CELLULAR AND MOBILE COMMUNICATIONS

(B.TECH - IV - I SEM)

(JNTUH-R15)

Prepared by:

Dr. P.G.Krishna Mohan, Professor
Mr. U Somanaidu, Asst Professor
Mr. Shashikanth Reddy, Asst Professor

ELECTRONICS AND COMMUNICATION ENGINEERING

INSTITUTE OF AERONAUTICAL ENGINEERING
(Autonomous)
Dundigal, Hyderabad - 500 043

UNIT-I
INTRODUCTION TO CELLULAR SYSTEMS

Introduction

Communication is one of the integral parts of science that has always been a focus point for exchanging information among parties at locations physically apart. After its discovery, telephones have replaced the telegrams and letters. Similarly, the term 'mobile' has completely revolutionized the communication by opening up innovative applications that are limited to one's imagination. Today, mobile communication has become the backbone of the society. All the mobile system technologies have improved the way of living. Its main plus point is that it has privileged a common mass of society. In this chapter, the evolution as well as the fundamental techniques of the mobile communication is discussed.

Evolution of Mobile Radio Communications

The first wireline telephone system was introduced in the year 1877. Mobile communication systems as early as 1934 were based on Amplitude Modulation (AM) schemes and only certain public organizations maintained such systems. With the demand for newer and better mobile radio communication systems during the World War II and the development of Frequency Modulation (FM) technique by Edwin Armstrong, the mobile radio communication systems began to witness many new changes. Mobile telephone was introduced in the year 1946. However, during its initial three and a half decades it found very less market penetration owing to high

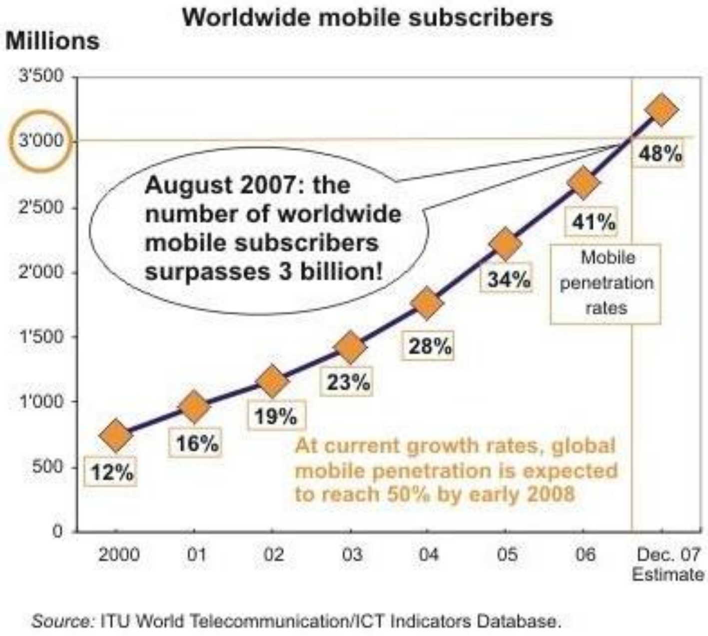


Figure 1.1: The worldwide mobile subscriber chart.

costs and numerous technological drawbacks. But with the development of the cellular concept in the 1960s at the Bell Laboratories, mobile communications began to be a promising field of expanse which could serve wider populations. Initially, mobile communication was restricted to certain official users and the cellular concept was never even dreamt of being made commercially available. Moreover, even the growth in the cellular networks was very slow. However, with the development of newer and better technologies starting from the 1970s and with the mobile users now connected to the Public Switched Telephone Network (PSTN), there has been an astronomical growth in the cellular radio and the personal communication systems. Advanced Mobile Phone System (AMPS) was the first U.S. cellular telephone system and it was deployed in 1983. Wireless services have since then been experiencing a 50% per year growth rate. The number of cellular telephone users grew from 25000 in 1984 to around 3 billion in the year 2007 and the demand rate is increasing day by day. A schematic of the subscribers is shown in Fig. 1.1.

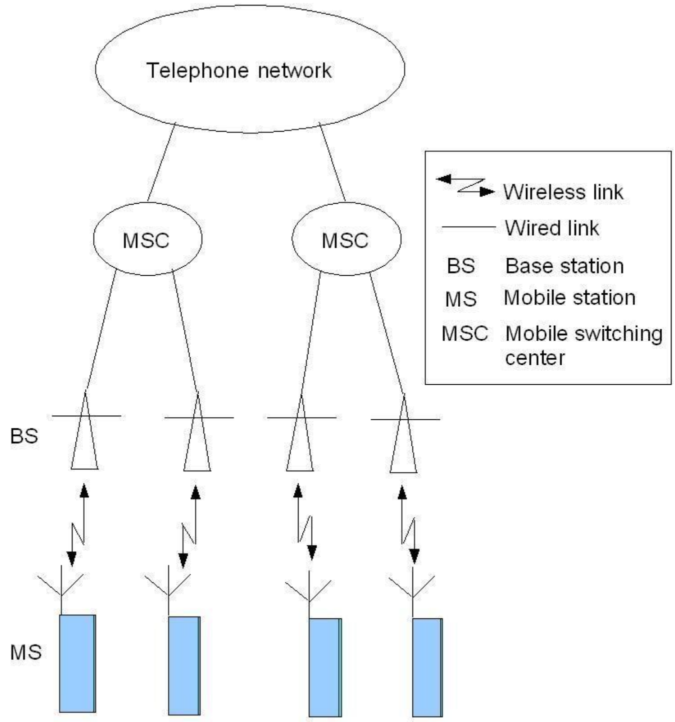


Figure 1.2: Basic mobile communication structure.

Present Day Mobile Communication

Since the time of wireless telegraphy, radio communication has been used extensively. Our society has been looking for acquiring mobility in communication since then. Initially the mobile communication was limited between one pair of users on single channel pair. The range of mobility was defined by the transmitter power, type of antenna used and the frequency of operation. With the increase in the number of users, accommodating them within the limited available frequency spectrum became a major problem. To resolve this problem, the concept of cellular communication was evolved. The present day cellular communication uses a basic unit called cell. Each cell consists of small hexagonal area with a base station located at the center of the cell which communicates with the user. To accommodate multiple users Time Division multiple Access (TDMA), Code Division Multiple Access (CDMA), Frequency Division Multiple Access (FDMA) and their hybrids are used. Numerous mobile radio standards have been deployed at various places such as AMPS, PACS,

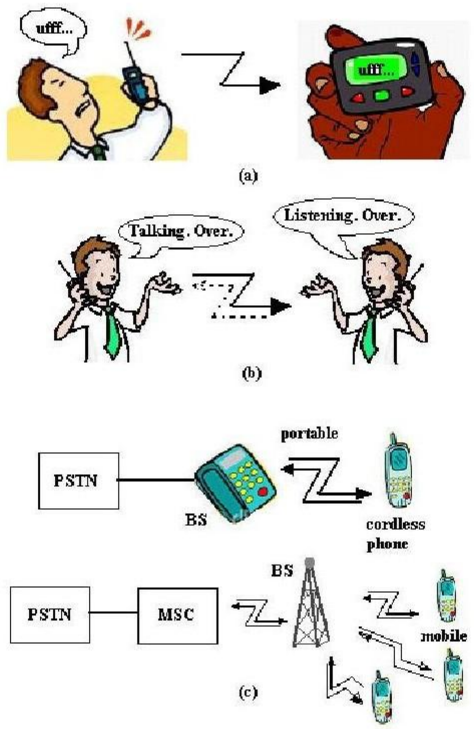


Figure 1.3: The basic radio transmission techniques: (a) simplex, (b) half duplex and (c) full duplex.

GSM, NTT, PHS and IS-95, each utilizing different set of frequencies and allocating different number of users and channels.

Fundamental Techniques

By definition, mobile radio terminal means any radio terminal that could be moved during its operation. Depending on the radio channel, there can be three differ- ent types of mobile communication. In general, however, a Mobile Station (MS) or subscriber unit communicates to a fixed Base Station (BS) which in turn com- municates to the desired user at the other end. The MS consists of transceiver, control circuitry, duplexer and an antenna while the BS consists of transceiver and channel multiplexer along with antennas mounted on the tower. The BS are also linked to a power source for the transmission of the radio signals for communication and are connected to a fixed backbone network. Figure 1.2 shows a basic mobile communication with low power transmitters/receivers at the BS, the MS and also the Mobile Switching Center (MSC). The MSC is sometimes also called Mobile Tele- phone Switching Office (MTSO). The radio signals emitted by the BS decay as the signals travel away from it. A minimum amount of signal strength is needed in order to be detected by the mobile stations or mobile sets which are the hand-held personal units (portables) or those installed in the vehicles (mobiles). The region over which the signal strength lies above such a threshold value is known as the coverage area of a BS. The fixed backbone network is a wired network that links all the base stations and also the landline and other telephone networks through wires.

Radio Transmission Techniques

Based on the type of channels being utilized, mobile radio transmission systems may be classified as the following three categories which is also shown in Fig. 1.3:

- **Simplex System:** Simplex systems utilize simplex channels i.e., the communication is unidirectional. The first user can communicate with the second user. However, the second user cannot communicate with the first user. One example of such a system is a pager.
- **Half Duplex System:** Half duplex radio systems that use half duplex radio channels allow for non-simultaneous bidirectional communication. The first user can communicate with the second user but the second user can communicate to the first user only after the first user has finished his conversation. At a time, the user can only transmit or receive information. A walkie-talkie is an example of a half duplex system which uses 'push to talk' and 'release to listen' type of switches.
- **Full Duplex System:** Full duplex systems allow two way simultaneous communications. Both the users can communicate to each other simultaneously. This can be done by providing two simultaneous but separate channels to both the users. This is possible by one of the two following methods.
  - **Frequency Division Duplexing (FDD):** FDD supports two-way radio communication by using two distinct radio channels. One frequency channel is transmitted downstream from the BS to the MS (forward channel).

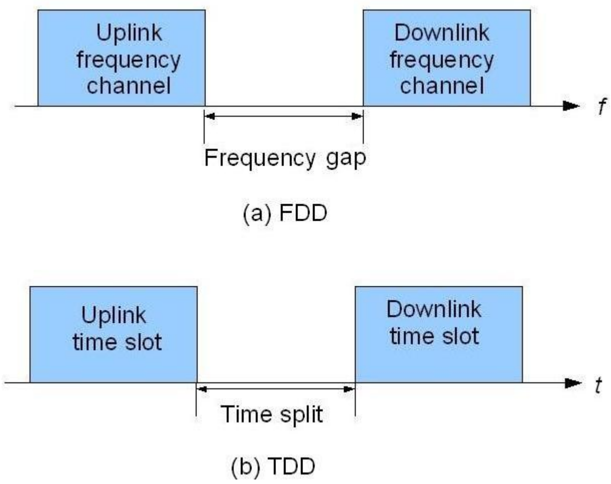


Figure 1.4: (a) Frequency division duplexing and (b) time division duplexing.

A second frequency is used in the upstream direction and supports transmission from the MS to the BS (reverse channel). Because of the pairing of frequencies, simultaneous transmission in both directions is possible. To mitigate self-interference between upstream and downstream transmis- sions, a minimum amount of frequency separation must be maintained between the frequency pair, as shown in Fig. 1.4.

– **Time Division Duplexing (TDD):** TDD uses a single frequency band to transmit signals in both the downstream and upstream directions. TDD operates by toggling transmission directions over a time interval. This toggling takes place very rapidly and is imperceptible to the user.

A full duplex mobile system can further be subdivided into two category: a single MS for a dedicated BS, and many MS for a single BS. Cordless telephone systems are full duplex communication systems that use radio to connect to a portable handset to a single dedicated BS, which is then connected to a dedi- cated telephone line with a specific telephone number on the Public Switched Telephone Network (PSTN). A mobile system, in general, on the other hand, is the example of the second category of a full duplex mobile system where many users connect among themselves via a single BS.

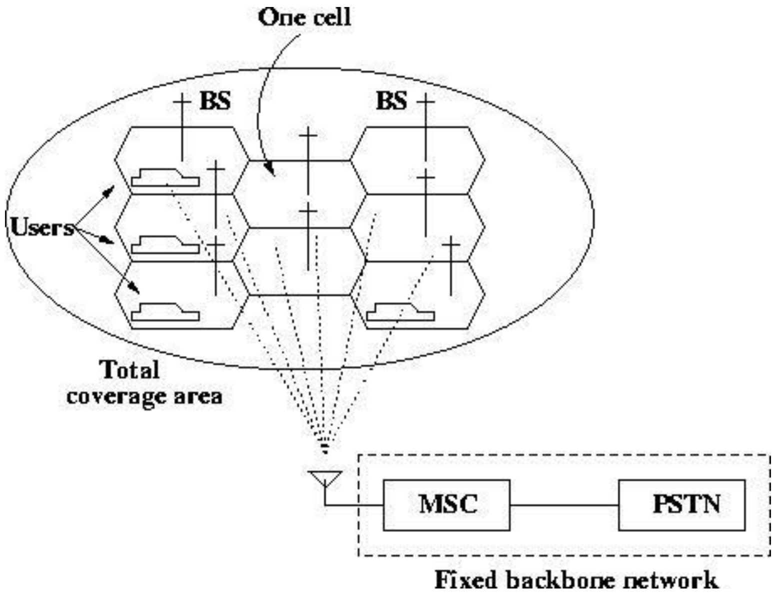


Figure 1.5: Basic Cellular Structure.

How a Mobile Call is Actually Made?

In order to know how a mobile call is made, we should first look into the basics of cellular concept and main operational channels involved in making a call. These are given below.

Cellular Concept

Cellular telephone systems must accommodate a large number of users over a large geographic area with limited frequency spectrum, i.e., with limited number of channels. If a single transmitter/ receiver is used with only a single base station, then sufficient amount of power may not be present at a huge distance from the BS. For a large geographic coverage area, a high powered transmitter therefore has to be used. But a high power radio transmitter causes harm to environment. Mobile communication thus calls for replacing the high power transmitters by low power transmitters by dividing the coverage area into small segments, called *cells*. Each cell uses a certain number of the available channels and a group of adjacent cells together use all the available channels. Such a group is called a *cluster*. This cluster can repeat itself and hence the same set of channels can be used again and again. Each cell has a low power transmitter with a coverage area equal to the area of the cell. This technique of substituting a single high powered transmitter by several low powered transmitters to support many users is the backbone of the cellular concept.

Operational Channels

In each cell, there are four types of channels that take active part during a mobile call. These are:

- **Forward Voice Channel (FVC):** This channel is used for the voice transmission from the BS to the MS.
- **Reverse Voice Channel (RVC):** This is used for the voice transmission from the MS to the BS.
- **Forward Control Channel (FCC):** Control channels are generally used for controlling the activity of the call, i.e., they are used for setting up calls and to divert the call to unused voice channels. Hence these are also called *setup channels*. These channels transmit and receive call initiation and service request messages. The FCC is used for control signaling purpose from the BS to MS.
- **Reverse Control Channel (RCC):** This is used for the call control purpose from the MS to the BS. Control channels are usually monitored by mobiles.

Making a Call

When a mobile is idle, i.e., it is not experiencing the process of a call, then it searches all the FCCs to determine the one with the highest signal strength. The mobile then monitors this particular FCC. However, when the signal strength falls below a particular threshold that is insufficient for a call to take place, the mobile again searches all the FCCs for the one with the highest signal strength. For a particular country or continent, the control channels will be the same. So all mobiles in that country or continent will search among the same set of control channels. However, when a mobile moves to a different country or continent, then the control channels for that particular location will be different and hence the mobile will not work.

Each mobile has a *mobile identification number* (MIN). When a user wants to make a call, he sends a call request to the MSC on the reverse control channel. He also sends the MIN of the person to whom the call has to be made. The MSC then sends this MIN to all the base stations. The base station transmits this MIN and all the mobiles within the coverage area of that base station receive the MIN and match it with their own. If the MIN matches with a particular MS, that mobile sends an acknowledgment to the BS. The BS then informs the MSC that the mobile is within its coverage area. The MSC then instructs the base station to access specific unused voice channel pair. The base station then sends a message to the mobile to move to the particular channels and it also sends a signal to the mobile for ringing.In order to maintain the quality of the call, the MSC adjusts the transmitted power of the mobile which is usually expressed in dB or dBm. When a mobile moves from the coverage area of one base station to the coverage area of another base sta- tion i.e., from one cell to another cell, then the signal strength of the initial base station may not be sufficient to continue the call in progress. So the call has to be transferred to the other base station. This is called *handoff*. In such cases, in order to maintain the call, the MSC transfers the call to one of the unused voice channels of the new base station or it transfers the control of the current voice channels to the new base station.

Ex. 1: Suppose a mobile unit transmits 10 W power at a certain place. Express this power in terms of dBm.

Solution: Usually, 1 mW power developed over a 100 Ω load is equivalently called 0 dBm power. 1 W is equivalent to 0 dB, i.e., $10 \log_{10}(1W) = 0dB$. Thus, $1W = 10^3 mW = 30dBm = 0dB$. This means, $xdB = (x + 30)dBm$. Hence, $10W = 10 \log_{10}(10W) = 10dB = 40dBm$.

Ex. 2: Among a pager, a cordless phone and a mobile phone, which device would have the (i) shortest, and, (ii) longest battery life? Justify.

Solution: The ‘pager’ would have the longest and the ‘mobile phone’ would have the shortest battery life. (justification is left on the readers)

Future Trends

Tremendous changes are occurring in the area of mobile radio communications, so much so that the mobile phone of yesterday is rapidly turning into a sophisticated mobile device capable of more applications than PCs were capable of only a few years ago. Rapid development of the Internet with its new services and applications has created fresh challenges for the further development of mobile communication systems. Further enhancements in modulation schemes will soon increase the Internet access rates on the mobile from current 1.8 Mbps to greater than 10 Mbps. Bluetooth is rapidly becoming a common feature in mobiles for local connections.

The mobile communication has provided global connectivity to the people at a lower cost due to advances in the technology and also because of the growing competition among the service providers. We would review certain major features as well as standards of the mobile communication till the present day technology in the next chapter.

Modern Wireless Communication Systems

At the initial phase, mobile communication was restricted to certain official users and the cellular concept was never even dreamt of being made commercially available. Moreover, even the growth in the cellular networks was very slow. However, with the development of newer and better technologies starting from the 1970s and with the mobile users now connected to the PSTN, there has been a remarkable growth in the cellular radio. However, the spread of mobile communication was very fast in the 1990s when the government throughout the world provided radio spectrum licenses for Personal Communication Service (PCS) in 1.8 - 2 GHz frequency band.

1G: First Generation Networks

The first mobile phone system in the market was AMPS. It was the first U.S. cellular telephone system, deployed in Chicago in 1983. The main technology of this first generation mobile system was FDMA/FDD and analog FM.

2G: Second Generation Networks

Digital modulation formats were introduced in this generation with the main technology as TDMA/FDD and CDMA/FDD. The 2G systems introduced three popular TDMA standards and one popular CDMA standard in the market. These are as follows:

TDMA/FDD Standards

(a) Global System for Mobile (GSM): The GSM standard, introduced by Groupe Special Mobile, was aimed at designing a uniform pan-European mobile system. It was the first fully digital system utilizing the 900 MHz frequency band. The initial GSM had 200 KHz radio channels, 8 full-rate or 16 half-rate TDMA channels per carrier, encryption of speech, low speed data services and support for SMS for which it gained quick popularity.

(b) Interim Standard 136 (IS-136): It was popularly known as North American Digital Cellular (NADC) system. In this system, there were 3 full-rate TDMA users over each 30 KHz channel. The need of this system was mainly to increase the capacity over the earlier analog (AMPS) system.

(c) Pacific Digital Cellular (PDC): This standard was developed as the counterpart of NADC in Japan. The main advantage of this standard was its low transmission bit rate which led to its better spectrum utilization.

CDMA/FDD Standard

Interim Standard 95 (IS-95): The IS-95 standard, also popularly known as CDMA- One, uses 64 orthogonally coded users and codewords are transmitted simultaneously on each of 1.25 MHz channels. Certain services that have been standardized as a part of IS-95 standard are: short messaging service, slotted paging, over-the-air activation (meaning the mobile can be activated by the service provider without any third party intervention), enhanced mobile station identities etc.

2.5G Mobile Networks

In an effort to retrofit the 2G standards for compatibility with increased throughput rates to support modern Internet application, the new data centric standards were developed to be overlaid on 2G standards and this is known as 2.5G standard.

Here, the main upgradation techniques are:

- supporting higher data rate transmission for web browsing
- supporting e-mail traffic
- enabling location-based mobile service

2.5G networks also brought into the market some popular application, a few of which are: Wireless Application Protocol (WAP), General Packet Radio Service (GPRS), High Speed Circuit Switched Dada (HSCSD), Enhanced Data rates for GSM Evolution (EDGE) etc.

3G: Third Generation Networks

3G is the third generation of mobile phone standards and technology, supersed- ing 2.5G. It is based on the International Telecommunication Union (ITU) family of standards under the International Mobile Telecommunications-2000 (IMT-2000). ITU launched IMT-2000 program, which, together with the main industry and standardization bodies worldwide, targets to implement a global frequency band that would support a single, ubiquitous wireless communication standard for all countries,to provide the framework for the definition of the 3G mobile systems.

Several radio access technologies have been accepted by ITU as part of the IMT-2000 frame- work.

3G networks enable network operators to offer users a wider range of more advanced services while achieving greater network capacity through improved spectral efficiency. Services include wide-area wireless voice telephony, video calls, and broadband wireless data, all in a mobile environment. Additional features also include HSPA data transmission capabilities able to deliver speeds up to 14.4Mbit/s on the down link and 5.8Mbit/s on the uplink.

3G networks are wide area cellular telephone networks which evolved to incorporate high-speed internet access and video telephony. IMT-2000 defines a set of technical requirements for the realization of such targets, which can be summarized as follows:

- high data rates: 144 kbps in all environments and 2 Mbps in low-mobility and indoor environments
- symmetrical and asymmetrical data transmission
- circuit-switched and packet-switched-based services
- speech quality comparable to wire-line quality
- improved spectral efficiency
- several simultaneous services to end users for multimedia services
- seamless incorporation of second-generation cellular systems
- global roaming
- open architecture for the rapid introduction of new services and technology.

3G Standards and Access Technologies

As mentioned before, there are several different radio access technologies defined within ITU, based on either CDMA or TDMA technology. An organization called 3rd Generation Partnership Project (3GPP) has continued that work by defining a mobile system that fulfills the IMT-2000 standard. This system is called Universal Mobile Telecommunications System (UMTS). After trying to establish a single 3G standard, ITU finally approved a family of five 3G standards, which are part of the 3G framework known as IMT-2000:

- W-CDMA
- CDMA2000
- TD-SCDMA

Europe, Japan, and Asia have agreed upon a 3G standard called the Universal Mobile Telecommunications System (UMTS), which is WCDMA operating at 2.1 GHz. UMTS and WCDMA are often used as synonyms. In the USA and other parts of America, WCDMA will have to use another part of the radio spectrum.

3G W-CDMA (UMTS)

WCDMA is based on DS-CDMA (direct sequencecode division multiple access) tech- nology in which user-information bits are spread over a wide bandwidth (much larger than the information signal bandwidth) by multiplying the user data with the spreading code. The chip (symbol rate) rate of the spreading sequence is 3.84 Mcps, which, in the WCDMA system deployment is used together with the 5-MHz carrier spacing. The processing gain term refers to the relationship between the signal bandwidth and the information bandwidth. Thus, the name wideband is derived to differentiate it from the 2G CDMA (IS-95), which has a chip rate of 1.2288 Mcps. In a CDMA system, all users are active at the same time on the same frequency and are separated from each other with the use of user specific spreading codes.

The wide carrier bandwidth of WCDMA allows supporting high user-data rates and also has certain performance benefits, such as increased multipath diversity. The actual carrier spacing to be used by the operator may vary on a 200-kHz grid between approximately 4.4 and 5 MHz, depending on spectrum arrangement and the interference situation.

In WCDMA each user is allocated frames of 10 ms duration, during which the user-data rate is kept constant. However, the data rate among the users can change from frame to frame. This fast radio capacity allocation (or the limits for variation in the uplink) is controlled and coordinated by the radio resource management (RRM) functions in the network to achieve optimum throughput for packet data services and to ensure sufficient quality of service (QoS) for circuit-switched users. WCDMA supports two basic modes of operation: FDD and TDD.

In the FDD mode, separate 5-MHz carrier frequencies with duplex spacing are used for the uplink and downlink, respectively, whereas in TDD only one 5-MHz carrier is time shared between the up- link and the downlink. WCDMA uses coherent detection based on the pilot symbols and/or common pilot. WCDMA allows many performance- enhancement methods to be used, such as transmit diversity or advanced CDMA receiver concepts.Table summaries the main WCDMA parameters.

The support for handovers (HO) between GSM and WCDMA is part of the first standard version. This means that all multi-mode WCDMA/GSM terminals will support measurements from the one system while camped on the other one. This allows networks using both WCDMA and GSM to balance the load between the networks and base the HO on actual measurements from the terminals for different radio conditions in addition to other criteria available.

Table 2.1: Main WCDMA parameters

Multiple access method	DS-CDMA
Duplexing method	Frequency division duplex/time division duplex
Base stationsynchronisation	Asynchronous operation
Chip rate	3.84 Mcps
Frame length	10 ms
Service multiplexing	Multiple services with different quality of service requirements multiplexed on one connection
Multi-rate concept	Variable spreading factor and multicode
Detection	Coherent using pilot symbols or common pilot
Multi-user detection, smart antennas	Supported by the standard, optional in the implementation

The world's first commercial W-CDMA service, FoMA, was launched by NTT DoCoMo in Japan in 2001. FoMA is the short name for Freedom of Mobile Mul- timedia Access, is the brand name for the 3G services being offered by Japanese mobile phone operator NTT DoCoMo. Elsewhere, W-CDMA deployments have been exclusively UMTS based.

UMTS or W-CDMA, assures backward compatibility with the second generation GSM, IS-136 and PDC TDMA technologies, as well as all 2.5G TDMA technologies. The network structure and bit level packaging of GSM data is retained by W-CDMA, with additional capacity and bandwidth provided by a new CDMA air interface.

3G CDMA2000

Code division multiple access 2000 is the natural evolution of IS-95 (cdmaOne). It includes additional functionality that increases its spectral efficiency and data rate capability.(code division multiple access) is a mobile digital radio technology where channels are defined with codes (PN sequences). CDMA permits many simultaneous transmitters on the same frequency channel. Since more phones can be served by fewer cell sites, CDMA-based standards have a significant economic advantage over TDMA- or FDMA-based standards. This standard is being developed byTelecommunications Industry Association (TIA) of US and is is standardized by 3GPP2.

The main CDMA2000 standards are: CDMA2000 1xRTT,CDMA2000 1xEV and CDMA2000 EV-DV. These are the approved radio interfaces for the ITU's IMT-2000 standard. In the following, a brief discussion about all these standards is given.

CDMA2000 1xRTT: RTT stands for Radio Transmission Technology and the designation "1x", meaning "1 times Radio Transmission Technology", indicates the same RF bandwidth as IS-95.The main features of CDMA2000 1X are as follows:

- Supports an instantaneous data rate upto 307kpbs for a user in packet mode and a typical throughput rates of 144kbps per user,depending on the number of user, the velociy of user and the propagating conditions.
- Supports up to twice as many voice users a the 2G CDMA standard
- Provides the subscriber unit with upto two times the standby time for longer lasting battery life.

CDMA2000 EV: This is an evolutionary advancement of CDMA with the following characteristics:

- Provides CDMA carriers with the option of installing radio channels with data only (CDMA2000 EV-DO) and with data and voice (CDMA2000 EV-DV)

.

- The cdma2000 1xEV-DO supports greater than 2.4Mbps of instantaneous high-speed packet throughput per user on a CDMA channel, although the user data rates are much lower and highly dependent on other factors.

- CDMA2000 EV-DV can offer data rates upto 144kbps with about twice as many voice channels as IS-95B.

CDMA2000 3x is (also known as EV-DO Rev B) is a multi-carrier evolution.

- It has higher rates per carrier (up to 4.9 Mbit/s on the downlink per carrier). Typical deployments are expected to include 3 carriers for a peak rate of 14.7 Mbit/s.Higher rates are possible by bundling multiple channels together. It enhances the user experience and enables new services such as high definition video streaming.

- Uses statistical multiplexing across channels to further reduce latency, enhancing the experience for latency-sensitive services such as gaming, video telephony, remote console sessions and web browsing.

- It provides increased talk-time and standby time.

- The interference from the adjacent sectors is reduced by hybrid frequency re-use and improves the rates that can be offered, especially to users at the edge of the cell.

- It has efficient support for services that have asymmetric download and upload requirements (i.e. different data rates required in each direction) such as file transfers, web browsing, and broadband multimedia content delivery.

3G TD-SCDMA

Time Division-Synchronous Code Division Multiple Access, or TD-SCDMA, is a 3G mobile telecommunications standard, being pursued in the People's Republic of China by the Chinese Academy of Telecommunications Technology (CATT). This proposal was adopted by ITU as one of the 3G options in late 1999. TD-SCDMA is based on spread spectrum technology.

TD-SCDMA uses TDD, in contrast to the FDD scheme used by W-CDMA. By dynamically adjusting the number of timeslots used for downlink and uplink, the system can more easily accommodate asymmetric traffic with different data rate requirements on downlink and uplink than FDD schemes. Since it does not require paired spectrum for downlink and uplink, spectrum allocation flexibility is also increased. Also, using the same carrier frequency for uplink and downlink means that the channel condition is the same on both directions, and the base station can deduce the downlink channel information from uplink channel estimates, which is helpful to the application of beamforming techniques.

TD-SCDMA also uses TDMA in addition to the CDMA used in WCDMA. This reduces the number of users in each timeslot, which reduces the implementation complexity of multiuser detection and beamforming schemes, but the non-continuous transmission also reduces coverage (because of the higher peak power needed), mo- bility (because of lower power control frequency) and complicates radio resource management algorithms.

The ”S” in TD-SCDMA stands for ”synchronous”, which means that uplink signals are synchronized at the base station receiver, achieved by continuous timing adjustments. This reduces the interference between users of the same timeslot using different codes by improving the orthogonality between the codes, therefore increasing system capacity, at the cost of some hardware complexity in achieving uplink synchronization.

Wireless Transmission Protocols

There are several transmission protocols in wireless manner to achieve different application oriented tasks. Below, some of these applications are given.

Wireless Local Loop (WLL) and LMDS

Microwave wireless links can be used to create a wireless local loop. The local loop can be thought of as the ”last mile” of the telecommunication network that resides between the central office (CO) and the individual homes and business in close proximity to the CO. An advantage of WLL technology is that once the wireless equipment is paid for, there are no additional costs for transport between the CO and the customer premises equipment. Many new services have been proposed and this includes the concept of Local Multipoint Distribution Service (LMDS), which provides broadband telecommunication access in the local exchange.

Bluetooth

- Facilitates ad-hoc data transmission over short distances from fixed and mobile devices as shown in Figure 2.1
- Uses a radio technology called frequency hopping spread spectrum. It chops up the data being sent and transmits chunks of it on up to 79 different frequencies.

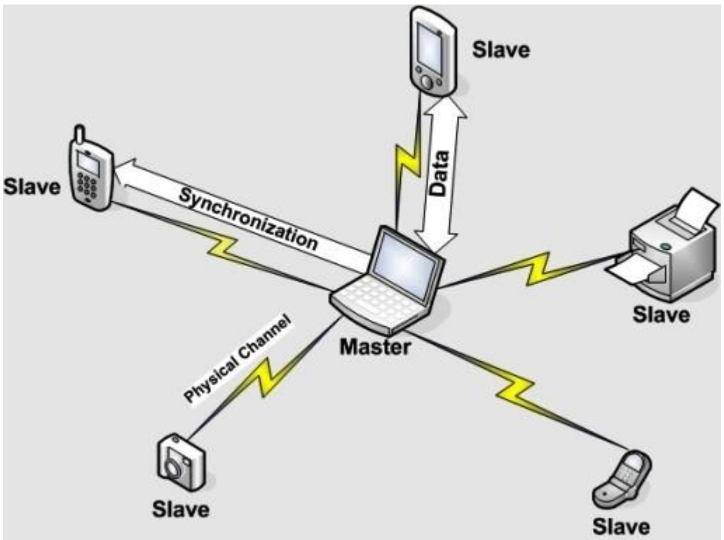


Figure 2.1: Data transmission with Bluetooth.

In its basic mode, the modulation is Gaussian frequency shift keying (GFSK). It can achieve a gross data rate of 1 Mb/s

- Primarily designed for low power consumption, with a short range (power-class-dependent: 1 meter, 10 meters, 100 meters) based on low-cost transceiver microchips in each device

Wireless Local Area Networks (W-LAN)

- IEEE 802.11 WLAN uses ISM band (5.275-5.825GHz)
- Uses 11Mcps DS-SS spreading and 2Mbps user data rates (will fallback to 1Mbps in noisy conditions)
- IEEE 802.11a stndard provides upto 54Mbps throughput in the 5GHz band. The DS-SS IEEE 802.11b has been called Wi-Fi. Wi-Fi networks have limited range. A typical Wi-Fi home router using 802.11b or 802.11g with a stock antenna might have a range of 32 m (120 ft) indoors and 95 m (300 ft) outdoors. Range also varies with frequency band.
- IEEE 802.11g uses Complementary Code Keying Orthogonal Frequency Divi- sion Multiplexing (CCK-OFDM) standards in both 2.4GHz and 5GHz bands.

WiMax

- Provides upto 70 Mb/sec symmetric broadband speed without the need for cables. The technology is based on the IEEE 802.16 standard (also called WirelessMAN)

- WiMAX can provide broadband wireless access (BWA) up to 30 miles (50 km) for fixed stations, and 3 - 10 miles (5 - 15 km) for mobile stations. In contrast, the WiFi/802.11 wireless local area network standard is limited in most cases to only 100 - 300 feet (30 - 100m)

- The 802.16 specification applies across a wide range of the RF spectrum, and WiMAX could function on any frequency below 66 GHz (higher frequencies would decrease the range of a Base Station to a few hundred meters in an urban environment).

Zigbee

- ZigBee is the specification for a suite of high level communication protocols using small, low-power digital radios based on the IEEE 802.15.4-2006 standard for wireless personal area networks (WPANs), such as wireless headphones connecting with cell phones via short-range radio.

- This technology is intended to be simpler and cheaper. ZigBee is targeted at radio-frequency (RF) applications that require a low data rate, long battery life, and secure networking.

- ZigBee operates in the industrial, scientific and medical (ISM) radio bands; 868 MHz in Europe, 915 MHz in countries such as USA and Australia, and 2.4 GHz in most worldwide.

Wibree

- Wibree is a digital radio technology (intended to become an open standard of wireless communications) designed for ultra low power consumption (button cell batteries) within a short range (10 meters / 30 ft) based around low-cost transceiver microchips in each device.

- Wibree is known as Bluetooth with low energy technology.
- It operates in 2.4 GHz ISM band with physical layer bit rate of 1 Mbps.

Conclusion: Beyond 3G Networks

Beyond 3G networks, or 4G (Fourth Generation), represent the next complete evolution in wireless communications. A 4G system will be able to provide a comprehensive IP solution where voice, data and streamed multimedia can be given to users at higher data rates than previous generations.There is no formal definition for 4G ; however, there are certain objectives that are projected for 4G. It will be capable of providing between 100 Mbit/s and 1 Gbit/s speeds both indoors and outdoors, with premium quality and high security. It would also support systems like multicarrier communication, MIMO and UWB.

UNIT 2

The Cellular Engineering Fundamentals

Introduction

In Chapter 1, we have seen that the technique of substituting a single high power transmitter by several low power transmitters to support many users is the backbone of the cellular concept. In practice, the following four parameters are most important while considering the cellular issues: system capacity, quality of service, spectrum efficiency and power management. Starting from the basic notion of a cell, we would deal with these parameters in the context of cellular engineering in this chapter.

What is a Cell?

The power of the radio signals transmitted by the BS decay as the signals travel away from it. A minimum amount of signal strength (let us say, x dB) is needed in order to be detected by the MS or mobile sets which may the hand-held personal units or those installed in the vehicles. The region over which the signal strength lies above this threshold value x dB is known as the coverage area of a BS and it must be a circular region, considering the BS to be isotropic radiator. Such a circle, which gives this actual radio coverage, is called the foot print of a cell (in reality, it is amorphous). It might so happen that either there may be an overlap between any two such side by side circles or there might be a gap between the

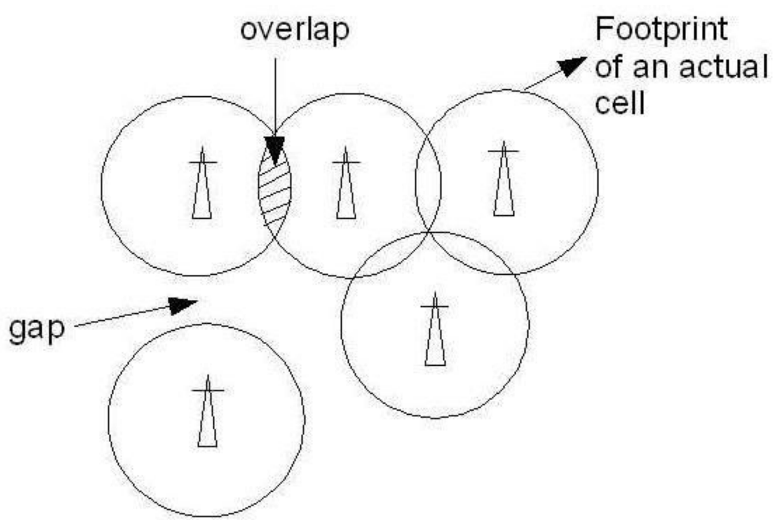


Figure 3.1: Footprint of cells showing the overlaps and gaps.

coverage areas of two adjacent circles. This is shown in Figure 3.1. Such a circular geometry, therefore, cannot serve as a regular shape to describe cells. We need a regular shape for cellular design over a territory which can be served by 3 regular polygons, namely, equilateral triangle, square and regular hexagon, which can cover the entire area without any overlap and gaps. Along with its regularity, a cell must be designed such that it is most reliable too, i.e., it supports even the weakest mobile with occurs at the edges of the cell. For any distance between the center and the farthest point in the cell from it, a regular hexagon covers the maximum area. Hence regular hexagonal geometry is used as the cells in mobile communication.

Frequency Reuse

Frequency reuse, or, frequency planning, is a technique of reusing frequencies and channels within a communication system to improve capacity and spectral efficiency. Frequency reuse is one of the fundamental concepts on which commercial wireless systems are based that involve the partitioning of an RF radiating area into cells. The increased capacity in a commercial wireless network, compared with a network with a single transmitter, comes from the fact that the same radio frequency can be reused in a different area for a completely different transmission.

Frequency reuse in mobile cellular systems means that frequencies allocated to

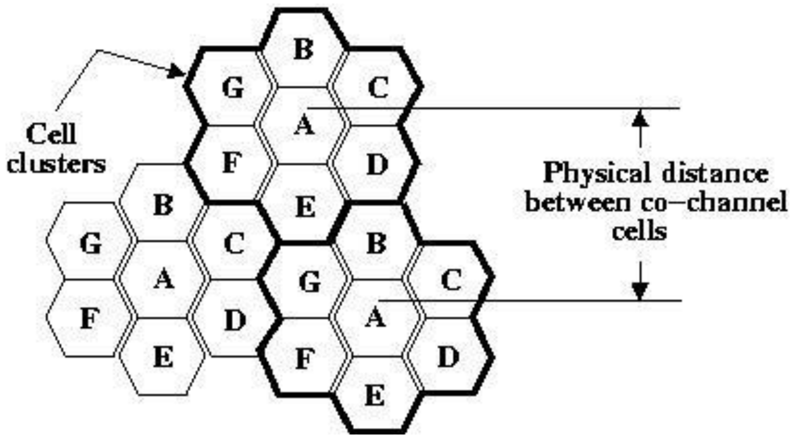


Figure 3.2: Frequency reuse technique of a cellular system.

the service are reused in a regular pattern of cells, each covered by one base station. The repeating regular pattern of cells is called cluster. Since each cell is designed to use radio frequencies only within its boundaries, the same frequencies can be reused in other cells not far away without interference, in another cluster. Such cells are called ‘co-channel’ cells. The reuse of frequencies enables a cellular system to handle a huge number of calls with a limited number of channels.

Figure 3.2 shows a frequency planning with cluster size of 7, showing the co-channels cells in different clusters by the same letter. The closest distance between the co-channel cells (in different clusters) is determined by the choice of the cluster size and the layout of the cell cluster. Consider a cellular system with S duplex channels available for use and let N be the number of cells in a cluster. If each cell is allotted K duplex channels with all being allotted unique and disjoint channel groups we have $S = KN$ under normal circumstances. Now, if the cluster are repeated M times within the total area, the total number of duplex channels, or, the total number of users in the system would be $T = MS = KMN$. Clearly, if K and N remain constant, then

$$T \propto M \tag{3.1}$$

and, if T and K remain constant, then

$$N \propto \frac{1}{M}. \tag{3.2}$$

Hence the capacity gain achieved is directly proportional to the number of times a cluster is repeated, as shown in (3.1), as well as, for a fixed cell size, small N decreases the size of the cluster with in turn results in the increase of the number of clusters (3.2) and hence the capacity. However for small N, co-channel cells are located much closer and hence more interference. The value of N is determined by calculating the amount of interference that can be tolerated for a sufficient quality communication. Hence the smallest N having interference below the tolerated limit is used. However, the cluster size N cannot take on any value and is given only by the following equation

$$N = i^2 + ij + j^2, \qquad i \geq 0, j \geq 0, \tag{3.3}$$

where i and j are integer numbers.

Ex. 1: Find the relationship between any two nearest co-channel cell distance D and the cluster size N.

Solution: For hexagonal cells, it can be shown that the distance between two adjacent cell centers = $\sqrt{3}R$, where R is the radius of any cell. The normalized co-channel cell distance D_n can be calculated by traveling 'i' cells in one direction and then traveling 'j' cells in anticlockwise 120^o of the primary direction. Using law of vector addition,

$$D_n^2 = j^2 \cos^2(30^o) + (i + j \sin(30^o))^2 \quad (3.4)$$

which turns out to be

$$D_n = \sqrt{i^2 + ij + j^2} = \sqrt{N}. \quad (3.5)$$

Multiplying the actual distance $\sqrt{3}R$ between two adjacent cells with it, we get

$$D = D_n \sqrt{3}R = \sqrt{3N}R. \quad (3.6)$$

Ex. 2: Find out the surface area of a regular hexagon with radius R, the surface area of a large hexagon with radius D, and hence compute the total number of cells in this large hexagon.

Hint: In general, this large hexagon with radius D encompasses the center cluster of N cells and one-third of the cells associated with six other peripheral large hexagons. Thus, the answer must be $N + 6(\frac{N}{3}) = 3N$.

Channel Assignment Strategies

With the rapid increase in number of mobile users, the mobile service providers had to follow strategies which ensure the effective utilization of the limited radio spectrum. With increased capacity and low interference being the prime objectives, a frequency reuse scheme was helpful in achieving this objectives. A variety of channel assignment strategies have been followed to aid these objectives. Channel assignment strategies are classified into two types: fixed and dynamic, as discussed below.

Fixed Channel Assignment (FCA)

In fixed channel assignment strategy each cell is allocated a fixed number of voice channels. Any communication within the cell can only be made with the designated unused channels of that particular cell. Suppose if all the channels are occupied, then the call is blocked and subscriber has to wait.

This is simplest of the channel assignment strategies as it requires very simple circuitry but provides worst channel utilization. Later there was another approach in which the channels were borrowed from adjacent cell if all of its own designated channels were occupied. This was named as *borrowing strategy*. In such cases the MSC supervises the borrowing pro- cess and ensures that none of the calls in progress are interrupted.

Dynamic Channel Assignment (DCA)

In dynamic channel assignment strategy channels are temporarily assigned for use in cells for the duration of the call. Each time a call attempt is made from a cell the corresponding BS requests a channel from MSC. The MSC then allocates a channel to the requesting the BS. After the call is over the channel is returned and kept in a central pool. To avoid co-channel interference any channel that in use in one cell can only be reassigned simultaneously to another cell in the system if the distance between the two cells is larger than minimum reuse distance. When compared to the FCA, DCA has reduced the likelihood of blocking and even increased the trunking capacity of the network as all of the channels are available to all cells, i.e., good quality of service. But this type of assignment strategy results in heavy load on switching center at heavy traffic condition.

Ex. 3: A total of 33 MHz bandwidth is allocated to a FDD cellular system with two 25 KHz simplex channels to provide full duplex voice and control channels. Compute the number of channels available per cell if the system uses (i) 4 cell, (ii) 7 cell, and (iii) 8 cell reuse technique. Assume 1 MHz of spectrum is allocated to control channels. Give a distribution of voice and control channels.

Solution: One duplex channel = 2 x 25 = 50 kHz of spectrum. Hence the total available duplex channels are = 33 MHz / 50 kHz = 660 in number. Among these channels, 1 MHz / 50 kHz = 20 channels are kept as control channels.

(a) For N = 4, total channels per cell = 660/4 = 165.

Among these, voice channels are 160 and control channels are 5 in number.

(b) For N = 7, total channels per cell are 660/7 ≈ 94. Therefore, we have to go for a more exact solution. We know that for this system, a total of 20 control channels and a total of 640 voice channels are kept. Here, 6 cells can use 3 control channels and the rest two can use 2 control channels each. On the other hand, 5 cells can use 92 voice channels and the rest two can use 90 voice channels each. Thus the total solution for this case is:

6 x 3 + 1 x 2 = 20 control channels, and,

5 x 92 + 2 x 90 = 640 voice channels.

This is one solution, there might exist other solutions too.

(c) The option N = 8 is not a valid option since it cannot satisfy equation (3.3) by two integers i and j.

Handoff Process

When a user moves from one cell to the other, to keep the communication between the user pair, the user channel has to be shifted from one BS to the other without interrupting the call, i.e., when a MS moves into another cell, while the conversation is still in progress, the MSC automatically transfers the call to a new FDD channel without disturbing the conversation. This process is called as *handoff*. A schematic diagram of handoff is given in Figure 3.3.

Processing of handoff is an important task in any cellular system. Handoffs must be performed successfully and be imperceptible to the users. Once a signal

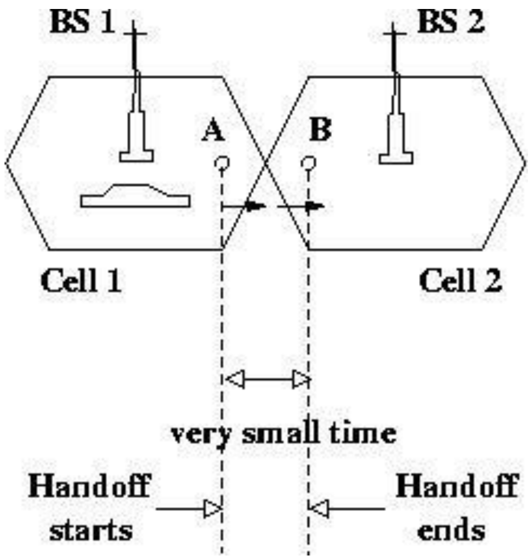


Figure 3.3: Handoff scenario at two adjacent cell boundary.

level is set as the minimum acceptable for good voice quality (P_{rmin}), then a slightly stronger level is chosen as the threshold (P_{rH})at which handoff has to be made, as shown in Figure 3.4. A parameter, called power margin, defined as

$$\Delta = P_{rH} - P_{rmin} \tag{3.7}$$

is quite an important parameter during the handoff process since this margin Δ can neither be too large nor too small. If Δ is too small, then there may not be enough time to complete the handoff and the call might be lost even if the user crosses the cell boundary.

If Δ is too high o the other hand, then MSC has to be burdened with unnecessary handoffs. This is because MS may not intend to enter the other cell. Therefore Δ should be judiciously chosen to ensure imperceptible handoffs and to meet other objectives.

Factors Influencing Handoffs

The following factors influence the entire handoff process:

(a) Transmitted power: as we know that the transmission power is different for different cells, the handoff threshold or the power margin varies from cell to cell.

(b) Received power: the received power mostly depends on the Line of Sight (LoS) path between the user and the BS. Especially when the user is on the boundary of

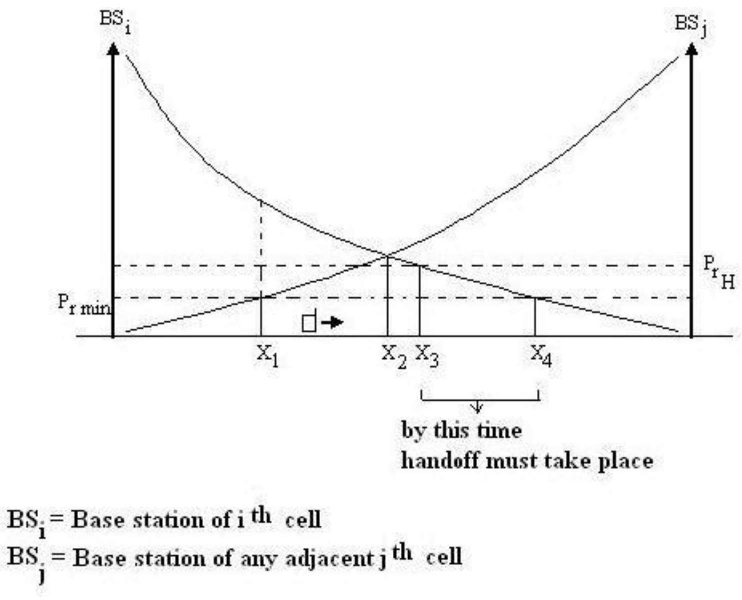


Figure 3.4: Handoff process associated with power levels, when the user is going from i-th cell to j-th cell.

the two cells, the LoS path plays a critical role in handoffs and therefore the power margin Δ depends on the minimum received power value from cell to cell.

(c) Area and shape of the cell: Apart from the power levels, the cell structure also a plays an important role in the handoff process.

(d) Mobility of users: The number of mobile users entering or going out of a particular cell, also fixes the handoff strategy of a cell.

To illustrate the reasons (c) and (d), let us consider a rectangular cell with sides R_1 and R_2 inclined at an angle θ with horizon, as shown in the Figure 3.5. Assume N_1 users are having handoff in horizontal direction and N_2 in vertical direction per unit length.

The number of crossings along R_1 side is : $(N_1 cos\theta + N_2 sin\theta)R_1$ and the number of crossings along R_2 side is : $(N_1 sin\theta + N_2 cos\theta)R_2$.

Then the handoff rate λ_H can be written as

$$\lambda_H = (N_1 cos\theta + N_2 sin\theta)R_1 + (N_1 sin\theta + N_2 cos\theta)R_2. \quad (3.8)$$

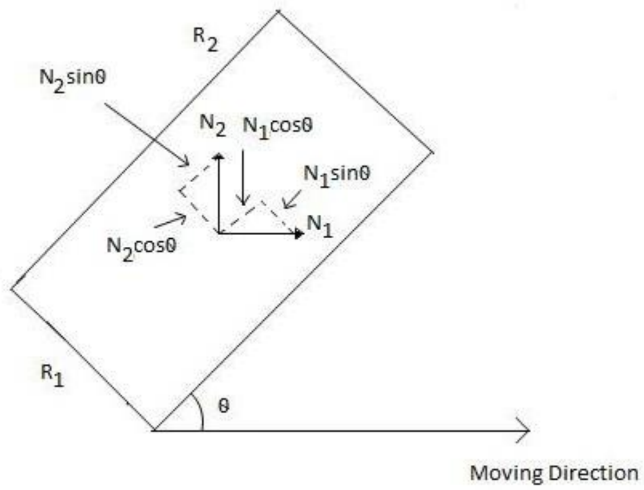


Figure 3.5: Handoff process with a rectangular cell inclined at an angle θ.

Now, given the fixed area $A = R_1 R_2$, we need to find λ_H^{min} for a given θ. Replacing R_1 by $\frac{A}{R_2}$ and equating $\frac{d\lambda_H}{dR_1}$ to zero, we get

$$R_1^2 = A\left(\frac{N_1 sin\theta + N_2 cos\theta}{N_1 cos\theta + N_2 sin\theta}\right). \quad (3.9)$$

Similarly, for R_2, we get

$$R_2^2 = A\left(\frac{N_1 cos\theta + N_2 sin\theta}{N_1 sin\theta + N_2 cos\theta}\right). \quad (3.10)$$

From the above equations, we have $\lambda_H = 2\sqrt{A(N_1 N_2 + (N_1^2 + N_2^2) cos\theta sin\theta)}$ which means it it minimized at $\theta = 0^o$. Hence $\lambda_H^{min} = 2\sqrt{AN_1 N_2}$. Putting the value of θ in (3.9) or (3.10), we have $\frac{R_1}{R_2} = \frac{N_1}{N_2}$. This has two implications: (i) that handoff is minimized if rectangular cell is aligned with X-Y axis, i.e., $\theta = 0^o$, and, (ii) that the number of users crossing the cell boundary is inversely proportional to the dimension of the other side of the cell. The above analysis has been carried out for a simple square cell and it changes in more complicated way when we consider a hexagonal cell.

Handoffs In Different Generations

In 1G analog cellular systems, the signal strength measurements were made by the BS and in turn supervised by the MSC. The handoffs in this generation can be termed as Network Controlled Hand-Off (NCHO). The BS monitors the signal strengths of voice channels to determine the relative positions of the subscriber. The special receivers located on the BS are controlled by the MSC to monitor the signal strengths of the users in the neighboring cells which appear to be in need of handoff. Based on the information received from the special receivers the MSC decides whether a handoff is required or not. The approximate time needed to make a handoff successful was about 5-10 s. This requires the value of Δ to be in the order of 6dB to 12dB.

In the 2G systems, the MSC was relieved from the entire operation. In this generation, which started using the digital technology, handoff decisions were mobile assisted and therefore it is called Mobile Assisted Hand-Off (MAHO). In MAHO, the mobile center measures the power changes received from nearby base stations and notifies the two BS. Accordingly the two BS communicate and channel transfer occurs. As compared to 1G, the circuit complexity was increased here whereas the delay in handoff was reduced to 1-5 s. The value of Δ was in the order of 0-5 dB. However, even this amount of delay could create a communication pause.

In the current 3G systems, the MS measures the power from adjacent BS and automatically upgrades the channels to its nearer BS. Hence this can be termed as Mobile Controlled Hand-Off (MCHO). When compared to the other generations, delay during handoff is only 100 ms and the value of Δ is around 20 dBm. The Quality Of Service (QoS) has improved a lot although the complexity of the circuitry has further increased which is inevitable.

All these types of handoffs are usually termed as hard handoff as there is a shift in the channels involved. There is also another kind of handoff, called soft handoff, as discussed below.

Handoff in CDMA: In spread spectrum cellular systems, the mobiles share the same channels in every cell. The MSC evaluates the signal strengths received from different BS for a single user and then shifts the user from one BS to the other without actually changing the channel. These types of handoffs are called as soft handoff as there is no change in the channel.

Handoff Priority

While assigning channels using either FCA or DCA strategy, a guard channel concept must be followed to facilitate the handoffs. This means, a fraction of total available channels must be kept for handoff requests. But this would reduce the carried traffic and only fewer channels can be assigned for the residual users of a cell. A good solution to avoid such a dead-lock is to use DCA with handoff priority (demand based allocation).

A Few Practical Problems in Handoff Scenario

(a) Different speed of mobile users: with the increase of mobile users in urban areas, microcells are introduced in the cells to increase the capacity (this will be discussed later in this chapter). The users with high speed frequently crossing the micro-cells become burdened to MSC as it has to take care of handoffs. Several schemes thus have been designed to handle the simultaneous traffic of high speed and low speed users while minimizing the handoff intervention from the MSC, one of them being the 'Umbrella Cell' approach. This technique provides large area coverage to high speed users while providing small area coverage to users traveling at low speed. By using different antenna heights and different power levels, it is possible to provide larger and smaller cells at a same location. As illustrated in the Figure 3.6, umbrella cell is co-located with few other microcells. The BS can measure the speed of the user by its short term average signal strength over the RVC and decides which cell to handle that call. If the speed is less, then the corresponding microcell handles the call so that there is good corner coverage. This approach assures that handoffs are minimized for high speed users and provides additional microcell channels for pedestrian users.

(b) Cell dragging problem: this is another practical problem in the urban area with additional microcells. For example, consider there is a LOS path between the MS and BS1 while the user is in the cell covered by BS2. Since there is a LOS with the BS1, the signal strength received from BS1 would be greater than that received from BS2. However, since the user is in cell covered by BS2, handoff cannot take place and as a result, it experiences a lot of interferences.

This problem can be solved by judiciously choosing the handoff threshold along with adjusting the coverage area.

(c) Inter-system handoff: if one user is leaving the coverage area of one MSC and is entering the area of another MSC, then the call might be lost if there is no handoff in this case too. Such a handoff is called inter-system handoff and in order to facilitate this, mobiles usually have roaming facility.

Interference & System Capacity

Susceptibility and interference problems associated with mobile communications equipment are because of the problem of time congestion within the electromagnetic spectrum. It is the limiting factor in the performance of cellular systems. This interference can occur from clash with another mobile in the same cell or because of a call in the adjacent cell. There can be interference between the base stations operating at same frequency band or any other non-cellular system's energy leaking inadvertently into the frequency band of the cellular system. If there is an interference in the voice channels, cross talk is heard will appear as noise between the users. The interference in the control channels leads to missed and error calls because of digital signaling. Interference is more severe in urban areas because of the greater RF noise and greater density of mobiles and base stations. The interference can be divided into 2 parts: co-channel interference and adjacent channel interference.

Co-channel interference (CCI)

For the efficient use of available spectrum, it is necessary to reuse frequency bandwidth over relatively small geographical areas. However, increasing frequency reuse also increases interference, which decreases system capacity and service quality. The cells where the same set of frequencies is used are call co-channel cells. Co-channel interference is the cross talk between two different radio transmitters using the same radio frequency as is the case with the co-channel cells. The reasons of CCI can be because of either adverse weather conditions or poor frequency planning or overly- crowded radio spectrum.

If the cell size and the power transmitted at the base stations are same then CCI will become independent of the transmitted power and will depend on radius of the cell (R) and the distance between the interfering co-channel cells (D). If D/R ratio is increased, then the effective distance between the co-channel cells will increase and interference will decrease. The parameter Q is called the frequency reuse ratio and is related to the cluster size. For hexagonal geometry

$$Q = D/R = \sqrt{3N}. \tag{3.11}$$

From the above equation, small of ‘Q’ means small value of cluster size ‘N’ and increase in cellular capacity. But large ‘Q’ leads to decrease in system capacity but increase in transmission quality. Choosing the options is very careful for the selection of ‘N’, the proof of which is given in the first section.

The Signal to Interference Ratio (SIR) for a mobile receiver which monitors the forward channel can be calculated as

$$\frac{S}{I} = \frac{S}{\sum_{i=1}^{i_0} I_i} \tag{3.12}$$

where i_0 is the number of co-channel interfering cells, S is the desired signal power from the baseband station and I_i is the interference power caused by the i-th interfering co-channel base station. In order to solve this equation from power calculations, we need to look into the signal power characteristics. The average power in the mobile radio channel decays as a power law of the distance of separation between transmitter and receiver. The expression for the received power P_r at a distance d can be approximately calculated as

$$P_r = P_0 \left(\frac{d}{d_0}\right)^{-n} \tag{3.13}$$

and in the dB expression as

$$P_r(dB) = P_0(dB) - 10n \log\left(\frac{d}{d_0}\right) \tag{3.14}$$

where P_0 is the power received at a close-in reference point in the far field region at a small distance do from the transmitting antenna, and ‘n’ is the path loss exponent. Let us calculate the SIR for this system. If D_i is the distance of the i-th interferer from the mobile, the received power at a given mobile due to i-th interfering cell is proportional to $(D_i)^{-n}$ (the value of ’n’ varies between 2 and 4 in urban cellular systems).

Let us take that the path loss exponent is same throughout the coverage area and the transmitted power be same, then SIR can be approximated as

$$\frac{S}{I} = \frac{R^{-n}}{\sum_{i=1}^{i_0} D_i^{-n}} \tag{3.15}$$

where the mobile is assumed to be located at R distance from the cell center. If we consider only the first layer of interfering cells and we assume that the interfering base stations are equidistant from the reference base station and the distance between the cell centers is 'D' then the above equation can be converted as

$$\frac{S}{I} = \frac{(D/R)^n}{i_0} = \frac{(\sqrt{3N})^n}{i_0} \tag{3.16}$$

which is an approximate measure of the SIR. Subjective tests performed on AMPS cellular system which uses FM and 30 kHz channels show that sufficient voice quality can be obtained by SIR being greater than or equal to 18 dB. If we take n=4, the value of 'N' can be calculated as 6.49. Therefore minimum N is 7. The above equations are based on hexagonal geometry and the distances from the closest interfering cells can vary if different frequency reuse plans are used.

We can go for a more approximate calculation for co-channel SIR. This is the example of a 7 cell reuse case. The mobile is at a distance of D-R from 2 closest interfering cells and approximately D+R/2, D, D-R/2 and D+R distance from other interfering cells in the first tier. Taking n = 4 in the above equation, SIR can be approximately calculated as

$$\frac{S}{I} = \frac{R^{-4}}{2(D-R)^{-4} + (D+R)^{-4} + (D)^{-4} + (D+R/2)^{-4} + (D-R/2)^{-4}} \tag{3.17}$$

which can be rewritten in terms frequency reuse ratio Q as

$$\frac{S}{I} = \frac{1}{2(Q-1)^{-4} + (Q+1)^{-4} + (Q)^{-4} + (Q+1/2)^{-4} + (Q-1/2)^{-4}}. \tag{3.18}$$

Using the value of N equal to 7 (this means Q = 4.6), the above expression yields that worst case SIR is 53.70 (17.3 dB). This shows that for a 7 cell reuse case the worst case SIR is slightly less than 18 dB. The worst case is when the mobile is at the corner of the cell i.e., on a vertex as shown in the Figure 3.6. Therefore N = 12 cluster size should be used. But this reduces the capacity by 7/12 times. Therefore, co-channel interference controls link performance, which in a way controls frequency reuse plan and the overall capacity of the cellular system. The effect of co-channel interference can be minimized by optimizing the frequency assignments of the base stations and their transmit powers. Tilting the base-station antenna to limit the spread of the signals in the system can also be done.

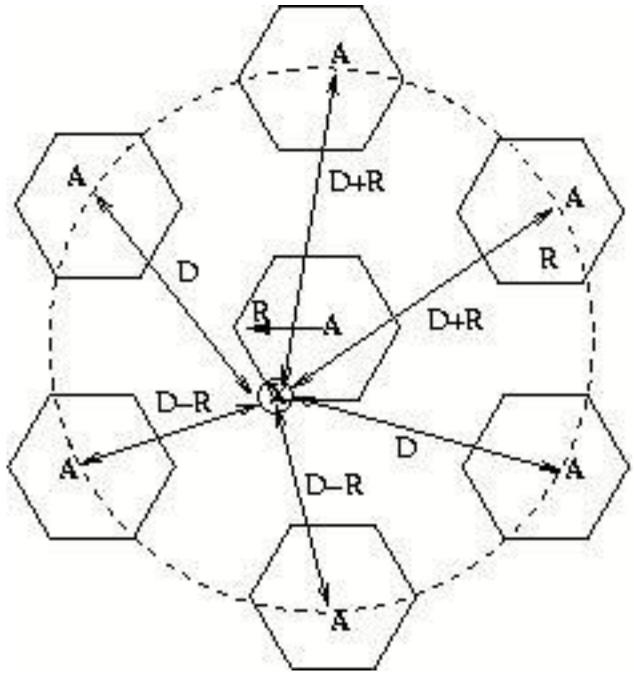


Figure 3.6: First tier of co-channel interfering cells

Adjacent Channel Interference (ACI)

This is a different type of interference which is caused by adjacent channels i.e. channels in adjacent cells. It is the signal impairment which occurs to one frequency due to presence of another signal on a nearby frequency. This occurs when imperfect receiver filters allow nearby frequencies to leak into the passband. This problem is enhanced if the adjacent channel user is transmitting in a close range compared to the subscriber’s receiver while the receiver attempts to receive a base station on the channel. This is called near-far effect. The more adjacent channels are packed into the channel block, the higher the spectral efficiency, provided that the performance degradation can be tolerated in the system link budget. This effect can also occur if a mobile close to a base station transmits on a channel close to one being used by a weak mobile. This problem might occur if the base station has problem in discriminating the mobile user from the ”bleed over” caused by the close adjacent channel mobile.

Adjacent channel interference occurs more frequently in small cell clusters and heavily used cells. If the frequency separation between the channels is kept large this interference can be reduced to some extent. Thus assignment of channels is given such that they do not form a contiguous band of frequencies within a particular cell and frequency separation is maximized. Efficient assignment strategies are very much important in making the interference as less as possible. If the frequency factor is small then distance between the adjacent channels cannot put the interference level within tolerance limits. If a mobile is 10 times close to the base station than other mobile and has energy spill out of its passband, then SIR for weak mobile is approximately

$$\frac{S}{I} = 10^{-n} \tag{3.19}$$

which can be easily found from the earlier SIR expressions. If n = 4, then SIR is −52 dB. Perfect base station filters are needed when close-in and distant users share the same cell. Practically, each base station receiver is preceded by a high Q cavity filter in order to remove adjacent channel interference. Power control is also very much important for the prolonging of the battery life for the subscriber unit but also reduces reverse channel SIR in the system. Power control is done such that each mobile transmits the lowest power required to maintain a good quality link on the reverse channel.

Enhancing Capacity And Cell Coverage

The Key Trade-off

Previously, we have seen that the frequency reuse technique in cellular systems allows for almost boundless expansion of geographical area and the number of mobile system users who could be accommodated. In designing a cellular layout, the two parameters which are of great significance are the cell radius R and the cluster size N, and we have also seen that co-channel cell distance $D = \sqrt{3N}R$. In the following, a brief description of the design trade-off is given, in which the above two parameters play a crucial role.

The cell radius governs both the geographical area covered by a cell and also the number of subscribers who can be serviced, given the subscriber density. It is easy to see that the cell radius must be as large as possible. This is because, every cell requires an investment in a tower, land on which the tower is placed, and radio transmission equipment and so a large cell size minimizes the cost per subscriber.

Eventually, the cell radius is determined by the requirement that adequate signal to noise ratio be maintained over the coverage area. The SNR is determined by several factors such as the antenna height, transmitter power, receiver noise figure etc. Given a cell radius R and a cluster size N, the geographic area covered by a cluster is

$$A_{cluster} = NA_{cell} = N3\sqrt{3}R^2/2. \tag{3.20}$$

If the total serviced area is A_{total}, then the number of clusters M that could be accommodated is given by

$$M = A_{total}/A_{cluster} = A_{total}/(N3\sqrt{3}R^2/2). \tag{3.21}$$

Note that all of the available channels N, are reused in every cluster. Hence, to make the maximum number of channels available to subscribers, the number of clusters M should be large, which, by Equation (3.21), shows that the cell radius should be small. However, cell radius is determined by a trade-off: R should be as large as possible to minimize the cost of the installation per subscriber, but R should be as small as possible to maximize the number of customers that the system can accommodate. Now, if the cell radius R is fixed, then the number of clusters could be maximized by minimizing the size of a cluster N. We have seen earlier that the size of a cluster depends on the frequency reuse ratio Q. Hence, in determining the value of N , another trade-off is encountered in that N must be small to accommodate large number of subscribers, but should be sufficiently large so as to minimize the interference effects.

Now, we focus on the issues regarding system expansion. The history of cellular phones has been characterized by a rapid growth and expansion in cell subscribers. Though a cellular system can be expanded by simply adding cells to the geographical area, the way in which user density can be increased is also important to look at. This is because it is not always possible to counter the increasing demand for cellular systems just by increasing the geographical coverage area due to the limitations in obtaining new land with suitable requirements. We discuss here two methods for dealing with an increasing subscriber density: Cell Splitting and Sectoring. The other method, microcell zone concept can treated as enhancing the QoS in a cellular system.

The basic idea of adopting the cellular approach is to allow space for the growth of mobile users. When a new system is deployed, the demand for it is fairly low and users are assumed to be uniformly distributed over the service area. However, as new users subscribe to the cellular service, the demand for channels may begin to exceed the capacity of some base stations. As discussed previously,the number of channels available to customers (equivalently, the channel density per square kilometer) could be increased by decreasing the cluster size. However, once a system has been initially deployed, a system-wide reduction in cluster size may not be necessary since user density does not grow uniformly in all parts of the geographical area. It might be that an increase in channel density is required only in specific parts of the system to support an increased demand in those areas. Cell-splitting is a technique which has the capability to add new smaller cells in specific areas of the system.

Cell-Splitting

Cell Splitting is based on the cell radius reduction and minimizes the need to modify the existing cell parameters. Cell splitting involves the process of sub-dividing a congested cell into smaller cells, each with its own base station and a corresponding reduction in antenna size and transmitting power. This increases the capacity of a cellular system since it increases the number of times that channels are reused. Since the new cells have smaller radii than the existing cells, inserting these smaller cells, known as microcells, between the already existing cells results in an increase of capacity due to the additional number of channels per unit area. There are few challenges in increasing the capacity by reducing the cell radius. Clearly, if cells are small, there would have to be more of them and so additional base stations will be needed in the system. The challenge in this case is to introduce the new base stations without the need to move the already existing base station towers. The other challenge is to meet the generally increasing demand that may vary quite rapidly between geographical areas of the system. For instance, a city may have highly populated areas and so the demand must be supported by cells with the smallest radius. The radius of cells will generally increase as we move from urban to sub urban areas, because the user density decreases on moving towards sub-urban areas. The key factor is to add as minimum number of smaller cells as possible

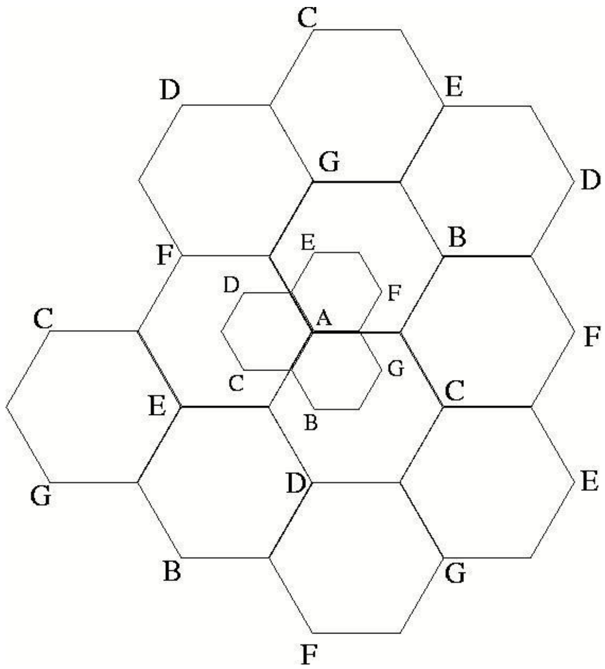


Figure 3.7: Splitting of congested seven-cell clusters.

wherever an increase in demand occurs. The gradual addition of the smaller cells implies that, at least for a time, the cellular system operates with cells of more than one size.

Figure 3.7 shows a cellular layout with seven-cell clusters. Consider that the cells in the center of the diagram are becoming congested, and cell A in the center has reached its maximum capacity. Figure also shows how the smaller cells are being superimposed on the original layout. The new smaller cells have half the cell radius of the original cells. At half the radius, the new cells will have one-fourth of the area and will consequently need to support one-fourth the number of subscribers. Notice that one of the new smaller cells lies in the center of each of the larger cells. If we assume that base stations are located in the cell centers, this allows the original base stations to be maintained even in the new system layout. However, new base stations will have to be added for new cells that do not lie in the center of the larger cells. The organization of cells into clusters is independent of the cell radius, so that the cluster size can be the same in the small-cell layout as it was in the large-cell layout. Also the signal-to-interference ratio is determined by cluster size and not by cell radius. Consequently, if the cluster size is maintained, the signal-to-interference ratio will be the same after cell splitting as it was before. If the entire system is replaced with new half-radius cells, and the cluster size is maintained, the number of channels per cell will be exactly as it was before, and the number of subscribers per cell will have been reduced.

When the cell radius is reduced by a factor, it is also desirable to reduce the transmitted power. The transmit power of the new cells with radius half that of the old cells can be found by examining the received power PR at the new and old cell boundaries and setting them equal. This is necessary to maintain the same frequency re-use plan in the new cell layout as well. Assume that PT1 and PT2 are the transmit powers of the larger and smaller base stations respectively. Then, assuming a path loss index n=4, we have power received at old cell boundary = P_{T1}/R^4 and the power received at new cell boundary = $P_{T2}/(R/2)^4$. On equating the two received powers, we get $P_{T2} = P_{T1} / 16$. In other words, the transmit power must be reduced by 12 dB in order to maintain the same S/I with the new system lay-out.

At the beginning of this channel splitting process, there would be fewer channels in the smaller power groups. As the demand increases, more and more channels need to be accommodated and hence the splitting process continues until all the larger cells have been replaced by the smaller cells, at which point splitting is complete within the region and the entire system is rescaled to have a smaller radius per cell. If a cellular layout is replaced entirety by a new layout with a smaller cell radius, the signal-to-interference ratio will not change, provided the cluster size does not change. Some special care must be taken, however, to avoid co-channel interference when both large and small cell radii coexist. It turns out that the only way to avoid interference between the large-cell and small-cell systems is to assign entirely different sets of channels to the two systems. So, when two sizes of cells co-exist in a system, channels in the old cell must be broken down into two groups, one that corresponds to larger cell reuse requirements and the other which corresponds to the smaller cell reuse requirements. The larger cell is usually dedicated to high speed users as in the umbrella cell approach so as to minimize the number of hand-offs.

Ex. 4: When the AMPS cellular system was first deployed, the aim of the system designers was to guarantee coverage. Initially the number of users was not significant. Consequently cells were configured with an eight-mile radius, and a 12-cell cluster size was chosen. The cell radius was chosen to guarantee a 17 dB

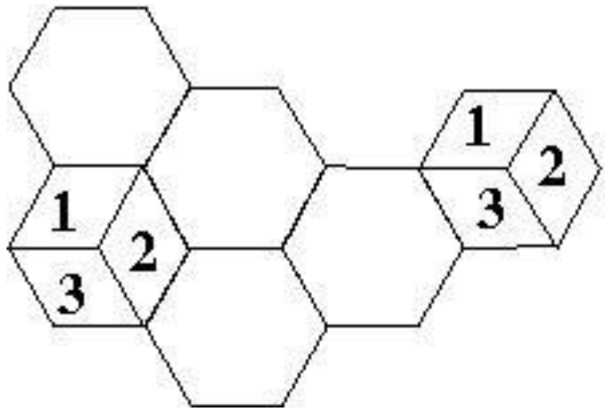


Figure 3.8: A cell divided into three 120^o sectors.

signal-to-noise ratio over 90% of the coverage area. Although a 12-cell cluster size provided more than adequate co-channel separation to meet a requirement for a 17 dB signal-to-interference ratio in an interference-limited environment, it did not provide adequate frequency reuse to service an explosively growing customer base. The system planners reasoned that a subsequent shift to a 7-cell cluster size would provide an adequate number of channels. It was estimated that a 7-cell cluster size should provide an adequate 18.7 dB signal-to-interference ratio. The margin, however, is slim, and the 17 dB signal-to-interference ratio requirement could not be met over 90 % of the coverage area.

Sectoring

Sectoring is basically a technique which can increase the SIR without necessitating an increase in the cluster size. Till now, it has been assumed that the base station is located in the center of a cell and radiates uniformly in all the directions behaving as an omni-directional antenna. However it has been found that the co-channel interference in a cellular system may be decreased by replacing a single omni-directional antenna at the base station by several directional antennas, each radiating within a specified sector. In the Figure 3.8, a cell is shown which has been split into three 120^o sectors. The base station feeds three 120^o directional antennas, each of which radiates into one of the three sectors. The channel set serving this cell has also been divided, so that each sector is assigned one-third of the available number cell of channels. This technique for reducing co-channel interference wherein by using suit-

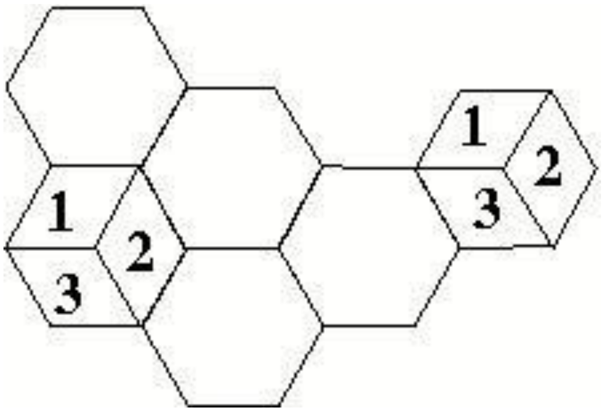


Figure 3.9: A seven-cell cluster with 60^o sectors.

able directional antennas, a given cell would receive interference and transmit with a fraction of available co-channel cells is called 'sectoring'. In a seven-cell-cluster layout with 120^o sectored cells, it can be easily understood that the mobile units in a particular sector of the center cell will receive co-channel interference from only two of the first-tier co-channel base stations, rather than from all six. Likewise, the base station in the center cell will receive co-channel interference from mobile units in only two of the co-channel cells. Hence the signal to interference ratio is now modified to

$$\frac{S}{I} = \frac{(\sqrt{3N})^n}{2} \tag{3.22}$$

where the denominator has been reduced from 6 to 2 to account for the reduced number of interfering sources. Now, the signal to interference ratio for a seven-cell cluster layout using 120^o sectored antennas can be found from equation (3.24) to be 23.4 dB which is a significant improvement over the Omni-directional case where the worst-case S/I is found to be 17 dB (assuming a path-loss exponent, n=4). Some cellular systems divide the cells into 60^o sectors. Similar analysis can be performed on them as well.

Ex. 5: A cellular system having a seven-cell cluster layout with omni-directional antennas has been performing satisfactorily for a required signal to interference ratio of 15 dB. However due to the need for increasing the number of available channels, a 60^o sectoring of the cells has been introduced. By what percentage can the number of channels N_{total} be increased assuming a path-loss component n=4?

Solution: The seven-cell cluster layout with 60^o sectoring is shown in the Figure 3.9.

It is easy to see that the shaded region in the center receives interference from just one first-tier cell and hence the signal to interference ratio can be obtained suitably as

$$\frac{S}{I} = \frac{(\sqrt{3N})^n}{1} = \frac{(\sqrt{(3)(7)})^4}{1} = 26.4dB. \tag{3.23}$$

Since the SIR exceeds 15 dB, one can try reducing the cluster size from seven to four. Now, the SIR for this reduced cluster size layout can be found to be

$$\frac{S}{I} = \frac{(\sqrt{3N})^n}{1} = \frac{(\sqrt{(3)(4)})^4}{1} = 21.6\ dB. \tag{3.24}$$

The S/I ratio is still above the requirement and so a further reduction in the cell cluster size is possible. For a 3-cell cluster layout, there are two interfering sources and hence the S/I ratio is found to be

$$\frac{S}{I} = \frac{(\sqrt{3N})^n}{1} = \frac{(\sqrt{33})^4}{2} = 16.07dB. \tag{3.25}$$

This is just above the adequate S/I ratio and further reduction in cluster size is not possible. So, a 3-cluster cell layout could be used for meeting the growth requirements. Thus, when the cluster size is reduced from 7 to 3, the total number of channels increased by a factor of 7/3.

The calculations in the above example are actually an idealization for several reasons. Firstly, practical antennas have side lobes and cannot be used to focus a transmitted beam into a perfect 120^o sector or 60^o sector. Due to this, additional interference will be introduced. Next, it is also a cause of concern that a given number of channels are not able to support as many subscribers when the pool of channels is divided into small groups. This is due to a reduction in Trunking Efficiency, a term which will be explained later on. Because sectoring involves using more than one antenna per base station, the available channels in the cell are divided and dedicated to a specific antenna. This breaks the available set of channels into smaller sets, thus reducing the trunking efficiency. Moreover, dividing a cell into sectors requires that a call in progress will have to be handed off (that is, assigned a new channel) when a mobile unit travels into a new sector. This increases the complexity of the system and also the load on the mobile switching center/base station.

3.7.4 Microcell Zone Concept

The increased number of handoffs required when sectoring is employed results in an increased load on the switching and control link elements of the mobile system. To overcome this problem, a new microcell zone concept has been proposed. As shown in Figure 3.10, this scheme has a cell divided into three microcell zones, with each of the three zone sites connected to the base station and sharing the same radio equipment. It is necessary to note that all the microcell zones, within a cell, use the same frequency used by that cell; that is no handovers occur between microcells. Thus when a mobile user moves between two microcell zones of the cell, the BS simply switches the channel to a different zone site and no physical re-allotment of channel takes place.

<u>Locating the mobile unit within the cell</u>: An active mobile unit sends a signal to all zone sites, which in turn send a signal to the BS. A zone selector at the BS uses that signal to select a suitable zone to serve the mobile unit - choosing the zone with the strongest signal.

<u>Base Station Signals</u>: When a call is made to a cellular phone, the system already knows the cell location of that phone. The base station of that cell knows in which zone, within that cell, the cellular phone is located. Therefore when it receives the signal, the base station transmits it to the suitable zone site. The zone site receives the cellular signal from the base station and transmits that signal to the mobile phone after amplification. By confining the power transmitted to the mobile phone, co-channel interference is reduced between the zones and the capacity of system is increased.

<u>Benefits of the micro-cell zone concept</u>: 1) Interference is reduced in this case as compared to the scheme in which the cell size is reduced.

2) Handoffs are reduced (also compared to decreasing the cell size) since the micro-cells within the cell operate at the same frequency; no handover occurs when the mobile unit moves between the microcells.

3) Size of the zone apparatus is small. The zone site equipment being small can be mounted on the side of a building or on poles.

4) System capacity is increased. The new microcell knows where to locate the mo-bile unit in a particular zone of the cell and deliver the power to that zone. Since

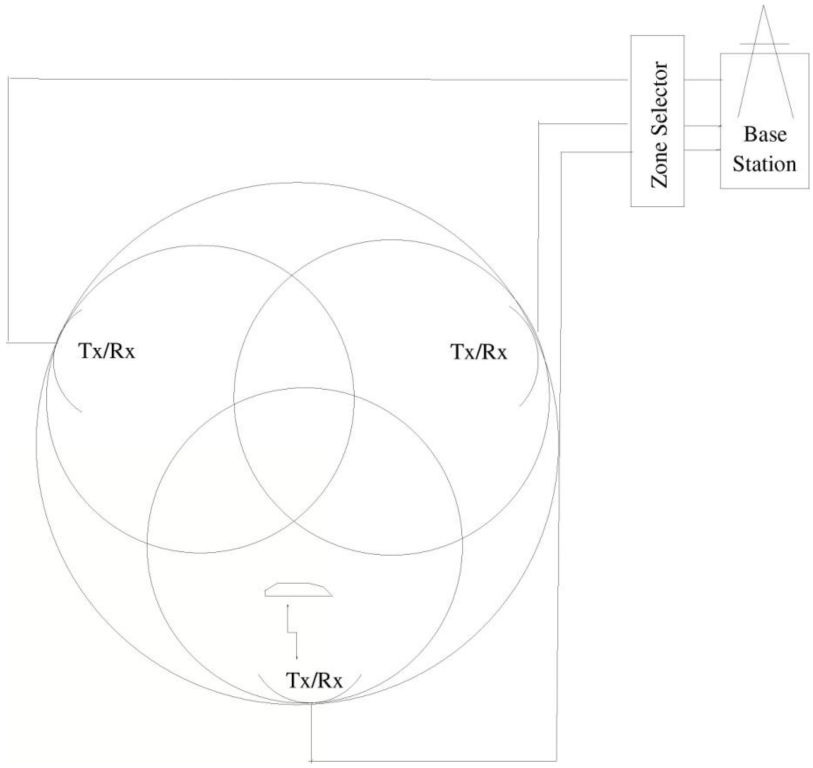


Figure 3.10: The micro-cell zone concept.

the signal power is reduced, the microcells can be closer and result in an increased system capacity. However, in a microcellular system, the transmitted power to a mobile phone within a microcell has to be precise; too much power results in interference between microcells, while with too little power the signal might not reach the mobile phone.This is a drawback of microcellular systems, since a change in the surrounding (a new building, say, within a microcell) will require a change of the transmission power.

Trunked Radio System

In the previous sections, we have discussed the frequency reuse plan, the design trade-offs and also explored certain capacity expansion techniques like cell-splitting and sectoring. Now, we look at the relation between the number of radio channels a cell contains and the number of users a cell can support. Cellular systems use the concept of trunking to accommodate a large number of users in a limited radio spectrum. It was found that a central office associated with say, 10,000 telephones requires about 50 million connections to connect every possible pair of users. However, a worst case maximum of 5000 connections need to be made among these telephones at any given instant of time, as against the possible 50 million connections. In fact, only a few hundreds of lines are needed owing to the relatively short duration of a call. This indicates that the resources are shared so that the number of lines is much smaller than the number of possible connections. A line that connects switching offices and that is shared among users on an as-needed basis is called a trunk.

The fact that the number of trunks needed to make connections between offices is much smaller than the maximum number that could be used suggests that at times there might not be sufficient facilities to allow a call to be completed. A call that cannot be completed owing to a lack of resources is said to be blocked. So one important to be answered in mobile cellular systems is: How many channels per cell are needed in a cellular telephone system to ensure a reasonably low probability that a call will be blocked?

In a trunked radio system, a channel is allotted on per call basis. The performance of a radio system can be estimated in a way by looking at how efficiently the calls are getting connected and also how they are being maintained at handoffs.

Some of the important factors to take into consideration are (i) Arrival statistics, (ii)Service statistics, (iii)Number of servers/channels.

Let us now consider the following assumptions for a bufferless system handling 'L' users as shown in Figure 3.11:

(i) The number of users L is large when compared to 1.

(ii) Arrival statistics is Poisson distributed with a mean parameter λ.

(iii) Duration of a call is exponentially distributed with a mean rate μ_1.

(iv) Residence time of each user is exponentially distributed with a rate parameter μ_2.

(v) The channel holding rate therefore is exponentially distributed with a parameter $\mu = \mu_1 + \mu_2$.

(vi) There is a total of 'J' number of channels ($J \leq L$).

To analyze such a system, let us recapitulate a queuing system in brief. Consider an M/M/m/m system which is an m-server loss system. The name M/M/m/m reflects

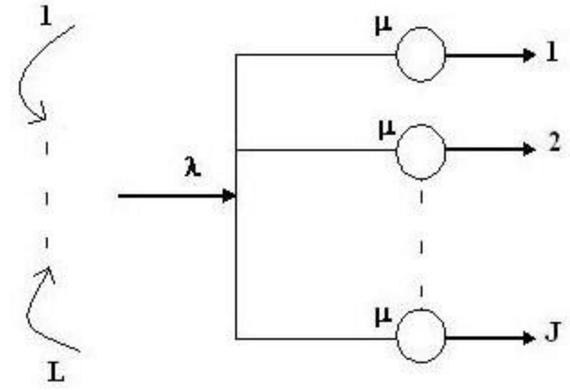


Figure 3.11: The bufferless J-channel trunked radio system.

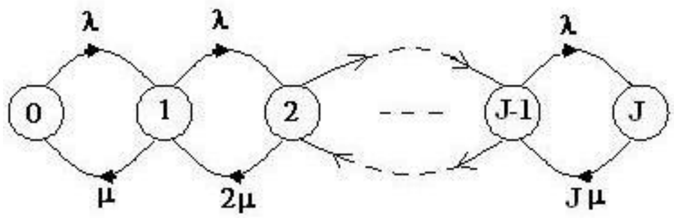


Figure 3.12: Discrete-time Markov chain for the M/M/J/J trunked radio system.

standard queuing theory nomenclature whereby:

(i) the first letter indicates the nature of arrival process(e.g. M stands for memory-less which here means a Poisson process).

(ii) the second letter indicates the nature of probability distribution of service times.(e.g M stands for exponential distribution). In all cases,successive inter arrival times and service times are assumed to be statistically independent of each other.

(iii) the third letter indicates the number of servers.

(iv) the last letter indicates that if an arrival finds all 'm' users to be busy, then it will not enter the system and is lost.

In view of the above, the bufferless system as shown in Figure 3.11 can be modeled as M/M/J/J system and the discrete-time Markov chain of this system is shown in Figure 3.12.

Trunking mainly exploits the statistical behavior of users so that a fixed number of channels can be used to accommodate a large, random user community. As the number of telephone lines decrease, it becomes more likely that all channels are busy for a particular user. As a result, the call gets rejected and in some systems, a queue may be used to hold the caller's request until a channel becomes available. In the telephone system context the term Grade of Service (GoS) is used to mean the probability that a user's request for service will be blocked because a required facility, such as a trunk or a cellular channel, is not available. For example, a GoS of 2 % implies that on the average a user might not be successful in placing a call on 2 out of every 100 attempts. In practice the blocking frequency varies with time. One would expect far more call attempts during business hours than during the middle of the night. Telephone operating companies maintain usage records and can identify a "busy hour", that is, the hour of the day during which there is the greatest demand for service. Typically, telephone systems are engineered to provide a specified grade of service during a specified busy hour.

User calling can be modeled statistically by two parameters: the average number of call requests per unit time λ_{user} and the average holding time H. The parameter λ_{user} is also called the average arrival rate, referring to the rate at which calls from a single user arrive. The average holding time is the average duration of a call. The

product:

$$A_{user} = \lambda_{user} H \tag{3.26}$$

that is, the product of the average arrival rate and the average holding time–is called the offered traffic intensity or offered load. This quantity represents the average traffic that a user provides to the system. Offered traffic intensity is a quantity that is traditionally measured in Erlangs. One Erlang represents the amount of traffic intensity carried by a channel that is completely occupied. For example, a channel that is occupied for thirty minutes during an hour carries 0.5 Erlang of traffic.

Call arrivals or requests for service are modeled as a Poisson random process. It is based on the assumption that there is a large pool of users who do not cooperate in deciding when to place calls. Holding times are very well predicted using an exponential probability distribution. This implies that calls of long duration are much less frequent than short calls. If the traffic intensity offered by a single user is A_{user}, then the traffic intensity offered by N users is $A = NA_{user}$.

The purpose of the statistical model is to relate the offered traffic intensity A, the grade of service P_b, and the number of channels or trunks C needed to maintain the desired grade of service.

Two models are widely used in traffic engineering to represent what happens when a call is blocked. The blocked calls cleared model assumes that when a channel or trunk is not available to service an arriving call, the call is cleared from the system. The second model is known as blocked calls delayed. In this model a call that cannot be serviced is placed on a queue and will be serviced when a channel or trunk becomes available.

Use of the blocked-calls-cleared statistical model leads to the Erlang B formula that relates offered traffic intensity A, grade of service P_b, and number of channels K.

Ex. 6: In a certain cellular system, an average subscriber places two calls per hour during a busy hour and the average holding time is 3 min. Each cell has 100 channels. If the blocked calls are cleared, how many subscribers can be serviced by each cell at 2 % GoS?

Solution: Using Erlang B table, it can be seen that for C = 100 and $ttoS = P_b$ = 2%, the total offered load A=87.972 Erlangs. Since an individual subscriber offers a load of A_{user} = (2 calls / 60 min)3 min = 0.1 Erlang, the maximum number of subscribers served is

$$N = A/A_{user} = 87.972/0.1 \approx 880. \tag{3.29}$$

Ex. 4: In the previous example, suppose that the channels have been divided into two groups of 50 channels each. Each subscriber is assigned to a group and can be served only by that group. How many subscribers can be served by the two group cell?

Solution: Using the Erlang B table with C = 50 and $ttOS = P_b$ = 2%, the total offered load per group is

$$A = 40.255 Erlangs \tag{3.30}$$

Thus the maximum number of users per group is

$$N_{group} = A/A_{user} \approx 403. \tag{3.31}$$

Thus, counting both the groups, maximum number of users in the two group cell is 806.

The above example indicates that the number of subscribers that can be supported by a given number of channels decreases as the pool of channels is sub-divided. We can express this in terms of the trunking efficiency, defined as the carrier load per channel, that is,

$$\xi = (1 - P_b)A/C. \tag{3.32}$$

This explains why the sectoring of a cell into either 120^o or 60^o sectors reduces the trunking efficiency of the system. Thus the system growth due to sectoring is impacted by trunking efficiency considerations.

UNIT 3

Free Space Radio Wave Propagation

Introduction

There are two basic ways of transmitting an electro-magnetic (EM) signal, through a guided medium or through an unguided medium. 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 to the presence of buildings, 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 a surface having sharp irregularities. Scattering occurs when the medium through the 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.

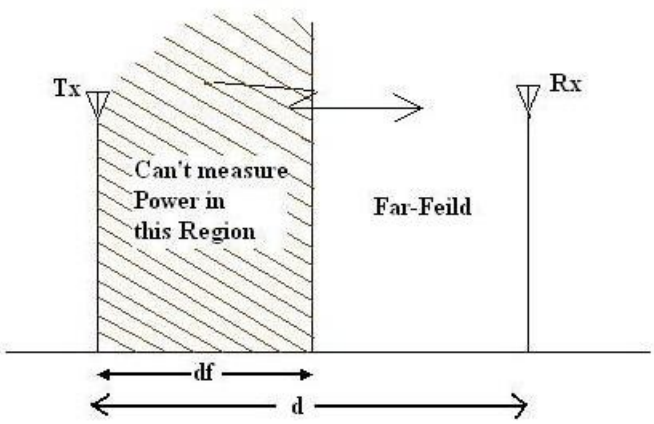


Figure 4.1: Free space propagation model, showing the near and far fields.

Free Space Propagation Model

Although EM signals when traveling through wireless channels experience fading effects due to various effects, but in some cases the transmission is with a direct line of sight such as in satellite communication. Free space model predicts that the received power decays as negative square root of the distance. Friis free space equation is given by

$$P_r(d) = \frac{P_t tt_t tt_r \lambda^2}{(4\pi)^2 d^2 L} \tag{4.1}$$

where P_t is the transmitted power, $P_r(d)$ is the received power, tt_t is the transmitter antenna gain, tt_r is the receiver antenna gain, d is the Tx-Rx separation and L is the system loss factor depended upon line attenuation, filter losses and antenna losses and not related to propagation. The gain of the antenna is related to the effective aperture of the antenna which in turn is dependent upon the physical size of the antenna as given below

$$\boldsymbol{tt} = 4\pi A_e/\lambda^2. \tag{4.2}$$

The path loss, representing the attenuation suffered by the signal as it travels through the wireless channel is given by the difference of the transmitted and received power in dB and is expressed as:

$$PL(dB) = 10 \log P_t/P_r. \tag{4.3}$$

The fields of an antenna can broadly be classified in two regions, the far field and the near field. It is in the far field that the propagating waves act as plane waves and the power decays inversely with distance. The far field region is also termed as Fraunhofer region and the Friis equation holds in this region. Hence, the Friis equation is used only beyond the far field distance, d_f, which is dependent upon the largest dimension of the antenna as

$$d_f = 2D^2/\lambda. \qquad (4.4)$$

Also we can see that the Friis equation is not defined for d=0. For this reason, we use a close in distance, d_o, as a reference point. The power received, $P_r(d)$, is then given by:

$$P_r(d) = P_r(d_o)(d_o/d)^2. \qquad (4.5)$$

Ex. 1: Find the far field distance for a circular antenna with maximum dimension of 1 m and operating frequency of 900 MHz.

Solution: Since the operating frequency f = 900 Mhz, the wavelength

$$\lambda = \frac{3 \times 10^8 m/s}{900 \times 10^6 Hz} m$$

. Thus, with the largest dimension of the antenna, D=1m, the far field distance is

$$d_f = \frac{2D^2}{\lambda} = \frac{2(1)^2}{0.33} = 6m$$

.

Ex. 2: A unit gain antenna with a maximum dimension of 1 m produces 50 W power at 900 MHz. Find (i) the transmit power in dBm and dB, (ii) the received power at a free space distance of 5 m and 100 m.

Solution:

(i) Tx power = 10log(50) = 17 dB = (17+30) dBm = 47 dBm

(ii) $d_f = \frac{2 \times D^2}{\lambda} = \frac{2 \times 1^2}{1/3} = 6m$

Thus the received power at 5 m can not be calculated using free space distance formula.

At 100 m ,

$$P_R = \frac{P_T G_T G_R \lambda^2}{4\pi d^2}$$

$$= \frac{50 \times 1 \times (1/3)^2}{4\pi 100^2}$$

$$= 3.5 \times 10^{-3} mW$$

$$P_R(dBm) = 10 log P_r(mW) = -24.5 dBm$$

Basic Methods of Propagation

Reflection, diffraction and scattering are the three fundamental phenomena that cause signal propagation in a mobile communication system, apart from LoS communication. The most important parameter, predicted by propagation models based on above three phenomena, is the received power. The physics of the above phenomena may also be used to describe small scale fading and multipath propagation. The following subsections give an outline of these phenomena.

Reflection

Reflection occurs when an electromagnetic wave falls on an object, which has very large dimensions as compared to the wavelength of the propagating wave. For example, such objects can be the earth, buildings and walls. When a radio wave falls on another medium having different electrical properties, a part of it is transmitted into it, while some energy is reflected back. Let us see some special cases. If the medium on which the e.m. wave is incident is a dielectric, some energy is reflected back and some energy is transmitted. If the medium is a perfect conductor, all energy is reflected back to the first medium. The amount of energy that is reflected back depends on the polarization of the e.m. wave.

Another particular case of interest arises in parallel polarization, when no reflection occurs in the medium of origin. This would occur, when the incident angle would be such that the reflection coefficient is equal to zero. This angle is the Brewster's angle. By applying laws of electro-magnetics, it is found to be

$$sin(\theta_B) = \sqrt{\frac{\epsilon_1}{\epsilon_1 + \epsilon_2}}. \tag{4.6}$$

Further, considering perfect conductors, the electric field inside the conductor is always zero. Hence all energy is reflected back. Boundary conditions require that

$$\theta_i = \theta_r \tag{4.7}$$

and

$$E_i = E_r \tag{4.8}$$

for vertical polarization, and

$$E_i = -E_r \tag{4.9}$$

for horizontal polarization.

Diffraction

Diffraction is the phenomenon due to which an EM wave can propagate beyond the horizon, around the curved earth's surface and obstructions like tall buildings. As the user moves deeper into the shadowed region, the received field strength decreases. But the diffraction field still exists an it has enough strength to yield a good signal. This phenomenon can be explained by the Huygen's principle, according to which, every point on a wavefront acts as point sources for the production of sec- ondary wavelets, and they combine to produce a new wavefront in the direction of propagation. The propagation of secondary wavelets in the shadowed region results in diffraction. The field in the shadowed region is the vector sum of the electric field components of all the secondary wavelets that are received by the receiver.

Scattering

The actual received power at the receiver is somewhat stronger than claimed by the models of reflection and diffraction. The cause is that the trees, buildings and lamp-posts scatter energy in all directions. This provides extra energy at the receiver. Roughness is tested by a Rayleigh criterion, which defines a critical height h_c of surface protuberances for a given angle of incidence θ_i, given by,

$$h_c = \frac{\lambda}{8 \sin\theta_i}. \tag{4.10}$$

A surface is smooth if its minimum to maximum protuberance h is less than h_c, and rough if protuberance is greater than h_c. In case of rough surfaces, the surface reflection coefficient needs to be multiplied by a scattering loss factor ρ_S, given by

$$\rho_S = exp(-8(\frac{\pi\sigma_h sin\theta_i}{\lambda})^2) \tag{4.11}$$

where σ_h is the standard deviation of the Gaussian random variable h. The following result is a better approximation to the observed value

$$\rho_S = exp(-8(\frac{\pi\sigma_h sin\theta_i}{\lambda})^2) I_0 [-8(\frac{\pi\sigma_h sin\theta_i}{\lambda})^2] \tag{4.12}$$

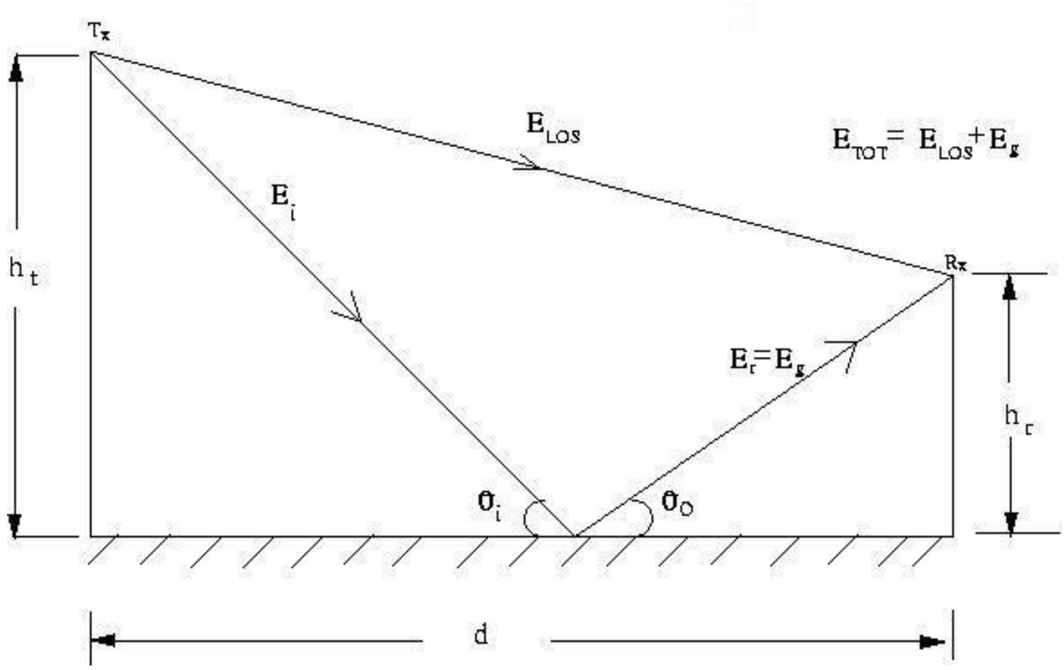


Figure 4.2: Two-ray reflection model.

which agrees very well for large walls made of limestone. The equivalent reflection coefficient is given by,

$$\Gamma_{rough} = \rho_S \Gamma. \tag{4.13}$$

Two Ray Reflection Model

Interaction of EM waves with materials having different electrical properties than the material through which the wave is traveling leads to transmitting of energy through the medium and reflection of energy back in the medium of propagation. The amount of energy reflected to the amount of energy incidented is represented by Fresnel reflection coefficient Γ, which depends upon the wave polarization, angle of incidence and frequency of the wave. For example, as the EM waves can not pass through conductors, all the energy is reflected back with angle of incidence equal to the angle of reflection and reflection coefficient $\Gamma = -1$. In general, for parallel and perpendicular polarizations, Γ is given by:

$$\Gamma_{\parallel} = E_r/E_i = \eta_2 \sin\theta_t - \eta_1 \sin\theta_i/\eta_2 \sin\theta_t + \eta_1 \sin\theta_i \tag{4.14}$$

$$\Gamma_{\perp} = E_r/E_i = \eta_2 \sin \theta_i - \eta_1 \sin \theta_t/\eta_2 \sin \theta_i + \eta_1 \sin \theta_t. \quad (4.15$$

) Seldom in communication systems we encounter channels with only LOS paths and hence the Friis formula is not a very accurate description of the communication link. A two-ray model, which consists of two overlapping waves at the receiver, one direct

path and one reflected wave from the ground gives a more accurate description as shown in Figure 4.2. A simple addition of a single reflected wave shows that power varies inversely with the forth power of the distance between the Tx and the Rx. This is deduced via the following treatment. From Figure 4.2, the total transmitted and received electric fields are

$$E_T^{T\ OT} = E_i + E_{LOS}, \tag{4.16}$$

$$E_R^{T\ OT} = E_g + E_{LOS}. \tag{4.17}$$

Let E_0 is the free space electric field (in V/m) at a reference distance d_0. Then

$$E(d,t) = \frac{E_0 d_0}{d} cos(\omega_c t - \varphi) \tag{4.18}$$

where

$$\varphi = \omega_c \frac{d}{c} \tag{4.19}$$

and $d > d_0$. The envelop of the electric field at d meters from the transmitter at any time t is therefore

$$|E(d, t)| = \frac{E_0 d_0}{d}. \tag{4.20}$$

This means the envelop is constant with respect to time.

Two propagating waves arrive at the receiver, one LOS wave which travels a distance of d^{j} and another ground reflected wave, that travels d^{jj} . Mathematically, it can be expressed as:

$$E(d^{j}, t) = \frac{E_0 d_0}{d^{j}} cos(\omega_c t - \varphi^{j}) \tag{4.21}$$

where

$$\varphi^{j} = \omega_c \frac{d^{j}}{c} \tag{4.22}$$

and

$$E(d^{jj}, t) = \frac{E_0 d_0}{d^{jj}} cos(\omega_c t - \varphi^{jj}) \tag{4.23}$$

where

$$\varphi^{j}_{j} = \omega_c \frac{d^{j}_{j}}{c}. \tag{4.24}$$

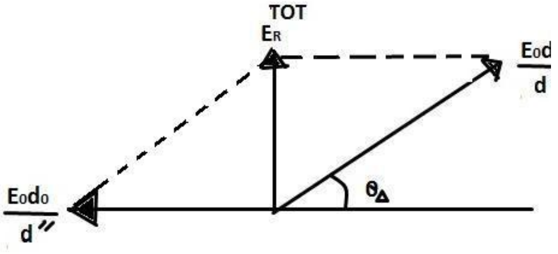


Figure 4.3: Phasor diagram of electric fields.

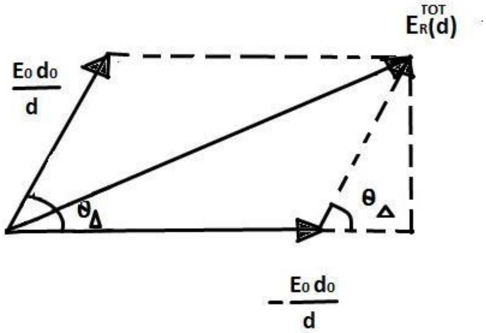


Figure 4.4: Equivalent phasor diagram of Figure 4.3.

According to the law of reflection in a dielectric, $\theta_i = \theta_0$ and $E_g = \Gamma E_i$ which means the total electric field,

$$E_t = E_i + E_g = E_i(1 + \Gamma). \tag{4.25}$$

For small values of θ_i, reflected wave is equal in magnitude and 180^o out of phase with respect to incident wave. Assuming perfect horizontal electric field polarization, i.e.,

$$\Gamma_\perp = -1 \Longrightarrow E_t = (1-1)E_i = 0, \tag{4.26}$$

the resultant electric field is the vector sum of E_{LOS} and E_g. This implies that,

$$E_R^{TOT} = |E_{LOS} + E_g|. \tag{4.27}$$

It can be therefore written that

$$E_R^{TOT}(d,t) = \frac{E0d0}{d^j} cos(\omega\, t - \varphi^j) + (-1)\frac{E0d0}{d^{jj}} cos(\omega\, t - \varphi^{jj}) \tag{4.28}$$

In such cases, the path difference is

$$\Delta = d^{jj} - d^j = \sqrt{(h_t + h_r)^2 + d^2} - \sqrt{(h_t - h_r)^2 + d^2}. \tag{4.29}$$

However, when T-R separation distance is very large compared to $(h_t + h_r)$, then

$$\Delta \approx \frac{2h_t h_r}{d} \tag{4.30}$$

Ex 3: Prove the above two equations, i.e., equation (4.29) and (4.30).

Once the path difference is known, the phase difference is and the time difference,

$$\tau_d = \frac{\Delta}{c} = \frac{\theta_\Delta}{2\pi f_c}. \tag{4.32}$$

When d is very large, then Δ becomes very small and therefore E_{LOS} and E_g are virtually identical with only phase difference, i.e.,

$$\left|\frac{E_0 d_0}{d}\right| \approx \left|\frac{E_0 d_0}{d'}\right| \approx \left|\frac{E_0 d_0}{d''}\right|. \tag{4.33}$$

Say, we want to evaluate the received E-field at any $t = \frac{d''}{c}$. Then,

$$E_R^{TOT}\left(d, t = \frac{d''}{c}\right) = \frac{E_0 d_0}{d'} \cos\left(\omega_c \frac{d''}{c} - \omega_c \frac{d'}{c}\right) - \frac{E_0 d_0}{d''} \cos\left(\omega_c \frac{d''}{c} - \omega_c \frac{d''}{c}\right) \tag{4.34}$$

$$= \frac{E_0 d_0}{d'} \cos\left(\frac{\Delta \omega_c}{c}\right) - \frac{E_0 d_0}{d''} \cos(0^o) \tag{4.35}$$

$$= \frac{E_0 d_0}{d'} \angle \theta_\Delta - \frac{E_0 d_0}{d''} \tag{4.36}$$

$$\approx \frac{E_0 d_0}{d} (\angle \theta_\Delta - 1). \tag{4.37}$$

Using phasor diagram concept for vector addition as shown in Figures 4.3 and 4.4, we get

$$|E_R^{TOT}(d)| = \sqrt{\left(\frac{E_0 d_0}{d} + \frac{E_0 d_0}{d} \cos(\theta_\Delta)\right)^2 + \left(\frac{E_0 d_0}{d} \sin(\theta_\Delta)\right)^2} \tag{4.38}$$

$$= \frac{E_0 d_0}{d} \sqrt{(\cos(\theta_\Delta) - 1)^2 + \sin^2(\theta_\Delta)} \tag{4.39}$$

$$= \frac{E_0 d_0}{d} \sqrt{2 - 2\cos\theta_\Delta} \tag{4.40}$$

$$= 2\frac{E_0 d_0}{d} \sin\left(\frac{\theta_\Delta}{2}\right). \tag{4.41}$$

For $\frac{\theta_\Delta}{2} < 0.5 rad$, $\sin(\frac{\theta_\Delta}{2}) \approx \frac{\theta_\Delta}{2}$. Using equation (4.31) and further equation (4.30), we can then approximate that

$$\sin\left(\frac{\theta_\Delta}{2}\right) \approx \frac{\pi}{\lambda} \Delta = \frac{2\pi h_t h_r}{\lambda d} < 0.5 rad. \tag{4.42}$$

This raises the wonderful concept of 'cross-over distance' d_c, defined as

$$d > d_c = \frac{20\pi h_t h_r}{5\lambda} = \frac{4\pi h_t h_r}{\lambda}. \tag{4.43}$$

The corresponding approximate received electric field is

$$E_R^{TOT}(d) \approx 2\frac{E_0 d_0}{d}\frac{2\pi h_t h_r}{\lambda d} = k\frac{h_t h_r}{d^2}. \tag{4.44}$$

Therefore, using equation (4.43) in (4.1), we get the received power as

$$P_r = \frac{P_{tttttr} h_t^2 h_r^2}{Ld^4}. \tag{4.45}$$

The cross-over distance shows an approximation of the distance after which the received power decays with its fourth order. The basic difference between equation (4.1) and (4.45) is that when $d < d_c$, equation (4.1) is sufficient to calculate the path loss since the two-ray model does not give a good result for a short distance due to the oscillation caused by the constructive and destructive combination of the two rays, but whenever we distance crosses the 'cross-over distance', the power falls off rapidly as well as two-ray model approximation gives better result than Friis equation.

Observations on Equation (4.45): The important observations from this equation are:

1. This equation gives fair results when the T-R separation distance crosses the cross-over distance.
1. In that case, the power decays as the fourth power of distance

$$P_r(d) = \frac{K}{d^4}, \tag{4.46}$$

with K being a constant.

2. Path loss is independent of frequency (wavelength).
3. Received power is also proportional to h_t^2 and h_r^2, meaning, if height of any of the antennas is increased, received power increases.

Diffraction

Diffraction is the phenomena that explains the digression of a wave from a straight line path, under the influence of an obstacle, so as to propagate behind the obstacle. It is an inherent feature of a wave be it longitudinal or transverse. For e.g the sound can be heard in a room, where the source of the sound is another room without having any line of sight. The similar phenomena occurs for light also but the diffracted light intensity is not noticeable. This is because the obstacle or slit need to be of the order of the wavelength of the wave to have a significant effect. Thus radiation from a point source radiating in all directions can be received at any

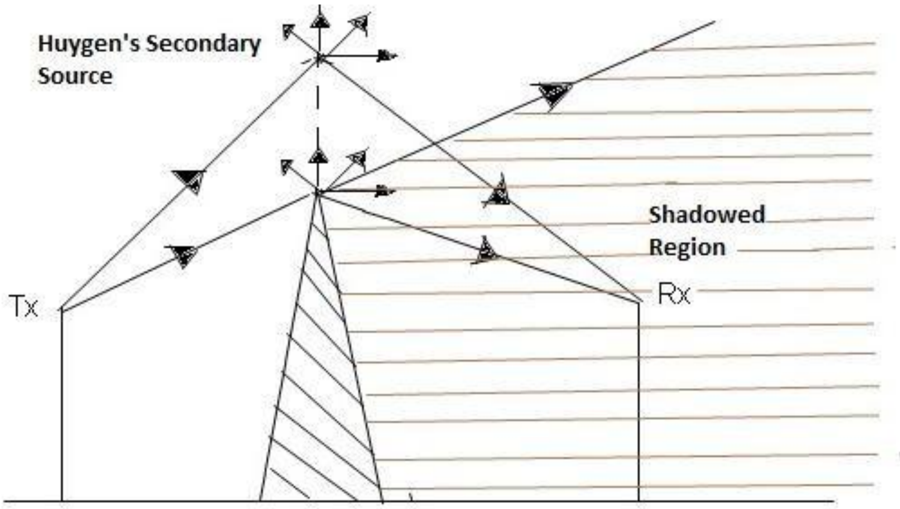


Figure 4.5: Huygen's secondary wavelets.

point, even behind an obstacle (unless it is not completely enveloped by it), as shown in Figure 4.5. Though the intensity received gets smaller as receiver is moved into the shadowed region. Diffraction is explained by Huygens-Fresnel principle which states that all points on a wavefront can be considered as the point source for secondary wavelets which form the secondary wavefront in the direction of the prorogation. Normally, in absence of an obstacle, the sum of all wave sources is zero at a point not in the direct path of the wave and thus the wave travels in the straight line. But in the case of an obstacle, the effect of wave source behind the obstacle cannot be felt and the sources around the obstacle contribute to the secondary wavelets in the shadowed region, leading to bending of wave. In mobile communication, this has a great advantage since, by diffraction (and scattering, reflection), the receiver is able to receive the signal even when not in line of sight of the transmitter. This we show in the subsection given below.

Knife-Edge Diffraction Geometry

As shown in Figure 4.6, consider that there's an impenetrable obstruction of hight h at a distance of d_1 from the transmitter and d_2 from the receiver. The path difference between direct path and the diffracted path is

$$\delta = \sqrt{d_1^2 + h^2} + \sqrt{d_2^2 + h^2} - (d_1 + d_2) \qquad (4.47$$

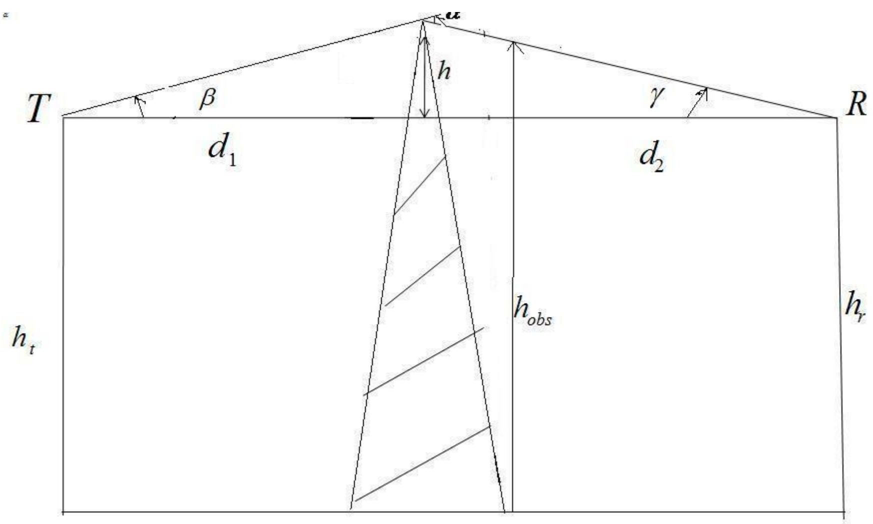


Figure 4.6: Diffraction through a sharp edge.

which can be further simplified as

$$\begin{aligned}\delta &= d_1(1 + h^2/2d_1^2) + d_2(1 + h^2/2d_2^2) - (d_1 + d_2)\\ &= h^2/(2d_1) + h^2/(2d_2) = h^2(d_1 + d_2)/(2d_1d_2).\end{aligned} \tag{4.48}$$

Thus the phase difference equals

$$\varphi = 2\pi\delta/\lambda = 2\pi h^2(d_1 + d_2)/\lambda 2(d_1d_2). \tag{4.49}$$

With the following considerations that

$$\alpha = \beta + \gamma \tag{4.50}$$

and

$$\alpha \approx \tan\alpha \tag{4.51}$$

we can write,

$$\alpha\tan\alpha = \tan\beta + \tan\gamma = h/d_1 + h/d_2 = h(d_1 + d_2)/d_1d_2. \tag{4.52}$$

In order to normalize this, we usually use a Fresnel-Kirchoff diffraction parameter v, expressed as

$$v = h\sqrt{2(d_1 + d_2)/(\lambda d_1d_2)} = \alpha\sqrt{(2d_1d_2)/(\lambda(d_1 + d_2))} \tag{4.53}$$

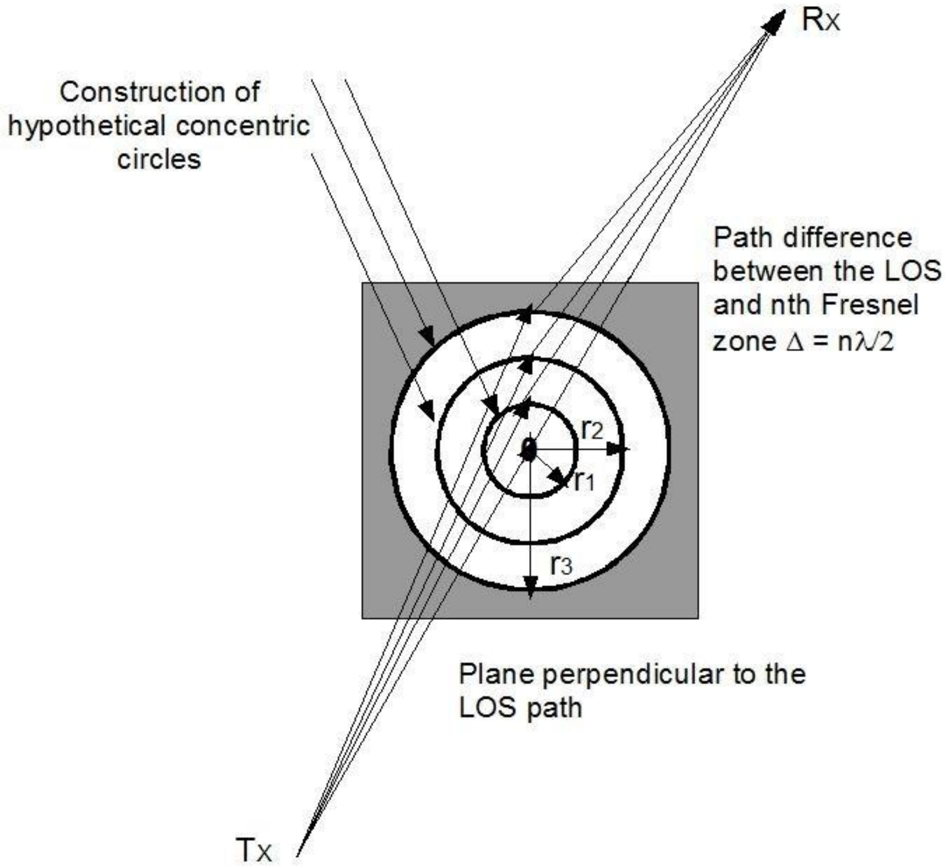


Figure 4.7: Fresnel

zones. and therefore the phase difference becomes

$$\varphi = \pi v^2/2. \tag{4.54}$$

From this, we can observe that: (i) phase difference is a function of the height of the obstruction, and also, (ii) phase difference is a function of the position of the obstruction from transmitter and receiver.

Fresnel Zones: the Concept of Diffraction Loss

As mentioned before, the more is the object in the shadowed region greater is the diffraction loss of the signal. The effect of diffraction loss is explained by Fresnel zones as a function of the path difference. The successive Fresnel zones are limited by the circular periphery through which the path difference of the secondary waves is $n\lambda/2$ greater than total length of the LOS path, as shown in Figure 4.7. Thus successive Fresnel zones have phase difference of π which means they alternatively provide constructive and destructive interference to the received the signal. The radius of the each Fresnel zone is maximum at middle of transmitter and receiver (i.e. when $d_1 = d_2$) and decreases as moved to either side. It is seen that the loci of a Fresnel zone varied over d1 and d2 forms an ellipsoid with the transmitter and receiver at its focii. Now, if there's no obstruction, then all Fresnel zones result in only the direct LOS prorogation and no diffraction effects are observed. But if an obstruction is present, depending on its geometry, it obstructs contribution from some of the secondary wavelets, resulting in diffraction and also the loss of energy, which is the vector sum of energy from unobstructed sources. please note that height of the obstruction can be positive zero and negative also. The diffraction losses are minimum as long as obstruction doesn't block volume of the 1st Fresnel zone. As a rule of thumb, diffraction effects are negligible beyond 55% of 1st Fresnel zone.

Ex 4: Calculate the first Fresnel zone obstruction height maximum for f = 800 MHz.

Solution:

$$\lambda = \frac{c}{f} = \frac{3 \times 10^8}{8 \times 10^2 \times 10^6} = \frac{3}{8}\,m$$

$$H = \frac{\lambda(d_1+d_2)}{d_1+d_2}$$

$$H_1 = \frac{\frac{3}{8}\,250\times 250}{500} = 6.89\,m$$

Thus $H_1 = 10 + 6.89 = 16.89m$

(b)

$$H_2 = \frac{\frac{3}{8}\times 100 \times 400}{500} = 10(0.3) = 5.48\,m$$

Thus

$$H_2 = 10 + 5.6 = 15.48m$$

. To have good power strength, obstacle should be within the 60% of the first fresnel zone.

Ex 5: Given f=900 MHz, $d_1 = d_2 = 1$ km, h = 25m, where symbols have usual meaning. Compute the diffraction loss. Also find out in which Fresnel zone the tip of the obstruction lies.

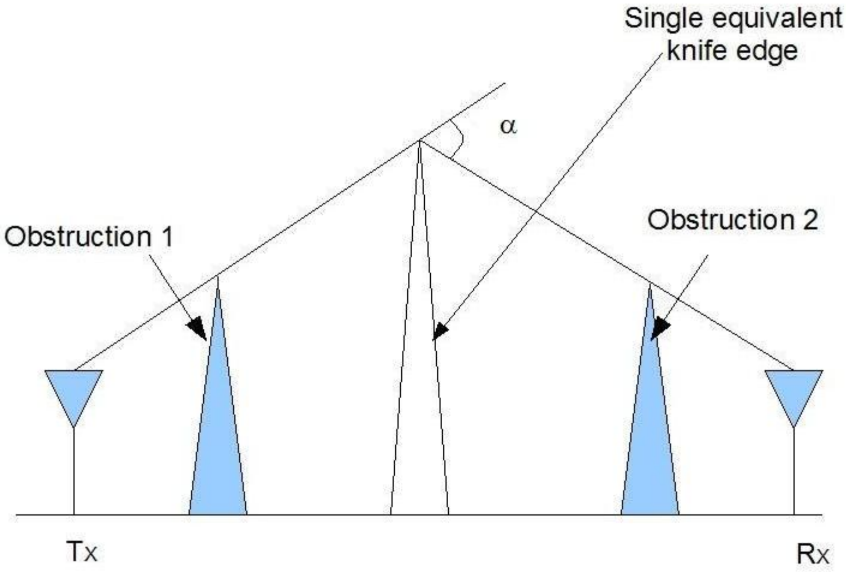


Figure 4.8: Knife-edge Diffraction Model

Given,

$$tt_d(dB) = 20\log(0.5 - 0.62v) \qquad -1 < v <= 0$$

$$tt_d(dB) = 20\log(0.225/v) \qquad v > 2.24$$

Solution:

$$v = h\sqrt{\frac{2(d_1 + d_2)}{\lambda d_1 d_2}} = 25\sqrt{\frac{2 \times 2000}{\frac{1}{3}\,10}} = 2.74$$

$$G_d(dB) = 20\log\left(\frac{225}{v}\right) = -21.7dB$$

Since loss = $-tt_d$ (dB) = 21.7 dB

$$n = \frac{(2.74)^2}{2} = 3.5$$

Thus n=4.

4.5.3 Knife-edge diffraction model

Knife-edge diffraction model is one of the simplest diffraction model to estimate the diffraction loss. It considers the object like hill or mountain as a knife edge sharp object. The electric field strength, E_d of a knife-edge diffracted wave is given by

$$E_d/E_o = F(v) = (1+j)/2 \int_v^{\infty} (exp((-j\pi t^2)/2)dt. \qquad (4.55)$$

The diffraction gain due to presence of knife edge can be given as

$$tt_d(db) = 20\log|F(v)| \qquad (4.56)$$

$$tt_d(db) = 0v <= -1 \qquad (4.57)$$

$$tt_d(db) = 20\log(0.5-0.62) \quad -1 <= v <= 0 \qquad (4.58)$$

$$tt_d(db) = 20\log(0.5exp(-0.95v)) \;\; 0 <= v <= 1 \qquad (4.59)$$

$$tt_d(db) = 20\log(0.4 - sqrt(0.1184-(0.38-0.1v^2))) \quad 1 <= v <= 2.4 \qquad (4.60)$$

$$tt_d(db) = 20\log(0.225/v) \quad v > 2.4 \qquad (4.61)$$

When there are more than one obstruction, then the equivalent model can be found by one knife-edge diffraction model as shown in Figure 4.8.

Link Budget Analysis

Log-distance Path Loss Model

According to this model the received power at distance *d* is given by,

$$PL(d)(\frac{d}{d_0})^n \Longrightarrow PL(dB) = PL(d_0) + 10nlog(\frac{d}{d_0}) \qquad (4.62)$$

The value of n varies with propagation environments. The value of n is 2 for free space. The value of n varies from 4 to 6 for obstruction of building, and 3 to 5 for urban scenarios. The important factor 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.

Limitations:

Surrounding environmental clutter may be different for two locations having the same transmitter to receiver separation. Moreover it does not account for the shadowing effects.

Log Normal Shadowing

The equation for the log normal shadowing is given by,

$$PL(dB) = \overline{PL}(dB) + X_\sigma = \overline{PL}(d_0) + 10n\log\left(\frac{d}{d_0}\right) + X_\sigma \tag{4.63}$$

where X_σ is a zero mean Gaussian distributed random variable in dB with standard deviation σ also in dB. In practice n and σ values are computed from measured data.

Average received power

The 'Q' function is given by,

$$Q(z) = 0.5\left(1 - erf\left(\frac{z}{\sqrt{2}}\right)\right) \tag{4.64}$$

and

$$Q(z) = 1 - Q(-z) \tag{4.65}$$

So the probability that the received signal level (in dB) will exceed a certain value γ is

$$P\,(P_d > \gamma) = Q\left(\frac{\gamma - \overline{P_r}}{\sigma}\right). \tag{4.66}$$

Outdoor Propagation Models

There are many empirical outdoor propagation models such as Longley-Rice model, Durkin's model, Okumura model, Hata model etc. Longley-Rice model is the most commonly used model within a frequency band of 40 MHz to 100 GHz over different terrains. Certain modifications over the rudimentary model like an extra urban factor (UF) due to urban clutter near the reciever is also included in this model. Below, we discuss some of the outdoor models, followed by a few indoor models too.

Okumura Model

The Okumura model is used for Urban Areas is a Radio propagation model that is used for signal prediction.The frequency coverage of this model is in the range of 200 MHz to 1900 MHz and distances of 1 Km to 100 Km.It can be applicable for base station effective antenna heights (h_t) ranging from 30 m to 1000 m.

Okumura used extensive measurements of base station-to-mobile signal attenuation throughout Tokyo to develop a set of curves giving median attenuation relative to free space (A_{mu}) of signal propogation in irregular terrain. The empirical path-loss formula of Okumura at distance d parameterized by the carrier frequency f_c is given by

$$P_L(d)dB = L(f_c, d) + A_{mu}(f_c, d) - tt(h_t) - tt(h_r) - tt_{AREA} \qquad (4.67)$$

where $L(f_c, d)$ is free space path loss at distance d and carrier frequency f_c, $A_{mu}(f_c, d)$ is the median attenuation in addition to free-space path loss across all environments, $tt(h_t)$ is the base station antenna height gain factor, $tt(h_r)$ is the mobile antenna height gain factor, tt_{AREA} is the gain due to type of environment. The values of $A_{mu}(f_c, d)$ and tt_{AREA} are obtained from Okumura's empirical plots. Okumura derived empirical formulas for $tt(h_t)$ and $tt(h_r)$ as follows:

$$tt(h_t) = 20 \log_{10}(h_t/200), \quad 30m < h_t < 1000m \qquad (4.68)$$

$$tt(h_r) = 10 \log_{10}(h_r/3), \quad h_r \leq 3m \qquad (4.69)$$

$$tt(h_r) = 20 \log_{10}(h_r/3), \quad 3m < h_r < 10m \qquad (4.70)$$

Correlation factors related to terrain are also developed in order to improve the models accuracy. Okumura's model has a 10-14 dB empirical standard deviation between the path loss predicted by the model and the path loss associated with one of the measurements used to develop the model.

Hata Model

The Hata model is an empirical formulation of the graphical path-loss data provided by the Okumura and is valid over roughly the same range of frequencies, 150-1500 MHz. This empirical formula simplifies the calculation of path loss because it is closed form formula and it is not based on empirical curves for the different parameters. The standard formula for empirical path loss in urban areas under the Hata model is

$$P_{L,urban}(d)dB = 69.55+26.16 \log_{10}(f_c)-13.82 \log_{10}(h_t)-a(h_r)+(44.9-6.55 \log_{10}(h_t)) \log_{10}(d) \qquad (4.71)$$

The parameters in this model are same as in the Okumura model,and $a(h_r)$ is a correction factor for the mobile antenna height based on the size of coverage area.For small to medium sized cities this factor is given by

$$a(h_r) = (1.11 \log_{10}(f_c) - 0.7)h_r - (1.56 \log_{10}(f_c) - 0.8)dB$$

and for larger cities at a frequencies f_c > 300 MHz by

$$a(h_r) = 3.2(\log_{10}(11.75h_r))^2 - 4.97dB$$

else it is

$$a(h_r) = 8.29(\log_{10}(1.54h_r))^2 - 1.1dB$$

Corrections to the urban model are made for the suburban, and is given by

$$P_{L,suburban}(d)dB = P_{L,urban}(d)dB - 2(\log_{10}(f_c/28))^2 - 5.4 \quad (4.72)$$

Unlike the Okumura model,the Hata model does not provide for any specific path-correlation factors. The Hata model well approximates the Okumura model for distances d > 1 Km. Hence it is a good model for first generation cellular systems, but it does not model propogation well in current cellular systems with smaller cell sizes and higher frequencies. Indoor environments are also not captured by the Hata model.

Indoor Propagation Models

The indoor radio channel differs from the traditional mobile radio channel in ways - the distances covered are much smaller ,and the variability of the environment is much greater for smaller range of Tx-Rx separation distances.Features such as layout of the building,the construction materials,and the building type strongly influence the propagation within the building.Indoor radio propagation is dominated by the same mechanisms as outdoor: reflection, diffraction and scattering with variable conditions. In general,indoor channels may be classified as either line-of-sight or obstructed.

Partition Losses Inside a Floor (Intra-floor)

The internal and external structure of a building formed by partitions and obstacles vary widely.Partitions that are formed as a part of building structure are called hard partitions , and partitions that 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.

Partition Losses Between Floors (Inter-floor)

The losses between floors of a building are determined by the external dimensions and materials of the building,as well as the type of construction used to create the floors and the external surroundings. Even the number of windows in a building and the presence of tinting can impact the loss between floors.

Log-distance Path Loss Model

It has been observed that indoor path loss obeys the distance power law given by

$$PL(dB) = PL(d_0) + 10n\log_{10}(d/d_0) + X_\sigma \tag{4.73}$$

where n depends on the building and surrounding type, and X_σ represents a normal random variable in dB having standard deviation of σ dB.

Summary

In this chapter, three principal propagation models have been identified: free-space propagation, reflection and diffraction, which are common terrestrial models and these mainly explains the large scale path loss. Regarding path-loss, one important factor introduced in this chapter is log-distance path loss model. These, however, may be insignificant when we consider the small-scale rapid path losses. This is discussed in the next chapter.

Multipath Wave Propagation and Fading

Multipath Propagation

In wireless telecommunications, multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. Causes of multipath include atmospheric ducting, ionospheric reflection and refraction, and reflection from water bodies and terrestrial objects such as mountains and buildings. The effects of multipath include constructive and destructive interference, and phase shifting of the signal. In digital radio communications (such as GSM) multipath can cause errors and affect the quality of communications. We discuss all the related issues in this chapter.

Multipath & Small-Scale Fading

Multipath signals are received in a terrestrial environment, i.e., where different forms of propagation are present and the signals arrive at the receiver from transmitter via a variety of paths. Therefore there would be multipath interference, causing multipath fading. Adding the effect of movement of either Tx or Rx or the surrounding clutter to it, the received overall signal amplitude or phase changes over a small amount of time. Mainly this causes the fading.

Fading

The term **fading**, or, small-scale fading, means rapid fluctuations of the amplitudes, phases, or multipath delays of a radio signal over a short period or short travel distance. This might be so severe that large scale radio propagation loss effects might be ignored.

Multipath Fading Effects

In principle, the following are the main multipath effects:

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

Factors Influencing Fading

The following physical factors influence small-scale fading in the radio propagation channel:

(1) Multipath propagation – Multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. The effects of multipath include constructive and destructive interference, and phase shifting of the signal.

(2) 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.

(3) Speed of surrounding objects – If objects in the radio channel are in motion, they induce a time varying Doppler shift on multipath components. If the surrounding objects move at a greater rate than the mobile, then this effect dominates fading.

(4) Transmission Bandwidth of the signal – If the transmitted radio signal bandwidth is greater than the "bandwidth" of the multipath channel (quantified by *coherence bandwidth*), the received signal will be distorted.

Types of Small-Scale Fading

The type of fading experienced by the signal through a mobile channel depends on the relation between the signal parameters (bandwidth, symbol period) and the channel parameters (rms delay spread and Doppler spread). Hence we have four different types of fading. There are two types of fading due to the time dispersive nature of the channel.

Fading Effects due to Multipath Time Delay Spread

Flat Fading

Such types of fading occurs when the bandwidth of the transmitted signal is less than the coherence bandwidth of the channel. Equivalently if the symbol period of the signal is more than the rms delay spread of the channel, then the fading is flat fading.

So we can say that flat fading occurs when

$$B_S \quad B_C \tag{5.1}$$

where B_S is the signal bandwidth and B_C is the coherence bandwidth. Also

$$T_S \quad \sigma_\tau \tag{5.2}$$

where T_S is the symbol period and σ_τ is the rms delay spread. And in such a case, mobile channel has a constant gain and linear phase response over its bandwidth.

Frequency Selective Fading

Frequency selective fading occurs when the signal bandwidth is more than the coherence bandwidth of the mobile radio channel or equivalently the symbols duration of the signal is less than the rms delay spread.

At the receiver, we obtain multiple copies of the transmitted signal, all attenuated and delayed in time. The channel introduces inter symbol interference. A rule of thumb for a channel to have flat fading is if

$$\frac{\sigma_\tau}{T_S} \leq 0.1 \tag{5.5}$$

Fading Effects due to Doppler Spread

Fast Fading

In a fast fading channel, the channel impulse response changes rapidly within the symbol duration of the signal. Due to Doppler spreading, signal undergoes frequency dispersion leading to distortion. Therefore a signal undergoes fast fading if

$$T_S \quad T_C \tag{5.6}$$

where T_C is the coherence time and

$$B_S \quad B_D \tag{5.7}$$

where B_D is the Doppler spread. Transmission involving very low data rates suffer from fast fading.

Slow Fading

In such a channel, the rate of the change of the channel impulse response is much less than the transmitted signal. We can consider a slow faded channel a channel in which channel is almost constant over atleast one symbol duration. Hence

$$T_S \quad T_C \tag{5.8}$$

and

B_S
B_D
5.9
)

We observe that the velocity of the user plays an important role in deciding whether the signal experiences fast or slow fading.

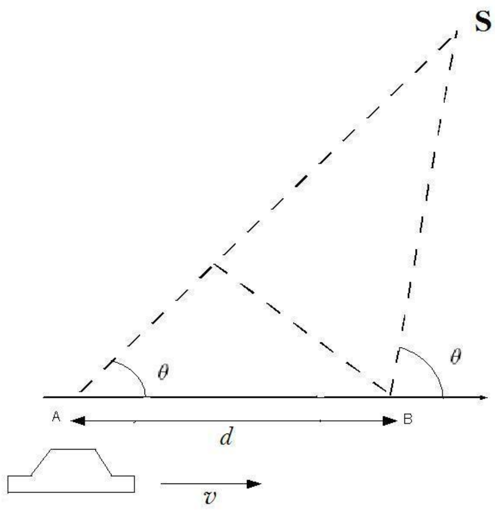


Figure 5.1: Illustration of Doppler effect.

Doppler Shift

The Doppler effect (or Doppler shift) is the change in frequency of a wave for an observer moving relative to the source of the wave. In classical physics (waves in a medium), the relationship between the observed frequency f and the emitted frequency f_o is given by:

$$f = \left(\frac{v \pm v_r}{v \pm v_s} \right) f_0 \tag{5.10}$$

where v is the velocity of waves in the medium, v_s is the velocity of the source relative to the medium and v_r is the velocity of the receiver relative to the medium.

In mobile communication, the above equation can be slightly changed according to our convenience since the source (BS) is fixed and located at a remote elevated level from ground. The expected Doppler shift of the EM wave then comes out to be $\pm \frac{v_r}{c} f_o$ or, $\pm \frac{v_r}{\lambda}$. As the BS is located at an elevated place, a $\cos \varphi$ factor would also be multiplied with this. The exact scenario, as given in Figure 5.1, is illustrated below.

Consider a mobile moving at a constant velocity v, along a path segment length d between points A and B, while it receives signals from a remote BS source S. The difference in path lengths traveled by the wave from source S to the mobile at points A and B is $\Delta l = d \cos \theta = v \Delta t \cos \theta$, where Δt is the time required for the mobile to travel from A to B, and θ is assumed to be the same at points A and B 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

$$\Delta\phi = \frac{2\pi\Delta l}{\lambda} = \frac{2\pi v\Delta t}{\lambda}\cos\theta \tag{5.11}$$

and hence the apparent change in frequency, or Doppler shift (f_d) is

$$f_d = \frac{1}{2\pi}.\frac{\Delta\phi}{\Delta t} = \frac{v}{\lambda}.\cos\theta. \tag{5.12}$$

Example 1

An aircraft is heading towards a control tower with 500 kmph, at an elevation of 20°. Communication between aircraft and control tower occurs at 900 MHz. Find out the expected Doppler shift.

Solution As given here,

$$v = 500\ kmph$$

the horizontal component of the velocity is

$$v^{j} = v\cos\theta = 500 \times \cos 20^\circ = 130\ m/s$$

Hence, it can be written that

$$\lambda = \frac{900 \times 10^6}{3 \times 10^8} = \frac{1}{3}m$$

$$f_d = \frac{130}{1/3} = 390\ Hz$$

If the plane banks suddenly and heads for other direction, the Doppler shift change will be 390 Hz to $-390\ Hz$.

Impulse Response Model of a Multipath Channel

Mobile radio channel may be modeled as a linear filter with time varying impulse response in continuous time. To show this, consider time variation due to receiver motion and time varying impulse response $h(d, t)$ and $x(t)$, the transmitted signal. The received signal $y(d, t)$ at any position d would be

$$y(d,t) = x(t) * h(d,t) = \int_{-\infty}^{\infty} x(\tau)\, h(d, t-\tau)\, d\tau \tag{5.13}$$

For a causal system: $h(d, t) = 0$, for $t < 0$ and for a stable system $\int_{-\infty}^{\infty} |h(d,t)|\, dt < \infty$

Applying causality condition in the above equation, $h(d, t-\tau) = 0$ for $t-\tau < 0$ $\Rightarrow \tau > t$, i.e., the integral limits are changed to

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

Since the receiver moves along the ground at a constant velocity v, the position of the receiver is $d = vt$, i.e.,

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

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

$$y(t) = \int_{-\infty}^{t} x(\tau)\, h(vt, t-\tau)\, d\tau = x(t) * h(vt, t) = x(t) * h(d, t) \tag{5.14}$$

It is useful to discretize the multipath delay axis τ of the impulse response into equal time delay segments called *excess delay bins*, each bin having a time delay width equal to $(\tau_{i+1} - \tau_i) = \Delta\tau$ and $\tau_i = i\Delta\tau$ for $i \in \{0, 1, 2, ..N-1\}$, where N represents the total number of possible equally-spaced multipath components, including the first arriving component. The useful frequency span of the model is $2/\Delta\tau$. The model may be used to analyze transmitted RF signals having bandwidth less than $2/\Delta\tau$.

If there are N multipaths, maximum excess delay is given by $N\Delta\tau$.

$$\{y(t) = x(t) * h(t, \tau_i) | i = 0, 1, ...N-1\} \tag{5.15}$$

Bandpass channel impulse response model is

$$x(t) \rightarrow h(t, \tau) = Re\{h_b(t, \tau)e^{j\omega ct}\} \rightarrow y(t) = Re\{r(t)e^{j\omega ct}\} \tag{5.16}$$

Baseband equivalent channel impulse response model is given by

$$c(t) \rightarrow \frac{1}{2} h_b(t, \tau) \rightarrow r(t) = c(t) * \frac{1}{2} h_b(t, \tau) \tag{5.17}$$

Average power is

$$\overline{x^2(t)} = \frac{1}{2}|c(t)|^2 \tag{5.18}$$

The baseband impulse response of a multipath channel can be expressed as

$$h_b(t,\tau) = \sum_{i=0}^{N-1} a_i(t,\tau)\exp[j(2\pi f_c\tau_i(t) + \phi_i(t,\tau))]\delta(\tau - \tau_i(t)) \qquad (5.19)$$

where $a_i(t,\tau)$ and $\tau_i(t)$ are the real amplitudes and excess delays, respectively, of the ith multipath component at time t. The phase term $2\pi f_c\tau_i(t) + \phi_i(t,\tau)$ in the above equation represents the phase shift due to free space propagation of the ith multipath component, plus any additional phase shifts which are encountered in the channel.

If the channel impulse response is wide sense stationary over a small-scale time or distance interval, then

$$h_b(\tau) = \sum_{i=0}^{N-1} a_i \exp[j\theta_i]\delta(\tau - \tau_i) \qquad (5.20)$$

For measuring $h_b(\tau)$, we use a probing pulse to approximate $\delta(t)$ i.e.,

$$p(t) \approx \delta(t-\tau) \qquad (5.21)$$

Power delay profile is taken by spatial average of $|h_b(t,\tau)|^2$ over a local area. The received power delay profile in a local area is given by

$$p(\tau) \approx \overline{k|h_b(t;\tau)|^2}. \qquad (5.22)$$

Relation Between Bandwidth and Received Power

In actual wireless communications, impulse response of a multipath channel is measured using channel sounding techniques. Let us consider two extreme channel sounding cases.

Consider a pulsed, transmitted RF signal

$$x(t) = Re\{p(t)e^{j2\pi f_c t}\} \qquad (5.23)$$

where $p(t) = \sqrt{\frac{4\tau_{max}}{T_{bb}}}$ for $0 \le t \le T_{bb}$ and 0 elsewhere. The low pass channel output is

$$r(t) = \frac{1}{2}\sum_{i=0}^{N-1} a_i \exp[j\theta_i]p(t-\tau_i)$$

$$= \sum_{i=0}^{N-1} a_i \exp[j\theta_i].\sqrt{\frac{\tau_{max}}{T_{bb}}}\, rect(t - \frac{T_b}{2} - \tau_i).$$

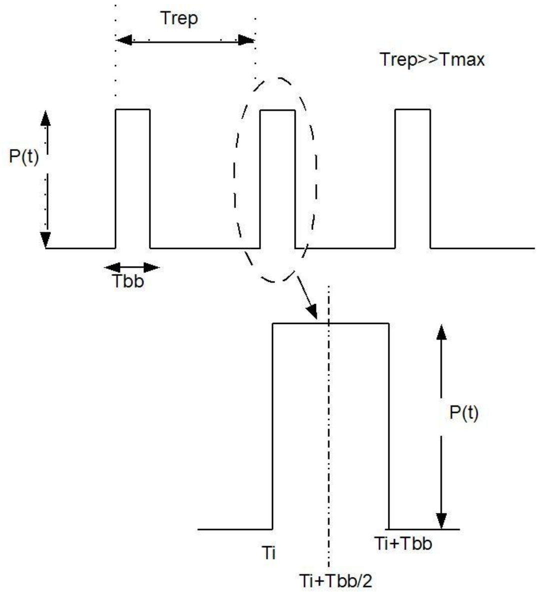


Figure 5.2: A generic transmitted pulsed RF signal.

The received power at any time t_0 is

$$
\begin{aligned}
|r(t_0)|^2 &= \frac{1}{\tau_{max}} \int_0^{\tau_{max}} r(t) r^*(t) dt \\
&= \frac{1}{\tau_{max}} \int_0^{\tau_{max}} \frac{1}{4} \left\{ \sum_{k=0}^{N-1} a_k^2(t_0) p^2(t - \tau_k) \right\} dt \\
&= \frac{1}{\tau_{max}} \sum_{k=0}^{N-1} a_k^2(t_0) \int_0^{\tau_{max}} \left\{ \sqrt{\frac{\tau_{max}}{T_{bb}}}\, rect\left(t - \frac{T_b}{2} - \tau_i\right) \right\}^2 dt \\
&= \sum_{k=0}^{N-1} a_k^2(t_0).
\end{aligned}
$$

Interpretation: If the transmitted signal is able to resolve the multipaths, then average small-scale receiver power is simply sum of average powers received from each multipath components.

$$
E_{a,\theta}[P_{WB}] = E_{a,\theta}\left[\sum_{i=0}^{N-1} |a_i \exp(j\theta_i)|^2\right] \approx \sum_{i=0}^{N-1} \overline{a_i^2} \tag{5.24}
$$

Now instead of a pulse, consider a CW signal, transmitted into the same channel and for simplicity, let the envelope be $c(t) = 2$. Then

$$
r(t) = \sum_{i=0}^{N-1} a_i \exp[j\theta_i(t, \tau)] \tag{5.25}
$$

and the instantaneous power is

$$|r(t)|^2 = \left|\sum_{i=0}^{N-1} a_i \exp[j\theta_i(t,\tau)]\right|^2 \tag{5.26}$$

Over local areas, a_i varies little but θ_i varies greatly resulting in large fluctuations.

$$\begin{aligned} E_{a,\theta}[P_{CW}] &= E_{a,\theta}\left[\left|\sum_{i=0}^{N-1} a_i \exp(j\theta_i)\right|^2\right] \\ &\approx \sum_{i=0}^{N-1} a_i^2 + 2\sum_{i=0}^{N-1}\sum_{i,j\neq i}^{N} r_{ij}\overline{\cos(\theta_i-\theta_j)} \end{aligned}$$

where $r_{ij} = E_a[a_i a_j]$.

If, $r_{ij} = \overline{\cos(\theta_i - \theta_j)} = 0$, then $E_{a,\theta}[P_{CW}] = E_{a,\theta}[P_{WB}]$. This occurs if multipath components are uncorrelated or if multipath phases are i.i.d over [0, 2π].

Bottomline:

1. If the signal bandwidth is greater than multipath channel bandwidth then fading effects are negligible
2. If the signal bandwidth is less than the multipath channel bandwidth, large fading occurs due to phase shift of unresolved paths.

Linear Time Varying Channels (LTV)

The time variant transfer function(TF) of an LTV channel is FT of $h(t,\tau)$ w.r.t. τ.

$$H(f,t) = FT[h(\tau,t)] = \int_{-\infty}^{\infty} h(\tau,t)e^{-j2\pi f\tau}\,d\tau \tag{5.27}$$

$$h(\tau,t) = FT^{-1}[H(f,t)] = \int_{-\infty}^{\infty} H(f,t)e^{j2\pi f\tau}\,df \tag{5.28}$$

The received signal

$$r(t) = \int_{-\infty}^{\infty} R(f,t)e^{j2\pi ft}\,df \tag{5.29}$$

where $R(f, t) = H(f, t)X(f)$.

For flat fading channel, $h(\tau,t) = Z(t)\delta(\tau-\tau_i)$ where $Z(t) = \sum \alpha_n(t)e^{-j2\pi f_c\tau_n(t)}$.
In this case, the received signal is

$$r(t) = \int_{-\infty}^{\infty} h(\tau,t)x(t-\tau)\,d\tau = Z(t)x(t-\tau_i) \tag{5.30}$$

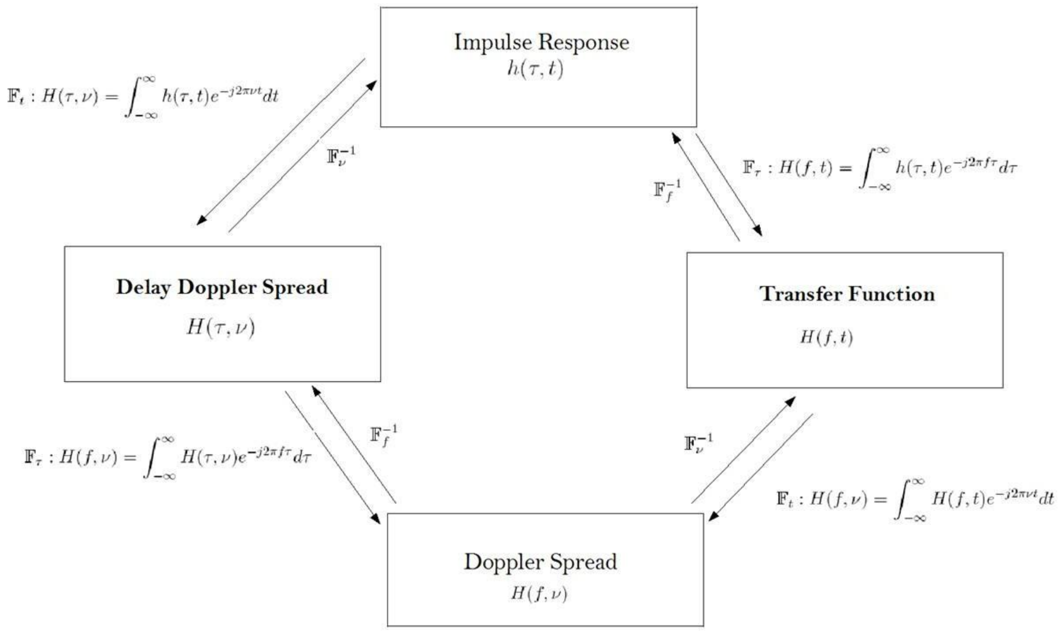


Figure 5.3: Relationship among different channel functions.

where the channel becomes multiplicative.

Doppler spread functions:

$$H(f,v) = FT[H(f,t)] = \int_{-\infty}^{\infty} H(f,t)e^{-j2\pi vt}dt \tag{5.31}$$

and

$$H(f,t) = FT^{-1}[H(f,v)] = \int_{-\infty}^{\infty} H(f,v)e^{j2\pi vt}dv \tag{5.32}$$

Delay Doppler spread:

$$H(\tau,v) = FT[h(\tau,t)] = \int_{-\infty}^{\infty} h(\tau,t)e^{-j2\pi vt}dt \tag{5.33}$$

Small-Scale Multipath Measurements

Direct RF Pulse System

A wideband pulsed bistatic radar usually transmits a repetitive pulse of width T_{bb} s, and uses a receiver with a wide bandpass filter ($BW = \frac{2}{T_{bb}}$ Hz). The signal is then amplified, envelope detected, and displayed and stored on a high speed oscilloscope. Immediate measurements of the square of the channel impulse response convolved with the probing pulse can be taken. If the oscilloscope is set on averaging mode, then this system provides a local average power delay profile.

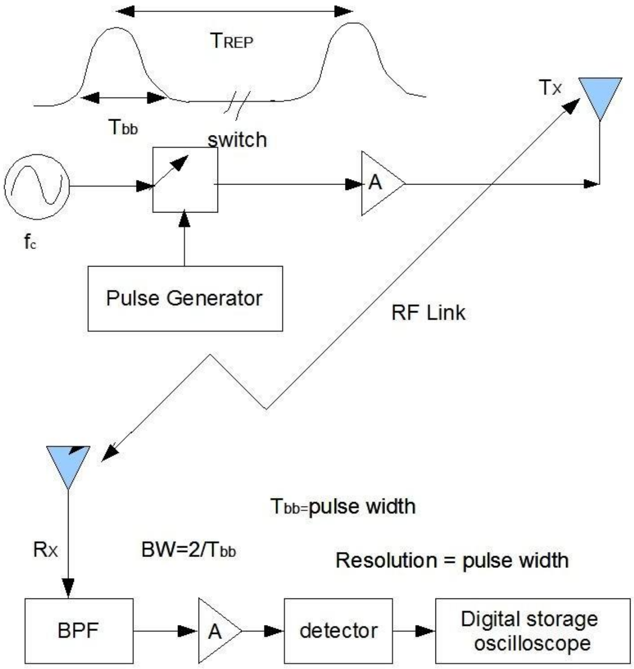


Figure 5.4: Direct RF pulsed channel IR measurement.

This system is subject to interference noise. If the first arriving signal is blocked or fades, severe fading occurs, and it is possible the system may not trigger properly.

Frequency Domain Channel Sounding

In this case we measure the channel in the frequency domain and then convert it into time domain impulse response by taking its inverse discrete Fourier transform (IDFT). A vector network analyzer controls a swept frequency synthesizer. An S-parameter test set is used to monitor the frequency response of the channel. The sweeper scans a particular frequency band, centered on the carrier, by stepping through discrete frequencies. The number and spacing of the frequency step impacts the time resolution of the impulse response measurement. For each frequency step, the S-parameter test set transmits a known signal level at port 1 and monitors the received signal at port 2. These signals allow the analyzer to measure the complex response, $S_{21}(\omega)$, of the channel over the measured frequency range. The $S_{21}(\omega)$ measure is the measure of the signal flow from transmitter antenna to receiver

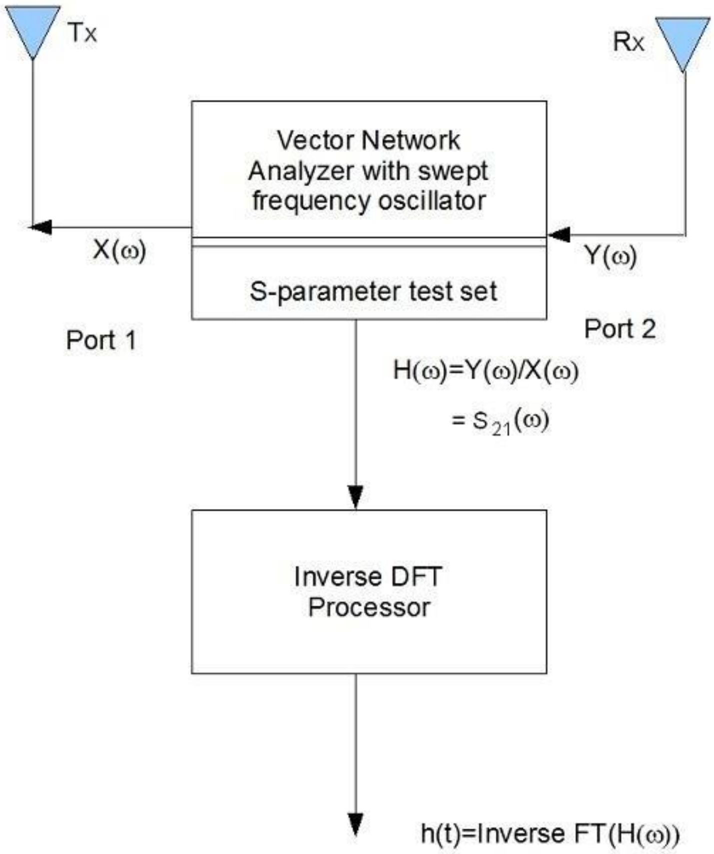


Figure 5.5: Frequency domain channel IR measurement.

antenna (i.e., the channel).

This system is suitable only for indoor channel measurements. This system is also non real-time. Hence, it is not suitable for time-varying channels unless the sweep times are fast enough.

Multipath Channel Parameters

To compare the different multipath channels and to quantify them, we define some parameters. They all can be determined from the power delay profile. These parameters can be broadly divided in to two types.

Time Dispersion Parameters

These parameters include the mean excess delay,rms delay spread and excess delay spread. The mean excess delay is the first moment of the power delay profile and is defined as

$$\bar{\tau} = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} = \frac{\sum_k P(\tau_k)\tau_k}{\sum_k P(\tau_k)} \tag{5.34}$$

where a_k is the amplitude, τ_k is the excess delay and $P(\tau_k)$ is the power of the individual multipath signals.

The mean square excess delay spread is defined as

$$\overline{\tau^2} = \frac{\sum_k P(\tau_k)\tau_k^2}{\sum_k P(\tau_k)} \tag{5.35}$$

Since the rms delay spread is the square root of the second central moment of the power delay profile, it can be written as

$$\sigma_\tau = \sqrt{\overline{\tau^2} - (\bar{\tau})^2} \tag{5.36}$$

As a rule of thumb, for a channel to be flat fading the following condition must be satisfied

$$\frac{\sigma_\tau}{T_S} \leq 0.1 \tag{5.37}$$

where T_S is the symbol duration. For this case, no equalizer is required at the receiver.

Example 2

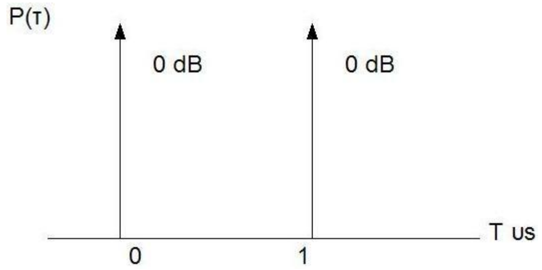
1. Sketch the power delay profile and compute RMS delay spread for the following:
   $P(\tau) = \sum_{n=0}^{1} \delta(\tau - n \times 10^{-6})$ (in watts)
2. If BPSK modulation is used, what is the maximum bit rate that can be sent through the channel without needing an equalizer?

Solution

1. $P(0) = 1$ watt, $P(1) = 1$ watt

$$\bar{\tau} = \frac{(1)(0) + (1)(1)}{1 + 1} = 0.5\mu s$$

$$\overline{\tau^2} = 0.5\mu s^2 \quad \sigma_\tau = 0.5\mu s$$

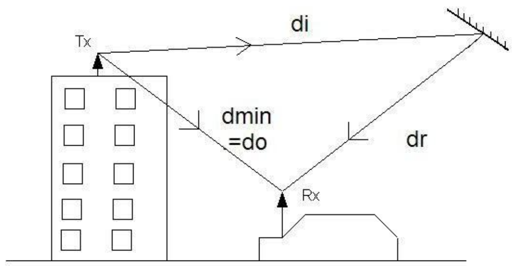


2. For flat fading channel, we need $\frac{\sigma_\tau}{T_s}$ 0.1 $\Rightarrow R_s = \frac{1}{T_s} = 0.2 \times 10^4 = 200\, kbps$

For BPSK we need $R_b = R_s = 200\ kbps$

Example 3 A simple delay spread bound: **Feher's upper bound**

Consider a simple worst-case delay spread scenario as shown in figure below.



Here $d_{min} = d_0$ and $d_{max} = d_i + d_r$

Transmitted power = P_T , Minimum received power = $P_{R\min} = P_{T\ hreshold}$

$$\frac{P_{R\min}}{P_T} = tt_T\, tt_R \left(\frac{\lambda}{4\pi d_{\max}}\right)^2$$

Put $tt_T = tt_R = 1$ i.e., considering omni-directional unity gain antennas

$$d_{\max} = \left(\frac{\lambda}{4\pi}\right)\left(\frac{P_T}{P_{R\min}}\right)^{\frac{1}{2}}$$

$$\tau_{\max} = \frac{d_{\max}}{c} = \left(\frac{\lambda}{4\pi c}\right)\left(\frac{P_T}{P_{R\min}}\right)^{\frac{1}{2}}$$

$$\tau_{\max} = \left(\frac{1}{4\pi f}\right)\left(\frac{P_T}{P_{R\min}}\right)^{\frac{1}{2}}$$

Frequency Dispersion Parameters

To characterize the channel in the frequency domain, we have the following parameters.

(1) Coherence bandwidth: it is a statistical measure of the range of frequencies over which the channel can be considered to pass all the frequency components with almost equal gain and linear phase. When this condition is satisfied then we say the channel to be flat.

Practically, coherence bandwidth is the minimum separation over which the two frequency components are affected differently. If the coherence bandwidth is considered to be the bandwidth over which the frequency correlation function is above 0.9, then it is approximated as

$$BC \approx \frac{1}{50\sigma_\tau}. \tag{5.38}$$

However, if the coherence bandwidth is considered to be the bandwidth over which the frequency correlation function is above 0.5, then it is defined as

$$B_C \approx \frac{1}{5\sigma_\tau}. \tag{5.39}$$

The coherence bandwidth describes the time dispersive nature of the channel in the local area. A more convenient parameter to study the time variation of the channel is the coherence time. This variation may be due to the relative motion between the mobile and the base station or the motion of the objects in the channel.

(2) Coherence time: this is a statistical measure of the time duration over which the channel impulse response is almost invariant. When channel behaves like this, it is said to be slow faded. Essentially it is the minimum time duration over which two received signals are affected differently. For an example, if the coherence time is considered to be the bandwidth over which the time correlation is above 0.5, then it can be approximated as

$$TC \approx \frac{9}{16\pi f_m} \tag{5.40}$$

where f_m is the maximum doppler spread given be $f_m = \frac{v}{\lambda}$

Another parameter is the Doppler spread (B_D) which is the range of frequencies over which the received Doppler spectrum is non zero.

Statistical models for multipath propagation

Many multipath models have been proposed to explain the observed statistical nature of a practical mobile channel. Both the first order and second order statistics

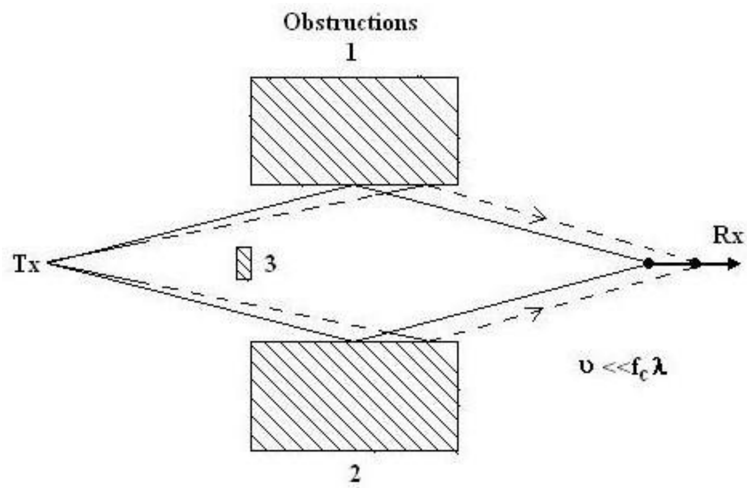


Figure 5.6: Two ray NLoS multipath, resulting in Rayleigh fading.

have been examined in order to find out the effective way to model and combat the channel effects. The most popular of these models are Rayleigh model, which describes the NLoS propagation. The Rayleigh model is used to model the statistical time varying nature of the received envelope of a flat fading envelope. Below, we discuss about the main first order and second order statistical models.

NLoS Propagation: Rayleigh Fading Model

Let there be two multipath signals S_1 and S_2 received at two different time instants due to the presence of obstacles as shown in Figure 5.6. Now there can either be constructive or destructive interference between the two signals.

Let E_n be the electric field and Θ_n be the relative phase of the various multipath signals.So we have

$$\tilde{E} = \sum_{n=1}^{N} E_n e^{j\theta_n} \tag{5.41}$$

Now if $N \rightarrow \infty$(i.e. are sufficiently large number of multipaths) and all the E_n are IID distributed, then by Central Limit Theorem we have,

$$\lim_{N\rightarrow\infty} \tilde{E} = \lim_{N\rightarrow\infty} \sum_{n=1}^{N} E_n e^{j\theta_n} \tag{5.42}$$

$$= Z_r + jZ_i = Re^{j\varphi} \tag{5.43}$$

where Z_r and Z_i are Gaussian Random variables. For the above case

$$R = \sqrt{Z_r^2 + Z_i^2} \tag{5.44}$$

and

$$\varphi = \tan^{-1}\frac{Z_i}{Z_r} \tag{5.45}$$

For all practical purposes we assume that the relative phase Θ_n is uniformaly distributed.

$$E[e^{j\theta_n}] = \frac{1}{2\pi}\int_0^{2\pi} e^{j\theta} d\theta = 0 \tag{5.46}$$

It can be seen that E_n and Θ_n are independent. So,

$$E[\tilde{E}] = E[\sum E_n e^{j\theta_n}] = 0 \tag{5.47}$$

$$E[|\tilde{E}|^2] = E[\sum E_n e^{j\theta_n} \sum E_n^{*} e^{-j\theta_n}] = E[\sum_m \sum_n E_n E_m e^{j(\theta_n-\theta_m)}] = \sum_{n=1}^{N} E_n^2 = P_0 \tag{5.48}$$

where P_0 is the total power obtained. To find the Cumulative Distribution Function(CDF) of R, we proceed as follows.

$$F_R(r) = P_r(R \le r) = \iint_A f_{Z_i,Z_r}(z_i, z_r)\,dz_i\,dz_r \tag{5.49}$$

where A is determined by the values taken by the dummy variable r. Let Z_i and Z_r be zero mean Gaussian RVs. Hence the CDF can be written as

$$F_R(r) = \iint_A \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(Z_r^2+Z_i^2)}{2\sigma^2}}\,dZ_i\,dZ_r \tag{5.50}$$

Let $Z_r = p\cos(\Theta)$ and $Z_i = p\sin(\Theta)$ So we have

$$F_R(r) = \int_0^{2\pi}\int_0^{2\pi} \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-p^2}{2\sigma^2}}\,p\,dp\,d\theta \tag{5.51}$$

$$= 1 - e^{\frac{-r^2}{2\sigma^2}} \tag{5.52}$$

Above equation is valid for all $r \geq 0$. The pdf can be written as

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{r^2}{2\sigma^2}} \tag{5.53}$$

and is shown in Figure 5.7 with different σ values. This equation too is valid for all $r \geq 0$. Above distribution is known as Rayleigh distribution and it has been derived

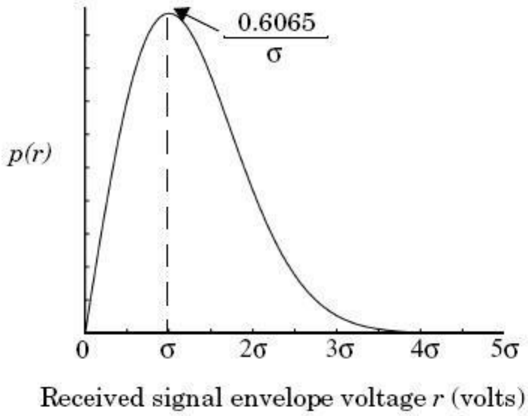


Figure 5.7: Rayleigh probability density function.

for slow fading. However, if f_D 1 Hz, we call it as Quasi-stationary Rayleigh fading. We observe the following:

$$E[R] = \sqrt{\frac{\pi}{2}}\sigma \tag{5.54}$$

$$E[R^2] = 2\sigma^2 \tag{5.55}$$

$$\mathrm{var}[R] = (2 - \frac{\pi}{2})\sigma^2 \tag{5.56}$$

$$median[R] = 1.77\sigma. \tag{5.57}$$

LoS Propagation: Rician Fading Model

Rician Fading is the addition to all the normal multipaths a direct LOS path.

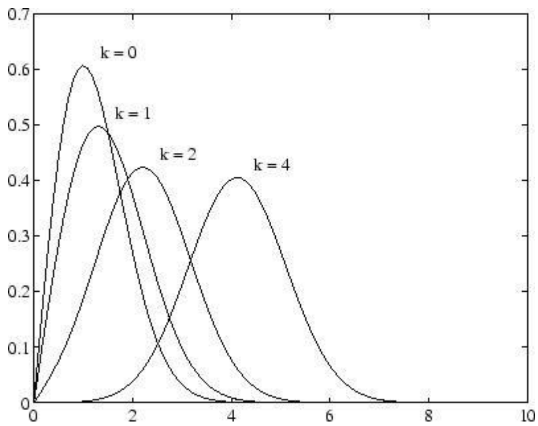


Figure 5.8: Ricean probability density function.

$$f_R(r) = \frac{r}{\sigma^2} e^{-\frac{(r^2+A^2)}{2\sigma^2}} I_0\left(\frac{Ar}{\sigma^2}\right) \tag{5.58}$$

for all $A \geq 0$ and $r \geq 0$. Here A is the peak amplitude of the dominant signal and $I_0(.)$ is the modified Bessel function of the first kind and zeroth order.

A factor K is defined as

$$K_{dB} = 10 \log \frac{A^2}{2\sigma^2} \tag{5.59}$$

As A $\rightarrow$ 0 then $K_{dB} \rightarrow \infty$.

Generalized Model: Nakagami Distribution

A generalization of the Rayleigh and Rician fading is the Nakagami distribution.

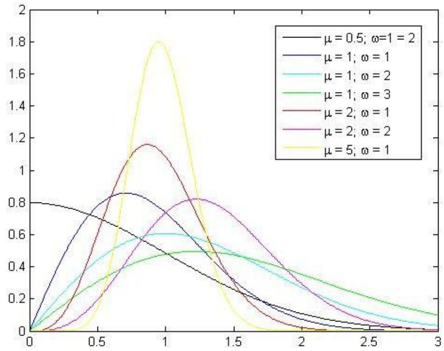


Figure 5.9: Nakagami probability density function.

Its pdf is given as,

$$f_R(r) = \frac{2r^{m-1}}{\Gamma(m)} \left(\frac{m^m}{\Omega^m}\right) e^{-\frac{mr^2}{\Omega}} \tag{5.60}$$

where,

$\Gamma(m)$ is the gamma function

Ω is the average signal power and

m is the fading factor.It is always greater than or equal to 0.5.

When m=1, Nakagami model is the Rayleigh model.

When

$$m = \frac{(M+1)^2}{2M+1}$$

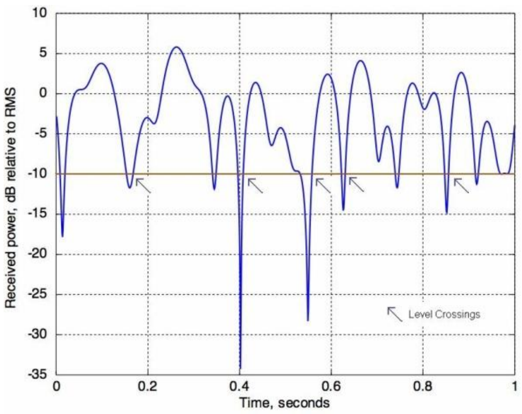


Figure 5.10: Schematic representation of level crossing with a Rayleigh fading envelope at 10 Hz Doppler spread.

where

$$M = \frac{A}{2\sigma}$$

Nakagami fading is the Rician fading.
As m $\rightarrow \infty$ Nakagami fading is the impulse channel and no fading occurs.

Second Order Statistics

To design better error control codes, we have two important second order parameters of fading model, namely the **level crossing rate** (LCR) and **average fade duration** (AFD). These parameters can be utilized to assess the speed of the user by measuring them through the reverse channel. The LCR is the expected rate at which the Rayleigh fading envelope normalized to the local rms amplitude crosses a specific level 'R' in a positive going direction.

$$N_R \int_0^{\infty} \quad = \sqrt{\quad} \; \rho e^{-\rho^2} \qquad (5.61)$$

D

where $\dot{r}$ is the time derivative of r(t), f_D is the maximum Doppler shift and ρ is the value of the specified level R, normalized to the local rms amplitude of the fading envelope.

The other important parameter, AFD, is the average period time for which the receiver power is below a specified level R.

$$\bar{\tau} = \frac{1}{N_r} P_r(r \leq R) \tag{5.62}$$

As

$$P_r(r \leq R) = \int_0^R p(r)dr = 1 - e^{-\rho^2}, \tag{5.63}$$

therefore,

$$\bar{\tau} = \frac{1 - e^{-\rho^2}}{\sqrt{2\pi} f_D \rho e^{-\rho^2}} \tag{5.64}$$

$$= \frac{e^{\rho^2} - 1}{\sqrt{2\pi} f_D \rho}. \tag{5.65}$$

Apart from LCR, another parameter is fading rate, which is defined as the number of times the signal envelope crosses the middle value (r_m) in a positive going direction per unit time. The average rate is expressed as

$$N(r_m) = \frac{2v}{\lambda}. \tag{5.66}$$

Another statistical parameter, sometimes used in the mobile communication, is called as depth of fading. It is defined as the ratio between the minimum value and the mean square value of the faded signal. Usually, an average value of 10% as depth of fading gives a marginal fading scenario.

Simulation of Rayleigh Fading Models

Clarke's Model: without Doppler Effect

In it, two independent Gaussian low pass noise sources are used to produce in-phase and quadrature fading branches. This is the basic model and is useful for slow fading channel. Also the Doppler effect is not accounted for.

Clarke and Gans' Model: with Doppler Effect

In this model, the output of the Clarke's model is passed through Doppler filter in the RF or through two initial baseband Doppler filters for baseband processing as shown in Figure 5.11. Here, the obtained Rayleigh output is flat faded signal but not frequency selective.

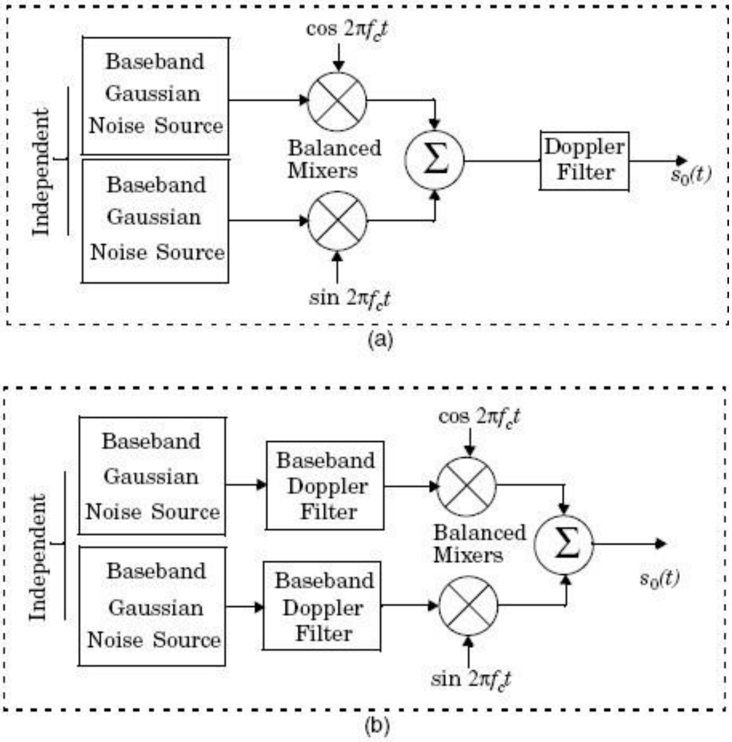


Figure 5.11: Clarke and Gan's model for Rayleigh fading generation using quadrature amplitude modulation with (a) RF Doppler filter, and, (b) baseband Doppler filter.

Rayleigh Simulator with Wide Range of Channel Conditions

To get a frequency selective output we have the following simulator through which both the frequency selective and flat faded Rayleigh signal may be obtained. This is achieved through varying the parameters a_i and τ_i, as given in Figure 5.12.

Two-Ray Rayleigh Faded Model

The above model is, however, very complex and difficult to implement. So, we have the two ray Rayleigh fading model which can be easily implemented in software as shown in Figure 5.13.

$$h_b(t) = \alpha_1 e^{j\varphi 1}\,\delta(t) + \alpha_2 e^{j\varphi 2}\,\delta(t-\tau) \tag{5.67}$$

where α_1 and α_2 are independent Rayleigh distributed and φ_1 and φ_2 are independent and uniformaly distributed over 0 to 2π. By varying τ it is possible to create a wide range of frequency selective fading effects.

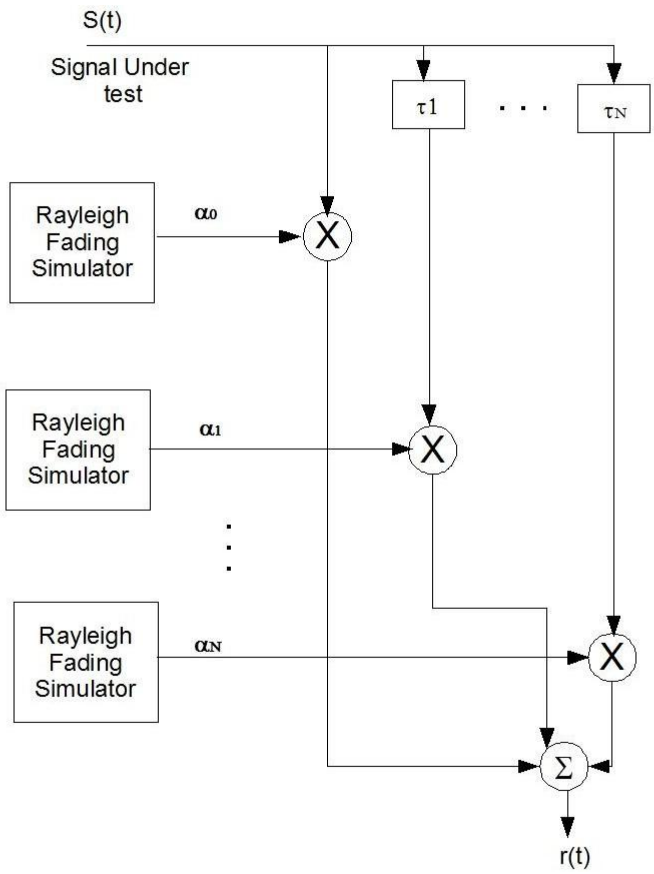


Figure 5.12: Rayleigh fading model to get both the flat and frequency selective channel conditions.

Saleh and Valenzuela Indoor Statistical Model

This method involved averaging the square law detected pulse response while sweeping the frequency of the transmitted pulse. The model assumes that the multipath components arrive in clusters. The amplitudes of the received components are independent Rayleigh random variables with variances that decay exponentially with cluster delay as well as excess delay within a cluster. The clusters and multipath components within a cluster form Poisson arrival processes with different rates.

SIRCIM/SMRCIM Indoor/Outdoor Statistical Models

SIRCIM (Simulation of Indoor Radio Channel Impulse-response Model) generates realistic samples of small-scale indoor channel impulse response measurements. Sub-

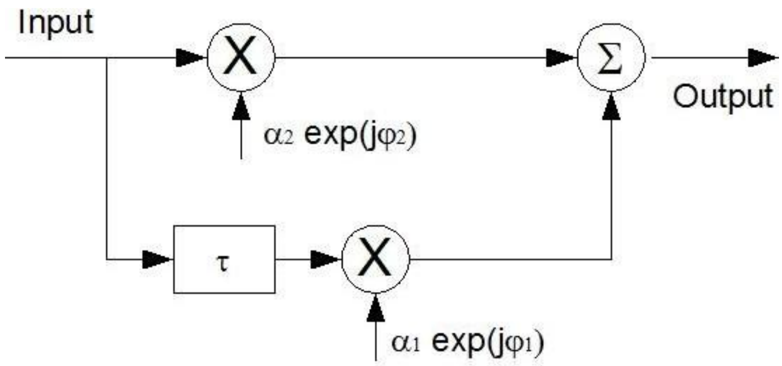


Figure 5.13: Two-ray Rayleigh fading model.

sequent work by Huang produced SMRCIM (Simulation of Mobile Radio Channel Impulse-response Model), a similar program that generates small-scale urban cellular and micro-cellular channel impulse responses.

Conclusion

In this chapter, the main channel impairment, i.e., fading, has been introduced which becomes so severe sometimes that even the large scale path loss becomes insignificant in comparison to it. Some statistical propagation models have been presented based on the fading characteristics. Mainly the frequency selective fading, fast fading and deep fading can be considered the major obstruction from the channel severity view point. Certain efficient signal processing techniques to mitigate these effects will be discussed in Chapter 7.

UNIT 4

TransmitterandReceiver Techniques

Introduction

Electrical communication transmitter and receiver techniques strive toward obtaining reliable communication at a low cost, with maximum utilization of the channel resources. The information transmitted by the source is received by the destination via a physical medium called a channel. This physical medium, which may be wired or wireless, introduces distortion, noise and interference in the transmitted information bearing signal. To counteract these effects is one of the requirements while designing a transmitter and receiver end technique. The other requirements are power and bandwidth efficiency at a low implementation complexity.

Modulation

Modulation is a process of encoding information from a message source in a manner suitable for transmission. It involves translating a baseband message signal to a passband signal. The baseband signal is called the modulating signal and the passband signal is called the modulated signal. Modulation can be done by varying certain characteristics of carrier waves according to the message signal. Demodulation is the reciprocal process of modulation which involves extraction of original baseband signal from the modulated passband signal.

Choice of Modulation Scheme

Several factors influence the choice of a digital modulation scheme. A desirable modulation scheme provides low bit error rates at low received signal to noise ratios, performs well in multipath and fading conditions, occupies a minimum of bandwidth, and is easy and cost-effective to implement. The performance of a modulation scheme is often measured in terms of its power efficiency and bandwidth efficiency. Power efficiency describes the ability of a modulation technique to preserve the fidelity of the digital message at low power levels. In a digital communication system, in order to increase noise immunity, it is necessary to increase the signal power. Bandwidth efficiency describes the ability of a modulation scheme to accommodate data within a limited bandwidth.

The system capacity of a digital mobile communication system is directly related to the bandwidth efficiency of the modulation scheme, since a modulation with a greater value of $\eta_b (= \frac{R}{B})$ will transmit more data in a given spectrum allocation.

There is a fundamental upper bound on achievable bandwidth efficiency. Shannon's channel coding theorem states that for an arbitrarily small probability of error, the maximum possible bandwidth efficiency is limited by the noise in the channel, and is given by the channel capacity formula

$$\eta B max = \frac{C}{B} = log_2(1 + \frac{S}{N}) \qquad (6.1)$$

Advantages of Modulation

1. Facilitates multiple access: By translating the baseband spectrum of signals from various users to different frequency bands, multiple users can be accommodated within a band of the electromagnetic spectrum.

2. Increases the range of communication: Low frequency baseband signals suffer from attenuation and hence cannot be transmitted over long distances. So translation to a higher frequency band results in long distance transmission.

3. Reduction in antenna size: The antenna height and aperture is inversely proportional to the radiated signal frequency and hence high frequency signal radiation result in smaller antenna size.

Linear and Non-linear Modulation Techniques

The mathematical relation between the message signal (applied at the modulator input) and the modulated signal (obtained at the modulator output) decides whether a modulation technique can be classified as linear or non-linear. If this input-output relation satisfies the principle of homogeneity and superposition then the modulation technique is said to be linear. The principle of homogeneity states that if the input signal to a system (in our case the system is a modulator) is scaled by a factor then the output must be scaled by the same factor. The principle of superposition states that the output of a linear system due to many simultaneously applied input signals is equal to the summation of outputs obtained when each input is applied one at a time.

For example an amplitude modulated wave consists of the addition two terms: the message signal multiplied with the carrier and the carrier itself. If $m(t)$ is the message signal and $s_{AM}(t)$ is the modulated signal given by:

$$s_{AM}(t) = A_c[1 + km(t)]\cos(2\pi f_c t) \tag{6.2}$$

Then,

1. From the principle of homogeneity: Let us scale the input by a factor a. So $m(t) = am_1(t)$ and the corresponding output becomes :

$$\begin{aligned} s_{AM1}(t) &= A_c[1 + am_1(t)]\cos(2\pi f_c t) \\ f&= as_{AM1}(t) \end{aligned} \tag{6.3}$$

2. From the principle of superposition: Let $m(t) = m_1(t) + m_2(t)$ be applied simultaneously at the input of the modulator. The resulting output is:

$$\begin{aligned} s_{AM}(t) &= A_c[1 + m_1(t) + m_2(t)]\cos(2\pi f_c t) \\ & s_{AM1}(t) + s_{AM2}(t) \\ &= A_c[2 + m_1(t) + m_2(t)]\cos(2\pi f_c t) \end{aligned} \tag{6.4}$$

   Here, $s_{AM1}(t)$ and $s_{AM2}(t)$ are the outputs obtained when $m_1(t)$ and $m_2(t)$ are applied one at a time.

Hence AM is a nonlinear technique but DSBSC modulation is a linear technique since it satisfies both the above mentioned principles.

Amplitude and Angle Modulation

Depending on the parameter of the carrier (amplitude or angle) that is changed in accordance with the message signal, a modulation scheme can be classified as an amplitude or angle modulation. Amplitude modulation involves variation of amplitude of the carrier wave with changes in the message signal. Angle modulation varies a sinusoidal carrier signal in such a way that the angle of the carrier is varied according to the amplitude of the modulating baseband signal.

Analog and Digital Modulation Techniques

The nature of the information generating source classifies a modulation technique as an analog or digital modulation technique. When analog messages generated from a source passe through a modulator, the resulting amplitude or angle modulation technique is called analog modulation. When digital messages undergo modulation the resulting modulation technique is called digital modulation.

Signal Space Representation of Digitally Modulated Signals

Any arbitrary signal can be expressed as the linear combination of a set of orthogonal signals or equivalently as a point in an M dimensional signal space, where M denotes the cardinality of the set of orthogonal signals. These orthogonal signals are normalized with respect to their energy content to yield an orthonormal signal set having unit energy. These orthonormal signals are independent of each other and form a basis set of the signal space.

Generally a digitally modulated signal *s(t)*, having a symbol duration *T*, is expressed as a linear combination of two orthonormal signals $\varphi_1(t)$ and $\varphi_2(t)$, constituting the two orthogonal axis in this two dimensional signal space and is expressed mathematically as,

$$s(t) = s_1\varphi_1(t) + s_2\varphi_2(t) \tag{6.5}$$

where $\varphi_1(t)$ and $\varphi_2(t)$ are given by,

$$\varphi_1(t) = \sqrt{\frac{2}{T}}\, cos(2\pi f_c t) \tag{6.6}$$

$$\varphi_2(t) = \sqrt{\frac{2}{T}}\, cos(2\pi f t) \qquad (6.7)$$

The coefficients s_1 and s_2 form the coordinates of the signal *s(t)* in the two dimensional signal space.

Complex Representation of Linear Modulated Sig- nals and Band Pass Systems

A band-pass signal s(t) can be resolved in terms of two sinusoids in phase quadrature as follows:

$$s(t) = s_I(t)cos(2\pi f_c t) - s_Q(t)sin(2\pi f_c t) \qquad (6.8)$$

Hence $s_I(t)$ and $s_Q(t)$ are known as the in-phase and quadrature-phase components respectively. When $s_I(t)$ and $s_Q(t)$ are incorporated in the formation of the following complex signal,

$$\tilde{s}(t) = s_I(t) + s_Q(t) \qquad (6.9)$$

then *s(t)* can be expressed in a more compact form as:

$$s(t) = Re\{\tilde{s}(t)e^{(j2\pi fct)}\}$$

(6.10

) where $\tilde{s}(t)$ is called the complex envelope of $s(t)$.

Analogously, band-pass systems characterized by an impulse response $h(t)$ can be expressed in terms of its in-phase and quadrature-phase components as:

$$h(t) = h_I(t)cos(2\pi f_c t) - h_Q(t)sin(2\pi f_c t)$$

(6.11

) The complex baseband model for the impulse response therefore becomes,

$$\tilde{h}(t) = h_I(t) + h_Q(t) \qquad (6.12)$$

$h(t)$ can therefore be expressed in terms of its complex envelope as

$$h(t) = Re\{\tilde{h}(t)e^{j2\pi fct}\}.$$

(6.13

) When *s(t)* passes through *h(t)*, then in the complex baseband domain, the output $\tilde{r}(t)$ of the bandpass system is given by the following convolution

$$\tilde{r}(t) = \frac{1}{2}\tilde{s}(t) \otimes \tilde{h}(t) \qquad (6.14)$$

Linear Modulation Techniques

6.5.1 Amplitude Modulation (DSBSC)

Generally, in amplitude modulation, the amplitude of a high frequency carrier signal, $\cos(2\pi f_c t)$, is varied in accordance to the instantaneous amplitude of the modulating message signal $m(t)$. The resulting modulated carrier or AM signal can be represented as:

$$s_{AM}(t) = A_c[1 + km(t)]\cos(2\pi f_c t). \qquad (6.15)$$

The modulation index k of an AM signal is defined as the ratio of the peak message signal amplitude to the peak carrier amplitude. For a sinusoidal modulating signal $m(t) = \frac{A_m}{A_c}\cos(2\pi f_m t)$, the modulation index is given by

$$k = \frac{A_m}{A_c}. \qquad (6.16)$$

This is a nonlinear technique and can be made linear by multiplying the carrier with the message signal.The resulting modulation scheme is known as DSBSC modulation. In DSBSC the amplitude of the transmitted signal, s(t), varies linearly with the modulating digital signal, m(t). Linear modulation techniques are bandwidth efficient and hence are very attractive for use in wireless communication systems where there is an increasing demand to accommodate more and more users within a limited spectrum. The transmitted signal DSBSC signal s(t) can be expressed as:

$$s(t) = Am(t)exp(j2\pi f_c t). \qquad (6.17)$$

If $m(t)$ is scaled by a factor of a, then s(t), the output of the modulator, is also scaled by the same factor as seen from the above equation. Hence the principle of homogeneity is satisfied. Moreover,

$$\begin{aligned} s_{12}(t) &= A[m_1(t) + m_2(t)]cos(2\pi f_c t) \qquad (6.18)\\ &= Am_1(t)cos(2\pi f_c t) + Am_2(t)cos(2\pi f_c t)\\ &= s_1(t) + s_2(t) \end{aligned}$$

where A is the carrier amplitude and f_c is the carrier frequency. Hence the principle of superposition is also satisfied. Thus DSBSC is a linear modulation technique.

AM demodulation techniques may be broadly divided into two categories: coherent and non-coherent demodulation. Coherent demodulation requires knowledge

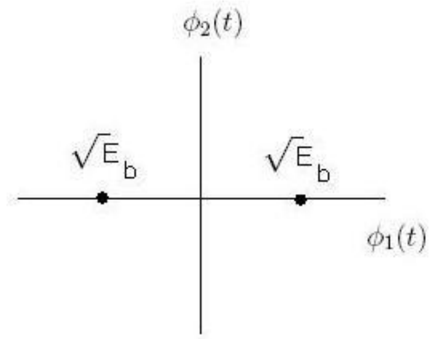


Figure 6.1: BPSK signal constellation.

of the transmitted carrier frequency and phase at the receiver, whereas non-coherent detection requires no phase information.

BPSK

In binary phase shift keying (BPSK), the phase of a constant amplitude carrier signal is switched between two values according to the two possible signals m1 and m2 corresponding to binary 1 and 0, respectively. Normally, the two phases are separated by 180^o. If the sinusoidal carrier has an amplitude A, and energy per bit $E_o = \frac{1}{2}A_c^2 T_b$ then the transmitted $BPSK$ signal is

$$s_{BPSK}(t) = m(t)\sqrt{\frac{2E_b}{T_b}}\cos(2\pi f_c t + \theta_c). \tag{6.19}$$

A typical BPSK signal constellation diagram is shown in Figure 6.1.

The probability of bit error for many modulation schemes in an $AWttN$ channel is found using the Q-function of the distance between the signal points. In case of $BPSK$,

$$P_{eBPSK} = Q(\sqrt{\frac{2E_b}{N_0}}). \tag{6.20}$$

QPSK

The Quadrature Phase Shift Keying (QPSK) is a 4-ary PSK signal. The phase of the carrier in the QPSK takes 1 of 4 equally spaced shifts. Although QPSK can be viewed as a quaternary modulation, it is easier to see it as two independently modulated quadrature carriers. With this interpretation, the even (or odd) bits are

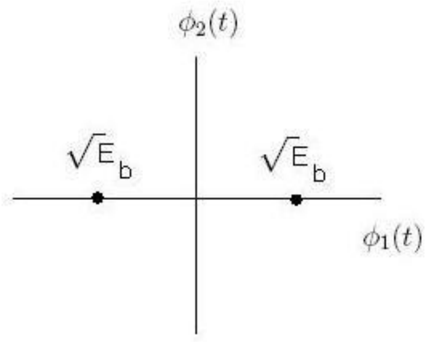


Figure 6.2: QPSK signal constellation.

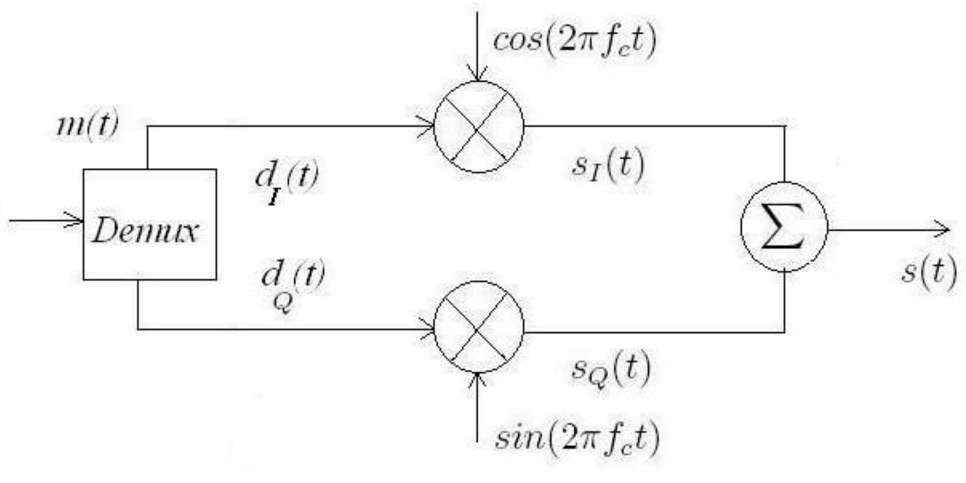


Figure 6.3: QPSK transmitter.

used to modulate the in-phase component of the carrier, while the odd (or even) bits are used to modulate the quadrature-phase component of the carrier.

The QPSK transmitted signal is defined by:

$$s_i(t) = A\cos(\omega t + (i-1)\pi/2), i = (1, 2, 3, 4) \tag{6.21}$$

and the constellation disgram is shown in Figure 6.2.

Offset-QPSK

As in QPSK, as shown in Figure 6.3, the NRZ data is split into two streams of odd and even bits. Each bit in these streams has a duration of twice the bit duration,

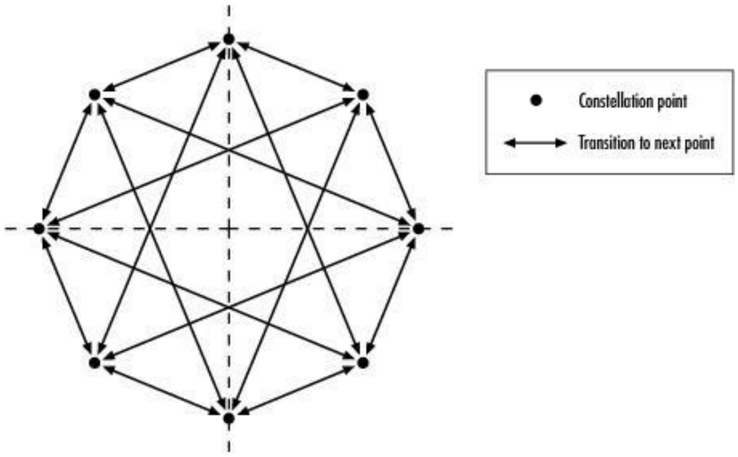


Figure 6.4: DQPSK constellation diagram.

T_b, of the original data stream. These odd ($d_1(t)$) and even bit streams ($d_2(t)$) are then used to modulate two sinusoidals in phase quadrature,and hence these data streams are also called the in-phase and and quadrature phase components. After modulation they are added up and transmitted. The constellation diagram of Offset-QPSK is the same as QPSK. Offset-QPSK differs from QPSK in that the $d_1(t)$ and $d_2(t)$ are aligned such that the timing of the pulse streams are offset with respect to each other by T_b seconds. From the constellation diagram it is observed that a signal point in any quadrant can take a value in the diagonally opposite quadrant only when two pulses change their polarities together leading to an abrupt 180 degree phase shift between adjacent symbol slots. This is prevented in O-QPSK and the allowed phase transitions are ± 90 degree.

Abrupt phase changes leading to sudden changes in the signal amplitude in the time domain corresponds to significant out of band high frequency components in the frequency domain. Thus to reduce these sidelobes spectral shaping is done at baseband. When high efficiency power amplifiers, whose non-linearity increases as the efficiency goes high, are used then due to distortion, harmonics are generated and this leads to what is known as spectral regrowth. Since sudden 180 degree phase changes cannot occur in OQPSK, this problem is reduced to a certain extent.

$\pi/4$ DQPSK

The data for $\pi/4$ DQPSK like QPSK can be thought to be carried in the phase of a single modulated carrier or on the amplitudes of a pair of quadrature carriers. The modulated signal during the time slot of $kT < t < (k + 1)T$ given by:

$$s(t) = cos(2\pi f_c t + \psi_{k+1}) \tag{6.22}$$

Here, $\psi_{k+1} = \psi_k + \Delta\psi_k$ and $\Delta\psi_k$ can take values $\pi/4$ for 00, $3\pi/4$ for 01, $-3\pi/4$ for 11 and $-\pi/4$ for 10. This corresponds to eight points in the signal constellation but at any instant of time only one of the four points are possible: the four points on axis or the four points off axis. The constellation diagram along with possible transitions are shown in Figure 6.4.

Line Coding

Specific waveforms are required to represent a zero and a one uniquely so that a sequence of bits is coded into electrical pulses. This is known as line coding. There are various ways to accomplish this and the different forms are summarized below.

1. Non-return to zero level (NRZ-L): 1 forces a a high while 0 forces a low.
2. Non-return to zero mark (NRZ-M): 1 forces negative and positive transitions while 0 causes no transitions.
3. Non-return to zero space (NRZ-S): 0 forces negative and positive transitions while 1 causes no transitions.
4. Return to zero (RZ): 1 goes high for half a period while 0 remains at zero state.
5. Biphase-L: Manchester 1 forces positive transition while 0 forces negative transition. In case of consecutive bits of same type a transition occurs in the beginning of the bit period.
6. Biphase-M: There is always a transition in the beginning of a bit interval. 1 forces a transition in the middle of the bit while 0 does nothing.

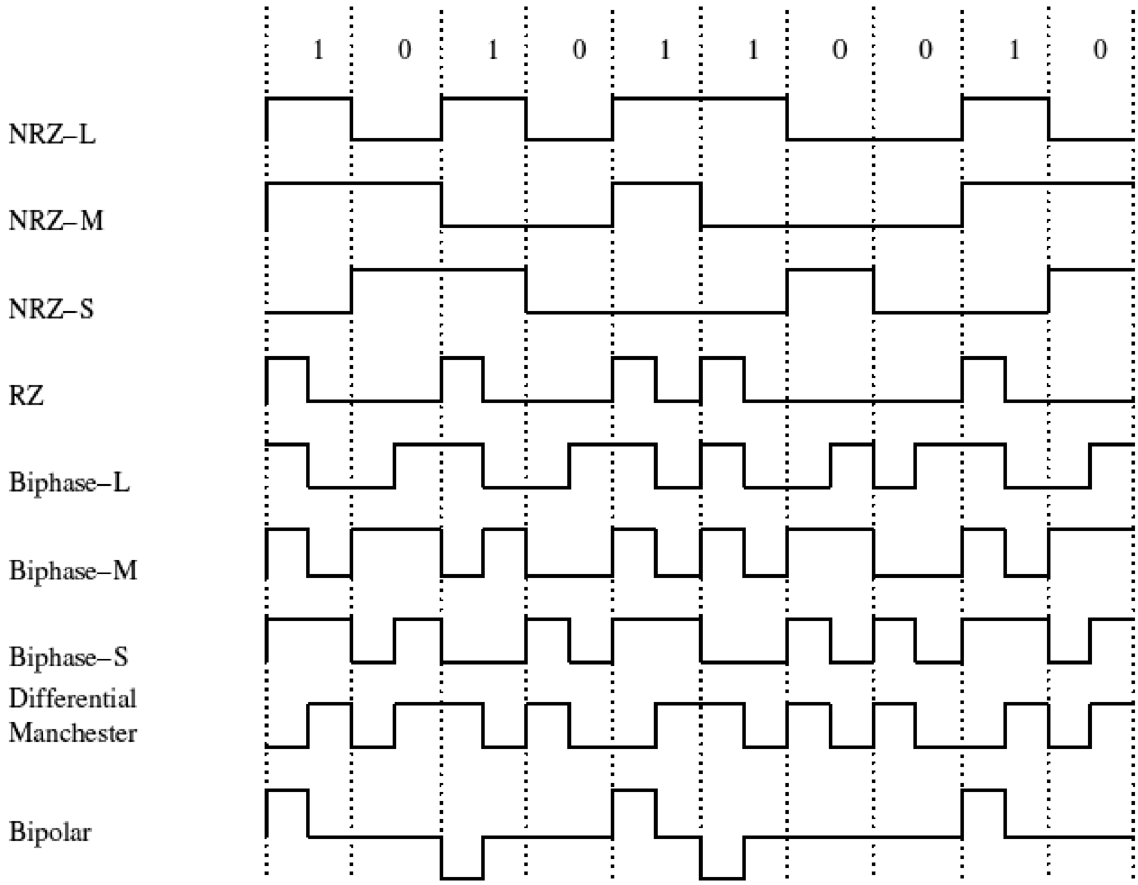


Figure 6.5: Scematic of the line coding techniques.

7. Biphase-S: There is always a transition in the beginning of a bit interval. 0 forces a transition in the middle of the bit while 1 does nothing.

8. Differential Manchester: There is always a transition in the middle of a bit interval. 0 forces a transition in the beginning of the bit while 1 does nothing.

9. Bipolar/Alternate mark inversion (AMI): 1 forces a positive or negative pulse for half a bit period and they alternate while 0 does nothing.

All these schemes are shown in Figure 6.5.

Pulse Shaping

Let us think about a rectangular pulse as defined in BPSK. Such a pulse is not desirable for two fundamental reasons:

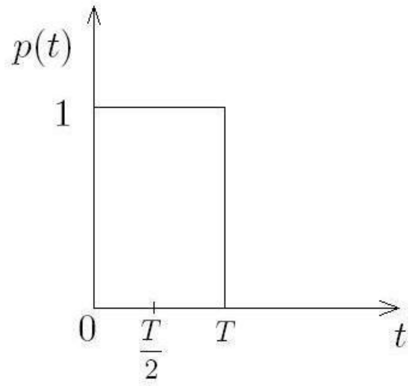


Figure 6.6: Rectangular Pulse

(a) the spectrum of a rectangular pulse is infinite in extent. Correspondingly, its frequency content is also infinite. But a wireless channel is bandlimited, means it would introduce signal distortion to such type of pulses,

(b) a wireless channel has memory due to multipath and therefore it introduces ISI.

In order to mitigate the above two effects, an efficient pulse shaping funtion or a premodulation filter is used at the Tx side so that QoS can be maintained to the mobile users during communication. This type of technique is called pulse shaping technique. Below, we start with the fundamental works of Nyquist on pulse shaping and subsequently, we would look into another type of pulse shaping technique.

Nyquist pulse shaping

There are a number of well known pulse shaping techniques which are used to simultaneously to reduce the inter-symbol effects and the spectral width of a modulated digital signal. We discuss here about the fundamental works of Nyquist. As pulse shaping is difficult to directly manipulate the transmitter spectrum at *RF* frequencies, spectral shaping is usually done through baseband or IF processing.

Let the overall frequency response of a communication system (the transmitter, channel and receiver) be denoted as $H_{eff}(f)$ and according to Nyquist it must be given by:

$$H_{eff}(f) = \frac{1}{f_s} rect\left(\frac{f}{f_s}\right) \quad (6.23)$$

Hence, the ideal pulse shape for zero ISI, given by $h_{eff}(t)$, such that,

$$H_{eff}(f) \leftrightarrow h_{eff}(t) \quad (6.24)$$

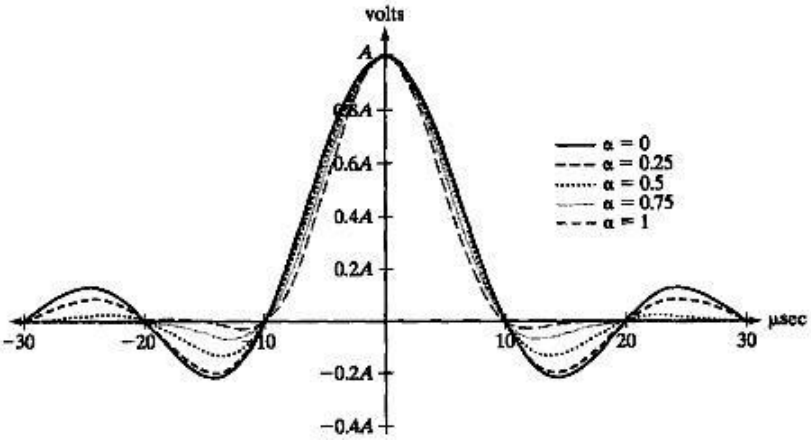


Figure 6.7: Raised Cosine Pulse.

is given by:

$$h_{eff}(t) = \frac{\sin(\frac{\pi t}{T_s})}{\frac{\pi t}{T_s}} \tag{6.25, 6.26}$$

Raised Cosine Roll-Off Filtering

If we take a rectangular filter with bandwidth $f_0 \geq \frac{1}{2T_s}$ and convolve it with any arbitrary even function Z(f) with zero magnitude outside the passband of the rectangular filter then a zero *ISI* effect would be achieved. Mathematically,

$$H_{eff}(f) = rect(\frac{f}{f_0}) * Z(f), \tag{6.27}$$

$$h_{eff}(t) = \frac{\sin(\frac{\pi t}{T_s})}{\frac{\pi t}{T_s}} z(t), \tag{6.28}$$

$$z(t) = \frac{\cos(\pi \rho t/T_s)}{1 - (\Delta \rho t/2T_s)^2}. \tag{6.29}$$

with ρ being the roll off factor $\in [0, 1]$. As ρ increases roll off in frequency domain increases but that in time domain decreases.

Realization of Pulse Shaping Filters

Since $h_{eff}(t)$ is non-causal, pulse shaping filters are usually truncated within $\pm 6T_s$ about $t = 0$ for each symbol. Digital communication systems thus often store several symbols at a time inside the modulator and then clock out a group of symbols by using a look up table that represents discrete time waveforms of stored symbols. This is the way to realize the pulse shaping filters using real time processors.

Non-Nyquist pulse shaping are also useful, which would be discussed later in this chapter while discussing GMSK.

Nonlinear Modulation Techniques

Many practical mobile radio communications use nonlinear modulation methods,where the amplitude of the carrier is constant,regardless of the variations in the modulating signal.The Constant envelope family of modulations has the following advantages :

1. Power efficient class C amplifiers without introducing degradation in the spectral occupancy of the transmitted signal.

2. Low out-of-band radiation of the order of -60 dB to -70dB can be achieved.

3. Limiter-discriminator detection can be used,which simplifies receiver design and provides high immunity against random FM noise and signal fluctuations due to Rayleigh fading.

However, even if constant envelope has many advantages it still uses more BW than linear modulation schemes.

Angle Modulation (FM and PM)

There are a number of ways in which the phase of a carrier signal may be varied in accordance with the baseband signal; the two most important classes of angle modulation being frequency modulation and phase modulation.

Frequency modulation (FM) involves changing of the frequency of the carrier signal according to message signal. As the information in frequency modulation is in the frequency of modulated signal, it is a nonlinear modulation technique. In this method, the amplitude of the carrier wave is kept constant (this is why FM is called constant envelope). FM is thus part of a more general class of modulation known as angle modulation.

Frequency modulated signals have better noise immunity and give better performance in fading scenario as compared to amplitude modulation.Unlike AM, in an

FM system, the modulation index, and hence bandwidth occupancy, can be varied to obtain greater signal to noise performance.This ability of an FM system to trade bandwidth for SNR is perhaps the most important reason for its superiority over AM. However, AM signals are able to occupy less bandwidth as compared to FM signals, since the transmission system is linear.

An FM signal is a constant envelope signal, due to the fact that the envelope of the carrier does not change with changes in the modulating signal. The constant envelope of the transmitted signal allows efficient Class C power amplifiers to be used for RF power amplification of FM. In AM, however, it is critical to maintain linearity between the applied message and the amplitude of the transmitted signal, thus linear Class A or AB amplifiers, which are not as power efficient, must be used.

FM systems require a wider frequency band in the transmitting media (generally several times as large as that needed for AM) in order to obtain the advantages of reduced noise and capture effect. FM transmitter and receiver equipment is also more complex than that used by amplitude modulation systems. Although frequency modulation systems are tolerant to certain types of signal and circuit nonlinearities, special attention must be given to phase characteristics. Both AM and FM may be demodulated using inexpensive noncoherent detectors. AM is easily demodulated using an envelope detector whereas FM is demodulated using a discriminator or slope detector. In FM the instantaneous frequency of the carrier signal is varied linearly with the baseband message signal m(t), as shown in following equation:

$$s_{FM}(t) = A_c \cos[2\pi f_c t + \theta(t)] = A_c \cos[2\pi f_c t + 2\pi k_f \int m(\eta)d\eta] \qquad (6.30)$$

where A_c, is the amplitude of the carrier, f_c is the carrier frequency, and k_f is the frequency deviation constant (measured in units of Hz/V).

Phase modulation (PM) is a form of angle modulation in which the angle $\theta(t)$ of the carrier signal is varied linearly with the baseband message signal $m(t)$, as shown in equation below.

$$s_{PM}(t) = A_c \cos(2\pi f_c t + k_\theta m(t)) \qquad (6.31)$$

The frequency modulation index β_f, defines the relationship between the message amplitude and the bandwidth of the transmitted signal, and is given by

$$\beta_f = \qquad (6.32)$$

where A_m is the peak value of the modulating signal, Δf is the peak frequency deviation of the transmitter and W is the maximum bandwidth of the modulating signal.

The phase modulation index β_p is given by

$$\beta_p = k_\theta A_m = \Delta\theta \tag{6.33}$$

where, $\Delta\theta$ is the peak phase deviation of the transmitter.

BFSK

In Binary Frequency Shift keying (BFSK),the frequency of constant amplitude car- rier signal is switched between two values according to the two possible message states (called high and low tones) corresponding to a binary 1 or 0. Depending on how the frequency variations are imparted into the transmitted waveform,the FSK signal will have either a discontinuous phase or continuous phase between bits. In general, an FSK signal may be represented as

$$S(t) = \sqrt{(2E_b/T)}\cos(2\pi f_i t). \tag{6.34}$$

where T is the symbol duration and E_b is the energy per bit.

$$S_i = \sqrt{(E_b)}\varphi(t). \tag{6.35}$$

$$\varphi(t) = \sqrt{(2/T)}\cos(2\pi f_i t). \tag{6.36}$$

There are two FSK signals to represent 1 and 0, i.e.,

$$S_1(t) = \sqrt{(2E_b/T)}\cos(2\pi f_1 t + \theta(0)) \quad \rightarrow 1 \tag{6.37}$$

$$S_2(t) = \sqrt{(2E_b/T)}\cos(2\pi f_2 t + \theta(0)) \quad \rightarrow 0 \tag{6.38}$$

where $\theta(0)$ sums the phase up to $t = 0$. Let us now consider a continuous phase FSK as

$$S(t) = \sqrt{(2E_b/T)}\cos(2\pi f_c t + \theta(t)). \tag{6.39}$$

Expressing $\theta(t)$ in terms of $\theta(0)$ with a new unknown factor h, we get

$$\theta(t) = \theta(0) \pm \pi h t/T \qquad 0 \le t \le T \tag{6.40}$$

116

and therefore

$$S(t) = \sqrt{\frac{2E_b}{T}} \cos(2\pi f_c t \pm \pi h t/T + \theta(0)) = \sqrt{\frac{2E_b}{T}} \cos(2\pi(f_c \pm h/2T)t + \theta(0)). \quad (6.41)$$

It shows that we can choose two frequencies f_1 and f_2 such that

$$f_1 = f_c + h/2T \quad (6.42)$$

$$f_2 = f_c - h/2T \quad (6.43)$$

for which the expression of FSK conforms to that of CPFSK. On the other hand, f_c and h can be expressed in terms of f_1 and f_2 as

$$f_c = [f_1 + f_2]/2 \quad (6.44)$$

$$h = \frac{(f_1 - f_2)}{1/T}. \quad (6.45)$$

Therefore, the unknown factor h can be treated as the difference between f_1 and f_2, normalized with respect to bit rate $1/T$. It is called the deviation ratio. We know that $\theta(t) - \theta(0) = \pm\pi ht/T$, $0 \leq t \leq T$. If we substitute $t = T$, we have

$$\theta(T) - \theta(0) = \pm\pi h \qquad \text{where} \quad (6.46)$$

$$= \pi h \qquad \rightarrow 1 \quad (6.47)$$

$$= -\pi h \qquad \rightarrow 0 \quad (6.48)$$

This type of CPFSK is advantageous since by looking only at the phase, the transmitted bit can be predicted. In Figure 6.8, we show a phase tree of such a CPFSK signal with the transmitted bit stream of 1101000.

A special case of CPFSK is achieved with $h = 0.5$, and the resulting scheme is called Minimum Shift Keying (MSK) which is used in mobile communications. In this case, the phase differences reduce to only $\pm\pi/2$ and the phase tree is called the

phase trellis. An MSK signal can also be thought as a special case of OQPSK where the baseband rectangular pulses are replaced by half sinusoidal pulses. Spectral characteristics of an MSK signal is shown in Figure 6.9 from which it is clear that ACI is present in the spectrum. Hence a pulse shaping technique is required. In order to have a compact signal spectrum as well as maintaining the constant envelope property, we use a pulse shaping filter with

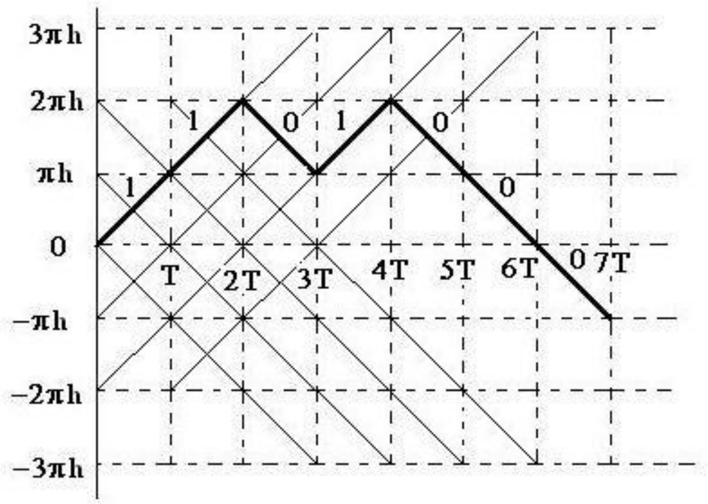


Figure 6.8: Phase tree of 1101000 CPFSK sequence.

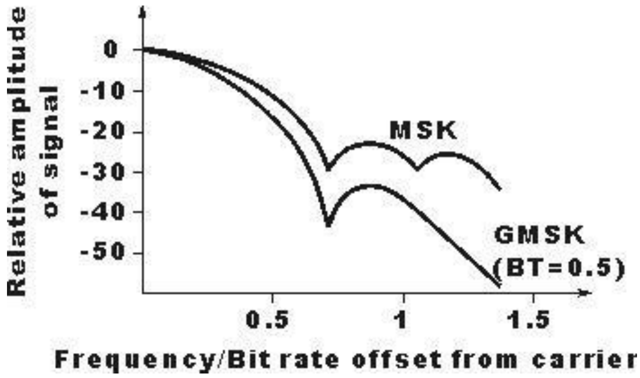


Figure 6.9: Spectrum of MSK

1. a narrow BW frequency and sharp cutoff characteristics (in order to suppress the high frequency component of the signal);
2. an impulse response with relatively low overshoot (to limit FM instant frequency deviation;
3. a phase trellis with $\pm\pi/2$ for odd T and 0 or π values for even T.

GMSK Scheme

GMSK is a simple modulation scheme that may be taken as a derivative of MSK. In GMSK, the sidelobe levels of the spectrum are further reduced by passing a non-

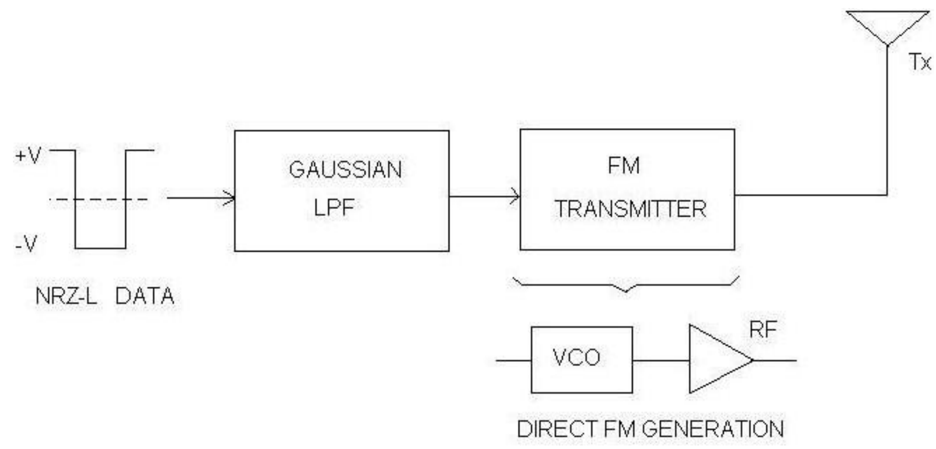


Figure 6.10: GMSK generation scheme.

return to zero (NRZ-L) data waveform through a premodulation Gaussian pulse shaping filter. Baseband Gaussian pulse shaping smoothes the trajectory of the MSK signals and hence stabilizes instantaneous frequency variations over time. This has the effect of considerably reducing the sidelobes in the transmitted spectrum. A GMSK generation scheme with NRZ-L data is shown in Figure 6.10 and a receiver of the same scheme with some MSI gates is shown in Figure 6.11.

GMSK Generator

The GMSK premodulation filter has characteristic equation given by

$$H(f) = \exp(-(\ln 2/2)(f/B)^2) \tag{6.49}$$

$$H(f) = \exp(-(\alpha f)^2)$$

where,

$$(\alpha)^2 = \ln 2/2(1/B)^2. \tag{6.50}$$

The premodulation Gaussian filtering introduces ISI in the transmitted signal, but it can be shown that the degradation is not that great if the 3dB bandwidth-bit duration product (BT) is greater than 0.5.

Spectrum of GMSK scheme is shown in Figure 6.12. From this figure, it is evident that when we are decreasing BT product, the out of band response decreases but

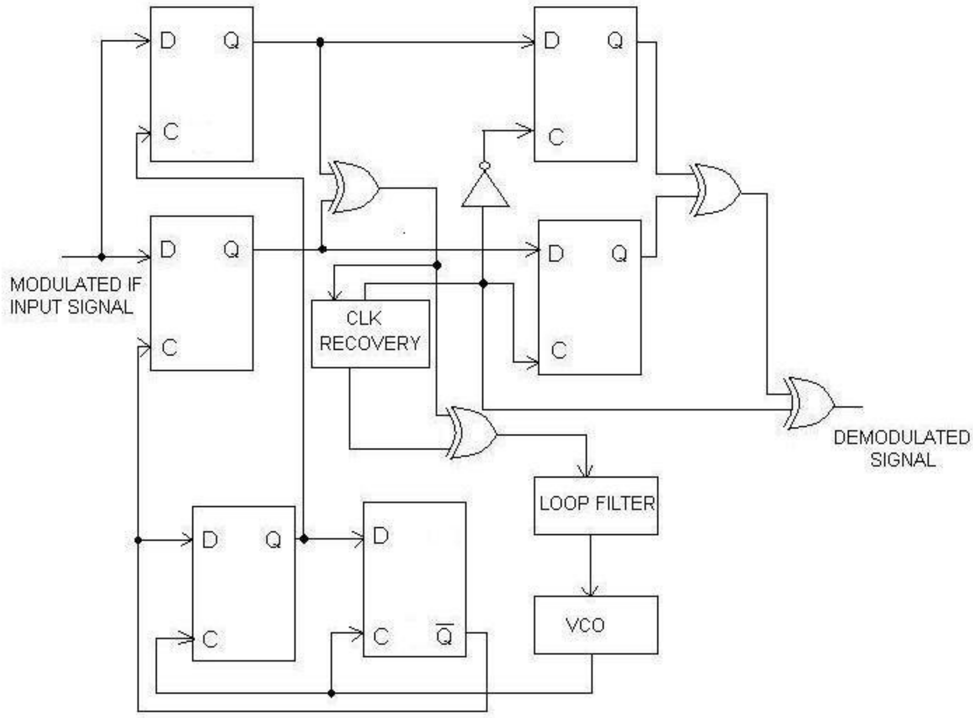


Figure 6.11: A simple GMSK receiver.

on the other hand irreducible error rate of the LPF for ISI increases. Therefore, a compromise between these two is required.

Problem: Find the 3dB BW for a Gaussian LPF used to produce 0.25 GMSK with a channel data rate Rb=270 kbps.What is the 90 percent power BW of the RF filter?

Solution: From the problem statement it is clear that

$$T = 1/R_b = 1/270 * (10^3) = 3.7\mu sec \tag{6.51}$$

Solving for B where $BT = 0.25$,

$$B = 0.25/T = 67.567kHz \tag{6.52}$$

Thus the 3 - dB bandwidth is 67.567 kHz. We use below table fig 6 to find out that 90 % power bandwidth is 0.57 R_b.

90 % RF $BW = 0.57R_b$ = 153.9 kHz.

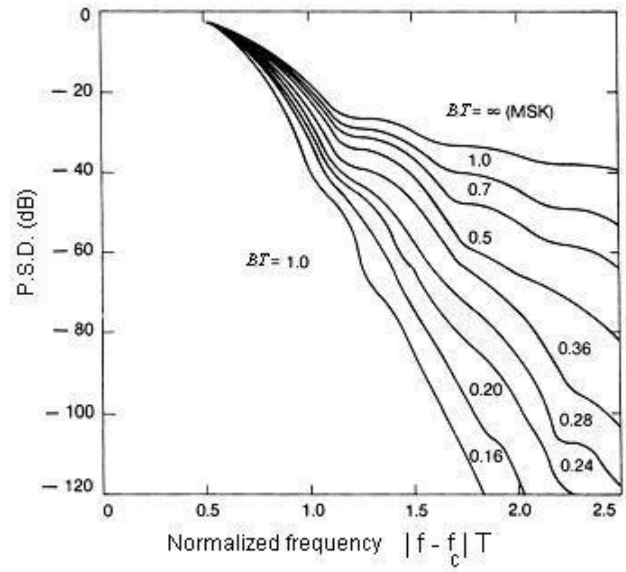


Figure 6.12: Spectrum of GMSK scheme.

Two Practical Issues of Concern

Inter Channel Interference

In FDMA, subscribers are allotted frequency slots called channels in a given band of the electromagnetic spectrum. The side lobes generated due to the transmission of a symbol in a particular channel overlaps with the channels placed adjacently. This is because of the fact that transmission of a time limited pulse leads to spectral spreading in the frequency domain. During simultaneous use of adjacent channels, when there is significant amount of power present in the side lobes, this kind of interference becomes so severe that the required symbol in a particular frequency slot is completely lost.

Moreover if two terminals transmit equal power then due to wave propagation through different distances to the receiver, the received signal levels in the two frequency slots will differ greatly. In such a case the side lobes of the stronger signal will severely degrade the transmitted signal in the next frequency slot having low power level. This is known as the near far problem.

Power Amplifier Nonlinearity

Power amplifiers may be designed as class A, class B, class AB, class C and class D. They form an essential section of mobile radio terminals. Due to power constraints on a transmitting terminal, an efficient power amplifier is required which can convert most of the input power to RF power. Class A amplifier is a linear amplifier but it has a power efficiency of only 25 %. As we go for subsequent amplifiers having greater power efficiency, the nonlinearity of the amplifier increases.

In general, an amplifier has linear input output characteristics over a range of input signal level, that is, it has a constant gain. However, beyond an input threshold level, the gain of the amplifier starts decreasing. Thus the amplitude of a signal applied at the input of an amplifier suffers from amplitude distortion and the resulting waveform obtained at the output of the amplifier is of the form of an amplitude modulated signal. Similarly, the phase characteristic of a practical amplifier is not constant over all input levels and results in phase distortion of the form of phase modulation.

The operating point of a practical amplifier is given in terms of either the input back-off or the output back-off.

$$Input\ \ back-off = 10\log_{10}\left(\frac{V_{in,rms}}{V_{out,rms}}\right) \tag{6.53}$$

$$Output\ \ back-off = 10\log_{10}\left(\frac{V_{out,rms}}{V_{out,rms}}\right) \tag{6.54}$$

Receiver performance in multipath channels

For a flat fading channel, the probability of error for coherent BPSK and coherent BFSK are respectively given as,

$$P_{e,BPSK} = \frac{1}{2}\left(1 - \sqrt{\frac{\gamma}{1+\gamma}}\right) \tag{6.55}$$

$$P_{e,BFSK} = \frac{1}{2}\left(1 - \sqrt{\frac{\gamma}{2+\gamma}}\right) \tag{6.56}$$

(6.57)

where γ is given by,

$$\gamma = \frac{E_b}{N_0}E(\alpha^2) \tag{6.58}$$

α^2 represents the instantaneous power values of the Rayleigh fading channel and E denotes the expectation operator.

Similarly, for differential BPSK and non coherent BFSK probability of error expressions are

$$P_{e,DPSK} = \frac{1}{2(1+\gamma)} \tag{6.59}$$

$$P_{e,NCFSK} = \frac{1}{(2+\gamma)}. \tag{6.60}$$

For large values of $SNR = \frac{Eb}{N0}$ the error probability given above have the simplified expression.

$$P_{e,BPSK} = \frac{1}{4\gamma} \tag{6.61}$$

$$P_{e,BFSK} = \frac{1}{2\gamma} \tag{6.62}$$

$$P_{e,DPSK} = \frac{1}{2\gamma} \tag{6.63}$$

$$P_{e,NCFSK} = \frac{1}{\gamma}. \tag{6.64}$$

From the above equations we observe that an inverse algebraic relation exists between the BER and SNR. This implies that if the required BER range is around 10^{-3} to 10^{-6}, then the SNR range must be around 30dB to 60dB.

Bit Error Rate and Symbol Error Rate

Bit error rate (P_{eb}) is the same as symbol error rate (P_{es}) when a symbol consists of a single bit as in BPSK modulation. For an MPSK scheme employing gray coded modulation, where N bits are mapped to a one of the M symbols, such that $2^N = M$, P_{eb} is given by

$$P_{eb} \approx \frac{P_{es}}{log_2 M} \tag{6.65}$$

And for M-ary orthogonal signalling P_{eb} is given by

$$P_{eb} = \frac{M/2}{M-1} P_{es}. \tag{6.66}$$

Example of a Multicarrier Modulation: OFDM

Multiplexing is an important signal processing operation in which a number of signals are combined and transmitted parallelly over a common channel. In order to avoid interference during parallel transmission, the signals can be separated in frequency and then the resulting technique is called Frequency Division Multiplexing (FDM). In FDM, the adjacent bands are non overlapping but if overlap is allowed by transmitting signals that are mutually orthogonal (that is, there is a precise mathematical relationship between the frequencies of the transmitted signals) such that one signal has zero effect on another, then the resulting transmission technique is known as Orthogonal Frequency Division Multiplexing (OFDM).

OFDM is a technique of transmitting high bit rate data into several parallel streams of low bit rate data. At any instant, the data transmitted simultaneously in each of these parallel data streams is frequency modulated by carriers (called subcarriers) which are orthogonal to each other. For high data rate communication the bandwidth (which is limited) requirement goes on increasing as the data rate increases or the symbol duration decreases. Thus in OFDM, instead of sending a particular number of symbols, say P, in T seconds serially, the P symbols can be sent in parallel with symbol duration now increased to T seconds instead of T/P seconds as was previously.

This offers many advantages in digital data transmission through a wireless time varying channel. The primary advantage of increasing the symbol duration is that the channel experiences flat fading instead of frequency selective fading since it is ensured that in the time domain the symbol duration is greater than the r.m.s. delay spread of the channel. Viewed in the frequency domain this implies that the bandwidth of the OFDM signal is less than coherent bandwidth of the channel.

Although the use of OFDM was initially limited to military applications due to cost and complexity considerations, with the recent advances in large-scale high-speed DSP, this is no longer a major problem. This technique is being used, in digital audio broadcasting (DAB), high definition digital television broadcasting (HDTV), digital video broadcasting terrestrial TV (DVB-T), WLAN systems based on IEEE 802.11(a) or HiperLan2, asymmetric digital subscriber lines (ADSL) and mobile communications. Very recently, the significance of the COFDM technique for UWA (underwater acoustic channel) has also been indicated. Moreover related or combined technology such as CDMA-OFDM, TDMA-OFDM, MIMO-OFDM, Vec- tor OFDM (V-OFDM), wide-band OFDM (W-OFDM), flash OFDM (F-OFDM), OFDMA, wavelet-OFDM have presented their great advantages in certain application areas.

Orthogonality of Signals

Orthogonal signals can be viewed in the same perspective as we view vectors which are perpendicular/orthogonal to each other. The inner product of two mutually orthogonal vectors is equal to zero. Similarly the inner product of two orthogonal signals is also equal to zero.

Let $\psi_k(t) = e^{j2\pi f_k t}$ and $\psi_n(t) = e^{j2\pi f_n t}$ be two complex exponential signals whose inner product, over the time duration of T_s, is given by:

$$N = \int_{iT_s}^{(i+1)T_s} \psi_k(t).\psi_n^*(t)dt \qquad (6.67)$$

When this integral is evaluated, it is found that if f_k and f_n are integer multiples of $1/T_s$ then N equals zero. This implies that for two harmonics of an exponential function having a fundamental frequency of $1/T_s$, the inner product becomes zero .But if $f_k = f_n$ then N equals T_s which is nothing but the energy of the complex exponential signal in the time duration of T_s.

Mathematical Description of OFDM

Let us now consider the simultaneous or parallel transmission of P number of complex symbols in the time slot of T_s second (OFDM symbol time duration) and a set of P orthogonal subcarriers, such that each subcarrier gets amplitude modulated by a particular symbol from this set of P symbols. Let each orthogonal carrier be of the form $exp\left(j2\pi n\frac{t}{T_s}\right)$ where n varies as 0, 1, 2..($P-1$). Here the variable 'n' denotes the n^{th} parallel path corresponding to the n^{th} subcarrier. Mathematically, we can obtain the transmitted signal in T_s seconds by summing up all the P number of amplitude modulated subcarriers, thereby yielding the following equation:

$$p(t) = \sum_{n=0}^{P-1} c_n g_n(t) exp\left(j2\pi n\frac{t}{T_s}\right) \quad for \quad 0 \le t \le T_s \qquad (6.68)$$

If $p(t)$ is sampled at $t = kT_s/P$, then the resulting waveform, is:

$$
\begin{aligned}
p(k) &= \sum_{n=0}^{P-1} c_n g_n(kT_s/P)\exp\left(j2\pi n \frac{kT_s/P}{T_s}\right) \\
&= \frac{1}{\sqrt{T_s}} \sum_{n=0}^{P-1} c_n \exp\left(j2\pi n \frac{k}{P}\right) \qquad for\ 0 \le k \le P-1 \qquad (6.69)
\end{aligned}
$$

This is nothing but the IDFT on the symbol block of P symbols. This can be realized using IFFT but the constraint is that P has to be a power of 2. So at the receiver, FFT can be done to get back the required block of symbols. This implementation is better than using multiple oscillators for subcarrier generation which is uneconomical and since digital technology has greatly advanced over the past few decades, IFFTs and FFTs can be implemented easily. The frequency spectrum, therefore consists of a set of P partially overlapping sinc pulses during any time slot of duration T_s. This is due to the fact that the Fourier Transform of a rectangular pulse is a sinc function. The receiver can be visualized as consisting of a bank of demodulators, translating each subcarrier down to DC, then integrating the resulting signal over a symbol period to recover the raw data.

But the OFDM symbol structure so generated at the transmitter end needs to be modified. Since inter symbol interference (ISI) is introduced by the transmission channel due to multipaths and also due to the fact that when the bandwidth of OFDM signal is truncated, its effect in the time domain is to cause symbol spreading such that a part of the symbol overlaps with the adjacent symbols. In order to cope with ISI as discussed previously the OFDM symbol duration can be increased. But this might not be feasible from the implementation point of view specifically in terms of FFT size and Doppler shifts.

A different approach is to keep a guard time interval between two OFDM symbols in which part of the symbol is copied from the end of the symbol to the front and is popularly known as the cyclic-prefix. If we denote the guard time interval as T_g and T_s be the useful symbol duration, then after this cyclical extension the total symbol duration becomes $T = T_g + T_s$. When the guard interval is longer than the length of the channel impulse response, or the multipath delay, then ISI can be eliminated. However the disadvantage is the reduction in data rate or throughput and greater power requirements at the transmitting end. The OFDM transmitter and receiver

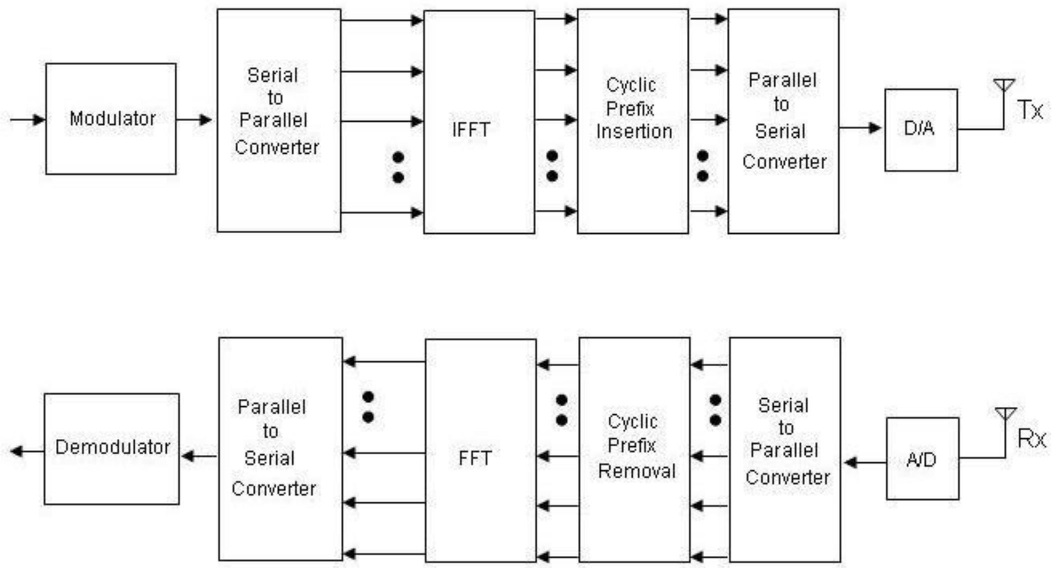


Figure 6.13: OFDM Transmitter and Receiver Block Diagram.

sections are as given in the following diagram.

Conclusion

In this chapter, a major chunk has been devoted to digital communication systems which obviously have certain distinction in comparison to their analog counterpart due to their signal-space representation. The important modulation techniques for wireless communication such as QPSK, MSK, GMSK were taken up at length. A relatively new modulation technology, OFDM, has also been discussed. Certain practical issues of concern are also discussed. It should be noted that albeit implementing these efficient modulation techniques, the channel still introduces fading in different ways. In order to prevent that, we need some additional signal processing techniques mainly at the receiver side. These techniques are discussed in the next chapter.

UNIT 5

HANDOFFS AND DROPPED CALLS

Introduction

Apart from the better transmitter and receiver technology, mobile communications require signal processing techniques that improve the link performance. Equalization, Diversity and channel coding are channel impairment improvement techniques. Equalization compensates for Inter Symbol Interference (ISI) created by multipath within time dispersive channels. An equalizer within a receiver compensates for the average range of expected channel amplitude and delay characteristics. In other words, an equalizer is a filter at the mobile receiver whose impulse response is inverse of the channel impulse response. As such equalizers find their use in frequency selective fading channels. Diversity is another technique used to compensate fast fading and is usually implemented using two or more receiving antennas. It is usually employed to reduce the depths and duration of the fades experienced by a receiver in a flat fading channel. Channel coding improves mobile communication link performance by adding redundant data bits in the transmitted message.At the baseband portion of the transmitter, a channel coder maps a digital message sequence in to another specific code sequence containing greater number of bits than original contained in the message. Channel Coding is used to correct deep fading or spectral null. We discuss all three of these techniques in this chapter. A general framework of the fading effects and their mitigation techniques is shown in Figure 7.1.

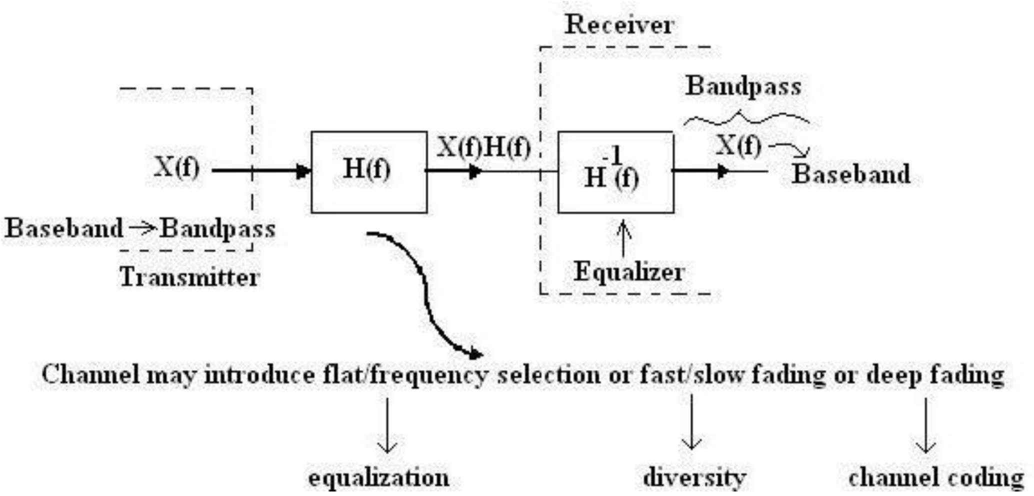


Figure 7.1: A general framework of fading effects and their mitigation techniques.

Equalization

ISI has been identified as one of the major obstacles to high speed data transmission over mobile radio channels. If the modulation bandwidth exceeds the coherence bandwidth of the radio channel (i.e., frequency selective fading), modulation pulses are spread in time, causing ISI. An equalizer at the front end of a receiver compensates for the average range of expected channel amplitude and delay characteristics. As the mobile fading channels are random and time varying, equalizers must track the time-varying characteristics of the mobile channel and therefore should be time-varying or adaptive. An adaptive equalizer has two phases of operation: training and tracking. These are as follows.

Training Mode:

- Initially a known, fixed length training sequence is sent by the transmitter so that the receiver equalizer may average to a proper setting.
- Training sequence is typically a pseudo-random binary signal or a fixed, of prescribed bit pattern.
- The training sequence is designed to permit an equalizer at the receiver to acquire the proper filter coefficient in the worst possible channel condition. An adaptive filter at the receiver thus uses a recursive algorithm to evaluate the channel and estimate filter coefficients to compensate for the channel.

Tracking Mode:

- When the training sequence is finished the filter coefficients are near optimal.
- Immediately following the training sequence, user data is sent.
- When the data of the users are received, the adaptive algorithms of the equalizer tracks the changing channel.
- As a result, the adaptive equalizer continuously changes the filter characteristics over time.

A Mathematical Framework

The signal received by the equalizer is given by

$$x(t) = d(t) * h(t) + n_b(t) \tag{7.1}$$

where $d(t)$ is the transmitted signal, $h(t)$ is the combined impulse response of the transmitter,channel and the RF/IF section of the receiver and $n_b(t)$ denotes the baseband noise.

If the impulse response of the equalizer is $h_{eq}(t)$, the output of the equalizer is

$$\hat{y}(t) = d(t) * h(t) * h_{eq}(t) + n_b(t) * h_{eq}(t) = d(t) * g(t) + n_b(t) * h_{eq}(t). \quad (7.2)$$

However, the desired output of the equalizer is $d(t)$ which is the original source data. Assuming $n_b(t)=0$, we can write $y(t) = d(t)$, which in turn stems the following equation:

$$g(t) = h(t) * h_{eq}(t) = \delta(t) \quad (7.3)$$

The main goal of any equalization process is to satisfy this equation optimally. In frequency domain it can be written as

$$H_{eq}(f) H(f) = 1 \quad (7.4)$$

which indicates that an equalizer is actually an inverse filter of the channel. If the channel is frequency selective, the equalizer enhances the frequency components with small amplitudes and attenuates the strong frequencies in the received frequency

spectrum in order to provide a flat, composite received frequency response and linear phase response. For a time varying channel, the equalizer is designed to track the channel variations so that the above equation is approximately satisfied.

Zero Forcing Equalization

In a zero forcing equalizer, the equalizer coefficients c_n are chosen to force the samples of the combined channel and equalizer impulse response to zero. When each of the delay elements provide a time delay equal to the symbol duration T, the frequency response $H_{eq}(f)$ of the equalizer is periodic with a period equal to the symbol rate 1/T. The combined response of the channel with the equalizer must satisfy Nyquist's criterion

$$H_{ch}(f)\, H_{eq}(f) = 1, \quad |f| < 1/2T \quad (7.5)$$

where $H_{ch}(f)$ is the folded frequency response of the channel. Thus, an infinite length zero-forcing ISI equalizer is simply an inverse filter which inverts the folded frequency response of the channel.

Disadvantage: Since $H_{eq}(f)$ is inverse of $H_{ch}(f)$ so inverse filter may excessively amplify the noise at frequencies where the folded channel spectrum has high attenuation, so it is rarely used for wireless link except for static channels with high SNR such as local wired telephone. The usual equalizer model follows a time varying or adaptive structure which is given next.

A Generic Adaptive Equalizer

The basic structure of an adaptive filter is shown in Figure 7.2. This filter is called the transversal filter, and in this case has N delay elements, N+1 taps and N+1 tunable complex multipliers, called weights. These weights are updated continuously by an adaptive algorithm. In the figure the subscript k represents discrete time index. The adaptive algorithm is controlled by the error signal e_k. The error signal is derived by comparing the output of the equalizer, with some signal d_k which is replica of transmitted signal. The adaptive algorithm uses e_k to minimize the cost function and uses the equalizer weights in such a manner that it minimizes the cost function iteratively. Let us denote the received sequence vector at the receiver and

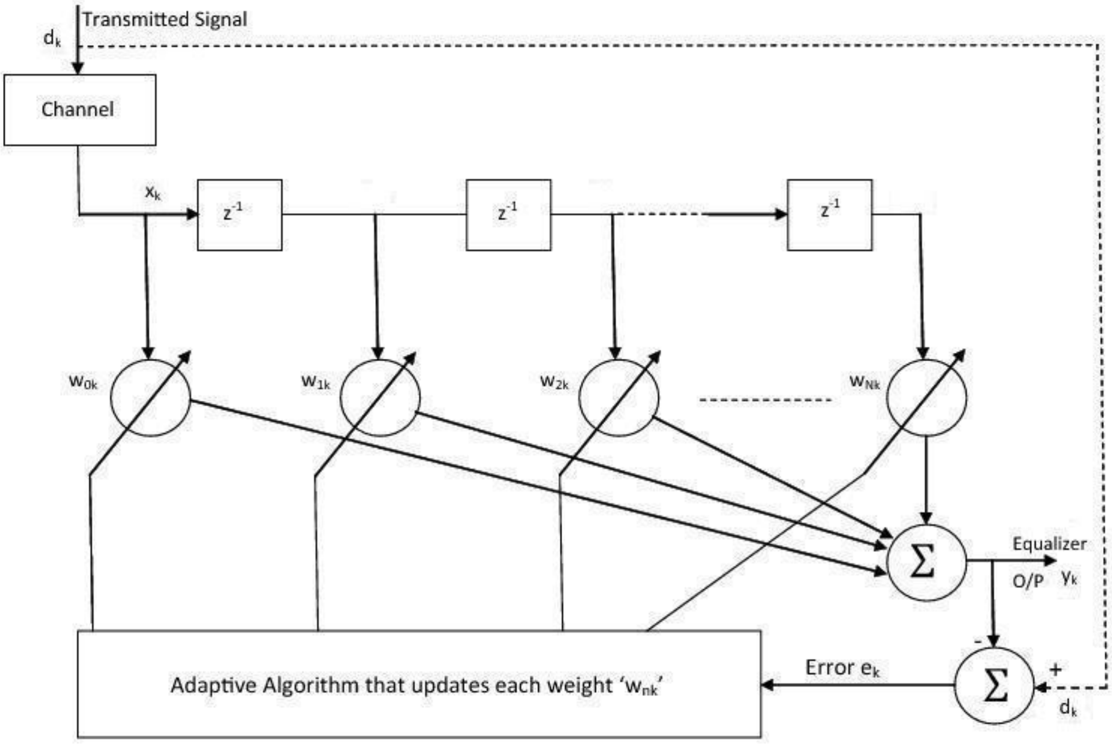


Figure 7.2: A generic adaptive equalizer.

the input to the equalizer as

$$\mathbf{x}_k = [x_k, x_{k-1}, ..., x_{k-N}]^T, \tag{7.6}$$

and the tap coefficient vector as

$$\mathbf{w}_k = [w_k^0, w_k^1, ..., w_k^N]^T. \tag{7.7}$$

Now, the output sequence of the equalizer y_k is the inner product of $\mathbf{x}_k$ and $\mathbf{w}_k$, i.e.,

$$y_k = \langle \mathbf{x}_k, \mathbf{w}_k \rangle = \mathbf{x}_k^T \mathbf{w}_k = \mathbf{w}_k^T \mathbf{x}_k. \tag{7.8}$$

The error signal is defined as

$$e_k = d_k - y_k = d_k - \mathbf{x}_k^T \mathbf{w}_k. \tag{7.9}$$

Assuming d_k and $\mathbf{x}_k$ to be jointly stationary, the Mean Square Error (MSE) is given as

$$\begin{aligned} MSE = E[e_k^2] &= E[(d_k - y_k)^2] \\ &= E[(d_k - \mathbf{x}_k^T \mathbf{w}_k)^2] \\ &= E[d_k^2] + \mathbf{w}_k^T E[\mathbf{x}_k \mathbf{x}_k^T]\mathbf{w}_k - 2E[d_k \mathbf{x}_k^T]\mathbf{w}_k \end{aligned} \tag{7.10}$$

where $\mathbf{w}_k$ is assumed to be an array of optimum values and therefore it has been taken out of the $E()$ operator. The MSE then can be expressed as

$$MSE = \xi = \sigma_d^2 + \mathbf{w}_k^T \mathbf{R} \mathbf{w}_k - 2\mathbf{p}^T \mathbf{w}_k \qquad (7.11)$$

where the signal variance $\sigma_d^2 = E[d_k^2]$ and the cross correlation vector $\mathbf{p}$ between the desired response and the input signal is defined as

$$p = E[d_k \mathbf{x}_k] = E\begin{bmatrix} d_k x_k & d_k x_{k-1} & d_k x_{k-2} & \cdots & d_k x_{k-N} \end{bmatrix}^T. \qquad (7.12)$$

The input correlation matrix $\mathbf{R}$ is defined as an $(N+1) \times (N+1)$ square matrix, where

$$\mathbf{R} = E\left[\mathbf{x}_k \mathbf{x}_k^T\right] = E\begin{bmatrix} x_k^2 & x_k x_{k-1} & x_k x_{k-2} & \cdots & x_k x_{k-N} \\ x_{k-1} x_k & x_{k-1}^2 & x_{k-1} x_{k-2} & \cdots & x_{k-1} x_{k-N} \\ x_{k-2} x_k & x_{k-2} x_{k-1} & x_{k-2}^2 & \cdots & x_{k-2} x_{k-N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{k-N} x_k & x_{k-N} x_{k-1} & x_{k-N} x_{k-2} & \cdots & x_{k-N}^2 \end{bmatrix}. \qquad (7.13)$$

Clearly, MSE is a function of $\mathbf{w}_k$. On equating $\frac{\partial \xi}{\partial w_k}$ to 0, we get the condition for minimum MSE (MMSE) which is known as Wiener solution:

$$\mathbf{w}_k = \mathbf{R}^{-1}\mathbf{p}. \qquad (7.14)$$

Hence, MMSE is given by the equation

$$MMSE = \xi_{min} = \sigma_d^2 - \mathbf{p}^T \mathbf{w}_k. \qquad (7.15)$$

Choice of Algorithms for Adaptive Equalization

Since an adaptive equalizer compensates for an unknown and time varying channel, it requires a specific algorithm to update the equalizer coefficients and track the channel variations. Factors which determine algorithm's performance are:

Rate of convergence: Number of iterations required for an algorithm, in response to a stationary inputs, to converge close enough to optimal solution. A fast rate of convergence allows the algorithm to adapt rapidly to a stationary environment of unknown statistics.

Misadjustment: Provides a quantitative measure of the amount by which the final value of mean square error, averaged over an ensemble of adaptive filters, deviates from an optimal mean square error.

Computational complexity: Number of operations required to make one complete iteration of the algorithm.

Numerical properties: Inaccuracies like round-off noise and representation errors in the computer, which influence the stability of the algorithm.

Three classic equalizer algorithms are primitive for most of today's wireless standards. These include the Zero Forcing Algorithm (ZF), the Least Mean Square Algorithm (LMS), and the Recursive Least Square Algorithm (RLS). Below, we discuss a few of the adaptive algorithms.

Least Mean Square (LMS) Algorithm

LMS algorithm is the simplest algorithm based on minimization of the MSE between the desired equalizer output and the actual equalizer output, as discussed earlier. Here the system error, the MSE and the optimal Wiener solution remain the same as given the adaptive equalization framework.

In practice, the minimization of the MSE is carried out recursively, and may be performed by use of the stochastic gradient algorithm. It is the simplest equalization algorithm and requires only 2N+1 operations per iteration. The filter weights are updated by the update equation. Letting the variable n denote the sequence of iteration, LMS is computed iteratively by

$$w_k(n+1) = w_k(n) + \mu e_k(n)\, x(n-k) \tag{7.16}$$

where the subscript k denotes the kth delay stage in the equalizer and μ is the step size which controls the convergence rate and stability of the algorithm.

The LMS equalizer maximizes the signal to distortion ratio at its output within the constraints of the equalizer filter length. If an input signal has a time dispersion characteristics that is greater than the propagation delay through the equalizer, then the equalizer will be unable to reduce distortion. The convergence rate of the LMS algorithm is slow due to the fact that there is only one parameter, the step size, that controls the adaptation rate. To prevent the adaptation from becoming unstable, the value of μ is chosen from

$$0 < \mu < 2 / \sum_{i=1}^{N} \lambda_i \tag{7.17}$$

where λ_i is the i-th eigenvalue of the covariance matrix R.

Normalized LMS (NLMS) Algorithm

In the LMS algorithm, the correction that is applied to $w_k(n)$ is proportional to the input sample $x(n-k)$. Therefore when $x(n-k)$ is large, the LMS algorithm experiences gradient noise amplification. With the normalization of the LMS step size by $\|\mathbf{x}(n)\|^2$ in the NLMS algorithm, this problem is eliminated. Only when $x(n-k)$ becomes close to zero, the denominator term $\|\mathbf{x}(n)\|^2$ in the NLMS equation becomes very small and the correction factor may diverge. So, a small positive number ε is added to the denominator term of the correction factor. Here, the step size is time varying and is expressed as

$$\mu(n) = \frac{\beta}{\|\mathbf{x}(n)\|^2 + \varepsilon}. \tag{7.18}$$

Therefore, the NLMS algorithm update equation takes the form of

$$w_k(n+1) = w_k(n) + \frac{\beta}{\|\mathbf{x}(n)\|^2 + \varepsilon} e_k(n)\, x(n-k). \tag{7.19}$$

Diversity

Diversity is a method used to develop information from several signals transmitted over independent fading paths. It exploits the random nature of radio propagation by finding independent signal paths for communication. It is a very simple concept where if one path undergoes a deep fade, another independent path may have a strong signal. As there is more than one path to select from, both the instantaneous and average SNRs at the receiver may be improved. Usually diversity decisions are made by receiver. Unlike equalization, diversity requires no training overhead as a training sequence is not required by transmitter. Note that if the distance between two receivers is a multiple of $\lambda/2$, there might occur a destructive interference between the two signals. Hence receivers in diversity technique are used in such a way that the signal received by one is independent of the other. Diversity can be of various forms, starting from space diversity to time diversity. We take up the types one by one in the sequel.

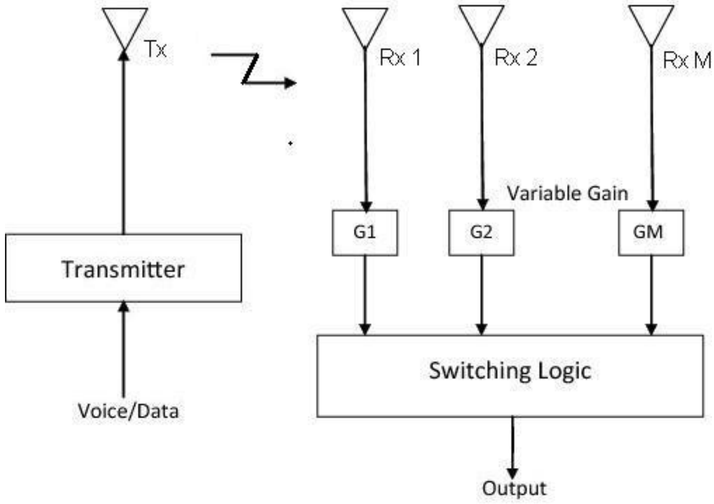


Figure 7.3: Receiver selection diversity, with M receivers.

Different Types of Diversity

Space Diversity

A method of transmission or reception, or both, in which the effects of fading are minimized by the simultaneous use of two or more physically separated antennas, ideally separated by one half or more wavelengths. Signals received from spatially separated antennas have uncorrelated envelopes.

Space diversity reception methods can be classified into four categories: selection, feedback or scanning, maximal ratio combining and equal gain combining.

(a) Selection Diversity:

The basic principle of this type of diversity is selecting the best signal among all the signals received from different branches at the receiving end. Selection Diversity is the simplest diversity technique. Figure 7.3 shows a block diagram of this method where 'M' demodulators are used to provide M diversity branches whose gains are adjusted to provide the same average SNR for each branch. The receiver branches having the highest instantaneous SNR is connected to the demodulator.

Let M independent Rayleigh fading channels are available at a receiver. Each channel is called a diversity branch and let each branch has the same average SNR. The signal to noise ratio is defined as

$$SNR = \Gamma = \frac{E_b}{N_0}\alpha^2 \tag{7.20}$$

where E_b is the average carrier energy, N_0 is the noise PSD, α is a random variable used to represent amplitude values of the fading channel.

The instantaneous SNR(γ_i) is usually defined as γ_i = instantaneous signal power per branch/mean noise power per branch. For Rayleigh fading channels, α has a Rayleigh distribution and so α^2 and consequently γ_i have a chi-square distribution with two degrees of freedom. The probability density function for such a channel is

$$p(\gamma_i) = \frac{1}{\Gamma} e^{\frac{-\gamma_i}{\Gamma}}. \tag{7.21}$$

The probability that any single branch has an instantaneous SNR less than some defined threshold γ is

$$\Pr[\gamma_i \leq \gamma] = \int_0^{\gamma} p(\gamma_i)\, d\gamma_i = \int_0^{\gamma} \frac{1}{\Gamma} e^{\frac{-\gamma_i}{\Gamma}} d\gamma_i = 1 - e^{\frac{-\gamma}{\Gamma}} = P(\Gamma). \tag{7.22}$$

Similarly, the probability that all M independent diversity branches receive signals which are simultaneously less than some specific SNR threshold γ is

$$\Pr[\gamma_1, \gamma_2, \ldots, \gamma_M \leq \gamma] = \left(1 - e^{\frac{-\gamma}{\Gamma}}\right)^M = P_M(\gamma) \tag{7.23}$$

where $P_M(\gamma)$ is the probability of all branches failing to achieve an instantaneous SNR = γ. Quite clearly, $P_M(\Gamma) < P(\Gamma)$. If a single branch achieves SNR > γ, then the probability that SNR > γ for one or more branches is given by

$$\Pr[\gamma_i > \gamma] = 1 - P_M(\gamma) = 1 - \left(1 - e^{\frac{-\gamma}{\Gamma}}\right)^M \tag{7.24}$$

which is more than the required SNR for a single branch receiver. This expression shows the advantage when a selection diversity is used.

To determine of average signal to noise ratio, we first find out the pdf of γ as

$$p_M(\gamma) = \frac{d}{d\gamma} P_M(\gamma) = \frac{M}{\Gamma}\left(1 - e^{-\gamma/\Gamma}\right)^{M-1} e^{-\gamma/\Gamma}. \tag{7.25}$$

The average SNR, $\bar{\gamma}$, can be then expressed as

$$\bar{\gamma} = \int_0^{\infty} \gamma p_M(\gamma)\, d\gamma = \Gamma \int_0^{\infty} Mx \left(1 - e^{-x}\right)^{M-1} e^{-x} dx \tag{7.26}$$

where $x = \gamma/\Gamma$ and Γ is the average SNR for a single branch, when no diversity is used.

This equation shows an average improvement in the link margin without requiring extra transmitter power or complex circuitry, and it is easy to implement as it needed a monitoring station and an antenna switch at the receiver. It is not an optimal diversity technique as it doesn't use all the possible branches simultaneously.

(b) Feedback or Scanning Diversity:

Scanning all the signals in a fixed sequence until the one with SNR more than a predetermined threshold is identified. Feedback or scanning diversity is very similar to selection diversity except that instead of always using the best of N signals, the N signals are scanned in a fixed sequence until one is found to be above a predetermined threshold. This signal is then received until it falls below threshold and the scanning process is again initiated. The resulting fading statistics are somewhat inferior, but the advantage is that it is very simple to implement(only one receiver is required).

(c) Maximal Ratio Combining:

Signals from all of the m branches are weighted according to their individual signal voltage to noise power ratios and then summed. Individual signals must be cophased before being summed, which generally requires an individual receiver and phasing circuit for each antenna element. Produces an output SNR equal to the sum of all individual SNR. Advantage of producing an output with an acceptable SNR even when none of the individual signals are themselves acceptable. Modern DSP techniques and digital receivers are now making this optimal form, as it gives the best statistical reduction of fading of any known linear diversity combiner. In terms of voltage signal,

$$r_m = \sum_{i=1}^{m} tt_i r_i \tag{7.27}$$

where tt_i is the gain and r_i is the voltage signal from each branch.

(d) Equal Gain Combining:

In some cases it is not convenient to provide for the variable weighting capability required for true maximal ratio combining. In such cases, the branch weights are all set unity, but the signals from each branch are co-phased to provide equal gain combining diversity. It allows the receiver to exploit signals that are simultaneously received on each branch. Performance of this method is marginally inferior to maximal ratio combining and superior to Selection diversity. Assuming all the tt_i to be

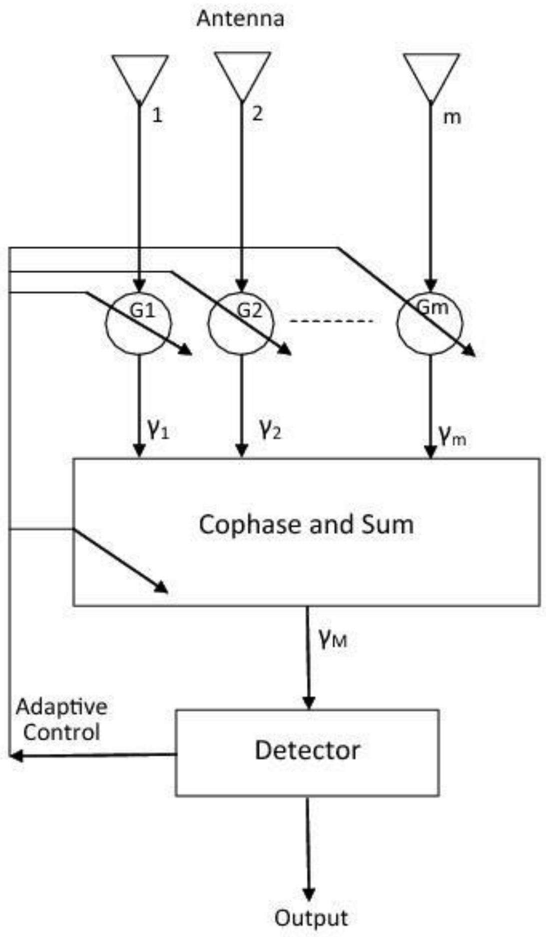


Figure 7.4: Maximal ratio combining technique.

unity, here,

$$r_m = \sum_{i=1}^{m} r_i. \tag{7.28}$$

Polarization Diversity

Polarization Diversity relies on the decorrelation of the two receive ports to achieve diversity gain. The two receiver ports must remain cross-polarized. Polarization Diversity at a base station does not require antenna spacing. Polarization diversity combines pairs of antennas with orthogonal polarizations (i.e. horizontal/vertical, ± slant 45^o, Left-hand/Right-hand CP etc). Reflected signals can undergo polarization changes depending on the channel. Pairing two complementary polarizations, this scheme can immunize a system from polarization mismatches that would otherwise cause signal fade. Polarization diversity has prove valuable at radio and mobile communication base stations since it is less susceptible to the near random orientations of transmitting antennas.

Frequency Diversity

In Frequency Diversity, the same information signal is transmitted and received simultaneously on two or more independent fading carrier frequencies. Rationale behind this technique is that frequencies separated by more than the coherence bandwidth of the channel will be uncorrelated and will thus not experience the same fades. The probability of simultaneous fading will be the product of the individual fading probabilities. This method is employed in microwave LoS links which carry several channels in a frequency division multiplex mode (FDM). Main disadvantage is that it requires spare bandwidth also as many receivers as there are channels used for the frequency diversity.

Time Diversity

In time diversity, the signal representing the same information are sent over the same channel at different times. Time diversity repeatedly transmits information at time spacings that exceeds the coherence time of the channel. Multiple repetition of the signal will be received with independent fading conditions, thereby providing for diversity. A modern implementation of time diversity involves the use of RAKE receiver for spread spectrum CDMA, where the multipath channel provides redundancy in the transmitted message. Disadvantage is that it requires spare bandwidth also as many receivers as there are channels used for the frequency diversity. Two important types of time diversity application is discussed below.

Application 1: RAKE Receiver

In CDMA spread spectrum systems, CDMA spreading codes are designed to provide very low correlation between successive chips, propagation delay spread in the radio channel provides multiple version of the transmitted signal at the receiver. Delaying multipath components by more than a chip duration, will appear like uncorrelated noise at a CDMA receiver. CDMA receiver may combine the time delayed versions of the original signal to improve the signal to noise ratio at the receiver. RAKE

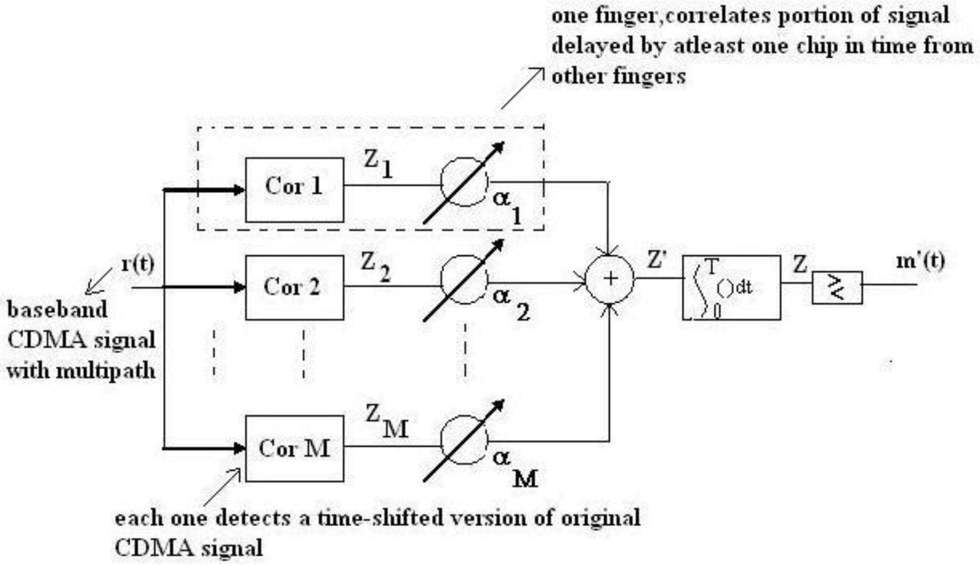


Figure 7.5: RAKE receiver.

receiver collect the time shifted versions of the original signal by providing a separate correlation receiver for M strongest multipath components. Outputs of each correlator are weighted to provide a better estimate of the transmitted signal than provided by a single component. Demodulation and bit decisions are based on the weighted output of the correlators. Schematic of a RAKE receiver is shown in Figure 7.5.

Application 2: Interleaver

In the encoded data bits, some source bits are more important than others, and must be protected from errors. Many speech coder produce several important bits in succession. Interleaver spread these bit out in time so that if there is a deep fade or noise burst, the important bits from a block of source data are not corrupted at the same time. Spreading source bits over time, it becomes possible to make use of error control coding. Interleaver can be of two forms, a block structure or a convolutional structure.

A block interleaver formats the encoded data into a rectangular array of m rows and n columns, and interleaves nm bits at a time. Each row contains a word of source data having n bits. an interleaver of degree m consists of m rows. source bits are placed into the interleaver by sequentially increasing the row number for each successive bit, and forming the columns. The interleaved source data is then read out row-wise and transmitted over the channel. This has the effect of separating the original source bits by m bit periods. At the receiver, de-interleaver stores the received data by sequentially increasing the row number of each successive bit, and then clocks out the data row-wise, one word at a time. Convolutional interleavers are ideally suited for use with convolutional codes.

Channel Coding

In channel coding, redundant data bits are added in the transmitted message so that if an instantaneous fade occurs in the channel, the data may still be recovered at the receiver without the request of retransmission. A channel coder maps the transmitted message into another specific code sequence containing more bits. Coded message is then modulated for transmission in the wireless channel. Channel Coding is used by the receiver to detect or correct errors introduced by the channel. Codes that used to detect errors, are error detection codes. Error correction codes can detect and correct errors.

Shannon's Channel Capacity Theorem

In 1948, Shannon showed that by proper encoding of the information, errors induced by a noise channel can be reduced to any desired level without sacrificing the rate of information transfer. Shannon's channel capacity formula is applicable to the AWGN channel and is given by:

$$C = B\log_2\left(1 + \frac{S}{N}\right) = B\log_2\left(1 + \frac{P}{N_0 B}\right) = B\log_2\left(1 + \frac{E_b R_b}{N_0 B}\right) \qquad (7.29)$$

where C is the channel capacity (bit/s), B is the channel bandwidth (Hz), P is the received signal power (W), N_0 is the single sided noise power density (W/Hz), E_b is the average bit energy and R_b is transmission bit rate.

Equation (7.29) can be normalized by the bandwidth B and is given as

$$\frac{C}{B} = \log_2\left(1 + \frac{E_b R_b}{N_0 B}\right) \qquad (7.30)$$

and the ratio C/B is denoted as bandwidth efficiency. Introduction of redundant bits increases the transmission bit rate and hence it increases the bandwidth requirement, which reduces the bandwidth efficiency of the link in high SNR conditions, but provides excellent BER performance at low SNR values. This leads to the following two inferences.

Corollary 1 : While dealing within maximum channel capacity, introduction of redundant bits increase the transmitter rate and hence bandwidth requirement also increases, while decreasing the bandwidth efficiency, but it also decreases the BER.

Corollary 2: If data redundancy is not introduced in a wideband noisy environment, error free performance in not possible (for example, CDMA communication in 3G mobile phones).

A channel coder operates on digital message (or source) data by encoding the source information into a code sequence for transmission through the channel. The error correction and detection codes are classified into three groups based on their structure.

1. Block Code
2. Convolution Code
3. Concatenated Code.

Block Codes

Block codes are *forward error correction* (FEC) codes that enable a limited number of errors to be detected and corrected without retransmission. Block codes can be used to improve the performance of a communications system when other means of improvement (such as increasing transmitter power or using a more sophisticated demodulator) are impractical.

In block codes, parity bits are added to blocks of message bits to make codewords or code blocks. In a block encoder, k information bits are encoded into n code bits. A total of $n-k$ redundant bits are added to the k information bits for the purpose of detecting and correcting errors. The block code is referred to as an (n, k) code, and the rate of the code is defined as $R_c = k/n$ and is equal to the rate of information divided by the raw channel rate.

Parameters in Block Code

(a) Code Rate (R_c): As defined above, $R_c = k/n$.

(b) Code Distance (d): Distance between two codewords is the number of elements in which two codewords C_i and C_j differs denoted by $d(C_i, C_j)$. If the code used is binary, the distance is known as 'Hamming distance'. For example d(10110, 11011) is 3. If the code 'C' consists of the set of codewords, then the minimum distance of the code is given by $d_{\min} = min\ \{d\ (C_i, C_j)\}$.

(c) Code Weight (w): Weight of a codeword is given by the number of nonzero elements in the codeword. For a binary code, the weight is basically the number of 1s in the codeword. For example weight of a code 101101 is 4.

Ex 1: The block code C = 00000, 10100, 11110, 11001 can be used to represent two bit binary numbers as:

- 00 – 00000
- 01 – 10100
- 10 – 11110
- 11 – 11001

Here number of codewords is 4, k = 2, and n = 5.

To encode a bit stream 1001010011

- First step is to break the sequence in groups of two bits, i.e., 10 01 01 00 11
- Next step is to replace each block by its corresponding codeword, i.e.,
  11110 10100 10100 00000 11001

Quite clearly, here, $d_{\min} = min\{d(C_i, C_j)\} = 2$.

Properties of Block Codes

(a) Linearity: Suppose C_i and C_j are two code words in an (n, k) block code. Let α_1 and α_2 be any two elements selected from the alphabet. Then the code is said to be linear if and only if $\alpha_1 C_1 + \alpha_2 C_2$ is also a code word. A linear code must contain the all-zero code word.

(b) Systematic: A systematic code is one in which the parity bits are appended to the end of the information bits. For an (n, k) code, the first k bits are identical to the information bits, and the remaining $n - k$ bits of each code word are linear combinations of the k information bits.

(c) Cyclic: Cyclic codes are a subset of the class of linear codes which satisfy the following cyclic shift property: If $C = [C_{n-1}, C_{n-2}, ..., C_0]$ is a code word of a cyclic code, then $[C_{n-2}, C_{n-3}, ..., C_0, C_{n-1}]$, obtained by a cyclic shift of the elements of C, is also a code word. That is, all cyclic shifts of C are code words.

In this context, it is important to know about **Finite Field or Galois Field**. Let F be a finite set of elements on which two binary operations – addition (+) and multiplication (.) are defined. The set F together with the two binary operations is called a *field* if the following conditions are satisfied:

1. F is a commutative group under addition.
2. The set of nonzero elements in F is a commutative group under multiplication.
3. Multiplication is distributive over addition; that is, for any three elements a, b, and c in F, $a(b + c) = ab + ac$
4. Identity elements 0 and 1 must exist in F satisfying $a + 0 = a$ and $a.1 = a$.
5. For any a in F, there exists an additive inverse $(-a)$ such that $a + (-a) = 0$.
6. For any a in F, there exists an multiplicative inverse a^{-1} such that $a.a^{-1} = 1$.

Depending upon the number of elements in it, a field is called either a finite or an infinite field. The examples of infinite field include Q (set of all rational numbers), R (set of all real numbers), C (set of all complex numbers) etc. A field with a finite number of elements (say q) is called a 'Galois Field' and is denoted by GF(q). A finite field entity p(x), called a polynomial, is introduced to map all symbols (with several bits) to the element of the finite field. A polynomial is a mathematical expression

$$p(x) = p_0 + p_1x + ... + p_mx_m \qquad (7.31)$$

where the symbol x is called the indeterminate and the coefficients $p_0, p_1, ..., p_m$ are the elements of GF(q). The coefficient p_m is called the leading coefficient. If p_m is not equal to zero, then m is called the degree of the polynomial, denoted as deg p(x). A polynomial is called monic if its leading coefficient is unity. The division algorithm states that for every pair of polynomials a(x) and b(x) in F(x), there exists a unique pair of polynomials q(x), the quotient, and r(x), the remainder, such that $a(x) = q(x)b(x) + r(x)$, where deg r(x)¡deg b(x). A polynomial p(x) in F(x) is said to be reducible if p(x)=a(x)b(x), otherwise it is called irreducible. A monic irreducible polynomial of degree at least one is called a prime polynomial.

An irreducible polynomial $p(x)$ of degree 'm' is said to be primitive if the smallest integer 'n' for which $p(x)$ divides $x^n + 1$ is $n = 2^m - 1$. A typical primitive polynomial is given by $p(x) = x^m + x + 1$.

A specific type of code which obeys both the cyclic property as well as polynomial operation is cyclic codes. Cyclic codes are a subset of the class of linear codes which satisfy the cyclic property. These codes possess a considerable amount of structure which can be exploited. A cyclic code can be generated by using a generator polynomial g(p) of degree (n-k). The generator polynomial of an (n,k) cyclic code is a factor of $p^n + 1$ and has the form

$$g(p) = p^{n-k} + g_{n-k-1}p^{n-k-1} + \cdots + g_1 p + 1. \quad (7.32)$$

A message polynomial x(p) can also be defined as

$$x(p) = x_{k-1}p^{k-1} + \cdots + x_1 p + x_0 \quad (7.33)$$

where $(x_{k-1}, \ldots, x_0)$ represents the k information bits. The resultant codeword c(p) can be written as

$$c(p) = x(p)\, g(p) \quad (7.34)$$

where c(p) is a polynomial of degree less than n. We would see an application of such codes in Reed-Solomon codes.

Examples of Block Codes

(a) Single Parity Check Code: In single parity check codes (example: ASCII code), an overall single parity check bit is appended to 'k' information bits. Let the infor- mation bit word be: $(b_1, b_2, ..., b_k)$, then parity check bit: $p = b_1 + b_2 +$
$+ b_k$
modulo 2 is appended at the (k+1)th position, making the overall codeword: $C = (b_1, b_2, ..., b_k, p)$. The parity bit may follow an even parity or an odd parity pattern. All error patterns that change an odd number of bits are detectable, and all even numbered error patterns are not detectable. However, such codes can only detect the error, it cannot correct the error.

Ex. 2: Consider a (8,7) ASCII code with information codeword (0, 1, 0, 1, 1, 0, 0) and encoded with overall even parity pattern. Thus the overall codeword is (0, 1, 0, 1, 1, 0, 0, 1) where the last bit is the parity bit. If there is a single error in bit 3: (0, 1, **1**, 1, 1, 0, 0, 1), then it can be easily checked by the receiver that now there are odd number of 1's in the codeword and hence there is an error. On the other hand, if there are two errors, say, errors in bit 3 and 5: (0, 1, **1**, 1, **0**, 0, 0, 1), then error

will not be detected.

After decoding a received codeword, let p_c be the probability that the decoder gives correct codeword C, p_e is the probability that the decoder gives incorrect codeword C^{j} C, and p_f is the probability that the decoder fails to give a codeword. In this case, we can write $p_c + p_e + p_f = 1$.

If in an n-bit codeword, there are j errors and p is the bit error probability, then the probability of obtaining j errors in this codeword is $P_j = {}^{n}C_j p^j (1 - p)^{n-j}$. Using this formula, for any $(n, n - 1)$ single parity check block code, we get

- $p_c = P_0$,
- $p_e = P_2 + P_4 + ... + P_{n^j}$ $(n^j = n$ if n is even, otherwise $n^j = n - 1)$,
- $p_f = P_1 + P_3 + ... + P_{n^j}$ $(n^j = n - 1$ if n is even, otherwise $n^j = n)$.

As an example, for a (5,4) single parity check block code, $p_c = P_0$, $p_e = P_2 + P_4$, and $p_f = P_1 + P_3 + P_5$.

(b) Product Codes: Product codes are a class of linear block codes which provide error detection capability using product of two block codes. Consider that nine information bits (1, 0, 1, 0, 0, 1, 1, 1, 0) are to be transmitted. These 9 bits can be divided into groups of three information bits and (4,3) single parity check codeword can be formed with even parity. After forming three codewords, those can be appended with a vertical parity bit which will form the fourth codeword. Thus the following codewords are transmitted:

$$C1 = [1\ 0\ 1\ 0]$$
$$C2 = [0\ 0\ 1\ 1]$$
$$C3 = [1\ 1\ 0\ 0]$$
$$C4 = [0\ 1\ 0\ 1].$$

Now if an error occurs in the second bit of the second codeword, the received codewords at the receiver would then be

$$C1 = [1\ 0\ 1\ 0]$$
$$C2 = [0\ 1\ 1\ 1] \leftarrow$$
$$C3 = [1\ 1\ 0\ 0]$$
$$C4 = [0\ 1\ 0\ 1]$$

$$\uparrow$$

and these would indicate the corresponding row and column position of the erroneous bit with vertical and horizontal parity check. Thus the bit can be corrected. Here we get a horizontal (4, 3) codeword and a vertical (4, 3) codeword and concatenating them we get a (16, 9) product code. In general, a product code can be formed as (n_1, k_1) & $(n_2, k_2) \rightarrow (n_1n_2, k_1k_2)$.

(c) Repetition Codes: In a (n,1) repetition code each information bit is repeated n times (n should be odd) and transmitted. At the receiver, the majority decoding principle is used to obtain the information bit. Accordingly, if in a group of n received bit, 1 occurs a higher number of times than 0, the information bit is decoded as 1. Such majority scheme works properly only if the noise affects less than n/2 number of bits.

Ex 3: Consider a (3,1) binary repetition code.

- For input bit 0, the codeword is (0 0 0) and for input bit 1, the codeword is (1 1 1).
- If the received codeword is (0 0 0), i.e. no error, it is decoded as 0.
- Similarly, if the received codeword is (1 1 1), i.e. no error, it is decoded as 1.
- If the received codeword is (0 0 1) or (0 1 0) or (1 0 0), then error is detected and it is decoded as 0 with majority decoding principle.
- If the received codeword is (0 1 1) or (1 1 0) or (1 0 1), once again error is detected and it is decoded as 1 with majority decoding principle.

For such a (3,1) repetition code, $p_c = P_0 + P_1$, $p_e = P_2 + P_3$, and $p_f = 0$.

(d) Hamming Codes: A binary Hamming code has the property that

$$(n, k) = (2^m - 1, 2^m - 1 - m) \tag{7.35}$$

where k is the number of information bits used to form a n bit codeword, and m is any positive integer. The number of parity symbols are $n - k = m$. Thus, a codeword is represented by $C = [i_1, \ldots i_n, p_1, \ldots, p_{n-k}]$. This is quite a useful code in communication which is illustrated via the following example.

Ex 4: Consider a (7, 4) Hamming code. With three parity bits we can correct exactly 1 error. The parity bits may follow such a modulo 2 arithmetic:

$$p_1 = i_1 + i_2 + i_3,$$
$$p_2 = i_2 + i_3 + i_4,$$
$$p_3 = i_1 + i_3 + i_4,$$

which is same as,

$$p_1 + i_1 + i_2 + i_3 = 0$$
$$p_2 + i_2 + i_3 + i_4 = 0$$
$$p_3 + i_1 + i_3 + i_4 = 0.$$

The transmitted codeword is then $C = [i_1, i_2, \ldots, i_4, p_1, p_2, p_3]$.
Syndrome Decoding: For this Hamming code, let the received codeword be $V = [v_1, v_2, \ldots, v_4, v_5, v_6, v_7]$. We define a syndrome vector S as

$$\begin{aligned} S &= [S_1\ S_2\ S_3] \\ S1 &= v_1 + v_2 + v_3 + v_5 \\ S2 &= v_2 + v_3 + v_4 + v_6 \\ S3 &= v_1 + v_2 + v4 + v7 \end{aligned}$$

It is obvious that in case of no error, the syndrome vector is equal to zero. Corresponding to this syndrome vector, there is an error vector e which can be obtained from a syndrome table and finally the required codeword is taken as $C = V + e$. In a nutshell, to obtain the required codeword, we perform the following steps:

1. Calculate S from decoder input V.
2. From syndrome table, obtain e corresponding to S.
3. The required codeword is then $C = V + e$.

A few cases are given below to illustrate the syndrome decoding.

1. Let C = [0 1 1 1 0 1 0] and V = [0 1 1 1 0 1 0]. This implies S = [0 0 0], and it corresponds to e = [0 0 0 0 0 0 0]. Thus, C = V + e = [0 1 1 1 0 1 0].

2. Let C = [1 1 0 0 0 1 0] and V = [1 1 0 1 0 1 0]. This means S = [0 1 1], from which we get e = [0 0 0 1 0 0 0] which means a single bit error is there in the received bit v_4. This will be corrected by performing the operation C = V + e.

3. Another interesting case is, let C = [0 1 0 1 1 0 0] and V = [0 0 1 1 1 0 1] (two errors at second and third bits). This makes S = [0 0 0] and as a result, e = [0 0 0 0 0 0 0]. However, $C \neq V$, and $C = V + e$ implies the double error cannot be

corrected. Therefore a (7,4) Hamming code can correct only single bit error.

(e) Golay Codes: Golay codes are linear binary (23,12) codes with a minimum distance of seven and a error correction capability of three bits. This is a special, one of a kind code in that this is the only nontrivial example of a perfect code. Every codeword lies within distance three of any codeword, thus making maximum likelihood decoding possible.

(f) BCH Codes: BCH code is one of the most powerful known class of linear cyclic block codes, known for their multiple error correcting ability, and the ease of encoding and decoding. It's block length is $n = 2^m - 1$ for $m \geq 3$ and number of errors that they can correct is bounded by $t < (2^m - 1)/2$. Binary BCH codes can be generalized to create classes of non binary codes which use m bits per code symbol.

(g) Reed Solomon (RS) Codes: Reed-Solomon code is an important subset of the BCH codes with a wide range of applications in digital communication and data storage. Typical application areas are storage devices (CD, DVD etc.), wireless communications, digital TV, high speed modems. It's coding system is based on groups of bits, such as bytes, rather than individual 0 and 1. This feature makes it particularly good at dealing with burst of errors: six consecutive bit errors. Block length of these codes is $n = 2^m - 1$, and can be extended to 2^m or $2^m + 1$. Number

of parity symbols that must be used to correct e errors is $n - k = 2e$. Minimum distance $d_{min} = 2e + 1$, and it achieves the largest possible d_{min} of any linear code.

For US-CDPD, the RS code is used with $m = 6$. So each of the 64 field elements is represented by a 6 bit symbol. For this case, we get the primitive polynomial as $p(x) = x^6 + x + 1$. Equating $p(x)$ to 0 implies $x^6 = x + 1$.

The 6 bit representation of the finite field elements is given in Table 7.1. The table elements continue up to α^{62}. However, to follow linearity property there should be

Table 7.1: Finite field elements for US-CDPD

	α^5	α^4	α^3	α^2	α^1	α^0
1	0	0	0	0	0	1
α^1	0	0	0	1	0	0
α^2	0	0	1	0	0	0
.	.	.	.	.	.	.
.	.	.	.	.	.	.
$\alpha^6 = \alpha + 1$	0	0	0	0	1	1
.	.	.	.	.	.	.
.	.	.	.	.	.	.

a zero codeword, hence α^{63} is assigned zero.

The encoding part of the RS polynomial is done as follows:

Information polynomial: $d(x) = C_{n-1}x^{n-1} + C_{n-2}x^{n-2} + ... + C_{2t}x^{2t}$,

Parity polynomial: $p(x) = C_{2t-1}x^{2t-1} + + C_0$,

Codeword polynomial: $c(x) = d(x) + p(x)$.

Since generating an information polynomial is difficult, so a generating polynomial is used instead. Information polynomial is then the multiple of generating polynomial. This process is given below.

Since this kind of codes are cyclic codes, we take a generating polynomial $g(x)$ such that $d(x) = g(x)q(x) + r(x)$ where $q(x)$ is the quotient polynomial and $r(x)$ is the remainder polynomial. The codeword polynomial would then be given as: $c(x) = g(x)q(x) + r(x) = p(x)$. If we assign a parity polynomial $p(x) = r(x)$, then the codeword polynomial $c(x) = g(x)p(x)$ and the entire process becomes easier.

On the decoder side one has to find a specific $r(x) = p(x)$ or vice-versa, but due to its complexity, it is mainly done using syndrome calculation. The details of such a syndrome calculation can be found in [1].

Convolutional Codes

A continuous sequence of information bits is mapped into a continuous sequence of encoder output bits. A convolutional code is generated by passing the information sequence through a finite state shift register. Shift register contains 'N' k-bit stages

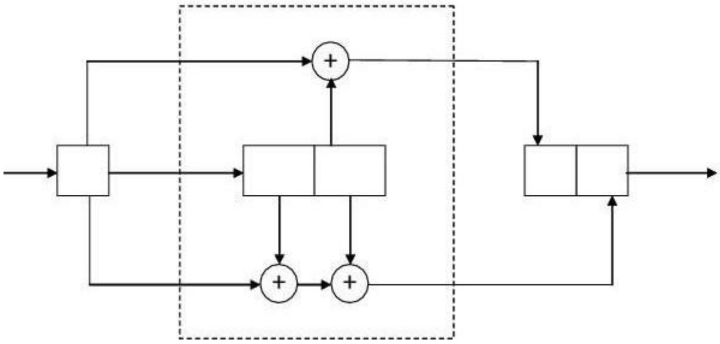


Figure 7.6: A convolutional encoder with n=2 and k=1.

and m linear algebraic function generators based on the generator polynomials. Input data is shifted into and along the shift register, k-bits at a time. Number of output bits for each k-bit user input data sequence is n bits, so the code rate $R_c = k/n$. The shift register of the encoder is initialized to all-zero-state before

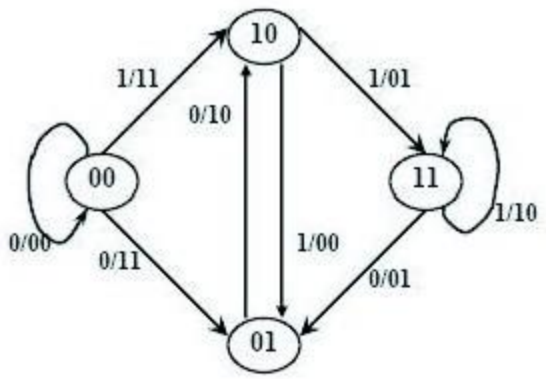


Figure 7.7: State diagram representation of a convolutional encoder.

encoding operation starts. It is easy to verify that encoded sequence is 00 11 10 00 01 . . . for an input message sequence of 01011 . . .. Convolution codes may be represented in various ways as given below.

State Diagram:

Since the output of the encoder is determined by the input and the current state of the encoder, a state diagram can be used to represent the encoding process. The state diagram is simply a graph of the possible states of the encoder and the possible transitions from one state to another. The path information between the states, denoted as b/c_1c_2, represents input information bit 'b' and the corresponding

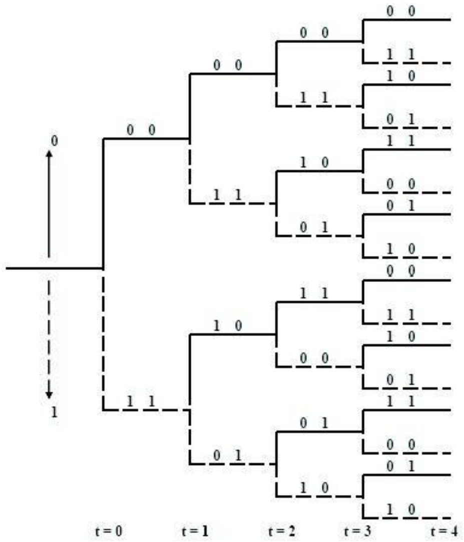


Figure 7.8: Tree diagram representation of a convolutional encoder.

output bits (c_1c_2). Again, it is not difficult to verify from the state diagram that an input information sequence b = (1011) generates an encoded sequence c = (11, 10, 00, 01).

Tree Diagram:

The tree diagram shows the structure of the encoder in the form of a tree with the branches representing the various states and the outputs of the coder. The encoded bits are labeled on the branches of the tree. Given an input sequence, the encoded sequence can be directly read from the tree. As an example, an input sequence (1011) results in the encoded sequence (11, 10, 00, 01).

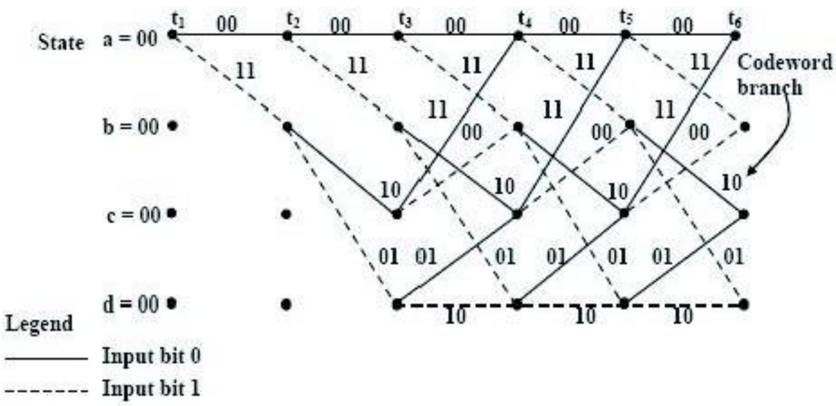


Figure 7.9: Trellis diagram of a convolutional encoder.

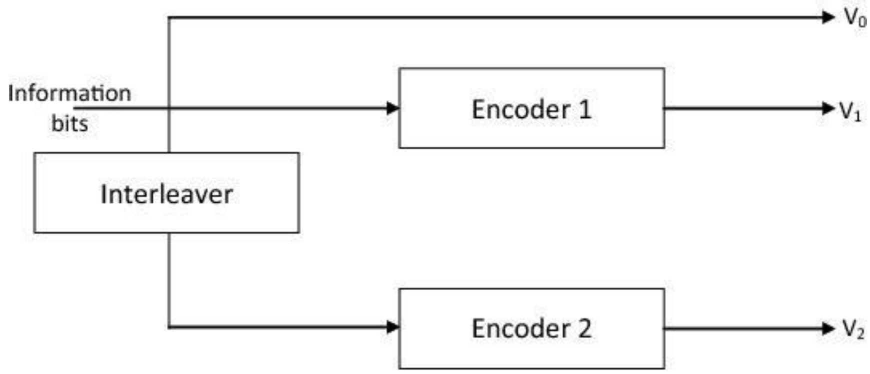


Figure 7.10: Block diagram of a turbo encoder.

Trellis Diagram:

Tree reveals that the structure repeats itself once the number of stages is greater than the constraint length. It is observed that all branches emanating from two nodes having the same state are identical in the sense that they generate identical output sequences. This means that the two nodes having the same label can be merged. By doing this throughout the tree diagram, we obtain another diagram called a Trellis Diagram which is more compact representation.

7.4.4 Concatenated Codes

Concatenated codes are basically concatenation of block and convolutional codes. It can be of two types: serial and parallel codes. Below, we discuss a popular parallel concatenated code, namely, turbo code.

Turbo Codes: A turbo encoder is built using two identical convolutional codes of special type with parallel concatenation. An individual encoder is termed a component encoder. An interleaver separates the two component encoders. The interleaver is a device that permutes the data sequence in some predetermined manner. Only one of the systematic outputs from the two component encoders is used to form a codeword, as the systematic output from the other component encoder is only a permuted version of the chosen systematic output. Figure 7.10 shows the block diagram of a turbo encoder using two identical encoders. The first encoder outputs the systematic V_0 and recursive convolutional V_1 sequences while the second encoder discards its systematic sequence and only outputs the recursive convolutional V_2 sequence. Depending on the number of input bits to a component encoder it may be binary or m-binary encoder. Encoders are also categorized as systematic or non-systematic. If the component encoders are not identical then it is called an asymmetric turbo code.

Conclusion

Although a lot of advanced powerful techniques for mitigating the fading effects such as space diversity in MIMO systems, space-time block coding scheme, MIMO equalization, BLAST architectures etc. have taken place in modern wireless communication, nevertheless, the discussed topics in this chapter are the basic building blocks for all such techniques and that stems the necessity for all these discussions. The effectiveness of the discussed topics would be more clear in the next chapter in the context of different multiple access techniques.

Multiple Access Techniques

Multiple access techniques are used to allow a large number of mobile users to share the allocated spectrum in the most efficient manner. As the spectrum is limited, so the sharing is required to increase the capacity of cell or over a geographical area by allowing the available bandwidth to be used at the same time by different users. And this must be done in a way such that the quality of service doesn't degrade within the existing users.

8.1 Multiple Access Techniques for Wireless Communi- cation

In wireless communication systems it is often desirable to allow the subscriber to send simultaneously information to the base station while receiving information from the base station.

A cellular system divides any given area into cells where a mobile unit in each cell communicates with a base station. The main aim in the cellular system design is to be able to increase the capacity of the channel i.e. to handle as many calls as possible in a given bandwidth with a sufficient level of quality of service. There are several different ways to allow access to the channel. These includes mainly the following:

1) Frequency division multiple-access (FDMA)
2) Time division multiple-access (TDMA)
3) Code division multiple-access (CDMA)

Table 8.1: MA techniques in different wireless communication systems

Advanced Mobile Phone Systems:	FDMA/FDD
Global System for Mobile:	TDMA/FDD
U.S. Digital Cellular:	TDMA/FDD
Japanese Digital Cellular:	TDMA/FDD
CT2 Cordless Telephone:	FDMA/TDD
Digital European Cordless Telephone:	FDMA/TDD
U.S. Narrowband Spread Spectrum (IS-95):	CDMA/FDD

4) Space Division Multiple access (SDMA)

FDMA,TDMA and CDMA are the three major multiple access techniques that are used to share the available bandwidth in a wireless communication system. Depending on how the available bandwidth is allocated to the users these techniques can be classified as narrowband and wideband systems.

Narrowband Systems

The term narrowband is used to relate the bandwidth of the single channel to the expected coherence bandwidth of the channel. The available spectrum is divided in to a large number of narrowband channels. The channels are operated using FDD. In narrow band FDMA, a user is assigned a particular channel which is not shared by other users in the vicinity and if FDD is used then the system is called FDMA/FDD. Narrow band TDMA allows users to use the same channel but allocated a unique time slot to each user on the channel, thus separating a small number of users in time on a single channel. For narrow band TDMA, there generally are a large number of channels allocated using either FDD or TDD, each channel is shared using TDMA. Such systems are called TDMA/FDD and TDMA/TDD access systems.

Wideband Systems

In wideband systems, the transmission bandwidth of a single channel is much larger than the coherence bandwidth of the channel. Thus, multipath fading doesnt greatly affect the received signal within a wideband channel, and frequency selective fades occur only in a small fraction of the signal bandwidth

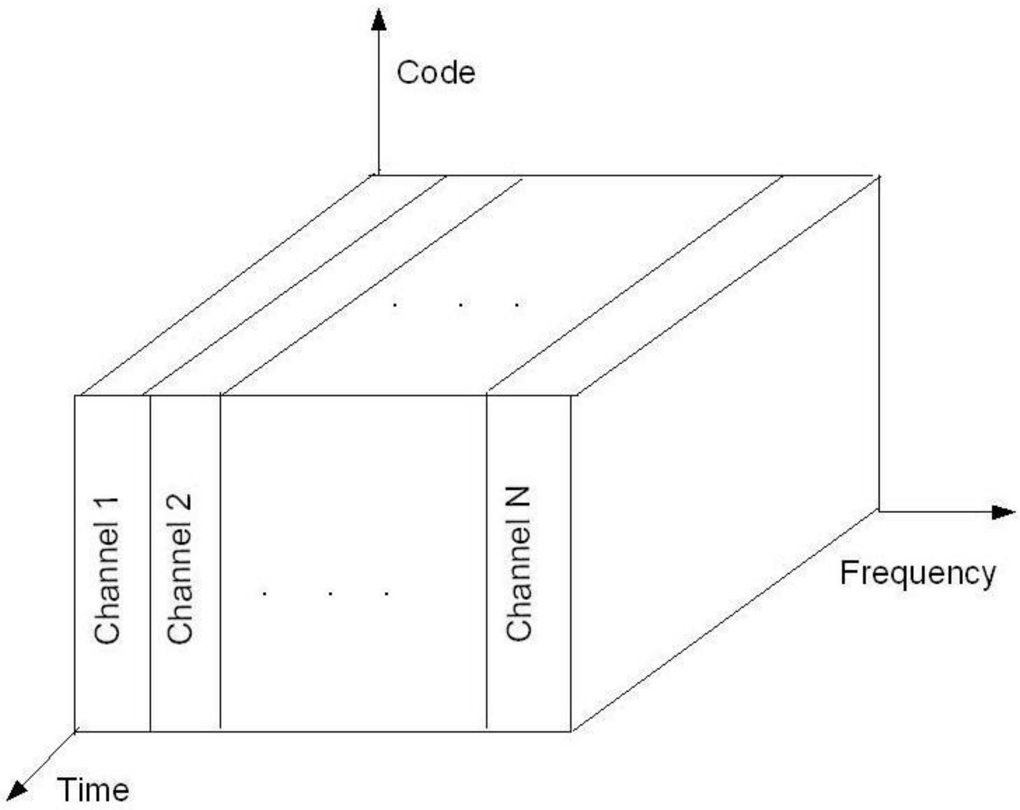


Figure 8.1: The basic concept of FDMA.

8.2 Frequency Division Multiple Access

This was the initial multiple-access technique for cellular systems in which each individual user is assigned a pair of frequencies while making or receiving a call as shown in Figure 8.1. One frequency is used for downlink and one pair for uplink. This is called frequency division duplexing (FDD). That allocated frequency pair is not used in the same cell or adjacent cells during the call so as to reduce the co channel interference. Even though the user may not be talking, the spectrum cannot be reassigned as long as a call is in place. Different users can use the same frequency in the same cell except that they must transmit at different times.

The features of FDMA are as follows: The FDMA channel carries only one phone circuit at a time. If an FDMA channel is not in use, then it sits idle and it cannot be used by other users to increase share capacity. After the assignment of the voice channel the BS and the MS transmit simultaneously and continuously. The bandwidths of FDMA systems are generally narrow i.e. FDMA is usually implemented in a narrow band system The symbol time is large compared to the average delay spread. The complexity of the FDMA mobile systems is lower than that of TDMA mobile systems. FDMA requires tight filtering to minimize the adjacent channel interference.

FDMA/FDD in AMPS

The first U.S. analog cellular system, AMPS (Advanced Mobile Phone System) is based on FDMA/FDD. A single user occupies a single channel while the call is in progress, and the single channel is actually two simplex channels which are frequency duplexed with a 45 MHz split. When a call is completed or when a handoff occurs the channel is vacated so that another mobile subscriber may use it. Multiple or simultaneous users are accommodated in AMPS by giving each user a unique signal. Voice signals are sent on the forward channel from the base station to the mobile unit, and on the reverse channel from the mobile unit to the base station. In AMPS, analog narrowband frequency modulation (NBFM) is used to modulate the carrier.

FDMA/TDD in CT2

Using FDMA, CT2 system splits the available bandwidth into radio channels in the assigned frequency domain. In the initial call setup, the handset scans the available channels and locks on to an unoccupied channel for the duration of the call. Using TDD(Time Division Duplexing), the call is split into time blocks that alternate between transmitting and receiving.

FDMA and Near-Far Problem

The near-far problem is one of detecting or filtering out a weaker signal amongst stronger signals. The near-far problem is particularly difficult in CDMA systems where transmitters share transmission frequencies and transmission time. In contrast, FDMA and TDMA systems are less vulnerable. FDMA systems offer different kinds of solutions to near-far challenge. Here, the worst case to consider is recovery of a weak signal in a frequency slot next to strong signal. Since both signals are present simultaneously as a composite at the input of a gain stage, the gain is set according to the level of the stronger signal; the weak signal could be lost in the noise floor. Even if subsequent stages have a low enough noise floor to provide

8.3 Time Division Multiple Access

In digital systems, continuous transmission is not required because users do not use the allotted bandwidth all the time. In such cases, TDMA is a complimentary access technique to FDMA. Global Systems for Mobile communications (GSM) uses the TDMA technique. In TDMA, the entire bandwidth is available to the user but only for a finite period of time. In most cases the available bandwidth is divided into fewer channels compared to FDMA and the users are allotted time slots during which they have the entire channel bandwidth at their disposal, as shown in Figure 8.2.

TDMA requires careful time synchronization since users share the bandwidth in the frequency domain. The number of channels are less, inter channel interference is almost negligible. TDMA uses different time slots for transmission and reception. This type of duplexing is referred to as Time division duplexing(TDD).

The features of TDMA includes the following: TDMA shares a single carrier fre- quency with several users where each users makes use of non overlapping time slots. The number of time slots per frame depends on several factors such as modulation technique, available bandwidth etc. Data transmission in TDMA is not continuous but occurs in bursts. This results in low battery consumption since the subscriber transmitter can be turned OFF when not in use. Because of a discontinuous trans- mission in TDMA the handoff process is much simpler for a subscriber unit, since it is able to listen to other base stations during idle time slots. TDMA uses different time slots for transmission and reception thus duplexers are not required. TDMA has an advantage that is possible to allocate different numbers of time slots per frame to different users. Thus bandwidth can be supplied on demand to different users by concatenating or reassigning time slot based on priority.

8.3.1 TDMA/FDD in GSM

As discussed earlier, GSM is widely used in Europe and other parts of the world. GSM uses a variation of TDMA along with FDD. GSM digitizes and compresses data, then sends it down a channel with two other streams of user data, each in its

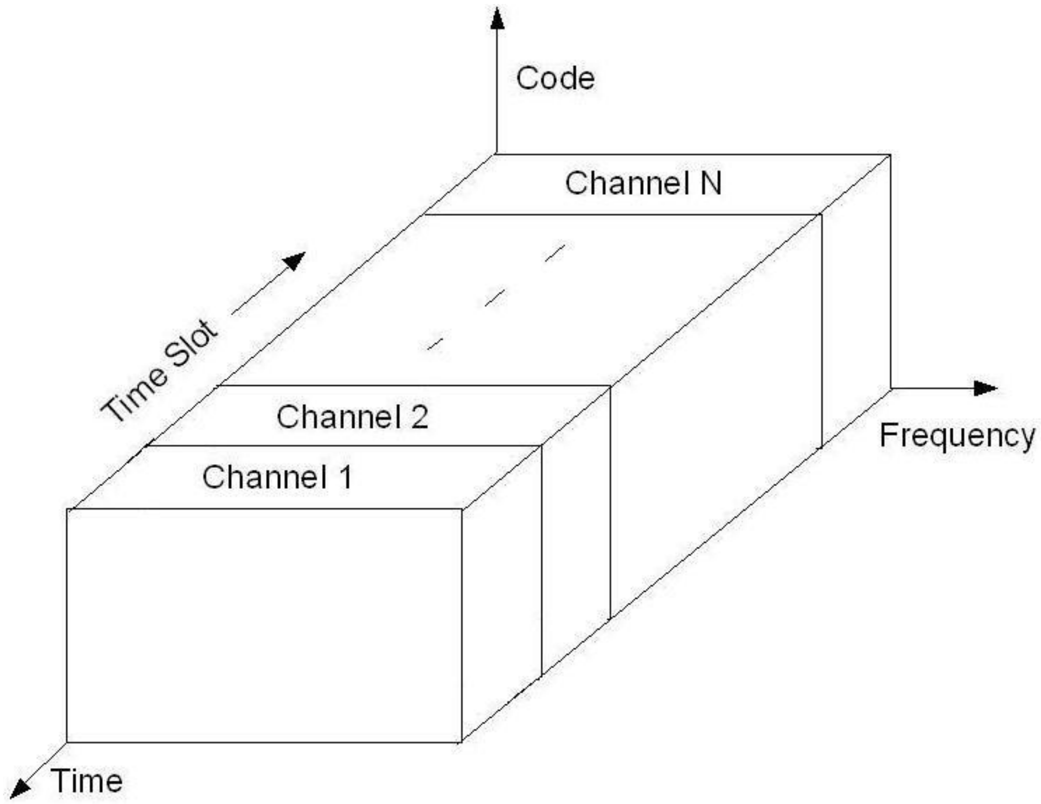


Figure 8.2: The basic concept of TDMA.

own time slot. It operates at either the 900 MHz or 1800 MHz frequency band. Since many GSM network operators have roaming agreements with foreign operators, users can often continue to use their mobile phones when they travel to other countries.

8.3.2 TDMA/TDD in DECT

DECT is a pan European standard for the digitally enhanced cordless telephony using TDMA/TDD. DECT provides 10 FDM channels in the band 1880-1990 Mhz. Each channel supports 12 users through TDMA for a total system load of 120 users. DECT supports handover, users can roam over from cell to cell as long as they remain within the range of the system. DECT antenna can be equipped with optional spatial diversity to deal with multipath fading.

Spread Spectrum Multiple Access

Spread spectrum multiple access (SSMA) uses signals which have a transmission bandwidth whose magnitude is greater than the minimum required RF bandwidth. A pseudo noise (PN) sequence converts a narrowband signal to a wideband noise like signal before transmission. SSMA is not very bandwidth efficient when used by a single user. However since many users can share the same spread spectrum bandwidth without interfering with one another, spread spectrum systems become bandwidth efficient in a multiple user environment.

There are two main types of spread spectrum multiple access techniques: Frequency hopped multiple access (FHMA)
Direct sequence multiple access (DSMA)
or Code division multiple access (CDMA).

Frequency Hopped Multiple Access (FHMA)

This is a digital multiple access system in which the carrier frequencies of the individual users are varied in a pseudo random fashion within a wideband channel. The digital data is broken into uniform sized bursts which is then transmitted on different carrier frequencies.

Code Division Multiple Access

In CDMA, the same bandwidth is occupied by all the users, however they are all assigned separate codes, which differentiates them from each other (shown in Figure 8.3). CDMA utilize a spread spectrum technique in which a spreading signal (which is uncorrelated to the signal and has a large bandwidth) is used to spread the narrow band message signal.

Direct Sequence Spread Spectrum (DS-SS)

This is the most commonly used technology for CDMA. In DS-SS, the message signal is multiplied by a Pseudo Random Noise Code. Each user is given his own codeword which is orthogonal to the codes of other users and in order to detect the user, the receiver must know the codeword used by the transmitter. There are, however, two problems in such systems which are discussed in the sequel.

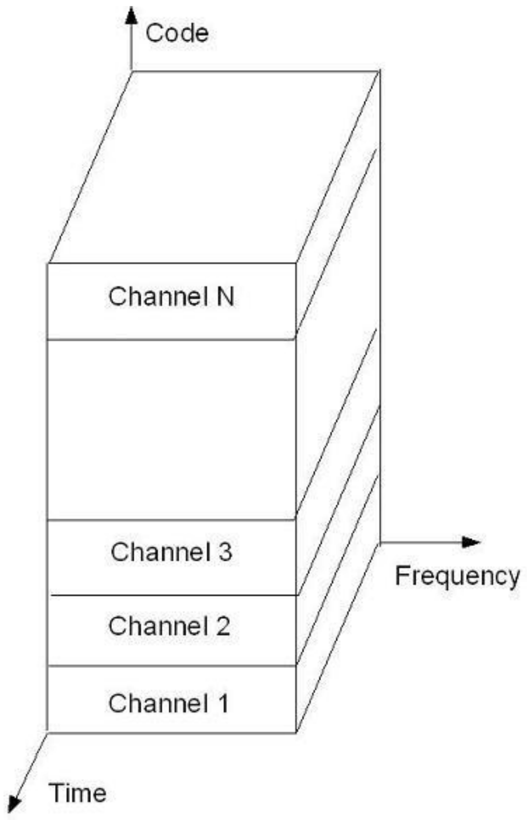


Figure 8.3: The basic concept of CDMA.

CDMA/FDD in IS-95

In this standard, the frequency range is: 869-894 MHz (for Rx) and 824-849 MHz (for Tx). In such a system, there are a total of 20 channels and 798 users per channel. For each channel, the bit rate is 1.2288 Mbps. For orthogonality, it usually combines 64 Walsh-Hadamard codes and a m-sequence.

8.4.3 CDMA and Self-interference Problem

In CDMA, self-interference arises from the presence of delayed replicas of signal due to multipath. The delays cause the spreading sequences of the different users to lose their orthogonality, as by design they are orthogonal only at zero phase offset. Hence in despreading a given user's waveform, nonzero contributions to that user's signal arise from the transmissions of the other users in the network. This is distinct from both TDMA and FDMA, wherein for reasonable time or frequency guardbands, respectively, orthogonality of the received signals can be preserved.

CDMA and Near-Far Problem

The near-far problem is a serious one in CDMA. This problem arises from the fact that signals closer to the receiver of interest are received with smaller attenuation than are signals located further away. Therefore the strong signal from the nearby transmitter will mask the weak signal from the remote transmitter. In TDMA and FDMA, this is not a problem since mutual interference can be filtered. In CDMA, however, the near-far effect combined with imperfect orthogonality between codes (e.g. due to different time sifts), leads to substantial interference. Accurate and fast power control appears essential to ensure reliable operation of multiuser DS-CDMA systems.

Hybrid Spread Spectrum Techniques

The hybrid combinations of FHMA, CDMA and SSMA result in hybrid spread spectrum techniques that provide certain advantages. These hybrid techniques are explained below,

Hybrid FDMA/CDMA (FCDMA):

An alternative to the CDMA technique in which the available wideband spectrum is divided into a smaller number of sub spectra with smaller bandwidths. The smaller sub channels become narrow band CDMA systems with processing gain lower than the original CDMA system. In this scheme the required bandwidth need not be contiguous and different user can be allotted different sub spectrum bandwidths depending on their requirements. The capacity of this hybrid FCDMA technique is given by the sum of the capacities of a system operating in the sub spectra.

Hybrid Direct Sequence/Frequency Hopped Multiple Access Techniques (DS/FHMA):

A direct sequence modulated signal whose center frequency is made to hop periodically in a pseudo random fashion is used in this technique. One of the advantages using this technique is they avoid near-far effect. However, frequency hopped CDMA systems are not adaptable to the soft handoff process since it is difficult to synchronize the frequency hopped base station receiver to the multiple hopped signals. Time and Code Division Multiple Access (TCDMA):

In this TCDMA method different cells are allocated different spreading codes. In each cell, only one user per cell is allotted a particular time slot. Thus at any time only one user is transmitting in each cell. When a handoff takes place the spreading code of that user is changed to the code of the new cell. TCDMA also avoids near-far effect as the number of users transmitting per cell is one.

Time Division Frequency Hopping (TDFH):

This technique has been adopted for the GSM standard, where the hopping sequence is predefined and the subscriber is allowed to hop only on certain frequencies which are assigned to a cell. The subscriber can hop to a new frequency at the start of a new TDMA frame, thus avoiding a severe fade or erasure event on a particular channel. This technique has the advantage in severe multipath or when severe channel interference occurs.

Space Division Multiple Access

SDMA utilizes the spatial separation of the users in order to optimize the use of the frequency spectrum. A primitive form of SDMA is when the same frequency is reused in different cells in a cellular wireless network. The radiated power of each user is controlled by Space division multiple access. SDMA serves different users by using spot beam antenna. These areas may be served by the same frequency or different frequencies. However for limited co-channel interference it is required that the cells be sufficiently separated. This limits the number of cells a region can be divided into and hence limits the frequency re-use factor. A more advanced approach can further increase the capacity of the network. This technique would enable frequency re-use within the cell. In a practical cellular environment it is improbable to have just one transmitter fall within the receiver beam width. Therefore it becomes imperative to use other multiple access techniques in conjunction with SDMA. When different areas are covered by the antenna beam, frequency can be re-used, in which case TDMA or CDMA is employed, for different frequencies FDMA can be used.

Conclusion

In this chapter, we have mainly discussed the fixed assignment type of MA techniques, namely, FDMA, TDMA and CDMA. We have, however, intensionally not covered the reservation-based MA schemes such as packet reservation MA or polling or token passing etc. The main idea to discuss only the basic MA techniques has been to grow up a fair idea about the resource sharing in a wireless media when there are many users, keeping the QoS view point in mind.