श्रीचन्द्रशेखरेन्द्रसरस्वतीविश्वमहाविद्यालयः

SRI CHANDRASEKHARENDRA SARASWATHI VISWA MAHAVIDYALAYA

Deemed to be University u/s 3 of UGC Act 1956 | Accredited with "A" grade by NAAC

Enathur, Kanchipuram - 631 561. Tamilnadu, India | www.kanchiuniv.ac.in

Sponsored and run by Sri Kanchi Kamakoti Peetam Charitable Trust

STUDY MATERIAL
of
MOBILE COMMUNICATION AND NETWORKS

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

Prepared by: Dr.M.Vinoth
Assistant Professor

UNIT-I

Pre-requisite:

- To understand the issues involved in mobile communication system design and analysis.

OUTCOMES:

- Analyze the basic concepts of mobile communication system design.

INTRODUCTION

The cellular concept was a major breakthrough in solving the problem of spectral congestion and user capacity. It offered very high capacity in a limited spectrum allocation without any major technological changes. The cellular concept is a system-level idea which calls for replacing a single, high power transmitter (large cell) with many low power transmitters (small cells), each providing coverage to only a small portion of the service area. Each base station is allocated a portion of the total number of channels available to the entire system, and nearby base stations are assigned different groups of channels so that all the available channels are assigned to a relatively small number of neighboring base stations. Neighboring base stations are assigned different groups of channels so that the interference between base stations (and the mobile users under their control) is minimized. By systematically spacing base stations and their channel groups throughout a market, the available channels are distributed throughout the geographic region and may be reused as many times as necessary so long as the interference between co channel stations is kept below acceptable levels. As the demand for service increases (i.e., as more channels are needed within a particular market), the number of base stations may be increased (along with a corresponding decrease in transmitter power to avoid added interference), thereby providing additional radio capacity with no additional increase in radio spectrum. This fundamental principle is the foundation for all modern wireless communication systems, since it enables a fixed number of channels to serve an arbitrarily large number of subscribers by reusing the channels throughout the coverage region. Furthermore, the cellular concept allows every piece of subscriber equipment within a country or continent to be manufactured with the same set of channels so that any mobile may be used anywhere within the region.

Frequency Reuse : -

Cellular radio systems rely on an intelligent allocation and reuse of channels throughout a coverage region [Oet83]. Each cellular base station is allocated a group of radio channels to be used within a small geographic area called a cell. Base stations in adjacent cells are assigned channel groups which contain completely different channels than neighboring cells. The base station antennas are designed to achieve the desired coverage within the particular cell. By limiting the coverage area to within the boundaries of a cell, the same group of channels may be used to cover different cells that are separated from one another by distances large enough to keep interference levels within tolerable limits. The design process of selecting and allocating channel groups for all of the cellular base stations within a system is called frequency reuse or frequency planning [Mac79]. The concept of cellular frequency reuse, where cells labeled with the same letter use the same group of channels. The frequency reuse plan is overlaid upon a map to indicate where different frequency channels are used. The hexagonal cell shape is conceptual and is a simplistic model of the radio coverage for each base station, but it has been universally adopted since the hexagon permits easy and manageable analysis of a cellular system.

The actual radio coverage of a cell is known as the footprint and is determined from field measurements or propagation prediction models. Although the real footprint is amorphous in nature, a regular cell shape is needed for systematic system design and adaptation for future growth. While it might seem natural to choose a circle to represent the coverage area of a base station, adjacent circles cannot be overlaid upon a map without leaving gaps or creating overlapping regions. Thus, when considering geometric shapes which cover an entire region without overlap and with equal area, there are three sensible choices a square, an equilateral triangle, and a hexagon. A cell must be designed to serve the weakest mobiles within the footprint, and these are typically located at the edge of the cell. For a given distance between the center of a polygon and its farthest perimeter points, the hexagon has the largest area of the three. Thus, by using the hexagon geometry, the fewest number of cells can cover a geographic region, and the hexagon closely approximates a circular radiation pattern which would occur for an omnidirectional base

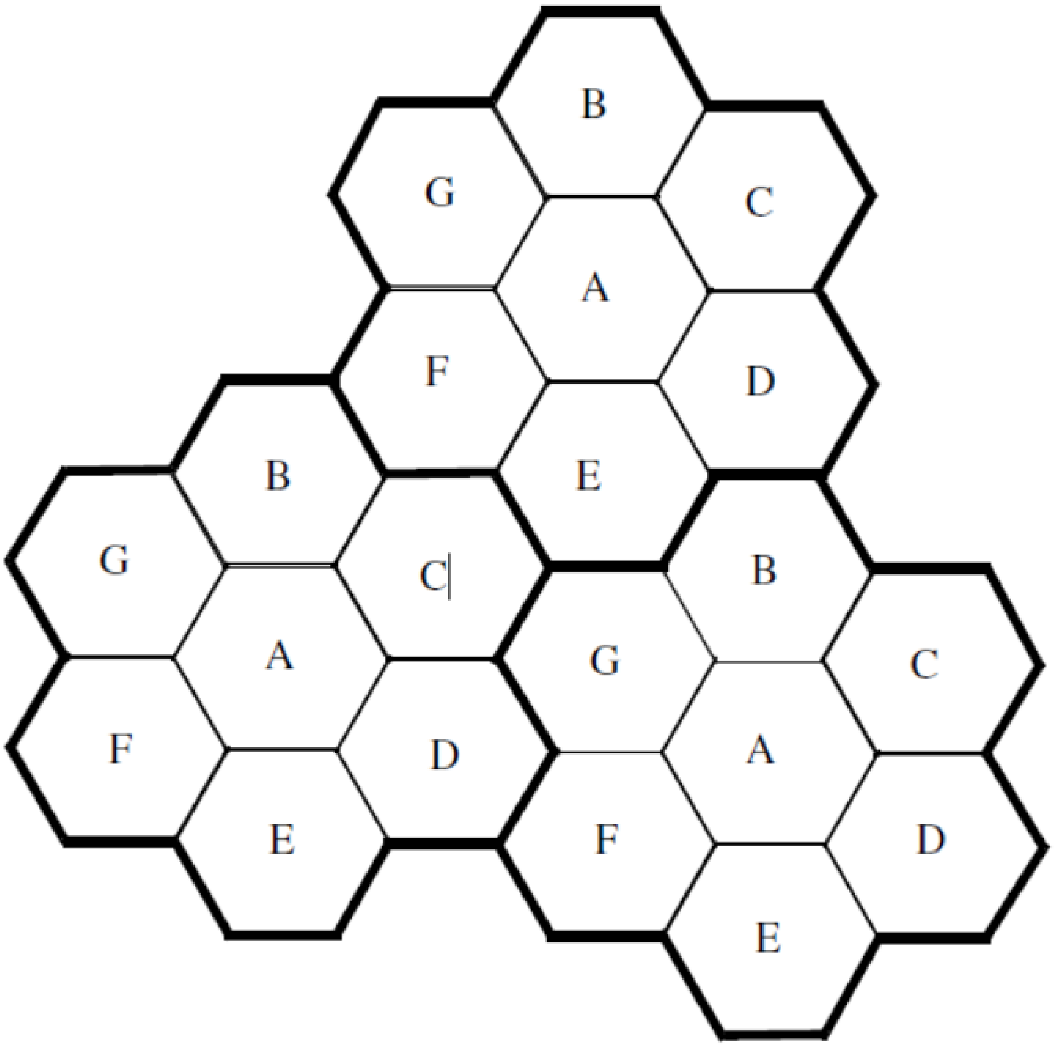


Illustration of the cellular frequency reuse concept. Cells with the same letter use the same set of frequencies. A cell cluster is outlined in bold and replicated over the coverage area. In this example, the cluster size, N, is equal to seven, and the frequency reuse factor is 1/7 since each cell contains one-seventh of the total number of available channels.

station antenna and free space propagation. Of course, the actual cellular footprint is determined by the contour in which a given transmitter serves the mobiles successfully.

When using hexagons to model coverage areas, base station transmitters are depicted as either being in the center of the cell (center-excited cells) or on three of the six cell vertices (edge-excited cells). Normally, omnidirectional antennas are used in center-excited cells and sectored directional antennas are used in corner-excited cells. Practical considerations usually do not allow base stations to be placed exactly as they appear in the hexagonal layout. Most system designs permit a base station to be positioned up to one-fourth the cell radius away from the ideal location.

To understand the frequency reuse concept, consider a cellular system which has a total of S duplex channels available for use. If each cell is allocated a group of k channels ($k < S$), and if the S channels are divided among N cells into unique and disjoint channel groups which each have the same number of channels, the total number of available radio channels can be expressed as

$$S = Kn$$

The N cells which collectively use the complete set of available frequencies is called a cluster. If a cluster is replicated M times within the system, the total number of duplex channels, C, can be used as a measure of capacity and is given by

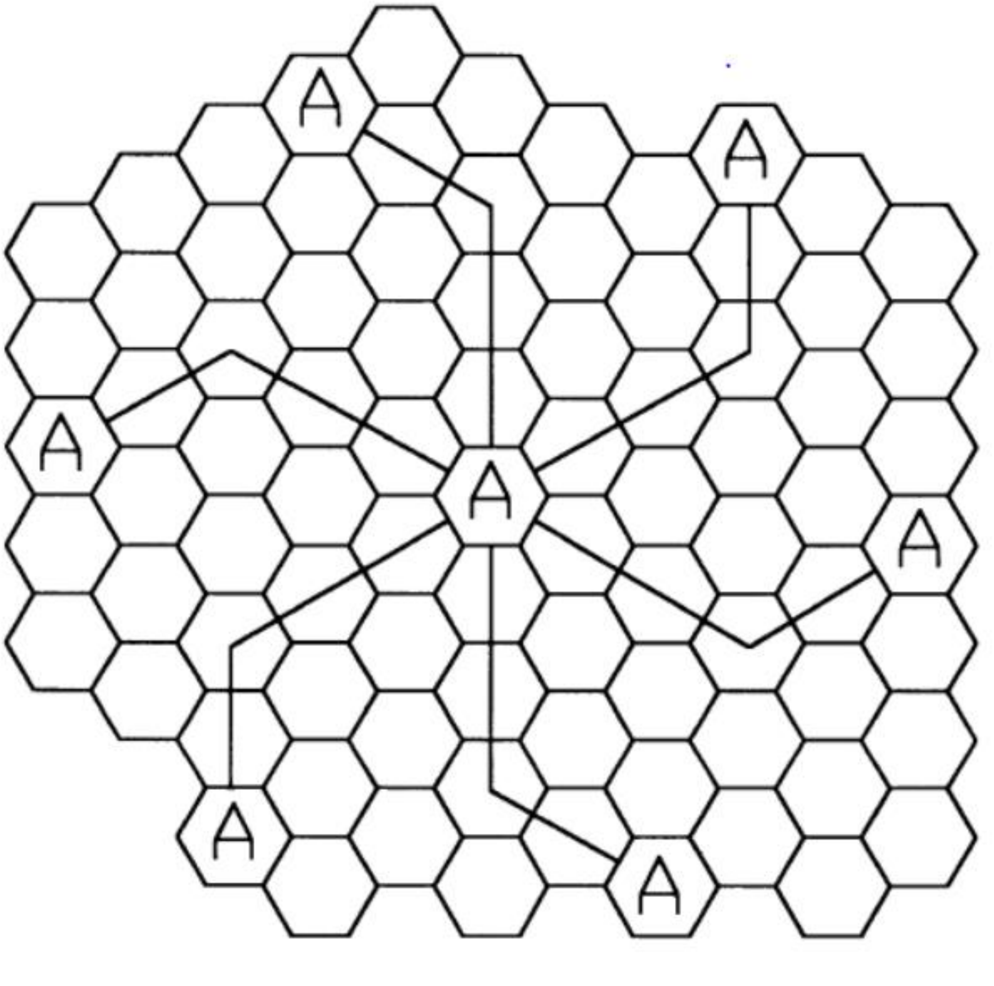
$$C = MkN = MS$$

the capacity of a cellular system is directly proportional to the number of times a cluster is replicated in a fixed service area. The factor N is called the cluster size and is typically equal to 4, 7, or 12. If the cluster size N is reduced while the cell size is kept constant, more clusters are required to cover a given area, and hence more capacity (a larger value of C) is achieved. A large cluster size indicates that the ratio between the cell radius and the distance between co-channel cells is small. Conversely, a small cluster size indicates that co-channel cells are located much closer together. The value for N is a function of how much interference a mobile or base station can tolerate while maintaining a sufficient quality of communications. From a design viewpoint, the smallest possible value of N is desirable in order to maximize capacity over a given coverage area (i.e., to maximize C . The frequency reuse factor of a cellular system is given by 1/N, since each cell within a cluster is only assigned 1/N of the total available channels in the system.

Due to the fact that the hexagonal geometry of has exactly six equidistant neighbors and that the lines joining the centers of any cell and each of its neighbors are separated by multiples of 60 degrees, there are only certain cluster sizes and cell layouts which are possible [Mac79]. In order to tessellate—to connect without gaps between adjacent cells—the geometry of hexagons is such that the number of cells per cluster, N, can only have values

$$N = i^2 + ij + j^2$$

where i and j are non-negative integers. To find the nearest co-channel neighbors of a particular cell, one must do the following: (1) move i cells along any chain of hexagons and then (2) turn 60 degrees counter-clockwise and move j cells. This is illustrated in Figure 3.2 for i = 3 and j = 2 (example, N = 19).



Method of locating co-channel cells in a cellular system. In this example, N = 19 (i.e., i = 3, j = 2).

Channel Assignment Strategies

For efficient utilization of the radio spectrum, a frequency reuse scheme that is consistent with the objectives of increasing capacity and minimizing interference is required. A variety of channel assignment strategies have been developed to achieve these objectives. Channel assignment strategies can be classified as either fixed or dynamic. In a fixed channel assignment strategy, each cell is allocated a predetermined set of voice channels. Any call attempt within the cell can only be served by the unused channels in that particular cell. If all the channels in that cell are occupied, the call is blocked and the subscriber does not receive service. Several variations of the fixed assignment strategy exist. In one approach, called the borrowing strategy, a cell is allowed to borrow channels from a neighboring cell if all of its own channels are already occupied. The mobile switching center (MSC) supervises such borrowing procedures and ensures that the borrowing of a channel does not disrupt or interfere with any of the calls in progress in the donor cell.

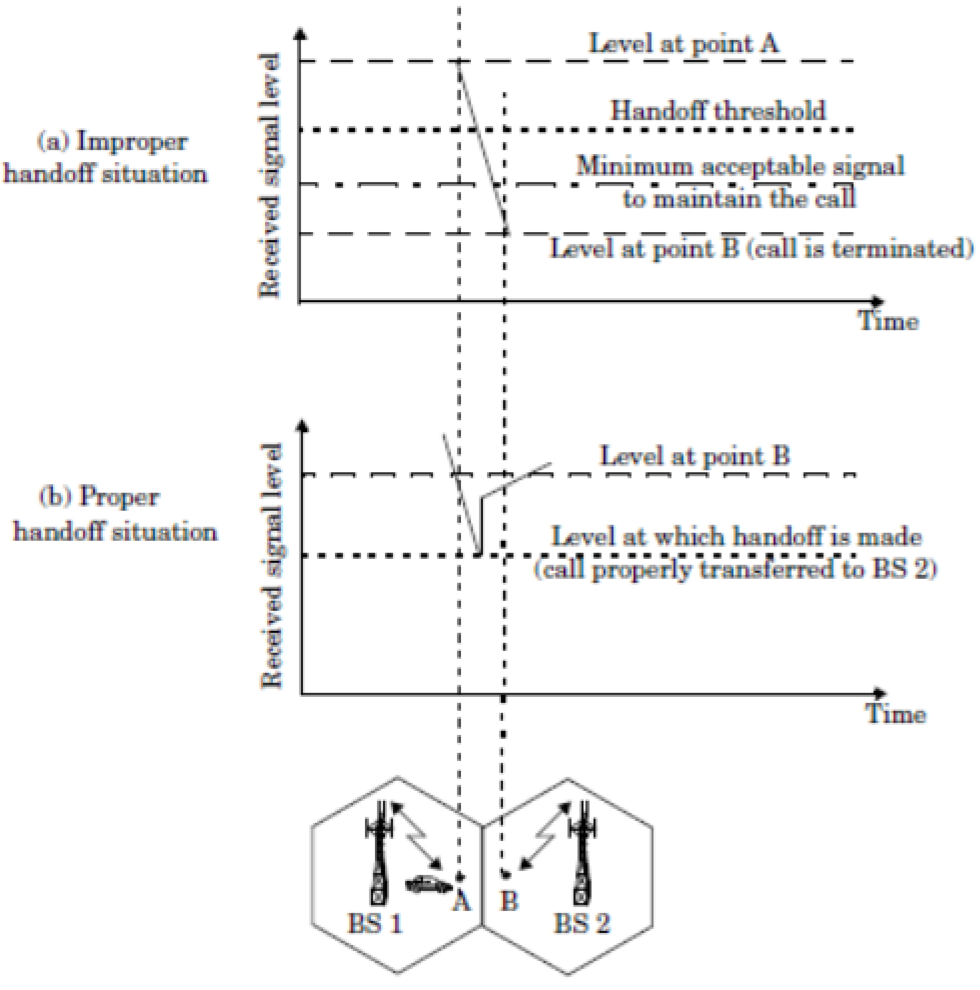
In a dynamic channel assignment strategy, voice channels are not allocated to different cells permanently. Instead, each time a call request is made, the serving base station requests a channel from the MSC. The switch then allocates a channel to the requested cell following an algorithm that takes into account the likelihood of future blocking within the cell, the frequency of use of the candidate channel, the reuse distance of the channel, and other cost functions. Accordingly, the MSC only allocates a given frequency if that frequency is not presently in use in the cell or any other cell which falls within the minimum restricted distance of frequency reuse to avoid co-channel interference. Dynamic channel assignment reduce the likelihood of blocking, which increases the trunking capacity of the system, since all the available channels in a market are accessible to all of the cells. Dynamic channel assignment strategies require the MSC to collect real-time data on channel occupancy, traffic distribution, and radio signal strength indications (RSSI) of all channels on a continuous basis. This increases the storage and computational load on the system but provides the advantage of increased channel utilization and decreased probability of a blocked call.

Handoff Strategies:-

When a mobile moves into a different cell while a conversation is in progress, the MSC automatically transfers the call to a new channel belonging to the new base station. This handoff operation not only involves identifying a new base station, but also requires that the voice and control signals be allocated to channels associated with the new base station.

Processing handoffs is an important task in any cellular radio system. Many handoff strategies prioritize handoff requests over call initiation requests when allocating unused channels in a cell site. Handoffs must be performed successfully and as infrequently as possible, and be imperceptible to the users. In order to meet these requirements, system designers must specify an optimum signal level at which to initiate a handoff. Once a particular signal level is specified as the minimum usable signal for acceptable voice quality at the base station receiver (normally taken as between –90 dBm and –100 dBm), a slightly stronger signal level is used as a threshold at which a handoff is made.

This margin, given by Δ = Pr handoff – Pr minimum usable, cannot be too large or too small. If Δ is too large, unnecessary handoffs which burden the MSC may occur, and if Δ is too small, there may be insufficient time to complete a handoff before a call is lost due to weak signal conditions. Therefore, Δ is chosen carefully to meet these conflicting requirements, a handoff situation. Demonstrates the case where a handoff is not made and the signal drops below the minimum acceptable level to keep the channel active. This dropped call event can happen when there is an excessive delay by the MSC in assigning a handoff or when the threshold Δ is set too small for the handoff time in the system. Excessive delays may occur during high traffic conditions due to computational loading at the MSC or due to the fact that no channels are available on any of the nearby base stations (thus forcing the MSC to wait until a channel in a nearby cell becomes free).



In deciding when to handoff, it is important to ensure that the drop in the measured signal level is not due to momentary fading and that the mobile is actually moving away from the serving base station. In order to ensure this, the base station monitors the signal level for a certain period of time before a handoff is initiated. This running average measurement of signal strength should be optimized so that unnecessary handoffs are avoided, while ensuring that necessary handoffs are completed before a call is terminated due to poor signal level. The length of time needed to decide if a handoff is necessary depends on the speed at which the vehicle is moving. If the slope of the short-term average received signal level in a given time interval is steep, the handoff should be made quickly. Information about the vehicle speed, which can be useful in handoff decisions, can also be computed from the statistics of the received short-term fading signal at the base station.

The time over which a call may be maintained within a cell, without handoff, is called the dwell time . The dwell time of a particular user is governed by a number of factors, including propagation, interference, distance between the subscriber and the base station, and other time varying effects. Chapter 5 shows that even when a mobile user is stationary, ambient motion in the vicinity of the base station and the mobile can produce fading; thus, even a stationary subscriber may have a random and finite dwell time. Analysis in indicates that the statistics of dwell time vary greatly, depending on the speed of the user and the type of radio coverage. For example, in mature cells which provide coverage for vehicular highway users, most users tend to have a relatively constant speed and travel along fixed and well-defined paths with good radio coverage. In such instances, the dwell time for an arbitrary user is a random variable with a distribution that is highly concentrated about the mean dwell time. On the other hand, for users in dense, cluttered microcell environments, there is typically a large variation of dwell time about the mean, and the dwell times are typically shorter than the cell geometry would otherwise suggest. It is apparent that the statistics of dwell time are important in the practical design of handoff algorithms

In first generation analog cellular systems, signal strength measurements are made by the base stations and supervised by the MSC. Each base station constantly monitors the signal strengths of all of its reverse voice channels to determine the relative location of each mobile user with respect to the base station tower.

In addition to measuring the RSSI of calls in progress within the cell, a spare receiver in each base station, called the locator receiver, is used to scan and determine signal strengths of mobile users which are in neighboring cells. The locator receiver is controlled by the MSC and is used to monitor the signal strength of users in neighboring cells which appear to be in need of handoff and reports all RSSI values to the MSC. Based on the locator receiver signal strength information from each base station, the MSC decides if a handoff is necessary or not. In today's second generation systems, handoff decisions are mobile assisted. In mobile assisted handoff (MAHO), every mobile station measures the received power from surrounding base stations and continually reports the results of these measurements to the serving base station. A handoff is initiated when the power received from the base station of a neighboring cell begins to exceed the power received from the current base station by a certain level or for a certain period of time. The MAHO method enables the call to be handed over between base stations at a much faster rate than in first generation analog systems since the handoff measurements are made by each mobile, and the MSC no longer constantly monitors signal strengths. MAHO is particularly suited for microcellular environments where handoffs are more frequent.

During the course of a call, if a mobile moves from one cellular system to a different cellular system controlled by a different MSC, an intersystem handoff becomes necessary. An MSC engages in an intersystem handoff when a mobile signal becomes weak in a given cell and the MSC cannot find another cell within its system to which it can transfer the call in progress. There are many issues that must be addressed when implementing an intersystem handoff. For instance, a local call may become a long-distance call as the mobile moves out of its home system and becomes a roamer in a neighboring system. Also, compatibility between the two MSCs must be determined before implementing an intersystem handoff. Different systems have different policies and methods for managing handoff requests. Some systems handle handoff requests in the same way they handle originating calls. In such systems, the probability that a handoff request will not be served by a new base station is equal to the blocking probability of incoming calls. However, from the user's point of view, having a call abruptly terminated while in the middle of a conversation is more annoying than being blocked occasionally on a new call attempt. To improve the quality of service as perceived by the users, various methods have been devised to prioritize handoff requests over call initiation requests when allocating voice channels.

Interference and System Capacity

Interference is the major limiting factor in the performance of cellular radio systems. Sources of interference include another mobile in the same cell, a call in progress in a neighboring cell, other base stations operating in the same frequency band, or any non-cellular system which inadvertently leaks energy into the cellular frequency band. Interference on voice channels causes cross talk, where the subscriber hears interference in the background due to an undesired transmission. On control channels, interference leads to missed and blocked calls due to errors in the digital signaling. Interference is more severe in urban areas, due to the greater RF noise floor and the large number of base stations and mobiles. Interference has been recognized as a major bottleneck in increasing capacity and is often responsible for dropped calls. The two major types of system-generated Cellular interference are co-channel interference and adjacent channel interference. Even though interfering signals are often generated within the cellular system, they are difficult to control in practice (due to random propagation effects). Even more difficult to control is interference due to out-of-band users, which arises without warning due to front end overload of subscriber equipment or intermittent intermodulation products. In practice, the transmitters from competing cellular carriers are often a significant source of out-of-band interference, since competitors often locate their base stations in close proximity to one another in order to provide comparable coverage to customers.

Co-channel Interference and System Capacity

Frequency reuse implies that in a given coverage area there are several cells that use the same set of frequencies. These cells are called co-channel cells, and the interference between signals from these cells is called co-channel interference. Unlike thermal noise which can be overcome by increasing the signal-to-noise ratio (SNR), co-channel interference cannot be combated by simply increasing the carrier power of a transmitter. This is because an increase in carrier transmit power increases the interference to neighboring co-channel cells. To reduce co-channel interference, co-channel cells must be physically separated by a minimum distance to provide sufficient isolation due to propagation.

When the size of each cell is approximately the same and the base stations transmit the same power, the co-channel interference ratio is independent of the transmitted power and becomes a function of the radius of the cell (R) and the distance between centers of the nearest co-channel cells (D). By increasing the ratio of D/R, the spatial separation between co-channel cells relative to the coverage distance of a cell is increased. Thus, interference is reduced from improved isolation of RF energy from the co-channel cell. The parameter Q, called the co-channel reuse ratio, is related to the cluster size . For a hexagonal geometry

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

A small value of *Q* provides larger capacity since the cluster size *N* is small, whereas a large value of *Q* improves the transmission quality, due to a smaller level of co-channel interference. A trade-off must be made between these two objectives in actual cellular design.

Channel Planning for Wireless Systems

Judiciously assigning the appropriate radio channels to each base station is an important process that is much more difficult in practice than in theory. While Equation is a valuable rule of thumb for determining the appropriate frequency reuse ratio (or cluster size) and the appropriate separation between adjacent co-channel cells, the wireless engineer must deal with the real-world difficulties of radio propagation and imperfect coverage regions of each cell. Cellular systems, in practice, seldom obey the homogenous propagation path loss assumption of Equation Generally, the available mobile radio spectrum is divided into channels, which are part of an air interface standard that is used throughout a country or continent. These channels generally are made up of control channels (vital for initiating, requesting, or paging a call), and voice channels (dedicated to carrying revenue-generating traffic). Typically, about 5% of the entire mobile spectrum is devoted to control channels, which carry data messages that are very brief and bursty in nature, while the remaining 95% of the spectrum is dedicated to voice channels. Channels may be assigned by the wireless carrier in any manner it chooses, since each market may have its own particular propagation conditions or particular services it wishes to offer and may wish to adopt its own particular frequency reuse scheme that fits its geographic conditions or air interface technology choice. However, in practical systems, the air interface standard ensures a distinction between voice and control channels, and thus control channels are generally not allowed to be used as voice channels, and vice versa. Furthermore, since control channels are vital in the successful launch of any call, the frequency reuse strategy applied to control channels is different and generally more conservative (e.g., is afforded greater S/I protection) than for the voice channels. This can be seen in Example 3.3, where the control channels are allocated using 21-cell reuse, whereas voice channels are assigned using seven-cell reuse. Typically, the control channels are able to handle a great deal of data such that only one control channel is needed within a cell. As described in Section 3.7.2, sectoring is often used to improve the signal-to-interference ratio which may lead to a smaller cluster size, and in such cases, only a single control channel is assigned to an individual sector of a cell.

Adjacent Channel Interference

Interference resulting from signals which are adjacent in frequency to the desired signal is called *adjacent channel interference*. Adjacent channel interference results from imperfect receiver filters which allow nearby frequencies to leak into the passband. The problem can be particularly serious if an adjacent channel user is transmitting in very close range to a subscriber's receiver, while the receiver attempts to receive a base station on the desired channel. This is referred to as the *near–far* effect, where a nearby transmitter (which may or may not be of the same type as that used by the cellular system) captures the receiver of the subscriber. Alternatively, the near– far effect occurs when a mobile close to a base station transmits on a channel close to one being used by a weak mobile. The base station may have difficulty in discriminating the desired mobile user from the "bleed over" caused by the close adjacent channel mobile.

Adjacent channel interference can be minimized through careful filtering and channel assignments. Since each cell is given only a fraction of the available channels, a cell need not be assigned channels which are all adjacent in frequency. By keeping the frequency separation between each channel in a given cell as large as possible, the adjacent channel interference may be reduced considerably.

Thus instead of assigning channels which form a contiguous band of frequencies within a particular cell, channels are allocated such that the frequency separation between channels in a given cell is maximized. By sequentially assigning successive channels in the frequency band to different cells, many channel allocation schemes are able to separate adjacent channels in a cell by as many as *N* channel bandwidths, where *N* is the cluster size.

Some channel allocation schemes also prevent a secondary source of adjacent channel interference by avoiding the use of adjacent channels in neighboring cell sites. If the frequency reuse factor is large (e.g., small *N*), the separation between adjacent channels
at the base station may not be sufficient to keep the adjacent channel interference level within tolerable limits. For example, if a close-in mobile is 20 times as close to the base station as another mobile and has energy spill out of its passband, the signal-to-interference ratio at the base station for the weak mobile (before receiver filtering) is approximately

$$\frac{S}{I} = (20)^{-n}$$

For a path loss exponent $n = 4$, this is equal to –52 dB. If the intermediate frequency (IF) filter of the base station receiver has a slope of 20 dB/octave, then an adjacent channel interferer must be displaced by at least six times the passband bandwidth from the center of the receiver frequency passband to achieve 52 dB attenuation. Here, a separation of approximately six channel bandwidths is required for typical filters in order to provide 0 dB SIR from a close-in adjacent channel user. This implies more than six channel separations are needed to bring the adjacent channel interference to an acceptable level. Tight base station filters are needed when close-in and distant users share the same cell. In practice, base station receivers are preceded by a high Q cavity filter in order to reject adjacent channel interference.

Power Control for Reducing Interference

In practical cellular radio and personal communication systems, the power levels transmitted by every subscriber unit are under constant control by the serving base stations. This is done to ensure that each mobile transmits the smallest power necessary to maintain a good quality link on the reverse channel. Power control not only helps prolong battery life for the subscriber unit, but also dramatically reduces the reverse channel S/I in the system.

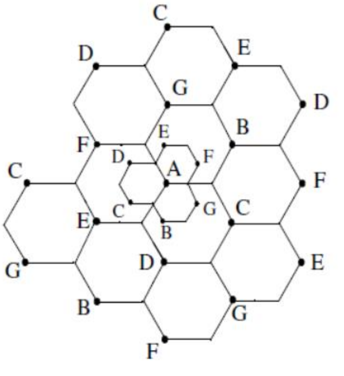
Improving Coverage and Capacity in Cellular Systems

As the demand for wireless service increases, the number of channels assigned to a cell eventually becomes insufficient to support the required number of users. At this point, cellular design techniques are needed to provide more channels per unit coverage area. Techniques such as cell splitting, sectoring, and coverage zone approaches are used in practice to expand the capacity of cellular systems. Cell splitting allows an orderly growth of the cellular system. Sectoring uses directional antennas to further control the interference and frequency reuse of channels. The zone microcell concept distributes the coverage of a cell and extends the cell boundary to hard to- reach places. While cell splitting increases the number of base stations in order to increase capacity, sectoring and zone microcells rely on base station antenna placements to improve capacity by reducing co-channel interference. Cell splitting and zone microcell techniques do not suffer the trunking inefficiencies experienced by sectored cells, and enable the base station to oversee all handoff chores related to the microcells, thus reducing the computational load at the MSC. These three popular capacity improvement techniques will be explained in detail.

Cell Splitting

Cell splitting is the process of subdividing a congested cell into smaller cells, each with its own base station and a corresponding reduction in antenna height and transmitter power. Cell splitting increases the capacity of a cellular system since it increases the number of times that channels are reused. By defining new cells which have a smaller radius than the original cells and by installing these smaller cells (called microcells) between the existing cells, capacity increases due to the additional number of channels per unit area.

Imagine if every cell in were reduced in such a way that the radius of every cell was cut in half. In order to cover the entire service area with smaller cells, approximately four times as many cells would be required. This can be easily shown by considering a circle with radius R. The area covered by such a circle is four times as large as the area covered by a circle with radius R/2. The increased number of cells would increase the number of clusters over the coverage region, which in turn would increase the number of channels, and thus capacity, in the coverage area. Cell splitting allows a system to grow by replacing large cells with smaller cells, while not upsetting the channel allocation scheme required to maintain the minimum co-channel reuse ratio Q between co-channel cells. An example of cell splitting is the base stations are placed at corners of the cells, and the area served by base station A is assumed to be saturated with traffic (i.e., the blocking of base station A exceeds acceptable rates). New base stations are therefore needed in the region to increase the number of channels in the area and to reduce the area served by the single base station. Note in the figure that the original base station A has been surrounded by six new microcells.



GENERATION

simply, the "G" stands for "GENERATION" . While you connected to internet, the speed of your internet is depends upon the signal strength that has been shown in alphabets like 2G, 3G, 4G etc. right next to the signal bar on your home screen. Each Generation is defined as a set of telephone network standards , which detail the technological implementation of a particular mobile phone system. The speed increases and the technology used to achieve that speed also changes. For eg, 1G offers 2.4 kbps, 2G offers 64 Kbps and is based on GSM, 3G offers 144 kbps-2 mbps whereas 4G offers 100 Mbps - 1 Gbps and is based on LTE technology .

The aim of wireless communication is to provide high quality, reliable communication just like wired communication(optical fibre) and each new generation of services represents a big step(a leap rather) in that direction. This evolution journey was started in 1979 from 1G and it is still continuing to 5G. Each of the Generations has standards that must be met to officially use the G terminology. There are institutions in charge of standardizing each generation of mobile technology. Each generation has requirements that specify things like throughput, delay, etc. that need to be met to be considered part of that generation. Each generation built upon the research and development which happened since the last generation. 1G was not used to identify wireless technology until 2G, or the second generation, was released. That was a major jump in the technology when the wireless networks went from analog to digital .

Features	1G	2G	3G	4G	5G
Start/Devlopment	1970/1984	1980/1999	1990/2002	2000/2010	2010/2015
Technology	AMPS, NMT, TACS	GSM	WCDMA	LTE, WiMax	MIMO, mm Waves
Frequency	30 KHz	1.8 Ghz	1.6 - 2 GHz	2 - 8 GHz	3 - 30 Ghz
Bandwidth	2 kbps	14.4 - 64 kbps	2 Mbps	2000 Mbps to 1 Gbps	1 Gbps and higher
AccessSystem	FDMA	TDMA/CDMA	CDMA	CDMA	OFDM/BDMA
Core Network	PSTN	PSTN	Packet Network	Internet	Internet

1G - First Generation

This was the first generation of cell phone technology . The very first generation of commercial cellular network was introduced in the late 70's with fully implemented standards being established throughout the 80's. It was introduced in 1987 by Telecom (known today as Telstra), Australia received its first cellular mobile phone network utilizing a 1G analog system. 1G is an analog technology and the phones generally had poor battery life and voice quality was large without much security, and would sometimes experience dropped calls . These are the analog telecommunications standards that were introduced in the 1980s and continued until being replaced by 2G digital telecommunications. The maximum speed of 1G is 2.4 Kbps

2G - Second Generation

Cell phones received their first major upgrade when they went from 1G to 2G. The main difference between the two mobile telephone systems (1G and 2G), is that the radio signals used by 1G network are analog, while 2G networks are digital . Main motive of this generation was to provide secure and reliable communication channel. It implemented the concept of CDMA and GSM . Provided small data service like sms and mms. Second generation 2G cellular telecom networks were commercially launched on the GSM standard in Finland by Radiolinja (now part of Elisa Oyj) in 1991. 2G capabilities are achieved by allowing multiple users on a single channel via multiplexing. During 2G Cellular phones are used for data also along with voice. The advance in technology from 1G to 2G introduced many of the fundamental services that we still use today, such as SMS, internal roaming , conference calls, call hold and billing based on services e.g. charges based on long distance calls and real time billing. The max speed of 2G with General Packet Radio Service (GPRS) is 50 Kbps or 1 Mbps with Enhanced Data Rates for GSM Evolution (EDGE). Before making the major leap from 2G to 3G wireless networks, the lesser-known 2.5G and 2.75G was an interim standard that bridged the gap.

3G - Third Generation

This generation set the standards for most of the wireless technology we have come to know and love. Web browsing, email, video downloading, picture sharing and other Smartphone technology were introduced in the third generation. Introduced commercially in 2001, the goals set out for third generation mobile communication were to facilitate greater voice and data capacity, support a wider range of applications, and increase data transmission at a lower cost .

The 3G standard utilises a new technology called UMTS as its core network architecture - Universal Mobile Telecommunications System. This network combines aspects of the 2G network with some new technology and protocols to deliver a significantly faster data rate. Based on a set of standards used for mobile devices and mobile telecommunications use services and networks that comply with the International Mobile Telecommunications-2000 (IMT-2000) specifications by the International Telecommunication Union. One of requirements set by IMT-2000 was that speed should be at least 200Kbps to call it as 3G service.

3G has Multimedia services support along with streaming are more popular. In 3G, Universal access and portability across different device types are made possible (Telephones, PDA's, etc.). 3G increased the efficiency of frequency spectrum by improving how audio is compressed during a call, so more simultaneous calls can happen in the same frequency range. The UN's International Telecommunications Union IMT-2000 standard requires stationary speeds of 2Mbps and mobile speeds of 384kbps for a "true" 3G. The theoretical max speed for HSPA+ is 21.6 Mbps.

Like 2G, 3G evolved into 3.5G and 3.75G as more features were introduced in order to bring about 4G. A 3G phone cannot communicate through a 4G network , but newer generations of phones are practically always designed to be backward compatible, so a 4G phone can communicate through a 3G or even 2G network .

4G - Fourth Generation

4G is a very different technology as compared to 3G and was made possible practically only because of the advancements in the technology in the last 10 years. Its purpose is to provide high speed , high quality and high capacity to users while improving security and lower the cost of voice and data services, multimedia and internet over IP. Potential and current applications include amended mobile web access, IP telephony , gaming services, high-definition mobile TV, video conferencing, 3D television, and cloud computing.

The key technologies that have made this possible are MIMO (Multiple Input Multiple Output) and OFDM (Orthogonal Frequency Division Multiplexing). The two important 4G standards are WiMAX (has now fizzled out) and LTE (has seen widespread deployment). LTE (Long Term Evolution) is a series of upgrades to existing UMTS technology and will be rolled out on Telstra's existing 1800MHz frequency band. The max speed of a 4G network when the device is moving is 100 Mbps or 1 Gbps for low mobility communication like when stationary or walking, latency reduced from around 300ms to less than 100ms, and significantly lower congestion. When 4G first became available, it was simply a little faster than 3G. 4G is not the same as 4G LTE which is very close to meeting the criteria of the standards. To download a new game or stream a TV show in HD, you can do it without buffering .

Newer generations of phones are usually designed to be backward-compatible , so a 4G phone can communicate through a 3G or even 2G network. All carriers seem to agree that OFDM is one of the chief indicators that a service can be legitimately marketed as being 4G. OFDM is a type of digital modulation in which a signal is split into several narrowband channels at different frequencies. There are a significant amount of infrastructure changes needed to be implemented by service providers in order to supply because voice calls in GSM , UMTS and CDMA2000 are circuit switched, so with the adoption of LTE, carriers will have to re-engineer their voice call network. And again, we have the fractional parts: 4.5G and 4.9G marking the transition of LTE (in the stage called LTE-Advanced Pro) getting us more MIMO, more D2D on the way to IMT-2020 and the requirements of 5G .

5G - Fifth Generation

5G is a generation currently under development , that's intended to improve on 4G. 5G promises significantly faster data rates, higher connection density, much lower latency, among other improvements. Some of the plans for 5G include device-to-device communication, better battery consumption, and improved overall wireless coverage. The max speed of 5G is aimed at being as fast as 35.46 Gbps , which is over 35 times faster than 4G.

Key technologies to look out for: Massive MIMO , Millimeter Wave Mobile Communications etc. Massive MIMO, milimetre wave, small cells, Li-Fi all the new technologies from the previous decade could be used to give 10Gb/s to a user, with an unseen low latency, and allow connections for at least 100 billion devices . Different estimations have been made for the date of commercial introduction of 5G networks. Next Generation Mobile Networks Alliance feel that 5G should be rolled out by 2020 to meet business and consumer demands.

UNIT-II

Pre-requisite:

- To understand the Fading's and signal propagation techniques.

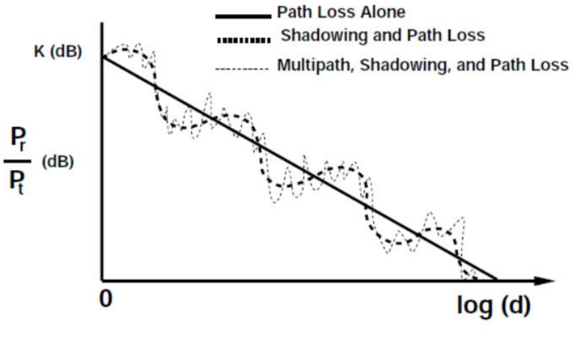
OUTCOMES:

- Analyze the different types of fading's in wireless channel.

INTRODUCTION

The wireless radio channel poses a severe challenge as a medium for reliable high-speed communication. It is not only susceptible to noise, interference, and other channel impediments, but these impediments change over time in unpredictable ways due to user movement. In this chapter we will characterize the variation in received signal power over distance due to path loss and shadowing. Path loss is caused by dissipation of the power radiated by the transmitter as well as effects of the propagation channel. Path loss models generally assume that path loss is the same at a given transmit receive distance1. Shadowing is caused by obstacles between the transmitter and receiver that attenuate signal power through absorption, reflection, scattering, and diffraction. When the attenuation is very strong, the signal is blocked. Variation due to path loss occurs over very large distances (100-1000 meters), whereas variation due to shadowing occurs over distances proportional to the length of the obstructing object (10-100 meters in outdoor environments and less in indoor environments). Since variations due to path loss and shadowing occur over relatively large distances, this variation is sometimes refered to as **large-scale propagation effects**.

small-scale propagation effects illustrates the ratio of the received-to-transmit power in dB versus log-distance for the combined effects of path loss, shadowing, and multipath. After a brief introduction and description of our signal model, we present the simplest model for signal propagation: free space path loss. A signal propagating between two points with no attenuation or reflection follows the free space propagation law. We then describe ray tracing propagation models. These models are used to approximate wave propagation according to Maxwell's equations, and are accurate models when the number of multipath components is small and the physical environment is known. Ray tracing models depend heavily on the geometry and dielectric properties of the region through which the signal propagates. We also described empirical models with parameters based on measurements for both indoor and outdoor channels. We also present a simple generic model with a few parameters that captures the primary impact of path loss in system analysis. A log-normal model for shadowing based on a large number of shadowing objects is also given. When the number of multipath components is large, or the geometry and dielectric properties of the propagation environment are unknown, statistical models must be used.



Radio Wave Propagation : -

Electromagnetic waves propagate through environments where they are reflected, scattered, and diffracted by walls, terrain, buildings, and other objects. The ultimate details of this propagation can be obtained by solving Maxwell's equations with boundary conditions that express the physical characteristics of these obstructing objects. This requires the calculation of the Radar Cross Section (RCS) of large and complex structures. Since these calculations are difficult, and many times the necessary parameters are not available, approximations have been developed to characterize signal propagation without resorting to Maxwell's equations.

The most common approximations use ray-tracing techniques. These techniques approximate the propagation of electromagnetic waves by representing the wave fronts as simple particles: the model determines the reflection and refraction effects on the wave front but ignores the more complex scattering phenomenon predicted by Maxwell's coupled differential equations. The simplest ray-tracing model is the two-ray model, which accurately describes signal propagation when there is one direct path between the transmitter and receiver and one reflected path. The reflected path typically bounces off the ground, and the two-ray model is a good approximation for propagation along highways, rural roads, and over water. We next consider more complex models with additional reflected, scattered, or diffracted components. Many propagation environments are not accurately reflected ray tracing models. In these cases it is common to develop analytical models based on empirical measurements, and we will discuss several of the most common of these empirical models.

Transmit and Receive Signal Models

Our models are developed mainly for signals in the UHF and SHF bands, from .3-3 GHz and 3-30 GHz, respectively. This range of frequencies is quite favorable for wireless system operation due to its propagation characteristics and relatively small required antenna size. We assume the transmission distances on the earth are small enough so as not to be affected by the earth's curvature.

All transmitted and received signals we consider are real. That is because modulators are built using oscillators that generate real sinusoids (not complex exponentials). While we model communication channels using a complex frequency response for analytical simplicity, in fact the channel just introduces an amplitude and phase change at each frequency of the transmitted signal so that the received signal is also real. Real modulated and demodulated signals are often represented as the real part of a complex signal to facilitate analysis.

We model the transmitted signal as

$$\begin{aligned} s(t) &= \Re\left\{u(t)e^{j2\pi f_c t}\right\} \\ &= \Re\{u(t)\}\cos(2\pi f_c t) - \Im\{u(t)\}\sin(2\pi f_c t) \\ &= x(t)\cos(2\pi f_c t) - y(t)\sin(2\pi f_c t), \end{aligned}$$

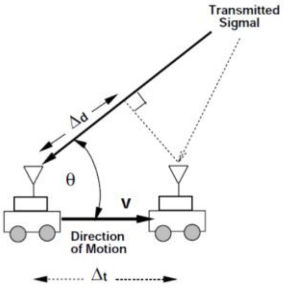
where u(t) = x(t) + jy(t) is a complex baseband signal with in-phase component x(t) = _{u(t)}, quadrature component y(t) = _{u(t)}, bandwidth Bu, and power Pu. The signal u(t) is called the complex envelope or complex lowpass equivalent signal of s(t). We call u(t) the complex envelope of s(t) since the magnitude of u(t) is the magnitude of s(t) and the phase of u(t) is the phase of s(t). This phase includes any carrier phase offset. This is a standard representation for bandpass signals with bandwidthB << fc, as it allows signal manipulation via u(t) irrespective of the carrier frequency. The power in the transmitted signal s(t) is Pt = Pu/2.
The received signal will have a similar form:

$$r(t) = \Re\left\{v(t)e^{j2\pi f_c t}\right\},$$

where the complex baseband signal v(t) will depend on the channel through which s(t) propagates. In particular, as discussed in Appendix A, if s(t) is transmitted through a time-invariant channel then v(t) = u(t) ∗ c(t), where c(t) is the equivalent lowpass channel impulse response for the channel. Time-varying channels will be treated in Chapter 3. The received signal may have a Doppler shift of fD = v cos θ / λ associated with it, where θ is the arrival angle of the received signal relative to the direction of motion, v is the receiver velocity towards the transmitter in the direction of motion, and λ = c/fc is the signal wavelength (c = 3 × 108 m/s is the speed of light). The geometry associated with the Doppler shift is shown in Fig. 2.2. The Doppler shift results from the fact that transmitter or receiver movement over a short time interval Δt causes a slight change in distance Δd = vΔt cos θ that the transmitted signal needs to travel to the receiver. The phase change due to this path length difference is Δϕ = 2π vΔt cos θ / λ . The Doppler frequency is then obtained from the relationship between signal frequency and phase:

$$f_D = \frac{1}{2\pi}\frac{\Delta\phi}{\Delta t} = v\cos\theta/\lambda.$$

If the receiver is moving towards the transmitter, i.e. −π/2 ≤ θ ≤ π/2, then the Doppler frequency is positive, otherwise it is negative. We will ignore the Doppler term in the free-space and ray tracing models of this chapter, since for typical vehicle speeds (75 Km/hr) and frequencies (around 1 GHz), it is on the order of 100 Hz



Suppose $s(t)$ of power P_t is transmitted through a given channel, with corresponding received signal $r(t)$ of power P_r, where P_r is averaged over any random variations due to shadowing. We define the **linear path loss** of the channel as the ratio of transmit power to receive power:

$$P_L = \frac{P_t}{P_r}.$$

We define the **path loss** of the channel as the dB value of the linear path loss or, equivalently, the difference in dB between the transmitted and received signal power:

$$P_L \text{ dB} = 10\log_{10}\frac{P_t}{P_r} \text{ dB}.$$

In general the dB path loss is a nonnegative number since the channel does not contain active elements, and thus can only attenuate the signal. The dB path gain is defined as the negative of the dB path loss: PG = −PL = 10 log10(Pr/Pt) dB, which is generally a negative number. With shadowing the received power will include the effects of path loss and an additional random component due to blockage from objects,

Free-Space Path Loss

Consider a signal transmitted through free space to a receiver located at distance d from the transmitter. Assume there are no obstructions between the transmitter and receiver and the signal propagates along a straight line between the two. The channel model associated with this transmission is called a line-of-sight (LOS) channel, and the corresponding received signal is called the LOS signal or ray. Free-space path loss introduces a complex scale factor [3], resulting in the received signal

$$r(t) = \Re\left\{\frac{\lambda\sqrt{G_l}e^{-j2\pi d/\lambda}}{4\pi d}u(t)e^{j2\pi f_c t}\right\}$$

where

G_l is the product of the transmit and receive antenna field radiation patterns in the LOS direction. The phase shift $e^{-j2\pi d/\lambda}$ is due to the distance d the wave travels.

The power in the transmitted signal $s(t)$ is P_t, so the ratio of received to transmitted power

$$\frac{P_r}{P_t} = \left[\frac{\sqrt{G_l}\lambda}{4\pi d}\right]^2$$

Thus, the received signal power falls off inversely proportional to the square of the distance d between the transmit and receive antennas. We will see in the next section that for other signal propagation models, the received signal power falls off more quickly relative to this distance. The received signal power is also proportional to the square of the signal wavelength, so as the carrier frequency increases, the received power decreases. This dependence of received power on the signal wavelength λ is due to the effective area of the receive antenna [3]. However, directional antennas can be designed so that receive power is an increasing function of frequency for highly directional links [4]. The received power can be expressed in dBm as

$$P_r \text{ dBm} = P_t \text{ dBm} + 10\log_{10}(G_l) + 20\log_{10}(\lambda) - 20\log_{10}(4\pi) - 20\log_{10}(d).$$

Free-space path loss is defined as the path loss of the free-space model:

$$P_L \text{ dB} = 10\log_{10}\frac{P_t}{P_r} = -10\log_{10}\frac{G_l\lambda^2}{(4\pi d)^2}.$$

The **free-space path gain** is thus

$$P_G = -P_L = 10\log_{10}\frac{G_l\lambda^2}{(4\pi d)^2}.$$

Shadow Fading

A signal transmitted through a wireless channel will typically experience random variation due to blockage from objects in the signal path, giving rise to random variations of the received power at a given distance. Such variations are also caused by changes in reflecting surfaces and scattering objects. Thus, a model for the random attenuation due to these effects is also needed. Since the location, size, and dielectric properties of the blocking objects as well as the changes in reflecting surfaces and scattering objects that cause the random attenuation are generally unknown, statistical models must be used to characterize this attenuation. The most common model for this additional attenuation is log-normal shadowing. This model has been confirmed empirically to accurately model the variation in received power in both outdoor and indoor radio propagation environments

In the log-normal shadowing model the ratio of transmit-to-receive power $\psi = P_t/P_r$ is assumed random with a log-normal distribution given by

$$p(\psi) = \frac{\xi}{\sqrt{2\pi}\sigma_{\psi_{dB}}\psi}\exp\left[-\frac{(10\log_{10}\psi - \mu_{\psi_{dB}})^2}{2\sigma^2_{\psi_{dB}}}\right],\ \psi > 0,$$

where $\xi = 10/\ln 10$, $\mu_{\psi_{dB}}$ is the mean of $\psi_{dB} = 10\log_{10}\psi$ in dB and $\sigma_{\psi_{dB}}$ is the standard deviation of ψ_{dB}, also in dB. The mean can be based on an analytical model or empirical measurements. For empirical measurements $\mu_{\psi_{dB}}$ equals the empirical path loss, since average attenuation from shadowing is already incorporated into the measurements. For analytical models, $\mu_{\psi_{dB}}$ must incorporate both the path loss (e.g. from free-space or a ray tracing model) as well as average attenuation from blockage. Alternatively, path loss can be treated separately from shadowing, as described in the next section. Note that if the ψ is log-normal, then the received power and receiver SNR will also be log-normal since these are just constant multiples of ψ. For received SNR the mean and standard deviation of this log-normal random variable are also in dB. For log-normal received power, since the random variable has units of power, its mean and standard deviation will be in dBm or dBW instead of dB. The mean of ψ (the linear average path gain) can be obtained

$$\mu_\psi = E[\psi] = \exp\left[\frac{\mu_{\psi_{dB}}}{\xi} + \frac{\sigma^2_{\psi_{dB}}}{2\xi^2}\right].$$

The conversion from the linear mean (in dB) to the log mean (in dB) is derived from

$$10\log_{10}\mu_\psi = \mu_{\psi_{dB}} + \frac{\sigma^2_{\psi_{dB}}}{2\xi}.$$

Performance in log-normal shadowing is typically parameterized by the log mean $\mu_{\psi_{dB}}$, which is refered to as the **average dB path loss** and is in units of dB. With a change of variables we see that the distribution of the dB value of ψ is Gaussian with mean $\mu_{\psi_{dB}}$ and standard deviation $\sigma_{\psi_{dB}}$:

$$p(\psi_{\text{dB}}) = \frac{1}{\sqrt{2\pi}\sigma_{\psi_{dB}}}\exp\left[-\frac{(\psi_{\text{dB}} - \mu_{\psi_{dB}})^2}{2\sigma^2_{\psi_{dB}}}\right].$$

The log-normal distribution is defined by two parameters: $\mu_{\psi dB}$ and $\sigma_{\psi dB}$. Since ψ = Pt/Pr is always greater than one, $\mu_{\psi dB}$ is always greater than or equal to zero. Note that the log-normal distribution (2.43) takes values for $0 \le \psi \le \infty$. Thus, for $\psi < 1$, Pr > Pt, which is physically impossible. However, this probability will be very small when $\mu_{\psi dB}$ is large and positive. Thus, the log-normal model captures the underlying physical model most accurately when $\mu_{\psi dB} >> 0$

Statistical Multipath Channel Models

If a single pulse is transmitted over a multipath channel the received signal will appear as a pulse train, with each pulse in the train corresponding to the LOS component or a distinct multipath component associated with a distinct scatterer or cluster of scatterers. An important characteristic of a multipath channel is the time delay spread it causes to the received signal. This delay spread equals the time delay between the arrival of the first received signal component (LOS or multipath) and the last received signal component associated with a single transmitted pulse. If the delay spread is small compared to the inverse of the signal bandwidth, then there is little time spreading in the received signal. However, when the delay spread is relatively large, there is significant time spreading of the received signal which can lead to substantial signal distortion.

Another characteristic of the multipath channel is its time-varying nature. This time variation arises because either the transmitter or the receiver is moving, and therefore the location of reflectors in the transmission path, which give rise to multipath, will change over time. Thus, if we repeatedly transmit pulses from a moving transmitter, we will observe changes in the amplitudes, delays, and the number of multipath components corresponding to each pulse. However, these changes occur over a much larger time scale than the fading due to constructive and destructive addition of multipath components associated with a fixed set of scatterers.

We will first use a generic time-varying channel impulse response to capture both fast and slow channel variations. We will then restrict this model to narrowband fading, where the channel bandwidth is small compared to the inverse delay spread. For this narrowband model we will assume a quasi-static environment with a fixed number of multipath components each with fixed path loss and shadowing. For this quasi-static environment we then characterize the variations over short distances (small-scale variations) due to the constructive and destructive addition of multipath components.

Narrowband Fading Models

Suppose the delay spread T_m of a channel is small relative to the inverse signal bandwidth B of the transmitted signal, i.e. $T_m \ll B^{-1}$. As discussed above, the delay spread T_m for time-varying channels is usually characterized by the rms delay spread, but can also be characterized in other ways. Under most delay spread characterizations $T_m \ll B^{-1}$ implies that the delay associated with the ith multipath component $\tau_i \leq T_m \forall i$, so $u(t-\tau_i) \approx u(t) \forall i$

$$r(t) = \Re\left\{u(t)e^{j2\pi f_c t}\left(\sum_n \alpha_n(t)e^{-j\phi_n(t)}\right)\right\}.$$

This scale factor is independent of the transmitted signal $s(t)$ or, equivalently, the baseband signal $u(t)$, as long as the narrowband assumption $T_m \ll 1/B$ is satisfied. In order to characterize the random scale factor caused by the multipath we choose $s(t)$ to be an unmodulated carrier with random phase offset ϕ_0:

$$s(t) = \Re\{e^{j(2\pi f_c t+\phi_0)}\} = \cos(2\pi f_c t - \phi_0),$$

which is narrowband for any T_m.

With this assumption the received signal becomes

$$r(t) = \Re\left\{\left[\sum_{n=0}^{N(t)} \alpha_n(t)e^{-j\phi_n(t)}\right] e^{j2\pi f_c t}\right\} = r_I(t)\cos 2\pi f_c t + r_Q(t)\sin 2\pi f_c t,$$

where the in-phase and quadrature components are given by

$$r_I(t) = \sum_{n=1}^{N(t)} \alpha_n(t)\cos\phi_n(t),$$

And $$r_Q(t) = \sum_{n=1}^{N(t)} \alpha_n(t) \sin \phi_n(t),$$

where the phase term $$\phi_n(t) = 2\pi f_c \tau_n(t) - \phi_{D_n} - \phi_0$$

now incorporates the phase offset ϕ_0 as well as the effects of delay and Doppler.

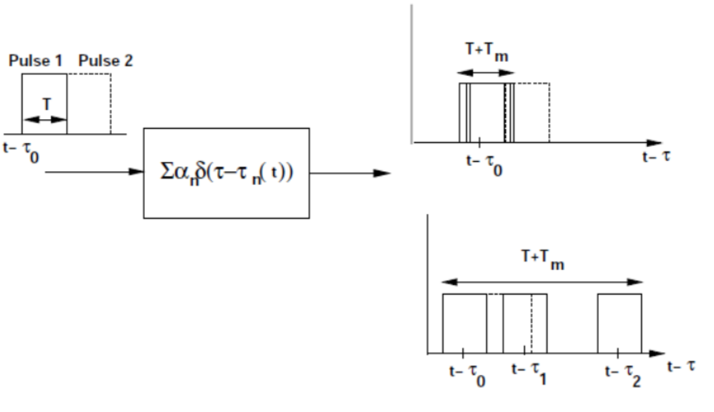
If N(t) is large we can invoke the Central Limit Theorem and the fact that $\alpha_n(t)$ and $\phi_n(t)$ are stationary and ergodic to approximate $r_I(t)$ and $r_Q(t)$ as jointly Gaussian random processes. The Gaussian property is also true for small N if the $\alpha_n(t)$ are Rayleigh distributed and the $\phi_n(t)$ are uniformly distributed on $[-\pi, \pi]$. This happens when the nth multipath component results from a reflection cluster with a large number of nonresolvable multipath components

Wideband Fading Models

When the signal is not narrowband we get another form of distortion due to the multipath delay spread. In this case a short transmitted pulse of duration T will result in a received signal that is of duration $T + T_m$, where T_m is the multipath delay spread. Thus, the duration of the received signal may be significantly increased.

linear modulation consists of a train of pulses where each pulse carries information in its amplitude and/or phase corresponding to a data bit or symbol5. If the multipath delay spread $T_m \ll T$ then the multipath components are received roughly on top of one another, as shown on the upper right of the figure. The resulting constructive and destructive interference causes narrowband fading of the pulse, but there is little time-spreading of the pulse and therefore little interference with a subsequently transmitted pulse. On the other hand, if the multipath delay spread $T_m \gg T$, then each of the different multipath components can be resolved, as shown in the lower right of the figure. However, these multipath components interfere with subsequently transmitted pulses. This effect is called intersymbol interference (ISI).

There are several techniques to mitigate the distortion due to multipath delay spread, including equalization, multicarrier modulation, and spread spectrum, which are discussed in Chapters 11-13. ISI migitation is not necessary if $T \gg T_m$, but this can place significant constraints on data rate. Multicarrier modulation and spread spectrum actually change the characteristics of the transmitted signal to mostly avoid intersymbol interference, however they still experience multipath distortion due to frequency-selective fading,



The difference between wideband and narrowband fading models is that as the transmit signal bandwidth B increases so that $T_m \approx B^{-1}$, the approximation $u(t - \tau_n(t)) \approx u(t)$ is no longer valid. Thus, the received signal is a sum of copies of the original signal, where each copy is delayed in time by τ_n and shifted in phase by $\phi_n(t)$. The signal copies will combine destructively when their phase terms differ significantly, and will distort the direct path signal when $u(t - \tau_n)$ differs from $u(t)$.

Although the approximation in no longer applies when the signal bandwidth is large relative to the inverse of the multipath delay spread, if the number of multipath components is large and the phase of each component is uniformly distributed then the received signal will still be a zero-mean complex Gaussian process with a Rayleigh-distributed envelope. However, wideband fading differs from narrowband fading in terms of the resolution of the different multipath components. Specifically, for narrowband signals, the multipath components have a time resolution that is less than the inverse of the signal bandwidth, so the multipath components characterized combine at the receiver to yield the original transmitted signal with amplitude and phase characterized by random processes. These random processes are characterized by their autocorrelation or PSD, and their instantaneous distributions, However, with wideband signals, the received signal experiences distortion due to the delay spread of the different multipath components, so the received signal can no longer be characterized by just the amplitude and phase random processes. The effect of multipath on wideband signals must therefore take into account both the multipath delay spread and the time-variations associated with the channel.

The starting point for characterizing wideband channels is the equivalent lowpass time-varying channel impulse response $c(\tau, t)$. Let us first assume that $c(\tau, t)$ is a continuous6 deterministic function of τ and t. Recall that τ represents the impulse response associated with a given multipath delay, while t represents time variations. We can take the Fourier transform of $c(\tau, t)$ with respect to t as

$$S_c(\tau, \rho) = \int_{-\infty}^{\infty} c(\tau, t) e^{-j2\pi\rho t} dt.$$

We call $S_c(\tau, \rho)$ the **deterministic scattering function** of the lowpass equivalent channel impulse response $c(\tau, t)$. Since it is the Fourier transform of $c(\tau, t)$ with respect to the time variation parameter t, the deterministic scattering function $S_c(\tau, \rho)$ captures the Doppler characteristics of the channel via the frequency parameter ρ.

In general the time-varying channel impulse response $c(\tau, t)$ given is random instead of deterministic due to the random amplitudes, phases, and delays of the random number of multipath components. In this case we must characterize it statistically or via measurements. As long as the number of multipath components is large, we can invoke the Central Limit Theorem to assume that $c(\tau, t)$ is a complex Gaussian process, so its statistical characterization is fully known from the mean, autocorrelation, and cross-correlation of its in-phase and quadrature components. As in the narrowband case, we assume that the phase of each multipath component is uniformly distributed. Thus, the in-phase and quadrature components of $c(\tau, t)$ are independent Gaussian processes with the same autocorrelation, a mean of zero, and a cross-correlation of zero. The same statistics hold for the in-phase and quadrature components if the channel contains only a small number of multipath rays as long as each ray has a Rayleigh-distributed amplitude and uniform phase. Note that this model does not hold when the channel has a dominant LOS component.

The statistical characterization of c(τ , t) is thus determined by its autocorrelation function, defined as

$$A_c(\tau_1, \tau_2; t, \Delta t) = E[c^*(\tau_1; t) c(\tau_2; t + \Delta t)].$$

Most channels in practice are wide-sense stationary (WSS), such that the joint statistics of a channel measured at two different times t and t+Δt depends only on the time difference Δt. For wide-sense stationary channels, the autocorrelation of the corresponding bandpass channel h(τ , t) = _{c(τ , t)ej2 π fct} can be obtained [16] from Ac(τ 1, τ 2; t,Δt) as7 Ah(τ 1, τ 2; t,Δt) = .5_{Ac(τ 1, τ 2; t,Δt)ej2 π fcΔt}. We will assume that our channel model is WSS, in which case the autocorrelation becomes indepedent of t:

$$A_c(\tau_1, \tau_2; \Delta t) = E[c^*(\tau_1; t) c(\tau_2; t + \Delta t)].$$

Moreover, in practice the channel response associated with a given multipath component of delay τ 1 is uncorrelated with the response associated with a multipath component at a different delay τ 2 _= τ 1, since the two components are caused by different scatterers.

Power Delay Profile

The **power delay profile** $A_c(\tau)$, also called the **multipath intensity profile**, is defined as the autocorrelation with $\Delta t = 0$: $A_c(\tau) \triangleq A_c(\tau, 0)$.

The power delay profile represents the average power associated with a given multipath delay, and is easily measured empirically. The average and rms delay spread are typically defined in terms of the power delay profile $A_c(\tau)$ as

$$\mu_{T_m} = \frac{\int_0^\infty \tau A_c(\tau)d\tau}{\int_0^\infty A_c(\tau)d\tau},$$

And $$\sigma_{T_m} = \sqrt{\frac{\int_0^\infty (\tau - \mu_{T_m})^2 A_c(\tau)d\tau}{\int_0^\infty A_c(\tau)d\tau}}.$$

Note that if we define the pdf p_{T_m} of the random delay spread T_m in terms of $A_c(\tau)$ as

$$p_{T_m}(\tau) = \frac{A_c(\tau)}{\int_0^\infty A_c(\tau)d\tau}$$

then μ_{T_m} and σ_{T_m} are the mean and rms values of T_m, respectively, relative to this pdf. Defining the pdf of T_m equivalently, defining the mean and rms delay spread by (3.54) and (3.55), respectively, weights the delay associated with a given multipath component by its relative power, so that weak multipath components contribute less to delay spread than strong ones. In particular, multipath components below the noise floor will not significantly impact these delay spread characterizations.

The time delay T where Ac(τ) ≈ 0 for τ ≥ T can be used to roughly characterize the delay spread of the channel, and this value is often taken to be a small integer multiple of the rms delay spread, i.e. Ac(τ) ≈ 0 for τ > 3σTm. With this approximation a linearly modulated signal with symbol period Ts experiences significant ISI if Ts << σTm. Conversely, when Ts >> σTm the system experiences negligible ISI. For calculations one can assume that Ts << σTm implies Ts < σTm/10 and Ts >> σTm implies Ts > 10σTm. When Ts is within an order of magnitude of σTm then there will be some ISI which may or may not significantly degrade performance, depending on the specifics of the system and channel.

While $\mu_{T_m} \approx \sigma_{T_m}$ in many channels with a large number of scatterers, the exact relationship between μ_{T_m} and σ_{T_m} depends on the shape of $A_c(\tau)$. A channel with no LOS component and a small number of multipath components with approximately the same large delay will have $\mu_{T_m} \gg \sigma_{T_m}$. In this case the large value of μ_{T_m} is a misleading metric of delay spread, since in fact all copies of the transmitted signal arrive at rougly the same time and the demodulator would synchronize to this common delay. It is typically assumed that the synchronizer locks to the multipath component at approximately the mean delay, in which case rms delay spread characterizes the time-spreading of the channel.

Coherence Bandwidth

We can also characterize the time-varying multipath channel in the frequency domain by taking the Fourier transform of $c(\tau, t)$ with respect to τ. Specifically, define the random process

$$C(f;t) = \int_{-\infty}^{\infty} c(\tau;t)e^{-j2\pi f\tau}d\tau.$$

Since $c(\tau; t)$ is a complex zero-mean Gaussian random variable in t, the Fourier transform above just represents the sum8 of complex zero-mean Gaussian random processes, and therefore $C(f; t)$ is also a zero-mean Gaussian random process completely characterized by its autocorrelation. Since $c(\tau; t)$ is WSS, its integral $C(f; t)$ is as well.

$$A_C(f_1, f_2; \Delta t) = \mathrm{E}[C^*(f_1;t)C(f_2;t+\Delta t)].$$

We can simplify $A_C(f_1, f_2; \Delta t)$ as

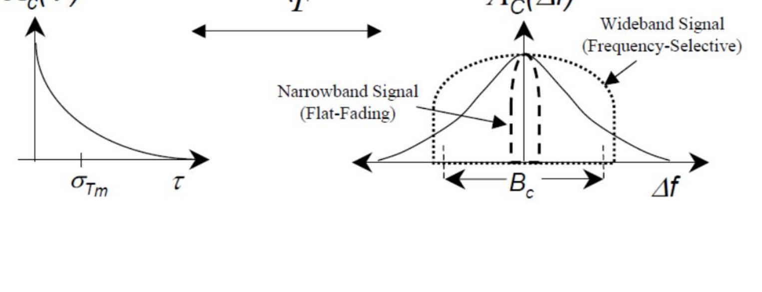
$$\begin{aligned} A_C(f_1, f_2; \Delta t) &= E\left[\int_{-\infty}^{\infty} c^*(\tau_1; t) e^{j2\pi f_1 \tau_1} d\tau_1 \int_{-\infty}^{\infty} c(\tau_2; t+\Delta t) e^{-j2\pi f_2 \tau_2} d\tau_2\right] \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} E[c^*(\tau_1; t) c(\tau_2; t+\Delta t)] e^{j2\pi f_1 \tau_1} e^{-j2\pi f_2 \tau_2} d\tau_1 d\tau_2 \\ &= \int_{-\infty}^{\infty} A_c(\tau, \Delta t) e^{-j2\pi(f_2 - f_1)\tau} d\tau. \\ &= A_C(\Delta f; \Delta t) \end{aligned}$$

where $\Delta f = f_2 - f_1$ and the third equality follows from the WSS and US properties of $c(\tau; t)$. Thus, the autocorrelation of $C(f; t)$ in frequency depends only on the frequency difference Δf. The function $A_C(\Delta f; \Delta t)$ can be measured in practice by transmitting a pair of sinusoids through the channel that are separated in frequency by Δf and calculating their cross correlation at the receiver for the time separation Δt.

$$A_C(\Delta f) = \int_{-\infty}^{\infty} A_c(\tau) e^{-j2\pi\Delta f \tau} d\tau.$$

So AC(Δf) is the Fourier transform of the power delay profile. Since AC(Δf) = E[C∗(f; t)C(f +Δf; t] is an autocorrelation, the channel response is approximately independent at frequency separationsΔf whereAC(Δf) ≈ 0. The frequency Bc where AC(Δf) ≈ 0 for all Δf > Bc is called the coherence bandwidth of the channel. By the Fourier transform relationship between Ac(τ) and AC(Δf), if Ac(τ) ≈ 0 for τ > T then AC(Δf) ≈ 0 for Δf > 1/T . Thus, the minimum frequency separation Bc for which the channel response is roughly independent is Bc ≈ 1/T , where T is typically taken to be the rms delay spread σTm of Ac(τ). A more general approximation is Bc ≈ k/σTm where k depends on the shape of Ac(τ) and the precise specification of coherence bandwidth. For example, Lee has shown that Bc ≈ .02/σTm approximates the range of frequencies over which channel correlation exceeds 0.9, while Bc ≈ .2/σTm approximates the range of frequencies over which this correlation exceeds 0.5

In general, if we are transmitting a narrowband signal with bandwidth$B \ll B_c$, then fading across the entire signal bandwidth is highly correlated, i.e. the fading is roughly equal across the entire signal bandwidth. This is usually referred to as **flat fading**. On the other hand, if the signal bandwidth$B \gg B_c$, then the channel amplitude values at frequencies separated by more than the coherence bandwidth are roughly independent. Thus, the channel amplitude varies widely across the signal bandwidth. In this case the channel is called **frequency-selective**. When B ≈ Bc then channel behavior is somewhere between flat and frequency-selective fading. Note that in linear modulation the signal bandwidth B is inversely proportional to the symbol time Ts, so flat fading corresponds to Ts ≈ 1/B >> 1/Bc ≈ σTm, i.e. the case where the channel experiences negligible ISI. Frequency-selective fading corresponds to Ts ≈ 1/B << 1/Bc = σTm, i.e. the case where the linearly modulated signal experiences significant ISI. Wideband signaling formats that reduce ISI, such as multicarrier modulation and spread spectrum, still experience frequency-selective fading across their entire signal bandwidth which causes performance degradation



Power Delay Profile, RMS Delay Spread, and Coherence Bandwidth.

Slow vs. Fast Fading

The terms *slow* and *fast* fading refer to the rate at which the magnitude and phase change imposed by the channel on the signal changes. The **coherence time** is a measure of the minimum time required for the magnitude change of the channel to become decorrelated from its previous value.

Slow fading arises when the coherence time of the channel is large relative to the delay constraint of the channel. In this regime, the amplitude and phase change imposed by the channel can be considered roughly constant over the period of use. Slow fading can be caused by events such as **shadowing**, where a large obstruction such as a hill or large building obscures the main signal path between the transmitter and the receiver. The amplitude change caused by shadowing is often modeled using a log-normal distribution with a standard deviation according to the Log Distance Path Loss Model.

Fast Fading occurs when the coherence time of the channel is small relative to the delay constraint of the channel. In this regime, the amplitude and phase change imposed by the channel varies considerably over the period of use.

In a fast-fading channel, the transmitter may take advantage of the variations in the channel conditions using time diversity to help increase robustness of the communication to a temporary deep fade. Although a deep fade may temporarily erase some of the information transmitted, use of an error-correcting code coupled with successfully transmitted bits during other time instances can allow for the erased bits to be recovered. In a slow-fading channel, it is not possible to use time diversity because the transmitter sees only a single realization of the channel within its delay constraint. A deep fade therefore lasts the entire duration of transmission and cannot be mitigated using coding.

The coherence time of the channel is related to a quantity known as the **Doppler spread** of the channel. When a user (or reflectors in its environment) is moving, the user's velocity causes a shift in the frequency of the signal transmitted along each signal path. This phenomenon is known as the Doppler shift. Signals travelling along different paths can have different Doppler shifts, corresponding to different rates of change in phase.

The difference in Doppler shifts between different signal components contributing to a single fading channel tap is known as the Doppler spread. Channels with a large Doppler spread have signal components that are each changing independently in phase over time. Since fading depends on whether signal components add constructively or destructively, such channels have a very short coherence time.

In general, coherence time is inversely related to Doppler spread, typically expressed as:

$$T_c = \frac{k}{D_s}$$

where T_c is the coherence time, D_s is the Doppler spread, and k is a constant taking on values in the range of 0.25 to 0.5.

Flat vs. Frequency-selective Fading

As the carrier frequency of a signal is varied, the magnitude of the change in amplitude will vary. The coherence bandwidth measures the minimum separation in frequency after which two signals will experience uncorrelated fading.

In **flat fading**, the coherence bandwidth of the channel is larger than the bandwidth of the signal. Therefore, all frequency components of the signal will experience the same magnitude of fading.

In **frequency-selective fading**, the coherence bandwidth of the channel is smaller than the bandwidth of the signal. Different frequency components of the signal therefore experience decorrelated fading.

In a frequency-selective fading channel, since different frequency components of the signal are affected independently, it is highly unlikely that all parts of the signal will be simultaneously affected by a deep fade. Certain modulation schemes such as OFDM and CDMA are well-suited to employing frequency diversity to provide robustness to fading. OFDM divides the wideband signal into many slowly modulated narrowband subcarriers, each exposed to flat fading rather than frequency selective fading. This can be combated by means of error coding, simple equalization or adaptive bit loading. Inter-symbol interference is avoided by introducing a guard interval between the symbols. CDMA uses the Rake receiver to deal with each echo separately.

Frequency-selective fading channels are also dispersive, in that the signal energy associated with each symbol is spread out in time. This causes transmitted symbols that are adjacent in time to interfere with each other. Equalizers are often deployed in such channels to compensate for the effects of the intersymbol interference.

UNIT-III

Pre-requisite:

- To understand the Antennas for Mobile Terminals.

OUTCOMES:

- Analyze the basic concepts of Fading and antenna design.

INTRODUCTION

The growing demand for wireless communication makes it important to determine the capacity limits of these channels. These capacity limits dictate the maximum data rates that can be transmitted over wireless channels with asymptotically small error probability, assuming no constraints on delay or complexity of the encoder and decoder. Channel capacity was pioneered by Claude Shannon in the late 1940s, using a mathematical theory of communication based on the notion of mutual information between the input and output of a channel Shannon defined capacity as the mutual information maximized over all possible input distributions. The significance of this mathematical construct was Shannon's coding theorem and converse, which proved that a code did exist that could achieve a data rate close to capacity with negligible probability of error, and that any data rate higher than capacity could not be achieved without an error probability bounded away from zero. Shannon's ideas were quite revolutionary at the time, given the high data rates he predicted were possible on telephone channels and the notion that coding could reduce error probability without reducing data rate or causing bandwidth expansion. In time sophisticated modulation and coding technology validated Shannon's theory such that on telephone lines today, we achieve data rates very close to Shannon capacity with very low probability of error.

We will consider flat-fading channel capacity where only the fading distribution is known at the transmitter and receiver. Capacity under this assumption is typically very difficult to determine, and is only known in a few special cases. Next we consider capacity when the channel fade level is known at the receiver only (via receiver estimation) or that the channel fade level is known at both the transmitter and the receiver (via receiver estimation and transmitter feedback). We will see that the fading channel capacity with channel side information at both the transmitter and receiver is achieved when the transmitter adapts its power, data rate, and coding scheme to the channel variation. The optimal power allocation in this case is a "water-filling" in time, where power and data rate are increased when channel conditions are favorable and decreased when channel conditions are not favorable.

We will also treat capacity of frequency-selective fading channels. For time-invariant frequency-selective channels the capacity is known and is achieved with an optimal power allocation that water-fills over frequency instead of time. The capacity of a time-varying frequency-selective fading channel is unknown in general. However, this channel can be approximated as a set of independent parallel flat-fading channels, whose capacity is the sum of capacities on each channel with power optimally allocated among the channels. The capacity of this channel is known and is obtained with an optimal power allocation that water-fills over both time and frequency.

Capacity in AWGN : -

Consider a discrete-time additive white Gaussian noise (AWGN) channel with channel input/output relationship y[i] = x[i] + n[i], where x[i] is the channel input at time i, y[i] is the corresponding channel output, and n[i] is a white Gaussian noise random process. Assume a channel bandwidth B and transmit power P. The channel SNR, the power in x[i] divided by the power in n[i], is constant and given by γ = P/(N0B), where N0 is the power spectral density of the noise. The capacity of this channel is given by Shannon's well-known formula [1]:

$$C = B \log2(1 + \gamma),$$

where the capacity units are bits/second (bps). Shannon's coding theorem proves that a code exists that achieves data rates arbitrarily close to capacity with arbitrarily small probability of bit error. The converse theorem shows that any code with rate $R > C$ has a probability of error bounded away from zero. The theorems are proved using the concept of mutual information between the input and output of a channel. For a memoryless time-invariant channel with random input X and random output Y, the channel's **mutual information** is defined as

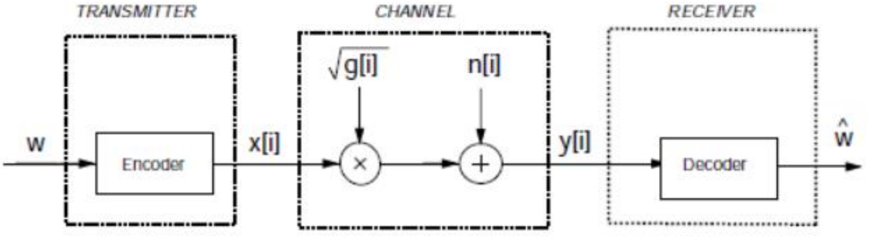
$$I(X;Y) = \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right),$$

where the sum is taken over all possible input and output pairs x ∈ X and y ∈ Y for X and Y the input and output alphabets. The log function is typically with respect to base 2, in which case the units of mutual information are bits per second. Mutual information can also be written in terms of the entropy in the channel output y and conditional output y|x as I(X; Y) = H(Y) − H(Y |X), where H(Y) = y∈Y p(y) log p(y) and H(Y |X) = x∈X,y∈Y p(x, y) log p(y|x). Shannon proved that channel capacity equals the mutual information of the channel maximized over all possible input distributions:

$$C = \max_{p(x)} I(X;Y) = \max_{p(x)} \sum_{x,y} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right).$$

Capacity of Flat-Fading Channels

We assume a discrete-time channel with stationary and ergodic time-varying gain $g[i]$, $0 \leq g[i]$, and AWGN $n[i]$, as shown in Figure 4.1. The channel power gain $g[i]$ follows a given distribution $p(g)$, e.g. for Rayleigh fading $p(g)$ is exponential. We assume that $g[i]$ is independent of the channel input. The channel gain $g[i]$ can change at each time i, either as an i.i.d. process or with some correlation over time. In a **block fading channel** $g[i]$ is constant over some blocklength T after which time $g[i]$ changes to a new independent value based on the distribution $p(g)$. Let P denote the average transmit signal power, $N_0/2$ denote the noise power spectral density of $n[i]$, and B denote the received signal bandwidth. The instantaneous received signal-to-noise ratio (SNR) is then $\gamma[i] = Pg[i]/(N_0B)$, $0 \leq \gamma[i] < \infty$, and its expected value over all time is $\gamma = Pg/(N_0B)$. Since $P/(N_0B)$ is a constant, the distribution of $g[i]$ determines the distribution of $\gamma[i]$ and vice versa. where an input message **w** is sent from the transmitter to the receiver. The message is encoded into the codeword **x**, which is transmitted over the time-varying channel as $x[i]$ at time i. The channel gain $g[i]$, also called the **channel side information** (CSI), changes during the transmission of the codeword.



The capacity of this channel depends on what is known about $g[i]$ at the transmitter and receiver. We will consider three different scenarios regarding this knowledge:

1. **Channel Distribution Information (CDI):** The distribution of $g[i]$ is known to the transmitter and receiver.
2. **Receiver CSI:** The value of $g[i]$ is known at the receiver at time i, and both the transmitter and receiver know the distribution of $g[i]$.
3. **Transmitter and Receiver CSI:** The value of $g[i]$ is known at the transmitter and receiver at time i, and both the transmitter and receiver know the distribution of $g[i]$.

Channel Distribution Information (CDI)

The capacity-achieving input distribution and corresponding fading channel capacity under CDI is known for two specific models of interest: i.i.d. Rayleigh fading channels and FSMCs. In i.i.d. Rayleigh fading the channel power gain is exponential and changes independently with each channel use. The optimal input distribution for this channel was shown in [10] to be discrete with a finite number of mass points, one of which is located at zero. This optimal distribution and its corresponding capacity must be found numerically. The lack of closed-form solutions for capacity or the optimal input distribution is somewhat surprising given the fact that the fading follows the most common fading distribution and has no correlation structure. For flat-fading channels that are not necessarily Rayleigh or i.i.d. upper and lower bounds on capacity have been determined in [11], and these bounds are tight at high SNRs

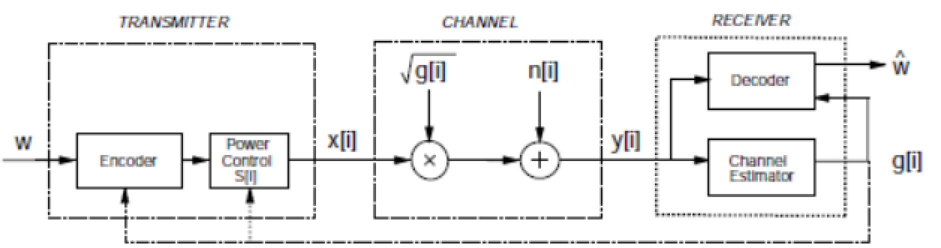
This model approximates the fading correlation as a Markov process. While the Markov nature of the fading dictates that the fading at a given time depends only on fading at the previous time sample, it turns out that the receiver must decode all past channel outputs jointly with the current output for optimal (i.e. capacity-achieving) decoding. This significantly complicates capacity analysis. The capacity of FSMCs has been derived for . inputs in [13, 14] and for general inputs in [15]. Capacity of the FSMC depends on the limiting distribution of the channel conditioned on all past inputs and outputs, which can be computed recursively. As with the i.i.d. Rayleigh fading channel, the complexity of the capacity analysis along with the final result for this relatively simple fading model is very high, indicating the difficulty of obtaining the capacity and related design insights into channels when only CDI is available.

Channel Side Information at Receiver

We now consider the case where the CSI $g[i]$ is known at the receiver at time i. Equivalently, $\gamma[i]$ is known at the receiver at time i. We also assume that both the transmitter and receiver know the distribution of $g[i]$. In this case there are two channel capacity definitions that are relevant to system design: Shannon capacity, also called **ergodic capacity**, and **capacity with outage**. As for the AWGN channel, Shannon capacity defines the maximum data rate that can be sent over the channel with asymptotically small error probability. Note that for Shannon capacity the rate transmitted over the channel is constant: the transmitter cannot adapt its transmission strategy relative to the CSI. Thus, poor channel states typically reduce Shannon capacity since the transmission strategy must incorporate the effect of these poor states. An alternate capacity definition for fading channels with receiver CSI is capacity with outage. Capacity with outage is defined as the maximum rate that can be transmitted over a channel with some outage probability corresponding to the probability that the transmission cannot be decoded with negligible error probability. The basic premise of capacity with outage is that a high data rate can be sent over the channel and decoded correctly except when the channel is in deep fading. By allowing the system to lose some data in the event of deep fades, a higher data rate can be maintained than if all data must be received correctly regardless of the fading state, as is the case for Shannon capacity. The probability of outage characterizes the probability of data loss or, equivalently, of deep fading.

Channel Side Information at Transmitter and Receiver

When both the transmitter and receiver have CSI, the transmitter can adapt its transmission strategy relative to this CSI, In this case there is no notion of capacity versus outage where the transmitter sends bits that cannot be decoded, since the transmitter knows the channel and thus will not send bits unless they can be decoded correctly. In this section we will derive Shannon capacity assuming optimal power and rate adaptation relative to the CSI, as well as introduce alternate capacity definitions and their power and rate adaptation strategies.



Capacity of Frequency-Selective Fading Channels

In this section we consider the Shannon capacity of frequency-selective fading channels. We first consider the capacity of a time-invariant frequency-selective fading channel. This capacity analysis is similar to that of a flat fading channel with the time axis replaced by the frequency axis. Next we discuss the capacity of time-varying frequency-selective fading channels.

Time-Varying Channels

The time-varying frequency-selective fading channel is similar to the model shown in Figure 4.9, except that $H(f) = H(f, i)$, i.e. the channel varies over both frequency and time. It is difficult to determine the capacity of time-varying frequency-selective fading channels, even when the instantaneous channel $H(f, i)$ is known perfectly at the transmitter and receiver, due to the random effects of self-interference (ISI). In the case of transmitter and receiver side information, the optimal adaptation scheme must consider the effect of the channel on the past sequence of transmitted bits, and how the ISI resulting from these bits will affect future transmissions [30]. The capacity of time-varying frequency-selective fading channels is in general unknown, however upper and lower bounds and limiting formulas exist

We can approximate channel capacity in time-varying frequency-selective fading by taking the channel bandwidth B of interest and divide it up into subchannels the size of the channel coherence bandwidth B_c, as shown in Figure 4.12. We then assume that each of the resulting subchannels is independent, time-varying, and flat-fading with $H(f, i) = H_j[i]$ on the jth subchannel.

Under this assumption, we obtain the capacity for each of these flat-fading subchannels based on the average power P_j that we allocate to each subchannel, subject to a total power constraint P. Since the channels are independent, the total channel capacity is just equal to the sum of capacities on the individual narrowband flatfading channels subject to the total average power constraint, averaged over both time and frequency:

$$C = \max_{\{\overline{P}_j\}:\sum_j \overline{P}_j \leq \overline{P}} \sum_j C_j(\overline{P}_j),$$

We will focus on Shannon capacity assuming perfect transmitter and receiver channel CSI, since this upperbounds capacity under any other side information assumptions or suboptimal power allocation strategies. We know that if we fix the average power per subchannel, the optimal power adaptation follows a water-filling formula. We also expect that the optimal average power to be allocated to each subchannel should also follow a water-filling, where more average power is allocated to better subchannels. Thus we expect that the optimal power allocation is a two-dimensional water-filling in both time and frequency. We now obtain this optimal two-dimensional water-filling and the corresponding Shannon capacity

It is interesting to note that in the two-dimensional water-filling the cutoff value for all subchannels is the same. This implies that even if the fading distribution or average fade power on the subchannels is different, all subchannels suspend transmission when the instantaneous SNR falls below the common cutoff value γ_0.

$$C = \sum_j \int_{\gamma_0}^{\infty} B_c \log_2\left(\frac{\gamma_j}{\gamma_0}\right) p(\gamma_j) d\gamma_j.$$

PLANAR INVERTED F ANTENNA.

The Inverted F antenna (IFA) Typically Consists of a rectangular planar element located above a ground plane ,a short circuiting plate or pin, and a feeding mechanism for the planar element. It’s shape appears like an Inverted F and hence named as PLANAR INVERTED F ANTENNA. PIFA is also referred to as short-circuited microstrip antenna due to the fact that it’s structure resembles to short –circuit MSA. It can be considered as a kind of linear inverted F antenna (IFA) with the wire radiator element replaced by a plate to expand the bandwidth.

In wireless communication a low profile antenna that supports multiband wideband operations is required. In order to meet these requirements PIFA is being used.These antennas are compact and support multiband and wideband operations therefore such antennas are suitable for the devices where space is a major issue.PIFA has a low backward radiation and hence it minimizes electromagnetic wave absorption or SAR.Thus PIFA is an antenna with many wide applications in today’s electronics world.

The working principle of PIFA is same as the microstrip antenna.

Lp +Wp=λ/4

Where Lp is the top patch length & Wp is top patch width. λ is wavelength corresponding to resonant frequency.

whenW/Lp =1 then

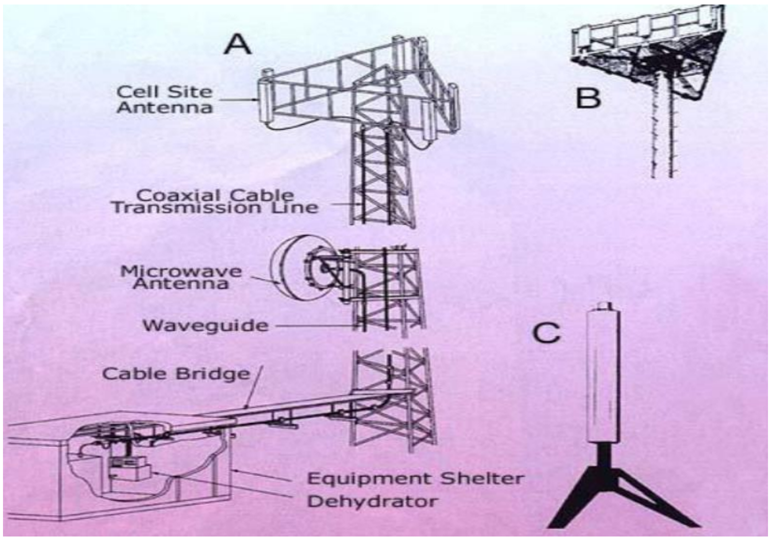
Lp+h= λ/4

When W=0 then

Lp+Wp+h= λ/4

Base station

Base station antennas for mobile communication systems have during the last two decades exploded in numbers in both our rural areas as well as in our city centers. These antennas are typically 1 to 2 meters long array antennas with gains between 15 and 21 dBi placed in towers between 25 and 75 meters above ground. They have high aperture efficiencies and are able to handle power of up to 500W or more without generating either 2nd or 3rd order passive intermodulation products greater than the system noise floor. In the mid 90s size reduction was an issue and manufacturers focused on designing flat panel antennas based on e.g. microstrip patches . However, it was soon discovered that it was the number of antennas rather than size that was the main issue for the operators and therefore both dual polarized (for diversity) and multi-band antennas were soon developed. With the introduction of digital radio systems in mobile telephony in the mid 90ties digital processing of the antenna signals also became an option. Diversity combining at the base station had been implemented already in analog systems but with the digital signal processing one could now also discuss the implementation of digital beamforming and even spatial multiplexing of signals . Today diversity is almost always implemented at the base station. However, recently diversity has also been considered at the mobile terminal. The resulting "Multiple Input Multiple Output", MIMO, system is perceived as a the next important technology step in increasing the data rates that are expected to grow up to and around 100Mbps in e.g. the long term evolution of the 3G systems .



Because of the high requirements on capacity, cellular networks are most often sectorized and the use of antennas that are omni-directional in the horizontal plane is very limited today. The most popular choices are base station antennas with horizontal half power beam widths of 65 or 90 degrees. The optimum horizontal and vertical beam width is decided by the network architecture and propagation environment. From this basic idea, the development of cellular systems has followed at least two different lines. In the U.S., systems are traditionally based on a hexagonal cellular layout with omni-directional antennas in the center of each cell. It was soon discovered that the use of directional antennas increased capacity [7], so each omni-directional antenna was replaced by three directional ones. Each directional antenna then illuminates a 120 wide sector and the original hexagon has been replaced with three rhombuses

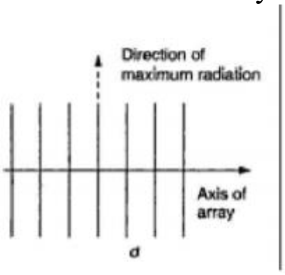
In GSM where the SNIR is allowed to be as low as 9 dB, a frequency re-use factor K=9 has been found possible. During the early deployment of cellular systems the load on the system was low and the interference of little or no importance. With the introduction of the portable phone the operating range started to become limited by the terminal antenna gain. However, today when urban cells are often in order of a few 100 meters or less, range is not an issue any more, but now the interference limits the performance. Any means of reducing its influence on the system, e.g. by shaping the beam of the base station antenna, is therefore more than welcomed by operators as well as handset manufacturers. Today a variety of sector antennas are offered with horizontal beam widths from about 30 to 120 degrees and gains from about 10 to 20 dBi. The use of omnidirectional antennas is very limited. The most popular choices are base station antennas with an horizontal half power beamwidth of 65 or 90degrees.

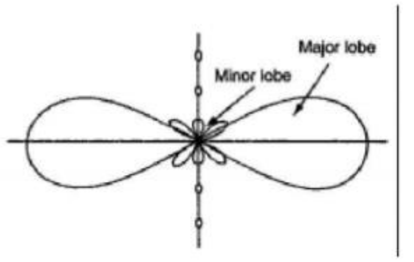
ANTENNA ARRAY

The study of a single small antenna indicates that the radiation fields are uniformly distributed and antenna provides wide beam width, but low directivity and gain. For example, the maximum radiation of dipole antenna takes place in the direction normal to its axis and decreases slowly as one moves toward the axis of the antenna. The antennas of such radiation characteristic may be preferred in broadcast services where wide coverage is required but not in point to point communication. Thus to meet the demands of point to point communication, it is necessary to design the narrow beam and high directive antennas, so that the radiation can be released in the preferred direction. The simplest way to achieve this requirement is to increase the size of the antenna, because a larger-size antenna leads to more directive characteristics. But from the practical aspect the method is inconvenient as antenna becomes bulky and it is difficult to change the size later. Another way to improve the performance of the antenna without increasing the size of the antenna is to arrange the antenna in a specific configuration, so spaced and phased that their individual contributions are maximum in desired direction and negligible in other directions. This way particularly, we get greater directive gain. This new arrangement of multi-element is referred to as an array of the antenna. The antenna involved in an array is known as element. The individual element of array may be of any form (wire. dipole. slot, aperture. etc.). Having identical element in an array is often simpler, convenient and practical, but it is not compulsory. The antenna array makes use of wave interference phenomenon that occurs between the radiations from the different elements of the array. Thus, the antenna array is one of the methods of combining the radiation from a group of radiators in such a way that the interference is constructive in the preferred direction and destructive in the remaining directions. The main function of an array is to produce highly directional radiation. The field is a vector quantity with both magnitude and phase. The total field (not power) of the array system at any point away from its centre is the vector sum of the field produced by the individual antennas. The relative phases of individual field components depend on the relative distance of the individual clement and in turn depend on the direction.

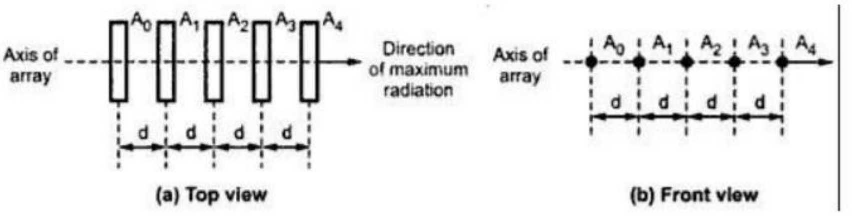
ARRAY CONFIGURATIONS Broadly, array antennas can be classified into four categories:
(a)Broadside array
(b) End-fire array
(c) Collinear array
(d) Parasitic array

Broadside Array- This is a type of array in which the number of identical elements is placed on a supporting line drawn perpendicular to their respective axes. Elements are equally spaced and fed with a current of equal magnitude and all in same phase. The advantage of this feed technique is that array fires in broad side direction (i.e. perpendicular to the line of array axis, where there are maximum radiation and small radiation in other direction). Hence the radiation pattern of broadside array is bidirectional and the array radiates equally well in either direction of maximum radiation. In Fig. 1 the elements are arranged in horizontal plane with spacing between elements and radiation is perpendicular to the plane of array (i.e. normal to plane of paper.) They may also be arranged in vertical and in this case radiation will be horizontal. Thus, it can be said that broadside array is a geometrical arrangement of elements in which the direction of maximum radiation is perpendicular to the array axis and to the plane containing the array clement. Radiation pattern of a broad side array is shown in Fig. 2. The bidirectional pattern of broadside array can be converted into unidirectional by placing an identical array behind this array at distance of λ/4 fed by current leading in phase by 90d



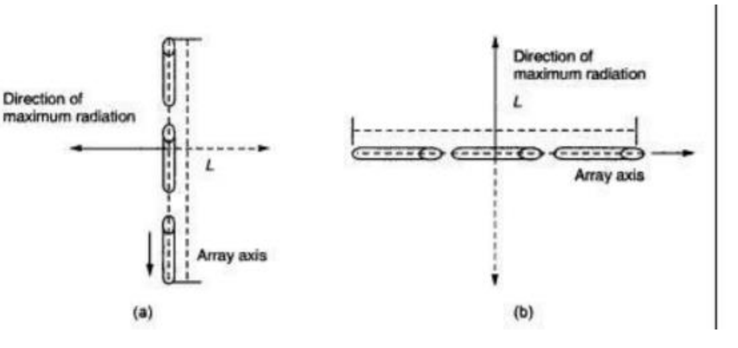


End Fire Array-The end fire array is very much similar to the broadside array from thepoint of view of arrangement. But the main difference is in the direction of maximum radiation. In broadside array, the direction of the maximum radiation is perpendicular to the axis of array; while in the end fire array, the direction of the maximum radiation is along the axis of array.



Thus in the end fire array number of identical antennas are spaced equally along a line. All the antennas are fed individually with currents of equal magnitudes but their phases vary progressively along the line to get entire arrangement unidirectional finally. i.e. maximum radiation along the axis of array. Thus end fire array can be defined as an array with direction of maximum radiation coincides with the direction of the axis of array to get unidirectional radiation

Collinear Array-In collinear array the elements are arranged co-axially, i.e., antennas areeither mounted end to end in a single line or stacked over one another. The collinear array is also a broadside array and elements are fed equally in phase currents. But the radiation pattern of a collinear array has circular symmetry with its main lobe everywhere normal to the principal axis. This is reason why this array is called broadcast or Omni-directional arrays. Simple collinear array consists of two elements: however, this array can also have more than two elements (Fig. 4). The performance characteristic of array does not depend directly on the number of elements in the array. For example, the power gain for collinear array of 2, 3, and 4 elements are respectively 2 dB, 3.2 dB and 4.4 dB respectively. The power gain of 4.4 dB obtained by this array is comparatively lower than the gain obtained by other arrays or devices. The collinear array provides maximum gain when spacing between elements is of the order of 0.3λ to 0.5λ; but this much spacing results in constructional and feeding difficulties. The elements are operated with their ends are much close to each other and joined simply by insulator



Increase in the length of collinear arrays increases the directivity: however, if the number of elements in an array is more (3 or 4), in order to keep current in phase in all the elements, it is essential to connect phasing stubs between adjacent elements. A collinear array is usually mounted vertically in order to increase overall gain and directivity in the horizontal direction. Stacking of dipole antennas in the fashion of doubling their number with proper phasing produces a 3 dB increase in directive gain.

Parasitic Arrays-In some way it is similar to broad side array, but only one element is fed directly from source, other element arc electromagnetically coupled because of its proximity to the feed element. Feed element is called driven element while other elements are called parasitic elements. A parasitic element lengthened by 5% to driven element act as reflector and another element shorted by 5% acts as director. Reflector makes the radiation maximum in perpendicular direction toward driven element and direction helps in making maximum radiation perpendicular to next parasitic element. The simplest parasitic array has three elements: reflector, driven element and director, and is used, for example in Yagi-Uda array antenna. The phase and amplitude of the current induced in a parasitic element depends upon its tuning and the spacing between elements and driven element to which it is coupled. Variation in spacing between driven element and parasitic elements changes the relative phases and this proves to be very convenient. It helps in making the radiation pattern unidirectional. A distance of λ/4 and phase difference of π/2 radian provides a unidirectional pattern. A properly designed parasitic array with spacing 0.1λ to 0.15λ provides a frequency bandwidth of the order of 2%, gain of the order of 8 dB and FBR of about 20 dB. It is of great practical importance, especially at higher frequencies between 150 and 100 MHz, for Yagi array used for TV reception

UNIT-IV

Pre-requisite:

- To know the fundamental limits on the diversity of wireless channels

Outcomes:

- Analyze mobile communication systems for improved performance..

INTRODUCTION

Diversity techniques that mitigate the effect of multipath fading are called **microdiversity**, and that is the focus of this chapter. Diversity to mitigate the effects of shadowing from buildings and objects is called **macrodiversity**. Macrodiversity is generally implemented by combining signals received by several base stations or access points. This requires coordination among the different base stations or access points. Such coordination is implemented as part of the networking protocols in infrastructure-based wireless networks. We will therefore defer discussion of macro diversity where we discuss the design of such networks.

There are many ways of achieving independent fading paths in a wireless system. One method is to use multiple transmit or receive antennas, also called an antenna array, where the elements of the array are separated in distance. This type of diversity is referred to as *space diversity*. Note that with receiver space diversity, independent fading paths are realized without an increase in transmit signal power or bandwidth. Moreover, coherent combining of the diversity signals leads to an increase in SNR at the receiver over the SNR that would be obtained with just a single receive antenna, which we discuss in more detail below. Conversely, to obtain independent paths through transmitter space diversity, the transmit power must be divided among multiple antennas. Thus, with coherent combining of the transmit signals the received SNR is the same as if there were just a single transmit antenna. Space diversity also requires that the separation between antennas be such that the fading amplitudes corresponding to each antenna are approximately independent

Directional antennas provide angle, or directional, diversity by restricting the receive antenna beamwidth to a given angle. In the extreme, if the angle is very small then at most one of the multipath rays will fall within the receive beam width, so there is no multipath fading from multiple rays. However, this diversity technique requires either a sufficient number of directional antennas to span all possible directions of arrival or a single antenna whose directivity can be steered to the angle of arrival of one of the multipath components (preferably the strongest one). Note also that with this technique the SNR may decrease due to the loss of multipath components that fall outside the receive antenna beam width, unless the directional gain of the antenna is sufficiently large to compensate for this lost power. Smart antennas are antenna arrays with adjustable phase at each antenna element: such arrays form directional antennas that can be steered to the incoming angle of the strongest multipath component

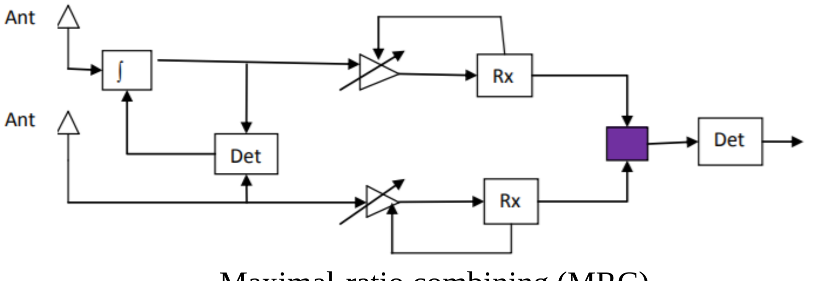
Receiver Diversity

It is important to combine the uncorrelated faded signals which were obtained from the diversity branches to get proper diversity benefit. The combing system should be in such a manner that improves the performance of the communication system. Diversity combing also increases the signal-to-noise ratio (SNR) or the power of received signal. Mainly, the combining should be applied in reception; however it is also possible to apply in transmission. There are many diversity combining methods available but only three of them are going to be discussed here.
Maximal ratio combining (MRC) ¬ Equal gain combining (EGC) ¬ Selection combining (SC)
The combining processes which use to combine multiple diversity branches in the reception, has two classes such as post-detection combing and pre-detection combining. The signals from diversity branches are combined coherently before detection in pre-detection combining. However, signals are detected individually before combining in post-detection. The performance of communication system is the same for both combining techniques for coherent detection. However, the performance of communication system is better by using pre-detection combining for non-coherent detection. It does mean that there is no effect in performance by the type of combining procedure for the coherent modulation case. The postdetection combining is not complex in non-coherent detection, results very common in use. There is a difference in system performance when used pre-detection combining and postdetection combining for non-coherent detection such as frequency modulation

(FM) discriminator or differential detection schemes. Moreover, the terms pre-detection and post-detection are also indicates the time of combining means when the combining is performed, before or after the hard decision. Squire-law non-coherent combining is employed frequently in diversity reception when noncoherent modulation methods are used. The demodulator outputs of all diversity branches are squired and summed to form a decision variable when used squire-law pre-detection combining. The system performance is decreased in non-coherent combining comparing to coherent combining and the degradation is called combining loss.

Maximal Ratio Combining (MRC)

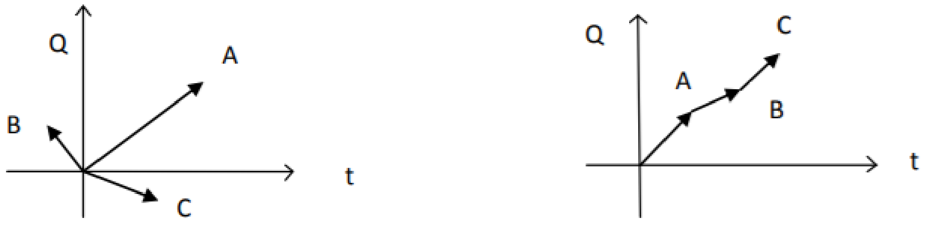
This is a very useful combining process to combat channel fading. This is the best combining process which achieves the best performance improvement comparing to other methods. The MRC is a commonly used combining method to improve performance in a noise limited communication systems where the AWGN and the fading are independent amongst the diversity branches. But the MRC employment needs summing circuits, weighting and co phasing. In the MRC combining technique, the signals from different diversity branches are co-phased and weighted before summing or combining. The weights have to be chosen as proportional to the respective signals level for maximizing the combined carrier-to-noise ratio (CNR). The applied weighting to the diversity branches has to be adjusted according to the SNR. For maximizing the SNR and minimizing the probability of error at the output combiner, signals of diversity branch is weighted before making sum with others by a factor, . Here is noise variance of diversity branch and is the complex conjugate of channel gain [1]. As a result the phase-shifts are compensated in the diversity channels and the signals coming from strong diversity branches which has low level noise, are weighted more comparing to the signals from the weak branches with high level of noise. The term in weighting can be neglected conditioning that has equal value for all d. Then the realization of the combiner needs the estimation of gains in complex channel and it does not need any estimation of the power of noise. It is feasible to employ MRC in transmission process of transmit diversity. But in this case the transmitter should get proper feedback information about the sub-channels state between single receive antenna and multiple transmit antennas. However, it is not feasible to weight transmissions from multiple antennas optimally for every receiving antenna, in a combined transmit-receive diversity channel. Moreover, if interference is limited in a communication system, then there is a scheme which combines the diversity branches in order to maximize the signal-to-interference-plus-noise ratio may allow much better performance than MRC provides. The assumption is valid for spatially white Gaussian noise if we can observe noise power at the receiver where just thermal noise is accounted. If we use the same type antenna elements then the thermal noise power is uncorrelated and equal for each branch.



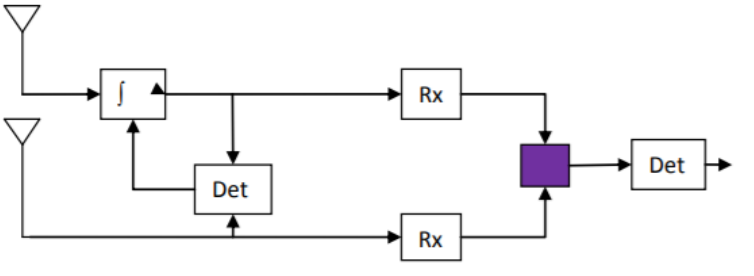
Maximal-ratio combining (MRC)

Equal-gain Combining (EGC):

MRC is the most ideal diversity combining but the scheme requires very expensive design at receiver circuit to adjust the gain in every branch. It needs an appropriate tracking for the complex fading, which very difficult to achieve practically. However, by using a simple phase lock summing circuit, it is very easy to implement an equal gain combining. The EGC is similar to MRC with an exception to omit the weighting circuits. The performance improvement is little bit lower in EGC than MRC because there is a chance to combine the signals with interference and noise, with the signals in high quality which are interference and noise free. EGC's normal procedure is coherently combined the individual signal branch but it non-coherently combine some noise components according to following figure:



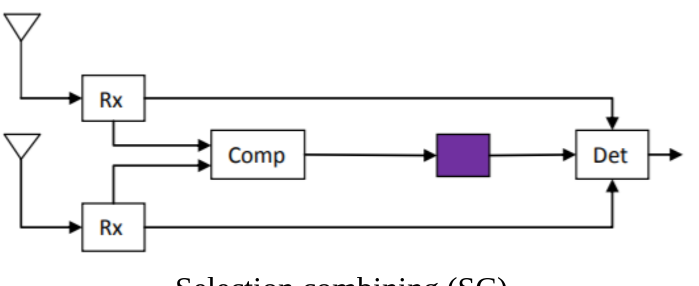
The EGC can employ in the reception of diversity with coherent modulation. The envelope gains of diversity channels are neglected in EGC and the diversity branches are combined here with equal weights but conjugate phase. The structure of equal-gain combining (EGC) is as following since there is no envelope gain estimation of the channel.



Equal gain combining (EGC)

Selection Combining (SC)

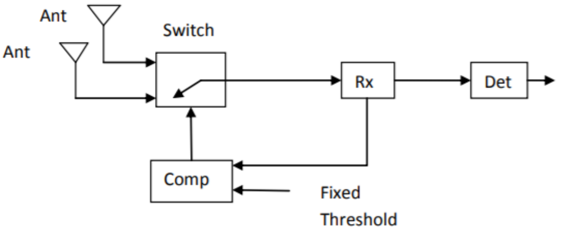
MRC and EGC are not suitable for very high frequency (VHF), ultra high frequency (UHF) or mobile radio applications. Realization of a co-phasing circuit with precise and stable tracking performance is not easy in a frequently changing, multipath fading and random-phase environment. SC is more suitable comparing to MRC and EGC in mobile radio application because of simple implementation procedure. In SC, the diversity branch which has the highest signal level has to be selected. Therefore, the main algorithm of this method is on the base of principle to select the signal amongst the all signals at the receiver end. If there is even a fast multipath fading environment, the stable operation easily can be achieved. It is experimentally proved that the performance improvement achieved by the selection combining is just little lower than performance improved achieved by an ideal MRC. As a result the SC is the most used diversity technique in wireless communication. The general form of selection combining is to monitor all the diversity branches and select the best one (the one which has the highest SNR) for detection. Therefore we can say that SC is not a combining method but a selection procedure at the available diversity. However, measuring SNR is quite difficult because the system has to select it in a very short time. But selecting the branch with the highest SNR is similar to select the branch with highest received power when average power of noise is the same on each branch. Therefore, it is practical to select the branch which has the largest signal composition, noise and interference. If there is an availability of feedback information about the channel state of the diversity branch the selection combining also can be used in transmission.



Selection combining (SC)

Switched Combining (SWC):

It is impractical to monitor the all diversity branches in selection combining. In addition, if we want to monitor the signals continuously then we need the same number of receivers and branches. Therefore, the form of switched combining is used to implement selection combining. According to the figure (a), switching from branch to branch occurs when the signal level falls under threshold. The value of threshold is fixed under a small area but the value is not the best necessarily over the total service area. As a result the threshold needs to be set frequently according to the movement of vehicle fig (b). It is very important to determine the optimal switching threshold in SWC. If the value of threshold is very high, then the rate of undesirable switching transient increases. However, if the threshold is very low then the diversity gain is also very low. The switching of switch combining can be performed periodically in the case of frequency hopping systems. Performance improvement obtained by the switching method leys on the value of threshold selection, the delay of time that creates from the loop of feedback of monitoring estimation, switching and decision. Moreover, phase transients and envelope of a carrier can reduce the improvement of performance. In the system of angle modulation, for example, GSM, the phase transient is responsible to create errors in detection stream of data. In this case, a predetection band pass filter may be used to remove envelope transients.



Performance of Maximal Ratio Combining:

If we consider that the fading is constant over one symbol period, the bit error probability of maximal-ratio combined the Quadrature Phase Shift Keying (QPSK) with gray coding over Rayleigh fading of D diversity branches. Which are corrupted by AWGN and having equal Signal to noise ratio. Then the equation of the bit error probability is:

$$Pb = \frac{1}{2}\left[1 - \frac{\mu}{\sqrt{2-\mu^2}} \sum_{d=0}^{D-1} \binom{2d}{d} \left(\frac{1-\mu^2}{4-2\mu^2}\right)^d\right]$$

Where is depends on the type of channel estimation and can be interpreted as cross correlation coefficient. cab be expressed using following equation in the case of coherent detection in conditioning perfect channel estimation.

$$\mu = \sqrt{\frac{\gamma_c}{1+\gamma_c}}$$

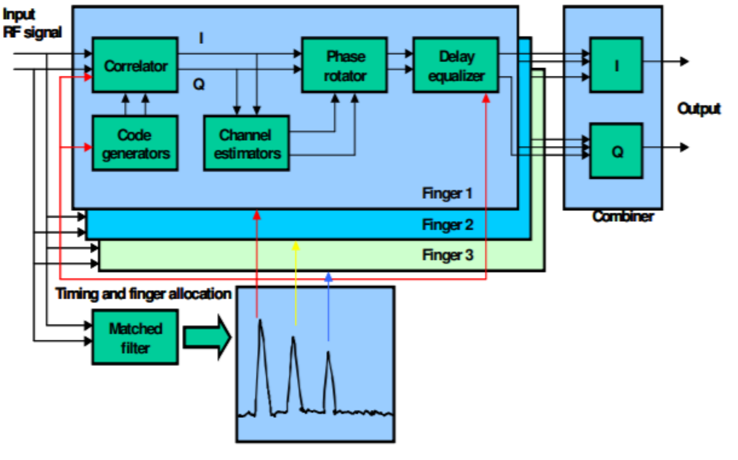
A QPSK signal can be expressed as two Binary Phase Shift Keying (BPSK) signals in phase quadrature [3]. As a result, if there is no cross talk or interference between signals on two quadratures then the bit error probability for QPSK and BPSK is similar. If there is no additional interference in an AWGN channel then the noise in the in-phase and quadrature components is independent. Therefore, the equation of bit error probability of coherent BPSK is used to find the bit error probability of coherent QPSK or the reverse way. On the other hand, if the bit error probability is a function of , the received SNR per channel in average must be normalized by the number of bits per symbol in QPSK.

RAKE Receiver

RAKE receiver is used in CDMA-based (Code Division Multiple Access) systems and can combine multipath components, which are time-delayed versions of the original signal transmission. Combining is done in order to improve the signal to noise ration at the receiver. RAKE receiver attempts to collect the time-shifted versions of the original signal by providing a separate correlation receiver for each of the multipath signals. This can be done due to multipath components are practically uncorrelated from another when their relative propagation delay exceeds a chip period

RAKE receiver, used specially in CDMA cellular systems, can combine multipath components, which are time-delayed versions of the original signal transmission. This combining is done in order to improve the signal to noise ratio (SNR) at the receiver. RAKE receiver attempts to collect the timeshifted versions of the original signal by providing a separate correlation receiver for each of the multipath signals. This can be done due to multipath components are practically uncorrelated from another when their relative propagation delay exceeds a chip period.

Due to reflections from obstacles a radio channel can consist of many copies of originally transmitted signals having different amplitudes, phases, and delays. If the signal components arrive more than duration of one chip apart from each other, a RAKE receiver can be used to resolve and combine them. The RAKE receiver uses a multipath diversity principle. It is like a rake that rakes the energy from the multipath propagated signal components



Block diagram of a RAKE receiver

When a signal is received in a matched filter over a multipath channel, the multiple delays appear at the receiver, as depicted in Figure 4. The RAKE receiver uses several baseband correlators to individually process several signal multipath components. The correlator outputs are combined to achieve improved communications reliability and performance. [2] Bit decisions based only a single correlation may produce a large bit error rate as the multipath component processed in that correlator can be corrupted by fading. In a RAKE receiver, if the output from one correlator is corrupted by fading, the others may not be, and the corrupted signal may be discounted through the weighting process.

Impulse response measurements of the multipath channel profile are executed through a matched filter to make a successful de-spreading. It reveals multipath channel peaks and gives timing and RAKE finger allocations to different receiver blocks. Later it tracks and monitors these peaks with a measurement rate depending on speeds of mobile station and on propagation environment. The number of available RAKE fingers depends on the channel profile and the chip rate. The higher the chip rate, the more resolvable paths there are, but higher chip rate will cause wider bandwidth. To catch all the energy from the channel more RAKE fingers are needed. A very large number of fingers lead to combining losses and practical implementation problems.

Equalization Techniques

To estimate and cater forthe transmitted signal symbols that has been impacted by Inter symbol Interference (ISI) and noisein the MIMO systemscommunication link, an equalization method is employed. A brief general overview ofthe four channel equalization techniques considered at the receiver in this work provided below. The receiver is assumed to possess a perfect knowledge of the channel state. Also, the weights of the equalizing filters are calculated dynamically.

One key efficient means to cater for the ISI multipath propagation environment is by employing a filtering method termed equalization . In this chapter, two linear and nonlinear channel equalization algorithms which includes the minimum-mean-square-error (MMSE), zero forcing (ZF), minimum-meansquare-error successive interference cancellation (MMSE-SIC) and Zero forcing- successive interference cancellation (ZF-SIC), are studied extensively through computer simulations to examine their impact on energy efficiency and data transmission capacity of large scale MIMO systems. In our previous chapters, the focus were solely on massive MIMOcommunication the downlink single cell scenarios. Even though the large-scale MIMO antenna systems proven to be very advantages in the previous chapters, it may be challenging to equip largescale MIMO antenna into mobile communication devices such as smart phones, palmtop computers and tablets.

Transmitter - The Alamouti Scheme

We now consider the same model as in the previous subsection but assume that the transmitter no longer knows the channel gains riej θ i , so there is no CSIT. In this case it is not obvious how to obtain diversity gain. Consider, for example, a naive strategy whereby for a two-antenna system we divide the transmit energy equally between the two antennas. Thus, the transmit signal on antenna i will be si(t) =√ 5s(t) for s(t) the transmit signal with energy per symbol Es. Assume each antenna has a complex Gaussian channel gain hi = riej θ i, i = 1, 2 with mean zero and variance one. The received signal is then

$$r(t) = \sqrt{.5}(h_1 + h_2)s(t).$$

so the received signal has the same distribution as if we had just used one antenna with the full energy per symbol. In other words, we have obtained no performance advantage from the two antennas, since we could not divide our energy intelligently between them or obtain coherent combining through co-phasing. Transmit diversity gain can be obtained even in the absence of channel information with an appropriate scheme to exploit the antennas. A particularly simple and prevalent scheme for this diversity that combines both space and time diversity was developed by Alamouti in [9]. Alamouti's scheme is designed for a digital communication system with two-antenna transmit diversity. The scheme works over two symbol periods where it is assumed that the channel gain is constant over this time. Over the first symbol period two different symbols s1 and s2 each with energy Es/2 are transmitted simultaneously from antennas 1 and 2, respectively. Over the next symbol period symbol −s∗ 2 is transmitted from antenna 1 and symbol s∗1 is transmitted from antenna 2, each with symbol energy Es/2.

$$\mathbf{y} = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} + \begin{bmatrix} n_1 \\ n_2^* \end{bmatrix} = \mathbf{H}_A\mathbf{s} + \mathbf{n},$$

where $\mathbf{s} = [s_1 s_2]^T$, $\mathbf{n} = [n_1 n_2]^T$, and

$$\mathbf{H}_A = \begin{bmatrix} h_1 & h_2 \\ h_2^* & -h_1^* \end{bmatrix}.$$

Let us define the new vector $\mathbf{z} = \mathbf{H}_A^H \mathbf{y}$. The structure of $\mathbf{H}_A$ implies that

$$H_A^H H_A = (|h_1^2| + |h_2^2|)\mathbf{I}_2,$$

is diagonal, and thus

$$\mathbf{z} = [z_1 \quad z_2]^T = (|h_1^2| + |h_2^2|)\mathbf{I}_2\mathbf{s} + \tilde{\mathbf{n}},$$

Where the factor of 2 comes from the fact that si is transmitted using half the total symbol energy Es. The received SNR is thus equal to the sum of SNRs on each branch, identical to the case of transmit diversity with MRC assuming that the channel gains are known at the transmitter. Thus, the Alamouti scheme achieves a diversity order of 2, the maximum possible for a two-antenna transmit system, despite the fact that channel knowledge is not available at the transmitter. However, it only achieves an array gain of 1, whereas MRC can achieve an array gain and a diversity gain of 2. The Alamouti scheme can be generalized for M > 2 when the constellations are real, but if the contellations are complex the generalization is only possible with a reduction in code rates

UNIT-V

Pre-requisite:

- To understand the multiple input multiple output (MIMO) systems design and analysis.

OUTCOMES:

- Analyze the diversity multiplexing and GSM, EDGE, GPRS, IS-95, CDMA 2000 and WCDMA.

INTRODUCTION

Wireless communication using multiple-input multiple-output (MIMO) systems enables increased spectral efficiency for a given total transmit power. Increased capacity is achieved by introducing additional spatial channels that are exploited by using space-time coding. In this article, we survey the environmental factors that affect MIMO capacity. These factors include channel complexity, external interference, and channel estimation error. We discuss examples of space-time codes, including space-time low-density parity-check codes and spacetime turbo codes, and we investigate receiver approaches, including multichannel multiuser detection (MCMUD). The 'multichannel' term indicates that the receiver incorporates multiple antennas by using space-time-frequency adaptive processing.

These spectral efficiency gains often require accurate knowledge of the channel at the receiver, and sometimes at the transmitter as well. In addition to spectral efficiency gains, ISI and interference from other users can be reduced using smart antenna techniques. The cost of the performance enhancements obtained through MIMO techniques is the added cost of deploying multiple antennas, the space and power requirements of these extra antennas (especially on small handheld units), and the added complexity required for multi-dimensional signal processing. In this chapter we examine these different uses for multiple antennas and find their performance advantages.

MIMO information-theoretic performance bounds in more detail in the next section. Capacity increases linearly with signal-to-noise ratio (SNR) at low SNR, but increases logarithmically with SNR at high SNR. In a MIMO system, a given total transmit power can be divided among multiple spatial paths (or modes), driving the capacity closer to the linear regime for each mode, thus increasing the aggregate spectral efficiency. As seen in Figure 1, which assumes an optimal high spectral-efficiency MIMO channel (a channel matrix with a flat singular-value distribution), MIMO systems enable high spectral efficiency at much lower required energy per information bit.

space time signal processing

In order to implement a MIMO communication system, we must first select a particular coding scheme. Most space-time coding schemes have a strong connection to well-known single-input single-output (SISO) coding approaches and assume an uninformed transmitter (UT). Later in the article we discuss space-time low-density parity-check codes, space-time turbo codes, and their respective receivers. Space-time coding can exploit the MIMO degrees of freedom to increase redundancy, spectral efficiency, or some combination of these characteristics.
There are a variety of extensions of LDPC codes to space-time codes, which are introduced and explained in the sidebar entitled "Space-Time Codes." For the experiments described below, only one type of extension was considered. Each space-time channel transmits one of several possible quadrature phase-shift keying (QPSK) waveforms with slightly offset carrier frequencies. The differential frequencies are sufficiently large to effectively decorrelate the transmitted waveforms over the length of a codeword (1024 bits) even if the data sequences in each channel are identical. These differential frequencies are also large compared to the expected Doppler spreads and small compared to the signal bandwidth. In the simplest example of such a code, the I and Q components of a transmitter represent, respectively, two different LDPC codewords. Each transmitter sends the same complex baseband sequence (QPSK) shifted in frequency. The transmitter outputs, viewed collectively as a vector at any instant, vary in time and thus effectively probe the environment characterized by the channel matrix. Since the transmitted vector varies significantly over the duration of a codeword, the coding provides spatial diversity.

Decoding occurs by forming likelihood ratios based on channel-matrix estimates and then using the iterative decoder described above. Note that the channel matrix can change during the codeword, in which case channel-matrix estimates can vary sample to sample. The LDPC space-time code just described exhibits full spatial redundancy among all transmitters. Less redundancy, and therefore higher data rates, can be achieved by dividing the transmitters into subsets, each of which is fully redundant yet different from any other subset. For example, the space-time code discussed later, in the section on experiments, uses four transmitters. The first two transmitters send two bits (redundant in I and Q) of a symbol of an LDPC code word over While the theoretical performance is determined by the channel phenomenology, practical MIMO performance requires the selection of a space-time code and an appropriate matched receiver. In this section we discuss the space-time turbo code used in this example. We develop a maximum-likelihood formulation of a multiple-antenna multiuser receiver, and we discuss suboptimal implementations of the receiver. We also introduce minimum-mean-squared-error extensions of the receiver, and we discuss the value and use of training data.

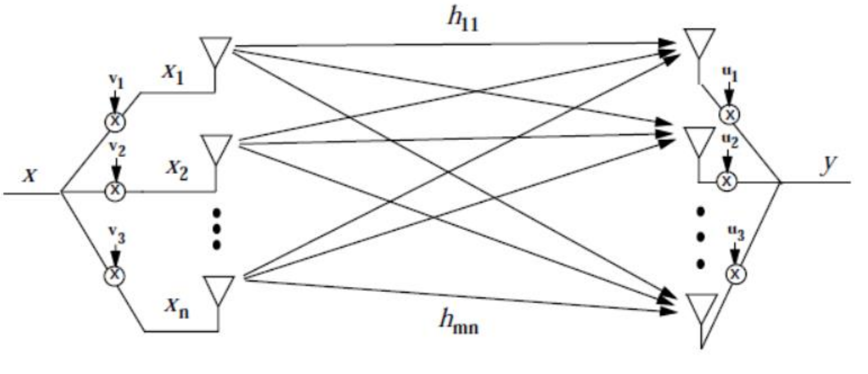
MIMO Diversity Gain: Beamforming

The multiple antennas at the transmitter and receiver can be used to obtain diversity gain instead of capacity gain. In this setting, the same symbol, weighted by a complex scale factor, is sent over each transmit antenna, so that the input covariance matrix has unit rank. This scheme is also referred to as **MIMO beamforming**. A beamforming strategy corresponds to the precoding and receiver matrices described in Section 10.2 being just column vectors:

V = **v** and **U** = **u**, the transmit symbol x is sent over the ith antenna with weight vi. On the receive side, the signal received on the ith antenna is weighted by ui. Both transmit and receive weight vectors are normalized so that $||u|| = ||v|| = 1$. The resulting received signal is given by

$$y = \mathbf{u}^*\mathbf{H}\mathbf{v}x + \mathbf{u}^*\mathbf{n},$$

where if $\mathbf{n} = (n1, \ldots, nMr)$ has i.i.d. elements, the statistics of $\mathbf{u}*\mathbf{n}$ are the same as the statistics for each of these elements.



Beamforming provides diversity gain by coherent combining of the multiple signal paths. Channel knowledge at the receiver is typically assumed since this is required for coherent combining. The diversity gain then depends on whether or not the channel is known at the transmitter. When the channel matrix **H** is known, the received SNR is optimized by choosing **u** and **v** as the principal left and right singular vectors of the channel matrix **H**. The corresponding received SNR can be shown to equal $\gamma = \lambda max\rho$, where λmax is the largest eigenvalue of the **Wishart matrixW**= **HH**H [21]. The resulting capacity is $C = B \log2(1 + \lambda max\rho)$, corresponding to the capacity of a SISO channel with channel power gain λmax. When the channel is not known at the transmitter, the transmit antenna weights are all equal, so the received SNR equals $\gamma = ||\mathbf{Hu}*||$, where **u** is chosen to maximize γ. Clearly the lack of transmitter CSI will result in a lower SNR and capacity than with optimal transmit weighting. While beamforming has a reduced capacity relative to optimizing the transmit precoding and receiver shaping matrices, the optimal demodulation complexity with beamforming is of the order of $|X|$ instead of $|X|RH$. An even simpler strategy is to use MRC at either the transmitter or receiver and antenna selection on the other end.

Diversity/Multiplexing Tradeoffs

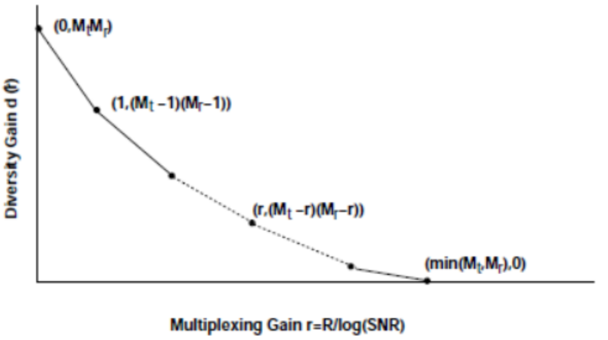
One option is to obtain capacity gain by decomposing the MIMO channel into parallel channels and multiplexing different data streams onto these channels. This capacity gain is also referred to as a **multiplexing gain**. However, the SNR associated with each of these channels depends on the singular values of the channel matrix.

In capacity analysis this is taken into account by assigning a relatively low rate to these channels. However, practical signaling strategies for these channels will typically have poor performance, unless powerful channel coding techniques are employed. Alternatively, beamforming can be used, where the channel gains are coherently combined to obtain a very robust channel with high diversity gain. It is not necessary to use the antennas purely for multiplexing or diversity. Some of the space-time dimensions can be used for diversity gain, and the remaining dimensions used for multiplexing gain.

The diversity/multiplexing tradeoff or, more generally, the tradeoff between data rate, probability of error, and complexity for MIMO systems has been extensively studied in the literature, from both a theoretical perspective and in terms of practical space-time code designs [50, 37, 38, 42]. This work has primarily focused on block fading channels with receiver CSI only since when both transmitter and receiver know the channel the tradeoff is relatively straightforward: antenna subsets can first be grouped for diversity gain and then the multiplexing gain corresponds to the new channel with reduced dimension due to the grouping. For the block fading model with receiver CSI only, as the blocklength grows asymptotically large, full diversity gain and full multiplexing gain (in terms of capacity with outage) can be obtained simultaneously with reasonable complexity by encoding diagonally across antennas [51, 52, 2]. An example of this type of encoding is D-BLAST, For finite blocklengths it is not possible to achieve full diversity and full multiplexing gain simultaneously, in which case there is a tradeoff between these gains. A simple characterization of this tradeoff is given in [37] for block fading channels with blocklength $T \geq Mt + Mr - 1$ in the limit of asymptotically high SNR. In this analysis a transmission scheme is said to achieve multiplexing gain r and diversity gain d if the data rate (bps) per unit Hertz R(SNR) and probability of error Pe(SNR) as functions of SNR satisfy

$$\lim_{\log_2 \mathrm{SNR} \to \infty} \frac{R(\mathrm{SNR})}{\log_2 \mathrm{SNR}} = r,$$

$$\lim_{\log \mathrm{SNR} \to \infty} \frac{\log P_e(\mathrm{SNR})}{\log \mathrm{SNR}} = -d,$$



It is also possible to adapt the diversity and multiplexing gains relative to channel conditions. Specifically, in poor channel states more antennas can be used for diversity gain, whereas in good states more antennas can be used for multiplexing. Adaptive techniques that change antenna use to trade off diversity and multiplexing based on channel conditions have been investigated.

Space-Time Modulation and Coding

Since a MIMO channel has input-output relationship $\mathbf{y} = \mathbf{Hx} + \mathbf{n}$, the symbol transmitted over the channel each symbol time is a vector rather than a scalar, as in traditional modulation for the SISO channel. Moreover, when the signal design extends over both space (via the multiple antennas) and time (via multiple symbol times), it is typically referred to as a **space-time code**.

Most space-time codes, including all codes discussed in this section, are designed for quasi-static channels where the channel is constant over a block of T symbol times, and the channel is assumed unknown at the transmitter.

Under this model the channel input and output become matrices, with dimensions corresponding to space (antennas) and time. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_T]$ denote the $M_t \times T$ channel input matrix with ith column $\mathbf{x}_i$ equal to the vector channel input over the ith transmission time. Let $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_T]$ denote the $M_r \times T$ channel output matrix with ith column $\mathbf{y}_i$ equal to the vector channel output over the ith transmission time, and let $\mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_T]$ denote the $M_r \times T$ noise matrix with ith column $\mathbf{n}_i$ equal to the receiver noise vector on the ith transmission time. With this matrix representation the input-output relationship over all T blocks becomes
$\mathbf{Y} = \mathbf{HX}+\mathbf{N}$.

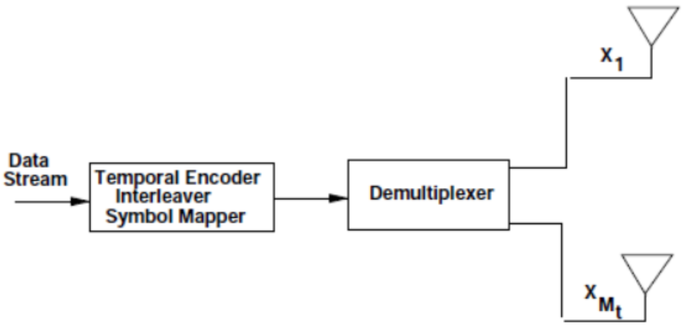
Space-Time Trellis and Block Codes

The rank and determinant criteria have been primarily applied to the design of space-time trellis codes (STTCs). STTCs are an extension of conventional trellis codes to MIMO systems [10, 44]. They are described using a trellis and decoded using ML sequence estimation via the Viterbi algorithm. STTCs can extract excellent diversity and coding gain, but the complexity of decoding increases exponentially with the diversity level and transmission rate Space-time block codes (STBCs) are an alternative space-time code that can also extract excellent diversity and coding gain with linear receiver complexity. Interest in STBCs were initiated by the Alamouti code described in which obtains full diversity order with linear receiver processing for a two-antenna transmit system.

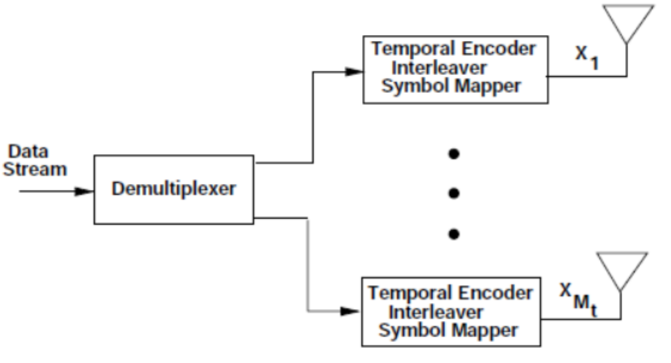
This scheme was generalized in to STBCs that achieve full diversity order with an arbitrary number of transmit antennas. However, while these codes achieve full diversity order, they do not provide coding gain, and thus have inferior performance to STTCs, which achieve both full diversity gain as well as coding gain. Added coding gain for both STTCs and STBCs can be achieved by concatenating these codes either in serial or in parallel with an outer channel code to form a turbo code . The linear complexity of the STBC designs result from making the codes orthogonal along each dimension of the code matrix. A similar design premise is used in to design **unitary space-time modulation** schemes for block fading channels when neither the transmitter nor the receiver have channel CSI. More comprehensive treatments of space-time coding can be found in and the references therein.

Spatial Multiplexing and BLAST Architectures

The basic premise of spatial multiplexing is to send M_t independent symbols per symbol period using the dimensions of space and time. In order to get full diversity order an encoded bit stream must be transmitted over all M_t transmit antennas. This can be done through a serial encoding, illustrated in Figure 10.10. With serial encoding the bit stream is temporally encoded over the channel blocklength T, interleaved, and mapped to a constellation point, then demultiplexed onto the different antennas. If each codeword is sufficiently long, it can be transmitted over all M_t transmit antennas and received by all M_r receive antennas, resulting in full diversity gain. However, the codeword length T required to achieve this full diversity is M_tM_r, and decoding complexity grows exponentially with this codeword length. This high level of complexity makes serial encoding impractical.



A simpler method to achieve spatial multiplexing, pioneered at Bell Laboratories as one of the Bell Labs Layered Space Time (BLAST) architectures for MIMO channels [2], is parallel encoding, With parallel encoding the data stream is demultiplexed into *Mt* independent streams. Each of the resulting substreams is passed through a SISO temporal encoder with blocklenth *T*, interleaved, mapped to a signal constellation point, and transmitted over its corresponding transmit antenna. This process can be considered to be the encoding of the serial data into a vertical vector, and hence is also referred to as vertical encoding or V-BLAST Vertical encoding can achieve at most a diversity order of *Mr*, since each coded symbol is transmitted from one antenna and received by *Mr* antennas. This system has a simple encoding complexity that is linear in the number of antennas. However, optimal decoding still requires joint detection of the code words from each of the transmit antennas, since all transmitted symbols are received by all the receive antennas. It that the receiver complexity can be significantly reduced through the use of symbol interference cancellation, as The symbol interference cancellation, which exploits the synchronicity of the symbols transmitted from each antenna, works as follows. First the *Mt* transmitted symbols are ordered in terms of their received SNR. An estimate of the received symbol with the highest SNR is made while treating all other symbols as noise. This estimated symbol is subtracted out, and the symbol with the next highest SNR estimated while treating the remaining symbols as noise. This process repeats until all *Mt* transmitted symbols have been estimated. After cancelling out interfering symbols, the coded substream associated with each transmit antenna can be individually decoded, resulting in a receiver complexity that is linear in the number of transmit antennas. In fact, coding is not even needed with this architecture, and data rates of 20-40 bps/Hz with reasonable error rates were reported in using uncoded V-BLAST



GSM, EDGE, GPRS, IS-95, CDMA 2000 and WCDMA

These systems were mostly deployed in the early 1990s. Due to incompatibilities in the first-generation analog systems, in 1982 the Groupe Sp´ecial Mobile (GSM) was formed to develop a uniform digital cellular standard for all of Europe. The TACS spectrum in the 900 MHz band was allocated for GSM operation across Europe to facilitate roaming between countries. In 1989 the GSM specification was finalized and the system was launched in 1991, although availability was limited until 1992. The GSM standard uses TDMA combined with slow frequency hopping to combat out-of-cell interference. Convolutional coding and parity check codes along with interleaving is used for error correction and detection. The standard also includes an equalizer to compensate for frequency-selective fading. The GSM standard is used in about 66 % of the world's cellphones, with more than 470 GSM operators in 172 countries supporting over a billion users. As the GSM standard became more global, the meaning of the acronym was changed to the Global System for Mobile Communications.

A competing standard for 2G systems based on CDMA was proposed by Qualcomm in the early 1990s. The standard, called IS-95 or IS-95a, was finalized in 1993 and deployed commercially under the name cdmaOne in 1995. Like IS-136, IS-95 was designed to be compatible with AMPS so that the two systems could coexist in the same frequency band. In CDMA all users are superimposed on top of each other with spreading codes that can separate out the users at the receiver. Thus, channel data rate does not apply to just one user, as in TDMA systems. The channel chip rate is 1.2288 Mchips/s for a total spreading factor of 128 for both the uplink and downlink. The spreading process in IS-95 is different for the downlink (DL) and the uplink (UL), with spreading on both links accomplished through a combination of spread spectrum modulation and coding. On the downlink data is first rate 1/2 convolutionally encoded and interleaved, then modulated by one of 64 orthogonal spreading sequences (Walsh functions). Then a synchronized scrambling sequence unique to each cell is superimposed on top of theWalsh function to reduce interference between cells. The scrambling requires synchronization between base stations. Uplink spreading is accomplished using a combination of a rate 1/3 convolutional code with interleaving, modulation by an orthogonal Walsh function, and modulation by a nonorthogonal user/base station specific code.

The IS-95 standard includes a parity check code for error detection, as well as power control for the reverse link to avoid the near-far problem. A 3-finger RAKE receiver is also specified to provide diversity and compensate for ISI. A form of base station diversity called soft handoff (SHO), whereby a mobile maintains a connection to both the new and old base stations during handoff and combines their signals, is also included in the standard. CDMA has some advantages over TDMA for cellular systems, including no need for frequency planning, SHO capabilities, the ability to exploit voice activity to increase capacity, and no hard limit on the number of users that can be accommodated in the system. There was much debate about the relative merits of the IS-54 and IS-95 standards throughout the early 1990s, with claims that IS-95 could achieve 20 times the capacity of AMPS whereas IS-54 could only achieve 3 times this capacity. In the end, both systems turned out to achieve approximately the same capacity increase over AMPS.

	GSM	IS-136	IS-95 (cdmaOne)	PDC
Uplink Frequencies (MHz)	890-915	824-849	824-849	810-830,1429-1453
Downlink Frequencies (MHz)	935-960	869-894	869-894	940-960, 1477-1501
Carrier Separation (KHz)	200	30	1250	25
Number of Channels	1000	2500	$\sim$ 2500	3000
Modulation	GMSK	$\pi/4$ DQPSK	BPSK/QPSK	$\pi/4$ DQPSK
Compressed Speech Rate (Kbps)	13	7.95	1.2-9.6 (Variable)	6.7
Channel Data Rate (Kbps)	270.833	48.6	(1.2288 Mchips/s)	42
Data Code Rate	1/2	1/2	1/2 (DL), 1/3 (UL)	1/2
ISI Reduction/Diversity	Equalizer	Equalizer	RAKE, SHO	Equalizer
Multiple Access	TDMA/Slow FH	TDMA	CDMA	TDMA

W-CDMA is the primary competing 3G standard to cdma2000. It has been selected as the 3G successor to GSM, and in this context is refered to as the Universal Mobile Telecommunications System (UMTS). W-CDMA is also used in the Japanese FOMA and J-Phone 3G systems. These different systems share the W-CDMA link layer protocol (air interface) but have different protocols for other aspects of the system such as routing and speech compression. W-CDMA supports peak rates of up to 2.4 Mbps, with typical rates anticipated in the 384 Kbps range. W-CDMA uses 5 MHz channels, in contrast to the 1.25 MHz channels of cdma2000. An enhancement to W-CDMA called High Speed Data Packet Access (HSDPC) provides data rates of around 9 Mbps, and this may be the precursor to 4th-generation systems.

3G Standard	cdma2000				W-CDMA		
Subclass	1X	1XEV-DO	1XEV-DV	3X	UMTS	FOMA	J-Phone
Channel Bandwidth (MHz)	1.25	1.25		3.75	5		
Chip Rate (Mchips/s)	1.2288			3.6864	3.84		
Peak Data Rate (Mbps)	.144	2.4	4.8	5-8	2.4 (8-10 with HSDPA)		
Modulation	QPSK (DL), BPSK (UL)						
Coding	Convolutional (low rate), Turbo (high rate)						
Power Control	800 Hz				1500 Hz		