbefoundinETSI(1998b).All data within the GPRS backbone, i.e., between the GSNs, is transferred using the **GPRStunnellingprotocol(GTP)**.GTPcanusetwodifferenttransportproto- cols, either the reliable**TCP**(needed for reliable transfer of X.25 packets) or the non-reliable **UDP** (used for IP packets). The network protocol for the GPRS backboneis**IP**(usinganylowerlayers).Toadapttothedifferentcharacteristicsof the underlying networks, the**subnetwork dependent convergence protocol (SNDCP)** is used between an SGSN and the MS. On top of SNDCP and GTP, user packetdataistunneledfromtheMStotheGGSNandviceversa.Toachievea high reliability of packet transfer between SGSN and MS, a special LLC is used,which comprises ARQ and FEC mechanisms for PTP (and later PTM) services.

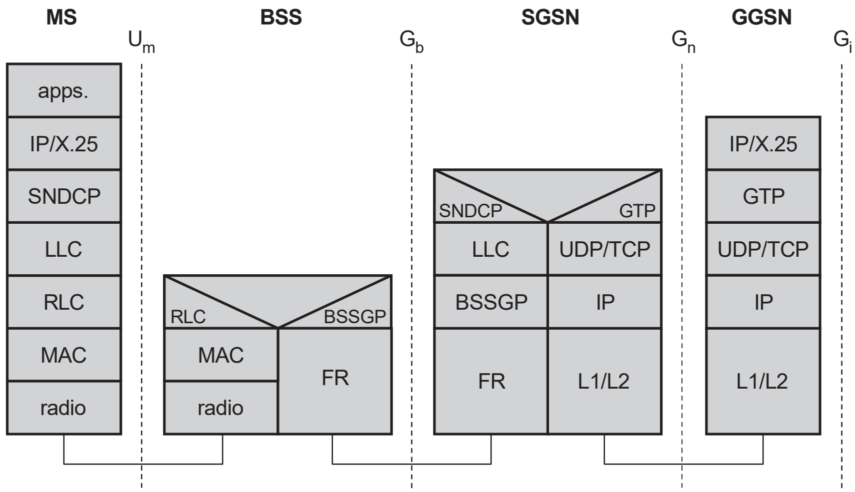


Figure4.17 GPRS transmission plane protocol reference model

A**base station subsystem GPRS protocol (BSSGP)**isusedtoconveyrouting and QoS-related information between the BSS and SGSN. BSSGP does not perform error correction and works on top of a**frame relay (FR)**network. Finally, radio link dependent protocols are needed to transfer data over the $U_m$interface. The**radio link protocol (RLC)**provides a reliable link, while the**MAC**controls access with signaling procedures for the radio channel and the mapping of LLC frames onto the GSM physical channels. The **radio interface** at $U_m$needed for GPRS does not require fundamental changes compared to standardGSM(Brasche,1997),(ETSI,1998d).However,severalnewlogicalchannels and their mapping onto physical resources have been defined. For example, one MS can allocate up to eight**packet data traffic channels (PDTCHs)**. Capacity can be allocated on demand and shared between circuit-switched channels and GPRS. This allocation can be done dynamically with load supervision or alternatively, capacity can be pre-allocated.

Averyimportantfactorforanyapplicationworkingend-to-endisthatit does not 'notice' any details from the GSM/GPRS-related infrastructure. The applicationuses,e.g.,TCPontopofIP,IPpacketsaretunneledtotheGGSN, whichforwardsthemintothePDN.AllPDNsforwardtheirpacketsforaGPRS user to the GGSN, the GGSN asks the current SGSN for tunnel parameters, and forwards the packets via SGSN to the MS. Although MSs using GPRS maybe considered as part of the internet, one should know that operators typically performanaddresstranslationintheGGSNusingNAT.AllMSsareassignedprivate IP addresses which are then translated into global addresses at the GGSN. The advantage of this approach is the inherent protection of MSs from attacks (the subscriber typically has to pay for traffic even if it originates from an attack!) – privateaddressesarenotroutedthroughtheinternetsoitisnotpossibleto reachanMSfromtheinternet.ThisisalsoadisadvantageifanMSwantsto offer a service using a fixed, globally visible IP address. This is difficult with IPv4andNATanditwillbeinterestingtoseehowIPv6isusedforthispurpose (while still protecting the MSs from outside attacks as air traffic is expensive).

4.2 DECT

Anotherfullydigitalcellularnetworkisthe**digitalenhancedcordlesstelecommu- nications (DECT)**system specified by ETSI (2002, 1998j, k), (DECT Forum, 2002). Formerly also called**digital European cordless telephone and digital European cordlesstelecommunications**,DECTreplacesolderanalogcordlessphonesystems such as CT1 and CT1+. These analog systems only ensured security to a limited extent as they did not use encryption for data transmission and only offered a relatively low capacity. DECT is also a more powerful alternative to the digital system CT2, which is mainly used in the UK (the DECT standard works throughoutEurope), and has even been selected as one of the 3G candidates in the IMT-2000 family (see section 4.4). DECT is mainly used in offices, on campus, at trade shows, or in the home. Furthermore, access points to the PSTN can be established within, e.g., railway stations, large government buildings and hospitals, offering a much cheaper telephone service compared to a GSM system. DECT could also be used to bridgethelastfewhundredmetersbetweenanewnetworkoperatorandcustomers.
Using this ‘small range’ local loop, new companies can offer their service without having their own lines installed in the streets. DECT systems offer many different interworking units, e.g., with GSM, ISDN, or data networks. Currently, over 100 million DECT units are in use (DECT, 2002).

A big difference between DECT and GSM exists in terms of cell diameterand cell capacity. While GSM is designed for outdoor use with a cell diameter of up to 70 km, the range of DECT is limited to about 300 m from the base station (only around 50 m are feasible inside buildings depending on the walls). Due to this limited range and additional multiplexing techniques, DECT can offer its service to some 10,000 people within one km^2. This is a typical scenario withina big city, where thousands of offices are located in skyscrapers close together. DECTalsousesbasestations,butthesebasestationstogetherwithamobilesta- tion are in a price range of€100 compared to several€10,000 for a GSM base station. GSM base stations can typically not be used by individuals for private networks. One reason is licensing as all GSM frequencies have been licensed to network operators. DECT can also handle handover, but it was not designed to work at a higher speed (e.g., up to 250 km/h like GSM systems). Devices handling GSM and DECT exist but have never been a commercial success.

DECTworksatafrequencyrangeof1880–1990MHzoffering120full duplexchannels.Timedivisionduplex(TDD)isappliedusing10msframes. The frequency range is subdivided into 10 carrier frequencies using FDMA, each framebeingdividedinto24slotsusingTDMA.FortheTDDmechanism, 12 slots are used as uplink, 12 slots as downlink (see Figure 3.4). The digital modulationschemeisGMSK–eachstationhasanaveragetransmissionpower ofonly10mWwithamaximumof250mW.

4.2.1 Systemarchitecture

A DECT system, may have various different physical implementation depending on its actual use. Different DECT entities can be integrated into one physical unit; entities can be distributed, replicated etc. However, all implementations are based on the same logical reference model of the system architecture as shown in Figure 4.18. A**global network**connects the local communication structure to the outside world and offers its services via the interface D_1. Global networks could be integrated services digital networks (ISDN), public switched telephone networks (PSTN), public land mobile networks (PLMN), e.g., GSM, or packetswitchedpublicdatanetwork(PSPDN).Theservicesofferedbythesenet- works include transportation of data and the translation of addresses and routing of data between the local networks.

Local networksin the DECT context offer local telecommunication ser- vices that can include everything from simple switching to intelligent call forwarding, address translation etc. Examples for such networks are analog or digitalprivatebranchexchanges(PBXs)orLANs,e.g.,thosefollowingtheIEEE 802.x family of LANs. As the core of the DECT system itself is quite simple, all typical network functions have to be integrated in the local or global network, where the databases**home data base (HDB)**and**visitor data base (VDB)**are also located. Both databases support mobility with functions that are similar to those in the HLR and VLR in GSM systems. Incoming calls are automatically for- wardedtothecurrentsubsystemresponsiblefortheDECTuser,andthecurrent VDB informs the HDB about changes in location.

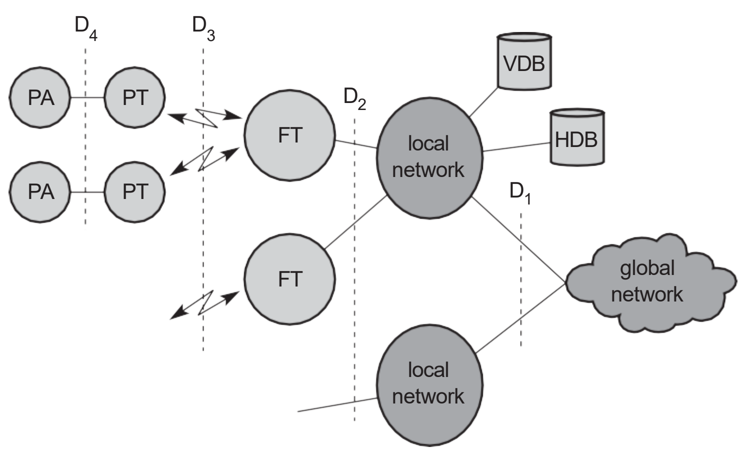


Figure4.18
DECTsystem architecturereference model

TheDECTcorenetworkconsistsofthe**fixedradiotermination(FT)**and the**portableradiotermination(PT)**,andbasicallyonlyprovidesamultiplex- ing service. FT and PT cover layers one to three at the fixed network side andmobile network side respectively. Additionally, several portable applications (PA) can be implemented on a device.

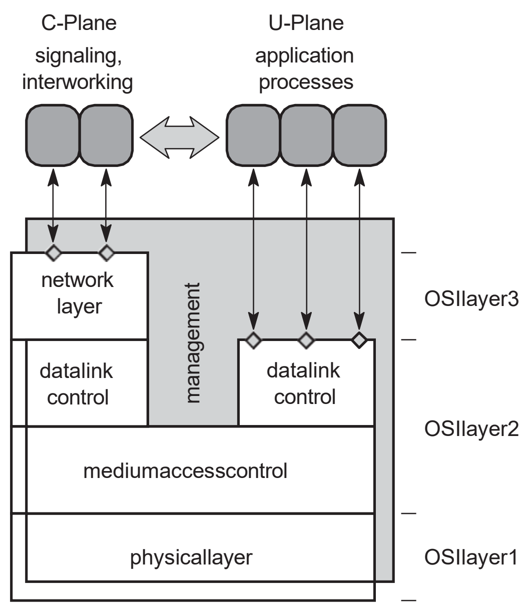
4.2.2 Protocol architecture

TheDECTprotocolreferencearchitecturefollowstheOSIreferencemodel.Figure 4.19showsthelayerscoveredbythestandard:thephysicallayer,mediumaccess control, and data link control[8]for both the**control plane (C-Plane)**and the **user plane (U-Plane)**. An additional network layer has been specified for theC-Plane,sothatuserdatafromlayertwoisdirectlyforwardedtotheU-Plane. A management plane vertically covers all lower layers of a DECT system.

4.2.2.1 Physical layer

As in all wireless networks, the **physical layer** comprises all functions for modulation/demodulation, incoming signal detection, sender/receiver synchronization, andcollectionofstatusinformationforthemanagementplane.Thislayergenerates the physical channel structure with a certain, guaranteed throughput. On request from the MAC layer, the physical layer assigns a channel for data transmission.

Figure4.19
DECTprotocol layers



8 Strictly speaking, the name "data link control" for the upper part of layer two is wrong in this architecture. According to the OSI reference model, the data link control (layer two) comprises the logicallink control (layer 2b) and the medium access control (layer 2a).

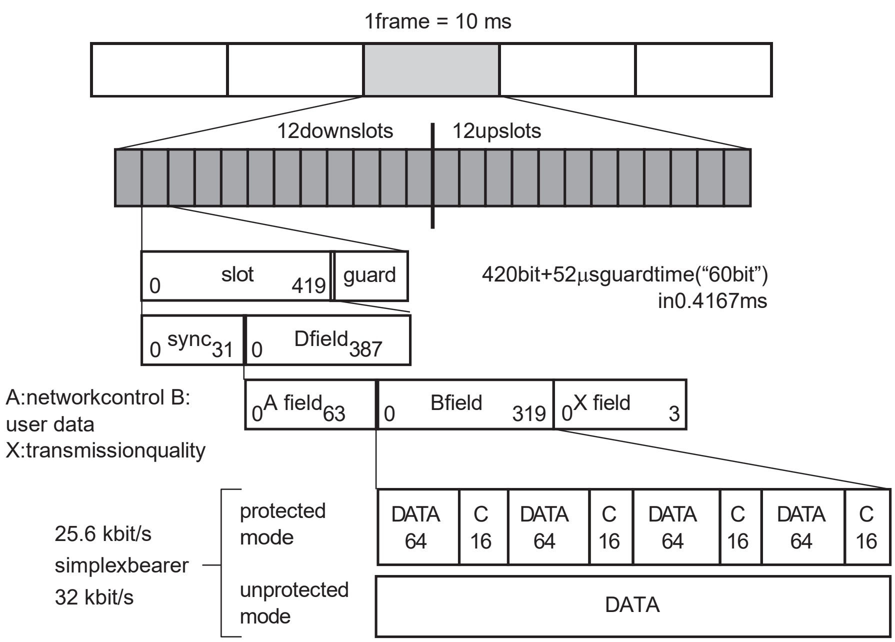


Figure4.20
DECTmultiplexand frame structure

Figure4.20showsthestandardTDMAframestructureusedinDECTand some typical data packets. Each frame has a duration of 10 ms and contains 12slotsforthedownlinkand12slotsfortheuplinkinthe**basicconnection** mode. If a mobile node receives data in slot s, it returns data in slot s+12. An **advancedconnection**modeallowsdifferentallocationschemes.Eachslothas a duration of 0.4167 ms and can contain several different physical packets. Typically, 420 bits are used for data; the remaining 52 μs are left as **guard space**. The 420 data bits are again divided into a 32 bit**synchronization pattern**fol- lowed by the **data** field D.

The fields for data transmission now use these remaining 388 bits for**net- work control**(A field), **user data** (B field), and the transfer of the **transmission quality**(Xfield).Whilenetworkcontrolistransmittedwithadatarateof
6.4 kbit/s (64 bit each 10 ms), the user data rate depends on additional error correctionmechanisms.The**simplexbearer**providesadatarateof32kbit/sin an**unprotectedmode**, while using a 16 bit CRC**checksum**C for a data blockof 64 bit in the**protected mode**reduces the data rate to 25.6 kbit/s. A**duplex bearer**serviceisproducedbycombiningtwosimplexbearers.DECTalsodefines bearer types with higher throughputs by combining slots, e.g., the**double duplex bearer** offers 80 kbit/s full-duplex.

4.2.2.2 Mediumaccesscontrollayer

The**medium access control (MAC**) layer establishes, maintains, and releases channels for higher layers by activating and deactivating physical channels. MAC multiplexesseverallogicalchannelsontophysicalchannels.Logicalchannels exist for signaling network control, user data transmission, paging, or sending broadcast messages. Additional services offered include segmentation/reassembly of packets and error control/error correction.

4.2.2.3 Datalinkcontrollayer

The**data link control (DLC)**layer creates and maintains reliable connections betweenthemobileterminalandthebasestation.Twoserviceshavebeen defined for the**C-Plane**: a**connectionlessbroadcast**service for paging (called **Lb**) and a**point-to-point**protocol similar to LAPD in ISDN, but adapted to the underlying MAC (called **LAPC+Lc**).

Severalservicesexistforthe**U-Plane**,e.g.,atransparentunprotectedservice (basicallyanullservice),aforwarderrorcorrectionservice,rateadaptationser- vices,andservicesforfutureenhancements.Ifservicesareused,e.g.,totransfer ISDN data at 64 kbit/s, then DECT also tries to transfer 64 kbit/s. However, in caseoferrors,DECTraisesthetransferrateto72kbit/s,andincludesFECanda buffer for up to eight blocks to perform ARQ. This buffer then introduces an additional delay of up to 80 ms.

4.2.2.4 Network layer

The **network laye**r of DECT is similar to those in ISDN and GSM and only exists for the**C-Plane.** This layer provides services to request, check, reserve, control, andreleaseresourcesatthefixedstation(connectiontothefixednetwork,wire- less connection) and the mobile terminal (wireless connection). The**mobility management (MM)**within the network layer is responsible for identity management, authentication, and the management of the location data bases.**Call control (CC)**handles connection setup, release, and negotiation. Two message services, the**connection oriented message service (COMS)**and the**connectionless message service (CLMS)**transfer data to and from the interworking unit that connects the DECT system with the outside world.

4.3 TETRA

Trunkedradiosystemsconstituteanothermethodofwirelessdatatransmission. These systems use many different radio carriers but only assign a specific carrier to acertainuserforashortperiodoftimeaccordingtodemand.While,forexample, taxi services, transport companies with fleet management systems and rescue teamsallhavetheirownuniquecarrierfrequencyintraditionalsystems,theycan shareawholegroupoffrequenciesintrunkedradiosystemsforbetterfrequency reuseviaFDMandTDMtechniques.Thesetypesofradiosystemstypicallyoffer interfaces to the fixed telephone network, i.e., voice and data services, but are not publicly accessible. These systems are not only simpler than most other networks, they are also reliable and relatively cheap to set up and operate, as they only haveto cover the region where the local users operate, e.g., a city taxi service.

To allow a common system throughout Europe, ETSI standardized the **TETRA**system(**terrestrialtrunkedradio**)[9]in1991(ETSI,2002),(TETRAMoU, 2002). This system should replace national systems, such as MODACOM, MOBI-TEX and COGNITO in Europe that typically connect to an X.25 packet network. (An example system from the US is ARDIS.) TETRA offers two standards:the **Voice+Data(V+D)**service(ETSI,1998l)andthe**packetdataoptimized(PDO)** service (ETSI, 1998m). While V+D offers circuit-switched voice and data trans-mission, PDO only offers packet data transmission, either connection-orientedto connect to X.25 or connectionless for the ISO CLNS (connectionless network service).Thelatterservicecanbepoint-to-pointorpoint-to-multipoint,thetyp- ical delay for a short message (128 byte) being less than 100 ms. V+D connection modes comprise unicast and broadcast connections, group commu- nication within a certain protected group, and a direct ad hoc mode without a base station. However, delays for short messages can be up to 500 ms or higher depending on the priority.

TETRAalsooffersbearerservicesofupto28.8kbit/sforunprotecteddata transmissionand9.6kbit/sforprotectedtransmission.Examplesforend-to-end services are call forwarding, call barring, identification, call hold, call priorities, emergencycallsandgroupjoins.ThesystemarchitectureofTETRAisverysimi- lar to GSM. Via the radio interface U_m, the **mobile station (MS)** connects to the **switching and management infrastructure (SwMI)**, which contains the user data bases (HDB, VDB), the base station, and interfaces to PSTN, ISDN, or PDN. The system itself, however, is much simpler in real implementation comparedto GSM, as typically no handover is needed. Taxis usually remain within a cer- tain area which can be covered by one TETRA cell.

Several frequencies have been specified for TETRA which uses FDD (e.g., 380–390 MHz uplink/390–400 MHz downlink, 410–420 MHz uplink/420–430MHzdownlink).Eachchannelhasabandwidthof25kHzandcancarry 36 kbit/s. Modulation is DQPSK. While V+D uses up to four TDMA voice or data channels per carrier, PDO performs statistical multiplexing. For accessing achannel, slotted Aloha is used.

Figure4.21showsthetypical**TDMAframestructure**ofTETRA.Each**frame** consistsoffourslots(fourchannelsintheV+Dservicepercarrier),withaframe durationof56.67ms.Each**slot**carries510bitswithin14.17ms,i.e.,36kbit/s.16 frames together with one**control frame**(CF) form a**multiframe**, and finally, a **hyperframe**contains60multiframes.Toavoidsendingandreceivingatthesame time,TETRAshiftstheuplinkforaperiodoftwoslotscomparedtothedownlink.

9 Formerlyknownastrans-Europeantrunkedradio,butworldwidemarketingisbetterwithout“Europe” in the name (see DECT).

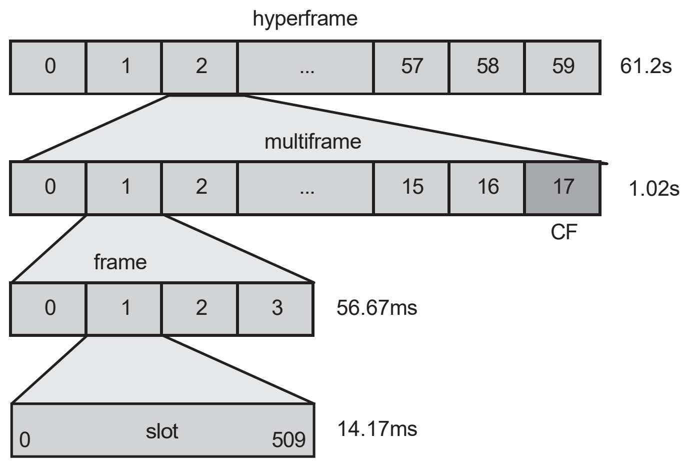


Figure4.21 TETRAframe structure

TETRA offers**traffic channels (TCH)**and**control channels (CCH)**similar to GSM. Typical TCHs are TCH/S for voice transmission, and TCH/7.2, TCH/4.8, TCH/2.4 for data transmission (depending on the FEC mechanisms required).

However,incontrasttoGSM,TETRAoffersadditionalserviceslikegroupcall, acknowledged group call, broadcast call, and discreet listening. Emergency services need a sub-second group-call setup in harsh environments which possibly lackallinfrastructure.ThesefeaturesarecurrentlynotavailableinGSMorother typicalmobiletelephonenetworks,soTETRAiscomplementarytoothersystems. TETRAhasbeenchosenbymanygovernmentorganizationsinEuropeandChina.

4.4 UMTSandIMT-2000

A lot has been written about third generation (or 3G) networks in the last few years. After a lot of hype and frustration these networks are currently deployed in many countries around the world. But how did it all start? First of all, the International Telecommunication Union (ITU) made a request for proposals for radio transmission technologies (RTT) for the**international mobile telecommunications(IMT)2000**program(ITU,2002),(Callendar,1997),(Shafi,1998). IMT-2000, formerly called future public land mobile telecommunication system (FPLMTS), tried to establish a common worldwide communication system that allowed for terminal and user mobility, supporting the idea of universal personal telecommunication (UPT). Within this context, ITU has created several recommendations for FPLMTS systems, e.g., network architectures for FPLMTS (M.817), Requirements for the Radio Interface(s) for FPLMTS (M.1034), or Framework for Services Supported by FPLMTS (M.816). The number 2000 in IMT-2000 should indicate the start of the system (year 2000+x) and the spectrum used (around 2000 MHz). IMT-2000 includes different environments such as indoor use, vehicles, satellites and pedestrians. The world radio conference (WRC) 1992 identified 1885–2025 and 2110–2200 MHz as the frequencybands that should be available worldwide for the new IMT-2000 systems (Recommendation ITU-R M.1036). Within these bands, two times 30 MHz have beenreservedformobilesatelliteservices(MSS).

Figure4.22showstheITUfrequencyallocation(fromtheworldadministra- tive radio conference, 1992) together with examples from several regionsthat alreadyindicatetheproblemofworldwidecommonfrequencybands.InEurope, some parts of the ITU's frequency bands for IMT-2000 are already allocated for DECT(seesection4.2).Theremainingfrequencieshavebeensplitintobandsfor UTRA-FDD (uplink: 1920–1980 MHz, downlink: 2110–2170 MHz) and UTRA-TDD(1900–1920MHzand2010–2025MHz).ThetechnologybehindUTRA-FDD and–TDDwillsubsequentlybeexplainedinmoredetailastheyformthebasisof UMTS. Currently, no other system is planned for IMT-2000 in Europe. More bandwidth is available in China for the Chinese 3G system TD-SCDMA or possibly other 3G technologies (such as W-CDMA or cdma2000 – it is still open whichsystemwilldominatetheChinesemarket;Chen,2002).Againslightlydifferentfrequenciesareusedbythe3GservicesinJapan,whicharebasedon W-CDMA(likeUTRA-FDD)orcdma2000.Anopenquestionisthefutureof3Gin theUSastheITU'sfrequencybandshavealreadybeenallocatedfor2Gnetworks or are reserved for other use. In addition to the original frequency allocations, the world radio conference (WRC) allocated new terrestrial IMT-2000 bands in the range of 800–1000 MHz, 1700–1900 MHz and 2500–2700 MHz in 2000.This approach includes the reuse of 2G spectrum (Evci, 2001).

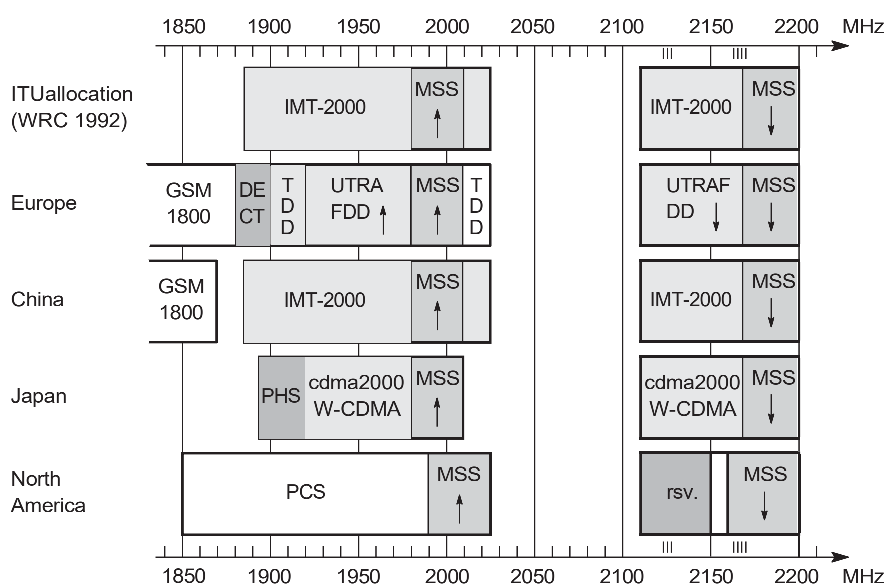


Figure4.22
IMT-2000frequencies

Now the reader might be confused by all the different technologies mentioned in the context of IMT-2000. Wasn't the plan to have a common global system? This was the original plan, but after many political discussions and fights about patents this idea was dropped and a so-called family of 3G standards was adopted.

FortheRTT,severalproposalswerereceivedin1998forindoor,pedestrian, vehicular, and satellite environments. These came from many different organizations, e.g.,**UWC-136**from the Universal Wireless Communications Consortium(US)thatextendstheIS-136standardintothethirdgenerationsystems,**cdma2000**that is based on the IS-95 system (US), and wideband packet CDMA (WP-CDMA) which tries to align to the European UTRA proposal. Basically, three big regions were submitting proposals to the ITU: ETSI for Europe, ARIB (Association of Radio Industries and Broadcasting) and TTC (Telecommunications Technology Council) for Japan, and ANSI (American National Standards Institute) for the US.

The European proposal for IMT-2000 prepared by ETSI is called**universal mobile telecommunications system (UMTS)** (Dasilva, 1997), (Ojanperä, 1998), the specific proposal for the radio interface RTT is**UMTS (now: universal) terrestrial radio access (UTRA)**(ETSI, 1998n), (UMTS Forum, 2002). UMTS as initially proposed by ETSI rather represents an evolution from the second generation GSM system to the third generation than a completely new system. In thisway, many solutions have been proposed for a smooth transition from GSM to UMTS, saving money by extending the current system rather than introducing anew one (GSMMoU, 1998).

One initial enhancement of GSM toward UMTS was**enhanced data rates for global (or: GSM) evolution (EDGE)**, which uses enhanced modulation schemes (8 PSK instead of GSM's GMSK, see chapter 2) and other techniques for dataratesofupto384kbit/susingthesame200kHzwidecarrierandthesame frequenciesasGSM(i.e.,adatarateof48kbit/spertimeslotisavailable).EDGE can be introduced incrementally offering some channels with EDGE enhancementthatcanswitchbetweenEDGEandGSM/GPRS.InEurope,EDGEwasnever usedasasteptowardUMTSbutoperatorsdirectlyjumpedontoUMTS.However, EDGEcanalsobeappliedtotheUSIS-136systemandmaybeachoiceforopera- tors that want to enhance their 2G systems (3G Americas, 2002).

Besides enhancing data rates, new additions to GSM, like**customized applicationformobileenhancedlogic(CAMEL)**introduceintelligentnet- work support. This system supports, for example, the creation of a **virtual home environment (VHE)**for visiting subscribers. GSMMoU (1998) provides many proposals covering QoS aspects, roaming, services, billing, accounting, radioaspects, core networks, access networks, terminal requirements, security, applicationdomains,operationandmaintenance,andseveralmigrationaspects.

UMTS fits into a bigger framework developed in the mid-nineties by ETSI, called**global multimedia mobility (GMM)**. GMM provides an architecture to integrate mobile and fixed**terminals**, many different**access network**s (GSM BSS,DECT,ISDN,UMTS,LAN,WAN,CATV,MBS),andseveral**coretransport**

networks(GSMNSS+IN,ISDN+IN,B-ISDN+TINA,TCP/IP)(ETSI,2002).Within thisframework,ETSIdeveloped**basicrequirements**forUMTSandforUTRA, the radio interface (ETSI, 1998h). Key requirements are minimum data rates of144 kbit/s for rural outdoor access (with the goal of 384 kbit/s) at a maximum speed of 500 km/h.[10]For suburban outdoor use a minimum of 384 kbit/s shouldbe achieved with the goal of 512 kbit/s at 120 km/h. For indoor or city use with relatively short ranges, up to 2 Mbit/s are required at 10 km/h (walking).

UMTS should also provide several bearer services, real-time and non real-timeservices,circuitandpacketswitchedtransmission,andmanydifferentdata rates.HandovershouldbepossiblebetweenUMTScells,butalsobetweenUMTS and GSM or satellite networks. The system should be compatible with GSM, ATM, IP, and ISDN-based networks. To reflect the asymmetric bandwidth needs of typical users, UMTS should provide a variable division of uplink and down- link data rates. Finally, UMTS has to fit into the IMT-2000 framework (this is probably the decisive factor for its success). As the global UMTS approach is rather ambitious, a more realistic alternative for the initial stages would be UMTS cells in cities providing a subset of services.

Several companies and interest groups have handed in proposals for UTRA (ETSI, 1998i), of which ETSI selected two for UMTS in January 1998. For the **paired band**(using FDD as a duplex mechanism), ETSI adopted the**W-CDMA** (Wideband CDMA) proposal, for the**unpaired band**(using TDD as duplex mechanism) the**TD-CDMA**(Time Division CDMA) proposal is used (Adachi, 1998), (Dahlman, 1998), (ETSI, 1998n). The paired band is typically used for publicmobilenetworkproviders(widearea,seeGSM),whiletheunpairedband is often used for local and indoor communication (see DECT). The following sections will present key properties of the initial UMTS system.

WhathappenedtotheIMT-2000family?Figure4.23givesanoverview.Asa single standard could not be found, the ITU standardized five groups of 3Gradio access technologies.

- **IMT-DS:**The**direct spread**technology comprises wideband CDMA (**W-CDMA**)systems.ThisisthetechnologyspecifiedforUTRA-FDDandused by all European providers and the Japanese NTT DoCoMo for 3G wide area services.ToavoidcompleteconfusionITU'snameforthetechnologyis IMT-DS,ETSIcalleditUTRA-FDDintheUMTScontext,andtechnology usediscalledW-CDMA(inJapanthisispromotedasFOMA,freedomof mobile multimedia access). Today, standardization of this technology takes placein3GPP(Thirdgenerationpartnershipproject,3GPP,2002a).Section 4.4.1providesmoredetailaboutthestandardizationprocess.
- **IMT-TC:** Initially, this family member, called **time code**, contained only the UTRA-TDD system which uses time-division CDMA (**TD-CDMA**). Later on, theChineseproposal,TD-synchronousCDMA(**TD-SCDMA**)wasadded.

10 Thisspeedisaproblemascurrently,onlyDABcanprovidehigherbitratesathighspeeds.

Both standards have been combined and 3GPP fosters the development of this technology. It is unclear when and to what extent this technology will be introduced. The initial UMTS installations are based on W-CDMA.

- **IMT-MC:**cdma2000 is a**multi-carrier**technology standardized by 3GPP2 (Third generation partnership project 2, 3GPP2, 2002), which was formed shortly after 3GPP to represent the second main stream in 3G technology. Version cdma2000 EV-DO has been accepted as the 3G standard.
- **IMT-SC:**The enhancement of the US TDMA systems, UWC-136, is a**single carrier**technology originally promoted by the Universal Wireless Communications Consortium (UWCC). It is now integrated into the 3GPP efforts. This technology applies EDGE, among others, to enhance the 2G IS-136 standard.
- **IMT-FT**: As **frequency time** technology, an enhanced version of the cordless telephone standard DECT has also been selected for applications that do not require high mobility. ETSI is responsible for the standardization of DECT.

The main driving forces in the standardization process are 3GPP and 3GPP2. ETSI has moved its GSM standardization process to 3GPP and plays a major role there. 3GPP tends to be dominated by European and Japanese manufacturers and standardizationbodies,while3GPP2isdominatedbythecompanyQualcomm and CDMA network operators. The quarrels between Qualcomm and European manufacturers (e.g., Nokia, Ericsson) regarding CDMA patents (UMTSand cdma2000 use CDMA) even escalated into the political arena back in 1998 (US vs EU).Everythingcooleddownwhenhundredsofpatentshadbeenexchanged and the systems had been harmonized (e.g., CDMA chipping rates).

Figure4.23
TheIMT-2000 family

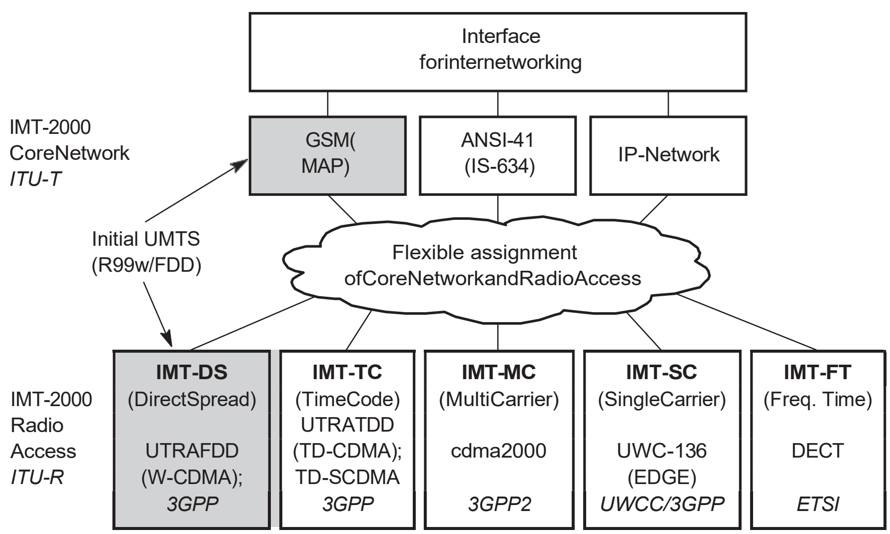


Figure 4.23 shows more than just the radio access technologies. One idea of IMT-2000 is the flexible assignment of a core network to a radio access system. The classical core network uses SS7 for signaling which is enhanced by ANSI-41 (cdmaOne, cdma2000, TDMA) or MAP (GSM) to enable roaming between different operators. The evolution toward 4G systems is indicated by the use of all-IP core networks (see Chapter 11). Obviously, internet-working functions have to be provided to enable cross-system data transfer, roaming, billing etc.

4.4.1 UMTSreleasesandstandardization

UMTS as discussed today and introduced in many countries relies on the initial releaseoftheUMTSstandardcalled**release99**or**R99**forshort.Thisreleaseof the specification describes the new radio access technologies UTRA FDD and UTRA TDD, and standardizes the use of a GSM/GPRS network as core within440 separate specifications. This enables a cost effective migration from GSM to UMTS.TheinitialinstallationswillevenoffertheFDDmodeonlyasindicated inFigure4.23.Thisreleasewas(almost)finalizedin1999–hencethenameR99. The following sections will focus on this release as it is unclear when, and to what extent, the following releases will be realized.

AfterR99therelease2000orR00followed.However,inSeptember20003GPP realized that it would be impossible to finalize the standard within the year 2000. 3GPP decided to split R2000 into two standards and call them release 4 (Rel-4) and release5(Rel-5).TheversionofallstandardsfinalizedforR99startwith3.x.y (a reason for renaming R99 into Rel-3), Rel-4 and Rel-5 versions start with 4.x.yand 5.x.y, respectively. The standards are grouped into series. For example, radio aspects are specified in series 25, technical realization in series 23, and codecs in series 26. The complete standard number (e.g., TS 25.401 V3.10.0) then identifies the series (25), the standard itself (401), the release (3), and the version within the release(10.0).All standards can be downloaded from`www.3gpp.org`(the example given is the UTRAN overall description, release 99, from June 2002).

Release 4 introduces quality of service in the fixed network plus several executionenvironments(e.g.,MExE,mobileexecutionenvironment,seechapter 10)andnewservicearchitectures.Furthermore,theChineseproposal,TD- SCDMA was added as low chiprate option to UTRA-TDD (only 1.28 Mchip/s occupying only 1.6 MHz bandwidth). This release already consists of over 500 specifications and was frozen in March 2001.

Release5specifiesaradicallydifferentcorenetwork.TheGSM/GPRSbased network will be replaced by an almost all-IP-core. While the radio interfaces remain the same, the changes in the core are tremendous for telecommunication network operators who have used traditional telephone technologies for many years. The content of this specification was frozen March 2002. This standard integrates IP-based multimedia services (IMS) controlled by the IETF's session initiation protocol (SIP, RFC 3261; Rosenberg, 2002; SIP Forum, 2002). Ahighspeeddownlinkpacketaccess(HSDPA)withspeedsintheorderof

8–10 Mbit/s was added as well as a wideband 16 kHz AMR codec for better audio quality. Additional features are end-to-end QoS messaging and several data compression mechanisms.

3GPP is currently working on **release 6**(and thinking of release 7) which is expected to be frozen in March 2003. This release comprises the use of multiple inputmultipleoutput(MIMO)antennas,enhancedMMS,securityenhancements,WLAN/UMTSinterworking,broadcast/multicastservices,enhancedIMS, IP emergency calls, and many more management features (3GPP, 2002a).

The reader should not forget that many companies still have to make any money from, release 99, so it is not clear at what time and to what extent the new releases will be implemented. The following describes the initial UMTS standard, release 99, which is currently deployed.

4.4.2 UMTSsystemarchitecture

Figure4.24showstheverysimplifiedUMTSreferencearchitecturewhichapplies to both UTRA solutions (3GPP, 2000). The**UTRA network (UTRAN)**handles celllevelmobilityandcomprisesseveral**radionetworksubsystems(RNS)**.The functionsoftheRNSincluderadiochannelcipheringanddeciphering,handovercontrol,radioresourcemanagement etc. The UTRAN is connected to the**user equipment(UE)**viatheradiointerface **U_u**(which is comparable to the $U_m$interfaceinGSM).Viathe**I_u**interface(which issimilartotheAinterfaceinGSM), UTRAN communicates with the **core network (CN).** The CN contains functions for inter-system handover, gateways to other networks (fixed or wireless), and performs location management if there is no dedicated connection between UE and UTRAN.

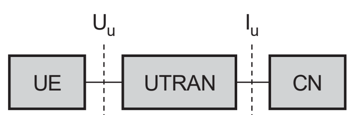


Figure4.24 Maincomponents oftheUMTS reference architecture

UMTS further subdivides the above simplified architecture into so-called **domains**(see Figure 4.25). The**user equipment**domain is assigned to a single user and comprises all the functions that are needed to access UMTS services. WithinthisdomainaretheUSIMdomainandthemobileequipmentdomain.The **USIM**domain contains the SIM for UMTS which performs functions for encryption and authentication of users, and stores all the necessary user-related data for UMTS. Typically, this USIM belongs to a service provider and contains a micro processor for an enhanced program execution environment (USAT, UMTS SIM applicationtoolkit).Theenddeviceitselfisinthe**mobileequipment**domain.All functionsforradiotransmissionaswellasuserinterfacesarelocatedhere.

The**infrastructure**domain is shared among all users and offers UMTS services to all accepted users. This domain consists of the **access network** domain, which contains the radio access networks (RAN), and the core network domain, whichcontainsaccessnetworkindependentfunctions.The**corenetwork** domain can be separated into three domains with specific tasks. The**serving network**domaincomprisesallfunctionscurrentlyusedbyauserforaccessing

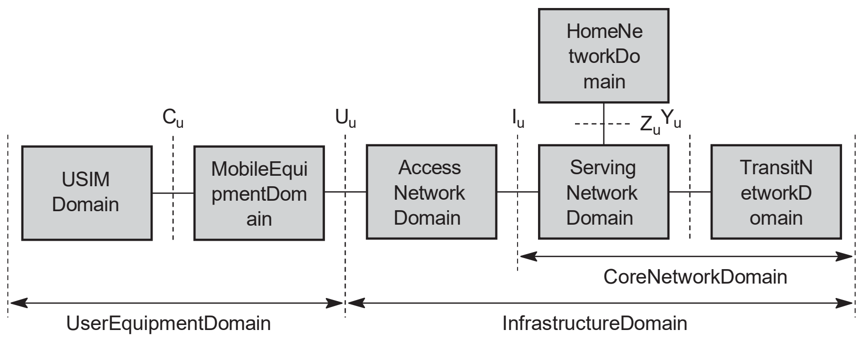


Figure 4.25 UMTSdomains andinterfaces

UMTS services. All functions related to the home network of a user, e.g., user data look-up, fall into the **home network** domain. Finally, the **transit network** domain may be necessary if, for example, the serving network cannot directly contact the home network. All three domains within the core network may bein fact the same physical network. These domains only describe functionalities.

4.4.3 UMTSradiointerface

The biggest difference between UMTS and GSM comes with the new radio interface (U_u). The duplex mechanisms are already well known from GSM (FDD) and DECT (TDD). However, the direct sequence (DS) CDMA used in UMTS is new(for European standards, not in the US where CDMA technology has been avail- able since the early nineties). DS-CDMA was introduced in chapters 2 and 3.This technology multiplies a stream of bits with a chipping sequence. This spreadsthesignaland,ifthechippingsequenceisunique,canseparatedifferent users.Allsignalsusethesamefrequencyband(inUMTS/IMT-20005MHz-wide bandshavebeenspecifiedandlicensedtonetworkoperators).Toseparatediffer- ent users, the codes used for spreading should be (quasi) orthogonal, i.e., their cross-correlation should be (almost) zero.

UMTS uses a constant **chipping rate** of 3.84 Mchip/s. Different user data rates canbesupportedusingdifferentspreadingfactors(i.e.,thenumberofchipsperbit). Figure4.26showsthebasicideasofspreadingandseparationofdifferentsendersin UMTS.Thefirststepinasenderisspreadingofuserdata(data$_i$)usingorthogonal **spreading**codes.Usingorthogonalcodesseparatesthedifferentdatastreamsofa sender. UMTS uses so-called **orthogonal variable spreading factor (OVSF)** codes. Figure 4.27 shows the basic idea of OVSF. Orthogonal codes are generated by doubling a chipping sequence X with and without flipping the sign of the chips. ThisresultsinXand–X,respectively.Doublingthechippingsequencealsoresults inspreadingabittwiceasmuchasbefore.ThespreadingfactorSF=nbecomes2n. Starting with a spreading factor of 1, Figure 4.27 shows the generation of orthogonalcodeswithdifferentspreadingfactors.Twocodesareorthogonalaslongasonecod eisneverapartoftheothercode.LookingatthecodingtreeinFigure4.27andcon- sideringtheconstructionofthecodes,orthogonalityisguaranteedifonecodehas not been generated based on another. For example, if a sender uses the code (1,–1)

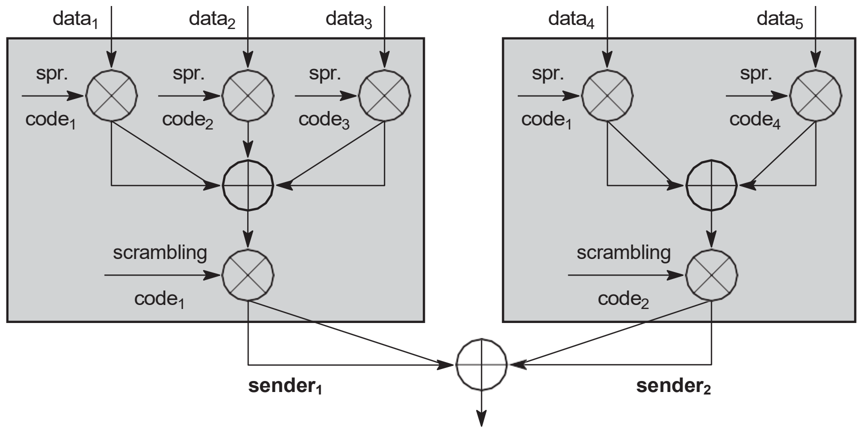


Figure4.26 Spreadingand scrambling of user date

with spreading factor 2, it is not allowed to use any of the codes located in the subtrees generated out of (1,–1). This means that, e.g., (1,–1,1,–1), (1,–1,–1,1,–1,1,1,–1), or (1,–1,–1,1,–1,1,1,–1,–1,1,1,–1,1,–1,–1,1) cannot be used anymore. However, it is no problem to use codes with different spreading factors if one code has not been generated using the other. Thus, (1,–1) block only the lower subtree in Figure 4.27, manyothercodesfromtheupperpartcanstillbeused.Anexampleforavalidcombinationin OVSFis(1,–1),(1,1,–1,–1),(1,1,1,1,1,1,1,1),(1,1,1,1,–1,–1,–1,–1,1,1,1,1,–1,–1,–1,–1),(1,1,1,1,–1,–1,–1,–1,–1,–1,–1,–1,1,1,1,1).Thiscombinationoccupies the whole code spaces and allows for the transmission of data with different spreadingfactors(2,4,8,and2*16).Thisexampleshowsthetightcouplingofavail- able spreading factors and orthogonal codes.

Now remember that UMTS uses a constant chipping rate (3.84 Mchip/s). Using different spreading factors this directly translates into the support of different data rates. If the chipping rate is constant, doubling the spreading factor means dividing the data rate by two. But this also means that UMTS can only support a single data stream with SF=1 as then no other code may be used. Using the example combination above, a stream with half the maximum data rate, one with a fourth, one with an eighth, and two with a sixteenth are supported at the same time.

Each sender uses OVSF to spread its data streams as Figure 4.26 shows. The spreading codes chosen in the senders can be the same. Using different spreading codes in all senders within a cell would require a lot of management and would increase the complexity. After spreading all chip streams are added and scrambled. **Scrambling**doesnotspreadthechipsequenceanyfurtherbutXORschipsbasedon a code. In the **FDD** mode, this scrambling code is unique for each sender and separates all senders (UE and base station) in a cell. After scrambling, the signals of different senders are quasi orthogonal. Quasi-orthogonal signals have the nice feature that they stay quasi-orthogonal even if they are not synchronized. Using orthogonalcodeswouldrequirechip-synchronousreceptionandtightsynchronization (this is done in other CDMA networks). For**TDD**the scrambling code is cell specific,i.e.,allstationsinacellusethesamescramblingcodeandcellsareseparated using different codes. The scrambled chips are **QPSK** modulated and transmitted.

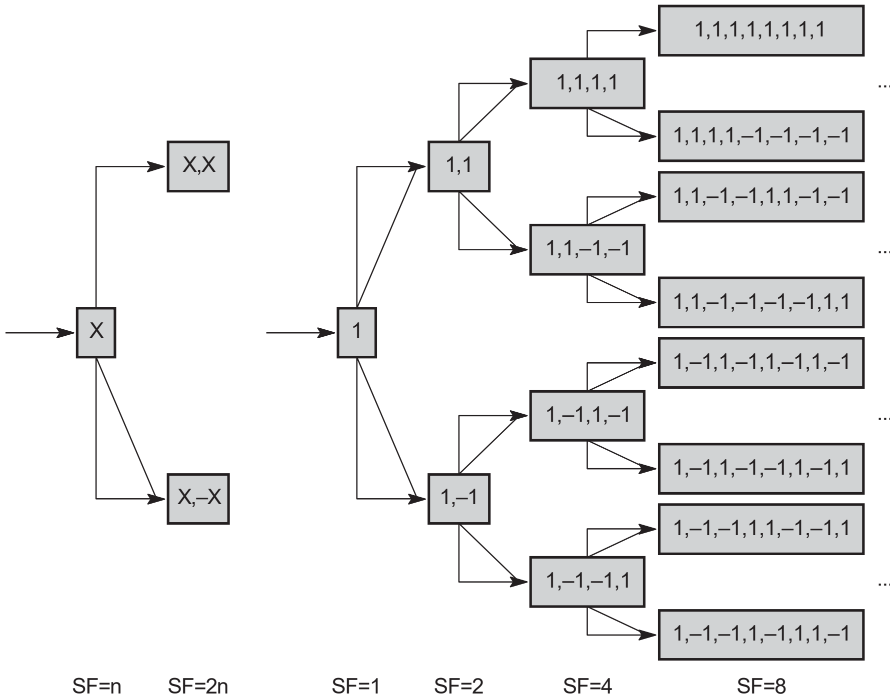


Figure4.27
OVSFcodetreeused fororthogonalspreading

4.4.3.1 UTRA-FDD(W-CDMA)

The FDD mode for UTRA uses**wideband CDMA (W-CDMA)**with direct sequence spreading. As implied by FDD, uplink and downlink use different frequencies. A mobile station in Europe sends via the uplink using a carrier between 1920 and 1980 MHz, the base station uses 2110 to 2170 MHz for the downlink(seeFigure4.22).Figure4.28showsaradioframecomprising15time slots. Time slots in W-CDMA are not used for user separation but to support periodic functions (note that this is in contrast to, e.g., GSM, where time slotsare used to separate users!). A radio frame consists of 38,400 chips and has a duration of 10 ms. Each time slot consists of 2,560 chips, which roughly equals 666.6μs.[11]TheoccupiedbandwidthperW-CDMAchannelis4.4to5MHz (channel spacing can be varied to avoid interference between channels of differ-ent operators). These 5 MHz bands of the spectrum have been sold in many countriesusinganauctionorabeautycontest.InGermany,theFDDspectrum wassoldforover50billionEurosduringanauction!Butthatwasatatime whenmarketingpeopletriedtoconvinceeveryonethatUMTSwouldbring

11 Early version of W-CDMA specified a chipping rate of 4.096 M chip/s and 16 time slots per frame.This was changed during the harmonization process which was necessary to avoid patent conflictsand to enable devices that can handle different CDMA standards. The harmonization process is fos-tered by the operators harmonization group (OHG), which is an informal steering group of wirelessoperator companies promoting 3G harmonization. The OHG was founded in 1999.

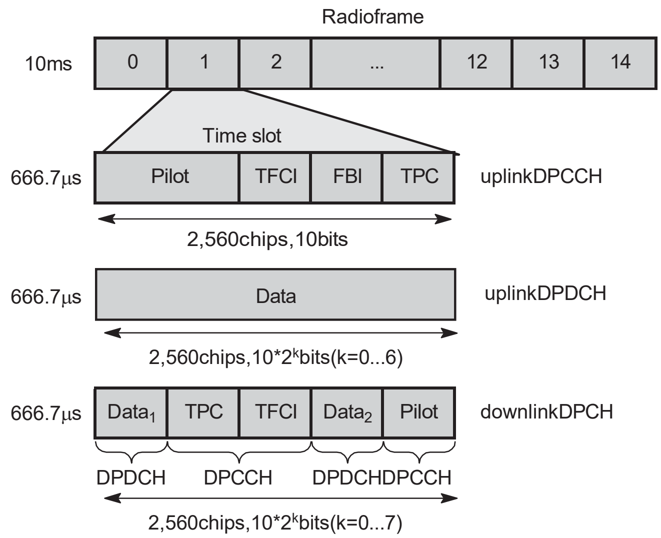


Figure4.28 UTRAFDD(W-CDMA) framestructure

high-bandwidthapplicationstoanymobiledevicewithhighprofitsforall. Today,mostpeoplearemuchmorerealisticandknowthatdatarateswillbe quitelowinthebeginning(150kbit/speruserarerealistic,2Mbit/sarenot). Thecapacityofacellunderrealisticassumptions(interferenceetc.),i.e.,the sum of all data rates, will rather be 2 Mbit/s. To provide high data rates a lot of moneyhastobeinvestedintheinfrastructure:UTRAFDDrequiresatleasttwice asmanybasestationsasGSM;celldiametersof500mwillbecommonplace. Thisshowsclearlythatthistechnologywillnotcoverwholecountriesinthe nearfuturebutcitiesandhighwaysonly.Peopleinthecountrysidewillhaveto rely on GSM/GPRS for many more years to come.

Back to the frame structure shown in Figure 4.28. Similar to GSM, UMTS defines many logical and physical channels, and their mapping. The figureshows three examples of physical channels as they are used for data trans- mission. Two physical channels are shown for the uplink.

- **Dedicated physical data channel (DPDCH):**This channel conveys user or signalingdata.Thespreadingfactorofthischannelcanvarybetween4and 256.Thisdirectlytranslatesintothedataratesthischannelcanoffer:960 kbit/s (spreading factor 4, 640 bits per slot, 15 slots per frame, 100 frames per second), 480, 240, 120, 60, 30, and 15 kbit/s (spreading factor 256). This also shows one of the problems of using OVSF for spreading: only certain mul- tiples of the basic data rate of 15 kbit/s can be used. If, for example, 250 kbit/s are needed the device has to choose 480 kbit/s, which wastes bandwidth. In each connection in layer 1 it can have between zero and six DPDCHs. This results in a theoretical maximum data rate of 5,740 kbit/s (UMTS describesUEs with a maximum of 1,920 kbit/s only). Table 4.7 shows typical user data rates together with the required data rates on the physical channels.

Userdatarate[kbit/s]	12.2 (voice)	64	144	384
DPDCH[kbit/s]	60	240	480	960
DPCCH[kbit/s]	15	15	15	15
Spreading	64	16	8	4

Table4.7 Typical UTRA-FDDuplink data rates

- **Dedicated physical control channel (DPCCH):** In each connection layer 1 needs exactly one DPCCH. This channel conveys control data for the physical layer only and uses the constant spreading factor 256. The**pilot**is used for channel estimation. The**transport format combination identifier (TFCI)** specifies the channels transported within the DPDCHs. Signaling for asofthandoverissupportedbythe**feedback information field (FBI)**.The last field,**transmit power control (TPC)**is used for controlling the transmission power of a sender. Power control is performed in each slot, thus 1,500 power control cycles are available per second. Tight power control is necessarytomitigatenear-far-effectsasexplainedinchapter2.Sixdifferent DPCCH bursts have been defined which differ in the size of the fields.
- **Dedicated physical channel (DPCH):** The downlink time multiplexes controlanduserdata.Spreadingfactorsbetween4and512areavailable.Again, manydifferentburstformats(17altogether)havebeendefinedwhichdiffer inthesizeofthefieldshowninFigure4.28.Theavailabledataratesfordata channels (DPDCH) within a DPCH are 6 (SF=512), 24, 51, 90, 210, 432, 912, and 1,872 kbit/s (SF=4).

While no collisions can occur on the downlink (only the base station sends on the downlink), medium access on the uplink has to be coordinated. A **physical random access channel (PRACH)** is used for this purpose. UTRA-FDD defines15randomaccessslotswithin20ms;withineachaccessslot16different access preambles can be used for random access. Using slotted Aloha, a UEcan accessanaccessslotbysendingapreamble.TheUEstartswiththelowestavail- able transmission power to avoid interfering with other stations. If no positive acknowledgement is received, the UE tries another slot and another preamble with the next higher power level (power ramping). The number of available access slots can be defined per cell and is transmitted via a broadcast channelto all UEs.

A UE has to perform the following steps during the**search for a cell**after power on:

- **Primary synchronization:**A UE has to synchronize with the help of a 256chipprimarysynchronizationcode.Thiscodeisthesameforallcellsand helps to synchronize with the time slot structure.

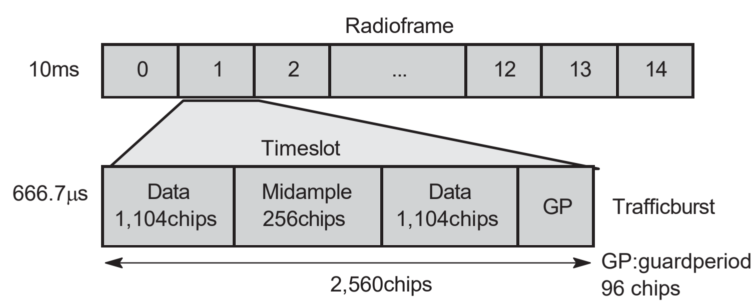
- **Secondary synchronization:** During this second phase the UE receives a secondarysynchronizationcodewhichdefinesthegroupofscrambling codes used in this cell. The UE is now synchronized with the frame structure.
- **Identificationofthescramblingcode:** TheUEtriesallscrambling codes within the group of codes to find the right code with the help of a correlator.AfterthesethreestepstheUEcanreceiveallfurtherdataover a broadcast channel.

4.4.3.2 UTRA-TDD(TD-CDMA)

The second UTRA mode, UTRA-TDD, separates up and downlink in time using a radio frame structure similar to FDD. 15 slots with 2,560 chips per slot form a radio frame with a duration of 10 ms. The chipping rate is also 3.84 Mchip/s. To reflectdifferentuserneedsintermsofdatarates,theTDDframecanbe**symmet-rical**or**asymmetrical**, i.e., the frame can contain the same number of uplink and downlink slots or any arbitrary combination. The frame can have only one **switching point**fromuplinktodownlinkorseveralswitchingpoints.However, at least one slot must be allocated for the uplink and downlink respectively.

Thesystemcanchangethespreadingfactor(1,2,4,8,16)asafunctionof the desired data rate. Using the burst type shown in Figure 4.29 results in datarates of 6,624, 3,312, 1,656, 828, and 414 kbit/s respectively (if all slots are usedfor data transmission). The figure shows a burst of type 2 which comprises two**data**fields of 1,104 chips each. Spreading is applied to these data fields only. Additionally,a**midample**isusedfortrainingandchannelestimation.AsTDD uses the same scrambling codes for all stations, the stations must be tightly syn-chronizedandthespreadingcodesareavailableonlyonceperslot.Thisresults in a maximum number of 16 simultaneous sending stations. To loosen the tight synchronizationalittlebit,a**guardperiod(GP)**hasbeenintroducedattheend ofeachslot.Duetothetightsynchronizationandtheuseoforthogonalcodes, a simpler power control scheme with less power control cycles (e.g., 100 per second) is sufficient.

Figure4.29
UTRATDD(TD-CDMA) framestructure



UTRA TDD occupies 5 MHz bandwidth per channel as UTRA FDD does per direction(FDDneeds2x5MHz).ComparedtothelicenseforFDD,TDDwas quitecheap.Germanypaidlessthan€300million.Figure4.22showstheloca- tion of the spectrum for this UMTS mode, but it is unclear to what extend this system will be deployed. The coverage per cell is even less than using FDD, UEs mustnotmovetoofast–thissoundslikethecharacteristicsofWLANswhich are currently deployed in many places.

4.4.4 UTRAN

Figure 4.30 shows the basic architecture of the UTRA network (UTRAN; 3GPP, 2002b). This consists of several**radio network subsystems (RNS)**. Each RNS is controlled by a**radio network controller (RNC)**and comprises several compo- nents that are called node B. An RNC in UMTS can be compared with the BSC; anodeBissimilartoaBTS.Each**nodeB**cancontrolseveralantennaswhich makearadiocell.Themobiledevice,UE,canbeconnectedtooneormore antennasaswillsubsequentlybeexplainedinthecontextofhandover.Each RNCisconnectedwiththecorenetwork(CN)overtheinterface**I_u**(similarto theroleoftheAinterfaceinGSM)andwithanodeBovertheinterface**I_{ub}**.A newinterface,whichhasnocounterpartinGSM,istheinterface**I_{ur}**connecting twoRNCswitheachother.Theuseofthisinterfaceisexplainedtogetherwith the UMTS handover mechanisms.

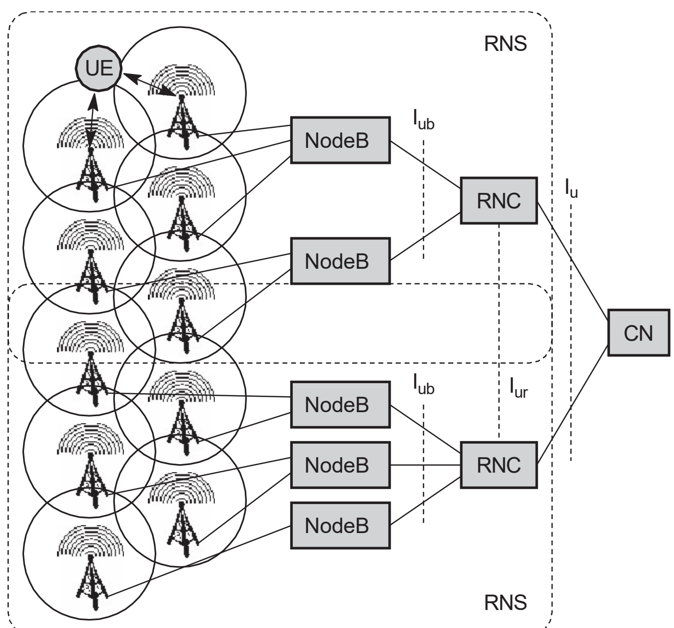


Figure4.30
Basic architecture
oftheUTRAnetwork

4.4.4.1 Radio networkcontroller

An **RNC** inUMTShasabroadspectrumoftasksaslistedinthe following:

- **Call admission control:** It is very important for CDMA systems to keep the interferencebelowacertainlevel.TheRNCcalculatesthetrafficwithin each cell and decides, if additional transmissions are acceptable or not.
- **Congestioncontrol:** Duringpacket-orienteddatatransmission,severalsta- tions share the available radio resources. The RNC allocates bandwidth to each station in a cyclic fashion and must consider the QoS requirements.
- **Encryption/decryption:** The RNC encrypts all data arriving from the fixed network before transmission over the wireless link (and vice versa).
- **ATM switching and multiplexing, protocol conversion:** Typically, the connectionsbetweenRNCs,nodeBs,andtheCNarebasedonATM.An RNC has to switch the connections to multiplex different data streams. Several protocols have to be converted – this is explained later.
- **Radio resource control:** The RNC controls all radio resources of the cells connected to it via a node B. This task includes interference and load measurements. The priorities of different connections have to be obeyed.
- **Radio bearer setup and release:** An RNC has to set-up, maintain, and release a logical data connection to a UE (the so-called UMTS radio bearer).
- **Code allocation:** The CDMA codes used by a UE are selected by the RNC. These codes may vary during a transmission.
- **Powercontrol:** TheRNConlyperformsarelativelyloosepowercontrol (the outer loop). This means that the RNC influences transmission powerbased on interference values from other cells or even other RNCs. But this isnotthetightandfastpowercontrolperformed1,500timespersecond.This iscarriedoutbyanodeB.Thisouterloopofpowercontrolhelpstomini- mize interference between neighbouring cells or controls the size of a cell.
- **Handover control and RNS relocation:** Dependingonthesignalstrengths received by UEs and node Bs, an RNC can decide if another cell would be better suited for a certain connection. If the RNC decides for handover it informs the new cell and the UE as explained in subsection 4.4.6. If a UE moves further out of the range of one RNC, a new RNC responsible for the UE has to be chosen. This is called RNS relocation.
- **Management:** Finally, the network operator needs a lot of information regardingthecurrentload,currenttraffic,errorstatesetc.tomanageitsnet- work. The RNC provides interfaces for this task, too.

4.4.4.2 NodeB

ThenamenodeBwaschosenduringstandardizationuntilanewandbetter namewasfound.However,noonecameupwithanythingbettersoit remained.AnodeBconnectstooneormoreantennascreatingoneormore cells(orsectorsinGSMspeak),respectively.ThecellscaneitheruseFDDorTDD orboth.AnimportanttaskofanodeBistheinnerlooppowercontroltomiti- gate near-far effects. This node also measures connection qualities and signal strengths.AnodeBcanevensupportaspecialcaseofhandover,aso-called softerhandoverwhichtakesplacebetweendifferentantennasofthesamenode B (see section 4.4.6).

4.4.4.3 Userequipment

TheUEshowninFigure4.30isthecounterpartofseveralnodesofthe architecture.

- AsthecounterpartofanodeB,theUEperformssignalqualitymeasurements, inner loop power control, spreading and modulation, and rate matching.
- As a counterpart of the RNC, the UE has to cooperate during handover and cell selection, performs encryption and decryption, and participates in the radio resource allocation process.
- As a counterpart of the CN, the UE has to implement mobility management functions, performs bearer negotiation, or requests certain servicesfrom the network.

This list of tasks of a UE, which is not at all exhaustive, already shows the complexity such a device has to handle. Additionally, users also want to have organizers,games,cameras,operatingsystemsetc.andthestand-bytimeshould be high.

4.4.5 Corenetwork

Figure 4.31 shows a high-level view of the UMTS release 99 core network architecturetogetherwithaUTRANRNSandaGSMBSS(seesection4.1).Thisshows the evolution from GSM/GPRS to UMTS. The core network (CN) shown here is basically the same as already explained in the context of GSM (see Figure 4.4) andGPRS(seeFigure4.16).The**circuit switched domain (CSD)**comprisesthe classicalcircuitswitchedservicesincludingsignaling.Resourcesarereservedat connectionsetupandtheGSMcomponentsMSC,GMSC,andVLRareused. The CSD connects to the RNS via a part of the $I_u$interface called **I_uCS**. The CSD components can still be part of a classical GSM network connected to a BSS but need additional functionalities (new protocols etc.).

The**packet switched domain (PSD)**uses the GPRS components SGSN and GGSN and connects to the RNS via the**I_uPS**part of the $I_u$interface. Bothdomainsneedthedata-basesEIRforequipmentidentificationandHLRforlocationmanagement(includingtheAuCforauthenticationandGRforuser specific GPRS data).

Reusing the existing infrastructure helps to save a lot of money and may convincemanyoperatorstouseUMTSiftheyalreadyuseGSM.TheUMTS industrypushestheirtechnologywiththehelpofthemarketdominanceof GSM.Thisisbasicallythesameascdma2000,whichisaevolutionofcdmaOne. Therealflexiblecorenetworkcomeswithreleases5and6,wheretheGSM

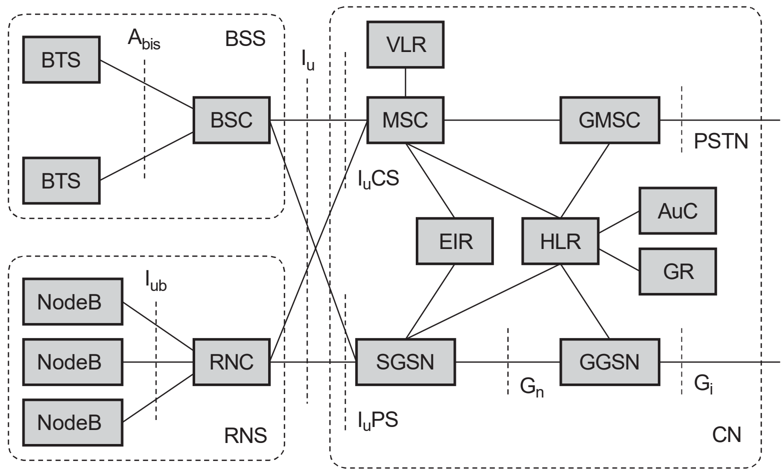


Figure4.31
UMTS core network together witha3GRNS anda2GBSS

circuit switched part is being replaced by an all-IP core. Chapter 11 presents this ideainthecontextof4Gnetworks.Itisnotyetclearwhenthisreplacementof GSMwilltakeplaceasmanyquestionsarestillopen(qualityofserviceand security being the most important).

Figure 4.32 shows the protocol stacks of the users planes of the circuit switched and packet switched domains, respectively. The**CSD**uses the**ATM adaptation layer 2**(AAL2) for user data transmission on top of ATM as transport technology. The RNC in the UTRAN implements the radio link control(RLC) and the MAC layer, while the physical layer is located in the node B. The AAL2**segmentation and reassembly**layer (SAR) is, for example, used to segment data packets received from the RLC into small chunks which can be transported in ATM. AAL2 and ATM has been chosen, too, because these protocols can transport and multiplex low bit rate voice data streams with low jitter and latency (compared to the protocols used in the PSD).

In the**PSD**several more protocols are needed. Basic data transport is performed by different lower layers (e.g., ATM with AAL5, frame relay). On top ofthese lower layers UDP/IP is used to create a UMTS internal IP network. Allpackets (e.g., IP, PPP) destined for the UE are encapsulated using the**GPRS tun- neling protocol (GTP)**. The RNC performs protocol conversion from the combination GTP/UDP/IP into the **packet data convergence protocol (PDCP)**. This protocol performs header compression to avoid redundant data transmis-sion using scarce radio resources. Comparing Figure 4.32 with Figure 4.17 (GPRS protocolreferencemodel)showsadifferencewithrespecttothetunnel.In UMTStheRNChandlesthetunnelingprotocolGTP,whileinGSM/GPRSGTPis used between an SGSN and GGSN only. The BSC in GSM is not involved in IPprotocol processing.

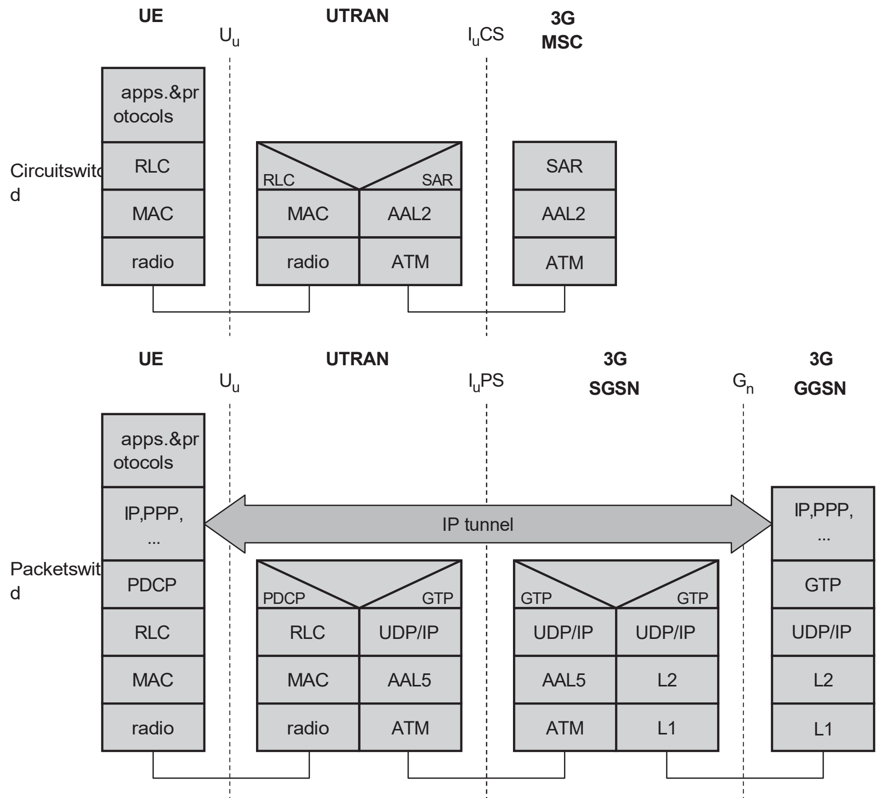


Figure4.32 Userplaneprotocol stacks (circuit and packet switched)

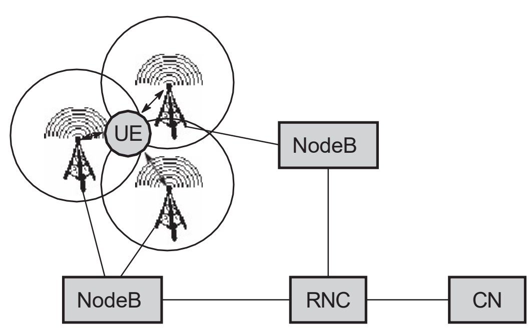
The**radio layer**(physical layer) depends on the UTRA mode (see sections 4.4.3.1and4.4.3.2).The**mediumaccesscontrol(MAC)**layercoordinates mediumaccessandmultiplexeslogicalchannelsontotransportchannels.The MAClayersalsohelptoidentifymobiledevicesandmayencryptdata.The **radio link control (RLC)**layer offers three different transport modes. The **acknowledged mode** transfer uses ARQ for error correction and guarantees one- timein-orderdeliveryofdatapackets.The**unacknowledgedmode**transfer doesnotperformARQbutguaranteesatleastone-timedeliveryofpacketswith the help of sequence numbers. The**transparent mode**transfer simply forwards MACdatawithoutanyfurtherprocessing.Thesystemthenhastorelyonthe FECwhichisalwaysusedintheradiolayer.TheRLCalsoperformssegmenta- tion and reassembly and flow control. For certain services the RLC also encrypts.

4.4.6 Handover

UMTSknowstwobasicclassesofhandovers:

- **Hard handover:** This handover type is already known from GSM and other TDMA/FDMA systems. Switching between different antennas or different systems is performed at a certain point in time.**UTRA TDD**can only usethis type. Switching between TDD cells is done between the slots of different frames.**Interfrequencyhandover**,i.e.,changingthecarrierfrequency, is a hard handover. Receiving data at different frequencies at the same time requiresamorecomplexreceivercomparedtoreceivingdatafromdifferent sourcesatthesamecarrierfrequency.Typically,all**inter system handovers** are hard handovers in UMTS. This includes handovers to and from GSM or other IMT-2000 systems. A special type of handover is the handover toa satellitesystem(inter-segmenthandover),whichisalsoahardhandover,as different frequencies are used. However, it is unclear what technology will be used for satellite links if it will ever come. To enable a UE to listen into GSM or other frequency bands, UMTS specifies a**compressed mode**transmission for UTRA FDD. During this mode a UE stops all transmission. To avoid data loss, either the spreading factor can be lowered before and after thebreakintransmission(i.e.,moredatacanbesentinshortertime)orless data is sent using different coding schemes.
- **Softhandover:**ThisistherealnewmechanisminUMTScomparedtoGSM and is only available in the FDD mode. Soft handovers are well known from traditionalCDMAnetworksastheyuse**macrodiversity**,abasicpropertyof CDMA. As shown in Figure 4.33, a UE can receive signals from up to three different antennas, which may belong to different node Bs. Towards the UE the RNC splits the data stream and forwards it to the node Bs. The UE combinesthereceiveddataagain.Intheotherdirection,theUEsimplysendsits data which is then received by all node Bs involved. The RNC combines the data streams received from the node Bs. The fact that a UE receives data from different antennas at the same time makes a handover soft. Moving from one cell to another is a smooth, not an abrupt process.

Figure4.33
Marco-diversity supportingsoft handovers



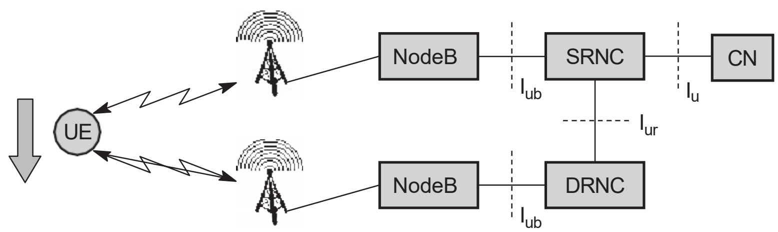


Figure 4.34 ServingRNCand drift RNC

Macro-diversity makes the transmission more robust with respect to fast fading,multi-pathpropagation,andshading.Ifonepathisblockedbyanob- staclethechancesaregoodthatdatacanstillbereceivedusinganother antenna.DuringasofthandoveraUEreceivespowercontrolcommandsfrom all involved node Bs. The UE then lowers transmission power as long as it still receives a command to lower the power. This avoids interference if, for example,the UE is in the transmission area of two antennas, one close, one further away. WithouttheabovemechanismtheUE'ssignalmaybetoostrongwhenlisten- ingtotheantennafurtheraway.ThelowertheinterferenceaUEintroduces intoacell,thehigherthecapacity.Withoutthiscontrol,cellbreathingwould beevenmoreproblematicthanitalreadyisinCDMAnetworks.

As soft handover is not supported by the CN, all mechanisms related to soft handovermustbelocatedwithinUTRAN.Figure4.34showsasituationwherea softhandoverisperformedbetweentwonodeBsthatdonotbelongtothesame RNC. In this case one RNC controls the connection and forwards all data to and from the CN. If the UE moves in the example from the upper cell to the lower cell, the upper RNC acts as a**serving RNC (SRNC)**while the other is the**drift RNC(DRNC)**.(IfthewholeRNSisconsidered,thetermsareservingRNSand driftRNS,respectively.)TheSRNCforwardsdatareceivedfromtheCNtoits nodeBandtotheDRNCviatheI_{ur}interface(splitting).Thismechanismdoes notexistin,e.g.,GSM.DatareceivedbythelowernodeBisforwardedbythe DRNC to the SRNC. The SRNC combines both data streams and forwards asingle stream of data to the CN. The CN does not notice anything from the simultaneous reception. If the UE moves further down and drops out of the transmission area of the upper node B, two RNCs reserve resources for data transmission, SRNC and DRNC, although none of SRNC's node Bs transmit data forthisUE.Toavoidwastingresources,SRNCrelocationcanbeperformed.This involves the CN so is a hard handover.

Figure 4.35 gives an overview of several common handover types in a com- bined UMTS/GSM network (UMTS specifies ten different types which include soft and hard handover). The combination of a UTRA-FDD/GSM device will be themostcommoncaseinthebeginningascoverageof3Gnetworkswillbe poor.

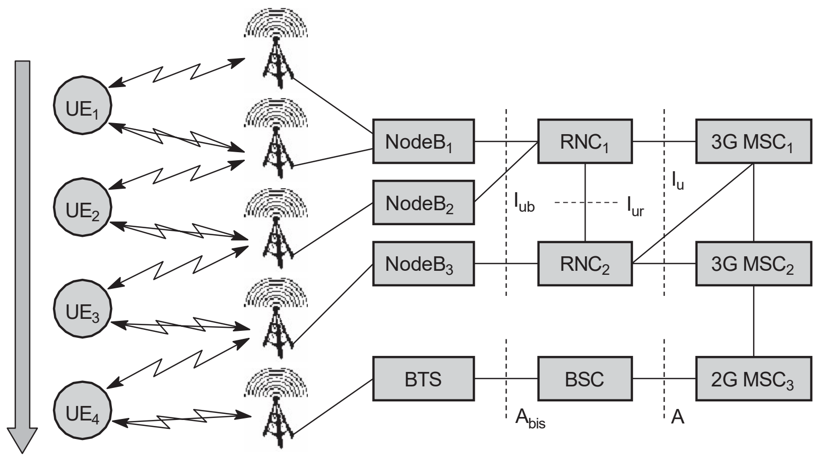


Figure4.35 Overview of different handover types

- **Intra-node B, intra-RNC:**UE_1 moves from one antenna of node B_1 to another antenna. This type of handover is called**softer handover**. In this case node B_1 performs combining and splitting of the data streams.
- **Inter-node B, intra-RNC:**UE_2 moves from node B_1 to node B_2. In this case RNC_1 supports the soft handover by combining and splitting data.
- **Inter-RNC:**WhenUE$_3$movesfromnodeB$_2$tonodeB$_3$twodifferenttypes ofhandovercantakeplace.The**internalinter-RNC**handoverisnotvisible for the CN, as described in Figure 4.34. RNC_1 can act as SRNC, RNC_2 will be the DRNC. The CN will communicate via the same interface $I_u$all the time.As soon as a relocation of the interface $I_u$takes place (relocation of the controlling RNC), the handover is called an**external inter-RNC**handover. Communication is still handled by the same MSC_1, but the externalhandover is now a hard handover.
- **Inter-MSC:**It could be also the case that MSC_2 takes over and performs a hard handover of the connection.
- **Inter-system:**UE_4 moves from a 3G UMTS network into a 2G GSM network. This hard handover is important for real life usability of the system due to the limited 3G coverage in the beginning.

4.5 Summary

This chapter has, for the most part, presented GSM as the most successful second generation digital cellular network. Although GSM was primarily designed for voice transmission, the chapter showed the evolution toward a more data-oriented transfer via HSCSD and GPRS. This evolution also includes thetransitionfromacircuit-switchednetworktoapacket-switchedsystemthat comes closer to the internet model. Other systems presented include DECT, the digitalstandardforcordlessphones,andTETRA,atrunkedradiosystem.DECT can be used for wireless data transmission on a campus or indoors, but also for wireless local loops (WLL). For special scenarios, e.g., emergencies, trunked radio systemssuchasTETRAcanbethebestchoice.Theyofferafastconnection setup(evenwithincommunicationgroups)andcanworkinanadhocnetwork, i.e., without a base station.

The situation in the US is different from Europe. Based on the analog AMPS system,theUSindustrydevelopedtheTDMAsystemIS-54thataddsdigitaltrafficchannels.IS-54usesdualmodemobilephonesandincorporatesseveralGSM ideas, such as, associated control channels, authentication procedures using encryption, and mobile assisted handover (called handoff). The Japanese PDC system was designed using many ideas in IS-54.

The next step, IS-136, includes digital control channels (IS-54 uses analog AMPS control channels) and is more efficient. Now fully digital phones can be used, several additional services are offered, e.g., voice mail, call waiting, identification,groupcalling,orSMS.IS-136isalsocalledNorthAmericanTDMA (NA-TDMA) or Digital AMPS (D-AMPS) and operates at 800 and 1,900 MHz. Enhancements of D-AMPS/IS-136 toward IMT-2000 include advanced modulation techniques for the 30 kHz radio carrier, shifting data rates up to 64 kbit/s (first phase, called 136+). The second phase, called 136HS (High Speed) comprises a new air interface specification based on the EDGE technology.

IS-95 (promoted as cdmaOne) is based on CDMA, which is a completely different medium access method. Before deployment, the system was proclaimed as having many advantages over TDMA systems, such as its much higher capacity of users per cell, e.g., 20 times the capacity of AMPS. Today, CDMA providers are making more realistic estimates of around five times as many users. IS-95 offers soft handover, avoiding the GSM ping-pong effect (Wong, 1997).However, IS-95 needs precise synchronization of all base stations (using GPS satellites which are military satellites, so are not under control of the network provider), frequent power control, and typically, dual mode mobile phones due to the limited coverage. The basic ideas of CDMA have been integrated intomost 3G systems.

Thischapteralsopresentedanoverviewofcurrentandfuturethirdgeneration systems. UMTS, a proposal of operators and companies involved in theGSMbusiness,wasdiscussedinmoredetail.Thisstandardismoreanevolutionary approach than a revolution. To avoid even higher implementation costs, UMTS tries to reuse as much infrastructure as possible while introducing new services and higher data rates based on CDMA technology. The initial installations will basically use the GSM/GPRS infrastructure and offer only moderate data rates. The initial capacity of a UMTS cell is approximately 2 Mbit/s; cell diameters are in the order of 500 m. UMTS will be used to offload GSM networks and to offer enhanced data rates in cities as a first step. Future releases aim to replace the infrastructure by an (almost) all-IP network. These ideas will be presented together with a look at fourth generation systems in chapter 11. It is quite clear that it will take a long time before 3G services are available in many places. It took GSM 10 years to become the most successful 2G mobile communication system. A similar period of time will be needed for 3G systems to succeed. Meanwhile, customers will need multiple mode phones offering,e.g., GSM 900/1800/1900 and UMTS UTRA-FDD services. It is not clear if and when UTRA-TDD will succeed. Providers already using cdmaOne will take the evolutionary path via cdma2000 1x toward the 3G system cdma2000 1x EV-DO. Several tests have already been conducted for 3G satellite services in the MSS spectrum (e.g., satellite based multicast, Nussli, 2002). However, right now manycompanieswillwaitbeforeinvestingmoneyinsatelliteservices(seechap- ter 5). The main problem of multi-mode systems is the inter-system handover. While this chapter introduces handover scenarios within UMTS and GSM, and betweenGSMandUMTS,evenmorecomplexscenarioscouldcomprisewireless LANs(seechapter7)orotherpacket-orientednetworks(Pahlavan,2000).

4.6 Reviewexercises

1 Name some key features of the GSM, DECT, TETRA, and UMTS systems. Which features do the systems have in common? Why have the three older different systems been specified? In what scenarios could one system replace another? What are the specific advantages of each system?

2 Whatarethemainproblemswhentransmittingdatausingwirelesssystemsthat were made for voice transmission? What are the possible steps to mitigate the problems and to raise efficiency? How can this be supported by billing?

3 WhichtypesofdifferentservicesdoesGSMoffer?Givesomeexamplesandrea- sons why these services have been separated.

4 ComparedtotheTCHsoffered,standardGSMcouldprovideamuchhigherdata rate(33.8kbit/s)whenlookingattheairinterface.Whatlowersthedatarates available to a user?

5 Name the main elements of the GSM system architecture and describe their functions.Whataretheadvantagesofspecifyingnotonlytheradiointerfacebut also all internal interfaces of the GSM system?

6 Describe the functions of the MS and SIM. Why does GSM separate the MS and SIM? How and where is user-related data represented/stored in the GSM system? How is user data protected from unauthorized access,especially over theairinterface?HowcouldthepositionofanMS(notonlythecurrentBTS)be localized? Think of the MS reports regarding signal quality.

7 Looking at the HLR/VLR database approach used in GSM – how does this architecture limit the scalability in terms of users,especially moving users?

8 WhyisanewinfrastructureneededforGPRS,butnotforHSCSD?Whichcompo- nents are new and what is their purpose?

9 WhatarethelimitationsofaGSMcellintermsofdiameterandcapacity(voice, data)forthetraditionalGSM,HSCSD,GPRS?Howcanthecapacitybeincreased?

10 What multiplexing schemes are used in GSM and for what purpose? Think of otherlayersapartfromthephysicallayer.

11 HowissynchronizationachievedinGSM?Whoisresponsibleforsynchronization and why is it so important?

12 What are the reasons for the delays in a GSM system for packet data traffic? Distinguishbetweencircuit-switchedandpacket-orientedtransmission.

13 Where and when can collisions occur while accessing the GSM system?Compare possible collisions caused by data transmission in standard GSM, HSCSD, and GPRS.

14 Whyandwhenaredifferentsignalingchannelsneeded?Whatarethedifferences?

15 Howislocalization,locationupdate,roaming,etc.doneinGSMandreflectedin the data bases? What are typical roaming scenarios?

16 Why are so many different identifiers/addresses (e.g., MSISDN, TMSI, IMSI) neededinGSM?Givereasonsanddistinguishbetweenuser-relatedandsystem-related identifiers.

17 Give reasons for a handover in GSM and the problems associated with it. What are the typical steps for handover, what types of handover can occur? Which resources need to be allocated during handover for data transmission using HSCSD or GPRS respectively? What about QoS guarantees?

18 WhatarethefunctionsofauthenticationandencryptioninGSM?Howissystem security maintained?

19 How can higher data rates be achieved in standard GSM, how is this possible with the additional schemes HSCSD, GPRS, EDGE? What are the main differences of the approaches, also in terms of complexity? What problems remain even if the data rate is increased?

20 What limits the data rates that can be achieved with GPRS and HSCSD using real devices (compared to the theoretical limit in a GSM system)?

21 Using the best delay class in GPRS and a data rate of 115.2 kbit/s – how many bytesareintransitbeforeafirstacknowledgementfromthereceivercouldreach the sender (neglect further delays in the fixed network and receiver system)? Now thinkoftypicalwebtransferwith10kbyteaveragetransmissionsize–howwould astandardTCPbehaveontopofGPRS(seechapters9and10)?Thinkofcongestionavoidanceanditsrelationtotheround-triptime.Whatchangesareneeded?

22 HowmuchoftheoriginalGSMnetworkdoesGPRSneed?Whichelementsofthe network perform the data transfer?

23 What are typical data rates in DECT? How are they achieved considering the TDMAframes?WhatmultiplexingschemesareappliedinDECTandforwhatpurposes? Compare the complexity of DECT with that of GSM.

24 Whowouldbethetypicalusersofatrunkedradiosystem?Whatmakestrunked radio systems particularly attractive for these user groups? What are the main differencestoexistingsystemsforthatpurpose?Whyaretrunkedradiosystems cheaper compared to, e.g., GSM systems for their main purposes?

25 Summarizethemainfeaturesofthirdgenerationmobilephonesystems.Howdo they achieve higher capacities and higher data rates? How does UMTS imple- ment asymmetrical communication and different data rates?

26 Compare the current situation of mobile phone networks in Europe, Japan, China,and North America. What are the main differences,what are efforts to find a common system or at least interoperable systems?

27 What disadvantage does OVSF have with respect to flexible data rates? How doesUMTSofferdifferentdatarates(distinguishbetweenFDDandTDDmode)?

28 How are different DPDCHs from different UEs within one cell distinguished in UTRA FDD?

29 Whichcomponentscanperformcombining/splittingatwhathandoversituation? WhatistheroleoftheinterfaceI_{ur}?WhycanCDMAsystemsoffersofthandover?

30 How does UTRA-FDD counteract the near-far effect? Why is this not a problem in GSM?

4.7 References

3GAmericas(2002)http://www.3gamericas.org/.

3GPP(2000)*GeneralUMTSarchitecture*,3rdGenerationPartnershipProject,3G TS 23.101 3.1.0 (2000-12).

3GPP(2002a)3rdGenerationPartnershipProject,http://www.3gpp.org/.

3GPP(2002b)*UTRANoveralldescription*,3rdGenerationPartnershipProject, 3GPP TS 25.401 V3.10.0 (2002-06).

3GPP2(2002)3rdGenerationPartnershipProject2,http://www.3gpp2.org/.

Adachi, F., Sawahashi, M., Suda, H. (1998) ‘Wideband DS-CDMA for next-generationmobile communications systems,’ *IEEE Communications Magazine*, 36(9).

Brasche, G., Walke, B. (1997) ‘Concepts, services, and protocols of the new GSM phase2+GeneralPacketRadioService,’ *IEEECommunicationsMagazine*,35(8).

Callendar,M.(1997)‘InternationalMobileTelecommunications-2000standards effortsoftheITU,’collectionofarticlesin*IEEEPersonalCommunications*,4(4).

Chen, H.-H., Fan, C.-X., Lu, W. (2002) ‘China’s Perspectives on 3G Mobile Communications and Beyond: TD-SCDMA Technology,’*IEEE Wireless Communications*, 9(2).

Dahlman, E., Gudmundson, B., Nilsson, M., Sköld, J. (1998) ‘UMTS/IMT-2000 based on wideband CDMA,’ *IEEE Communications Magazine*, 36(9).

Dasilva,J.,Ikonomou,D.,Erben,H.(1997)‘EuropeanR&Dprogramson third-generation mobile communication systems,’*IEEE Personal Communications*, 4(1).

DECT(2002),http://www.dect.ch/,http://www.dectweb.com/.

ETSI (1991a) *Bearer services supported by a GSM PLMN*, European Telecommunications Standards Institute, GSM recommendations 02.02.

ETSI(1991b)*GeneraldescriptionofaGSMPLMN*,EuropeanTelecommunications Standards Institute, GSM recommendations 01.02.

ETSI (1991c) *Subscriber Identity Modules, Functional Characteristics*, European Telecommunications Standards Institute, GSM recommendations 02.17.

ETSI (1993a)*Multiplexing and multiple access on the radio path*, European Telecommunications Standards Institute, GSM recommendations 05.02.

ETSI (1993b) *MS-BSS data link layer – general aspects*, European Telecommunications Standards Institute, GSM recommendations 04.05.

ETSI(1993c)*MS-BSSdatalinklayerspecification*,EuropeanTelecommunications Standards Institute, GSM recommendations 04.06.

ETSI (1997a)*High Speed Circuit Switched Data (HSCSD)*, Stage 1, European Telecommunications Standards Institute, GSM 02.34, V5.2.1.

ETSI(1998a)*GeneralPacketRadioService(GPRS)*;*Requirementsspecificationof GPRS*, European Telecommunications Standards Institute, TR 101 186, V6.0.0 (1998–04).

ETSI (1998b)*General Packet Radio Service (GPRS); Service description; Stage 2*, European Telecommunications Standards Institute, EN 301 344, V6.1.1 (1998–08).

ETSI (1998c)*General Packet Radio Service (GPRS); Service description; Stage 1*, European Telecommunications Standards Institute, EN 301 113, V6.1.1 (1998–11).

ETSI (1998d)*General Packet Radio Service (GPRS); Overall description of the GPRS radiointerface;Stage2*,EuropeanTelecommunicationsStandardsInstitute, TS 101 350, V6.0.1 (1998–08).

ETSI (1998e)*High Speed Circuit Switched Data (HSCSD); Stage 2*, European Telecommunications Standards Institute, TS 101 038, V5.1.0 (1998–07).

ETSI (1998f)*Universal Mobile Telecommunications System (UMTS); Concept groups for the definition of the UMTS Terrestrial Radio Access (UTRA)*, European Telecommunications Standards Institute, TR 101 397, V3.0.1 (1998–10).

ETSI (1998h)*High level requirements relevant for the definition of the UMTS Terrestrial Radio Access (UTRA) concept*, European Telecommunications Standards Institute, TR 101 398, V3.0.1 (1998–10).

ETSI (1998i)*Concept groups for the definition of the UMTS Terrestrial Radio Access (UTRA)*, European Telecommunications Standards Institute, TR 101 397,V3.0.1 (1998–10).

ETSI (1998j)*Digital Enhanced Cordless Telecommunications (DECT), Generic Access Profile(GAP)*,EuropeanTelecommunicationsStandardsInstitute,EN300 444, V1.3.2 (1998–03).

ETSI (1998k)*Digital Enhanced Cordless Telecommunications (DECT), Common Interface(CI)*,EuropeanTelecommunicationsStandardsInstitute,EN300 175, V1.4.1 (1998–02).

ETSI(1998l)*TerrestrialTrunkedRadio(TETRA),VoiceplusData(V+D)*,European TelecommunicationsStandardsInstitute,ETS300392seriesofstandards.

ETSI (1998m)*Terrestrial Trunked Radio (TETRA), Packet Data Optimized (PDO)*, EuropeanTelecommunicationsStandardsInstitute,ETS300393series of standards.

ETSI (1998n) *The ETSI UMTS Terrestrial Radio Access (UTRA) ITU-R Radio Transmission Technologies (RTT) Candidate Submission*, European Telecommunications Standards Institute.

ETSI (2002) European Telecommunications Standards Institute, http://www.etsi.org/.

Evci,C.(2001)'Optimizingandlicensingtheradiofrequencyspectrumforter-restrial3Gusers,'*AlcatelTelecommunicationsReview*,1/2001.

Goodman, D. (1997)*Wireless Personal Communications Systems*. Addison-Wesley Longman.

GSMAssociation(2002),http://www.gsmworld.com/.

GSMMoU(1998)*VisionfortheevolutionfromGSMtoUMTS*,GSMMoU Association, Permanent Reference Document, V 3.0.0.

GSM-R(2002),TheGSM-RIndustryGroup,http://www.gsm-rail.com/.

Halsall,F.(1996)*Datacommunications,computernetworksandopensystem*s. Addison-WesleyLongman.

ITU (2002) *International Mobile Telecommunications*, International Telecommunication Union, http://www.itu.int/imt/.

Nussli,C;Bertout,A.(2002)'Satellite-basedmulticastarchitectureformultimedia services in 3G mobile networks,' *Alcatel Telecommunications Review*, 2/2002.

Ojanperä,T.;Prasad,R.(1998)'Anoverviewofthird-generationwireless personal communications: A European perspective,'*IEEE Personal Communications*, 5(6).

Pahlavan, K., Krishnamurthy, P., Hatami, A., Ylianttila, M., Makela, J.-P., Pichna,R., Vallström, J. (2000) 'Handoff in Hybrid Mobile Data Networks,'*IEEE Personal Communications*, 7(2).

Pahlavan,K.,Krishnamurthy,P.(2002)*PrinciplesofWirelessNetworks*. PrenticeHall.

Rosenberg,J.,Schulzrinne,H.,Camarillo,G.,Johnston,A.,Peterson,J.,Sparks, R.,Handley,M.,Schooler,E.(2002)SIP:*SessionInitiationProtocol*,RFC3261, updated by RFC 3265.

SIPForum(2002)http://www.sipforum.com/.

Shafi, M., Sasaki, A., Jeong, D. (1998) 'IMT-2000 developments in the AsiaPacificregion,'collectionofarticles,*IEEECommunicationsMagazine*,36(9).

Stallings,W.(2002)*WirelessCommunicationsandNetworks*.PrenticeHall.

TETRA MoU (2002) TETRA Memorandum of Understanding,http://www.tetramou.com/.

Tripathi, N.D., Reed, J.H., VanLandingham, H.F. (1998) 'Handoffs in cellular systems,' *IEEEPersonalCommunications*,5(6).
UMTSForum(2002)http://www.umts-forum.org/.
Wong, D., Lim, T. (1997) 'Soft handoffs in CDMA mobile systems,' *IEEE Personal Communications*, 4(6).