

MODULE-5

CELLULAR CONCEPT

The design objective of early mobile radio systems was to achieve a large coverage area by using a single, high powered transmitter with an antenna mounted on a tall tower. While this approach achieved very good coverage, it also meant that it was impossible to reuse those same frequencies throughout the system, since any attempts to achieve frequency reuse would result in interference. Faced with the fact that government regulatory agencies could not make spectrum allocations in proportion to the increasing demand for mobile services, it became imperative to restructure the radio telephone system to achieve high capacity with limited radio spectrum while at the same time covering very large areas.

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. 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*.

Figure 1 illustrates 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 shown in Figure 1 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 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.

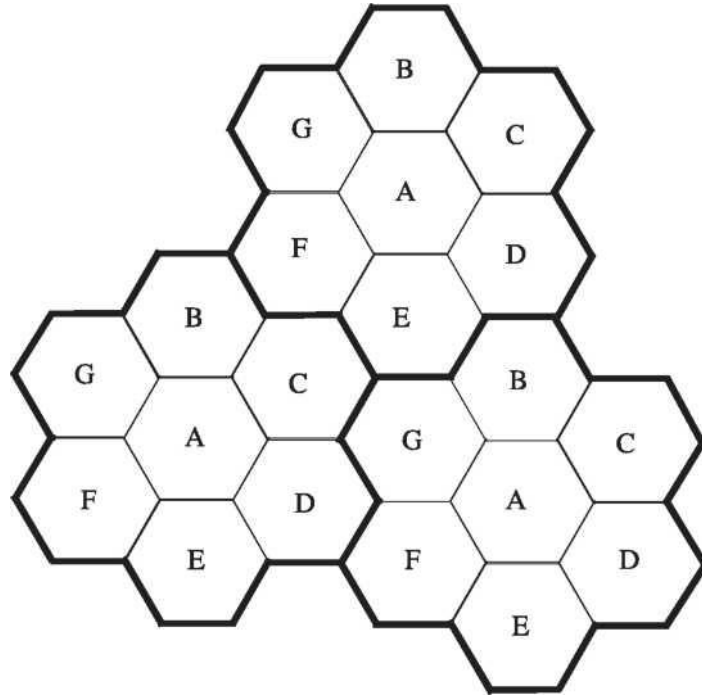


Figure 1 Illustration of the cellular frequency reuse concept. Cells with the same letter use the same set of frequencies. A cell cluster is outlined in bold and replicated over the coverage area. In this example, the cluster size, N , is equal to seven, and the frequency reuse factor is $1/7$ since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

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

$$S = k N \quad (1)$$

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

$$C = M k N = M S \quad (2)$$

As seen from Equation (2), 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 larger cluster size

causes the ratio between the cell radius and the distance between co-channel cells to decrease, leading to weaker co-channel interference. 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 in Equation (2)). 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 Figure 1 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. 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 which satisfy Equation (3).

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

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
- (2) Turn 60 degrees counter-clockwise and move j cells.

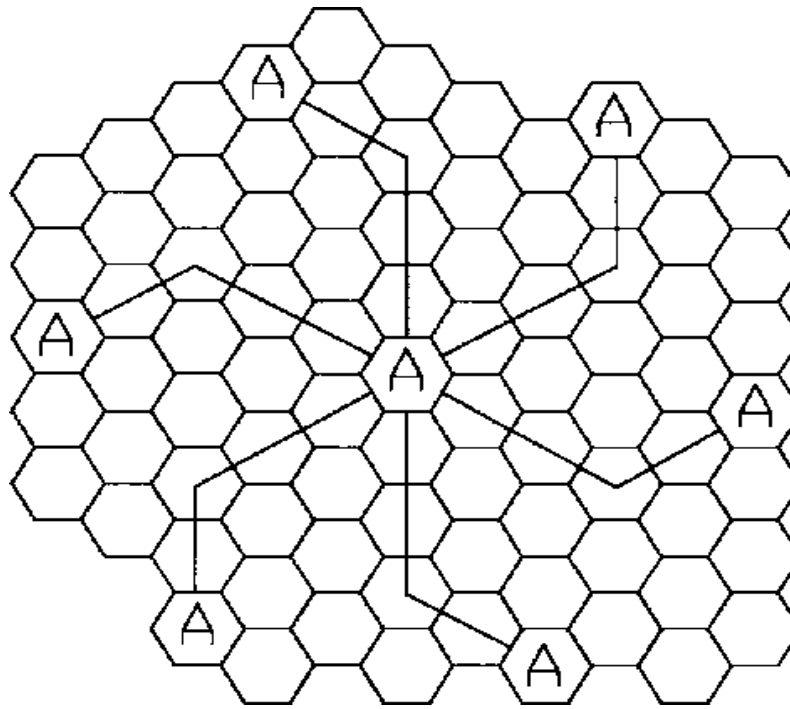


Figure 2 Method of locating co-channel cells in a cellular system. In this example, $N = 19$ (i.e., $i = 3, j = 2$).

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***. The choice of channel assignment strategy impacts the performance of the system, particularly as to how calls are managed when a mobile user is handed off from one cell to another .

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 :

$A = P_{r \text{ handoff}} - P_{r \text{ minimum usable}}$, cannot be too large or too small , unnecessary handoffs which burden the MSC may occur, and if A is too small, there may be insufficient time to complete a handoff before a call is lost due to weak signal conditions. Therefore, A is chosen carefully to meet these conflicting requirements. Figure 3 illustrates a handoff situation. Figure 3(a) 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 A 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. 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 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..

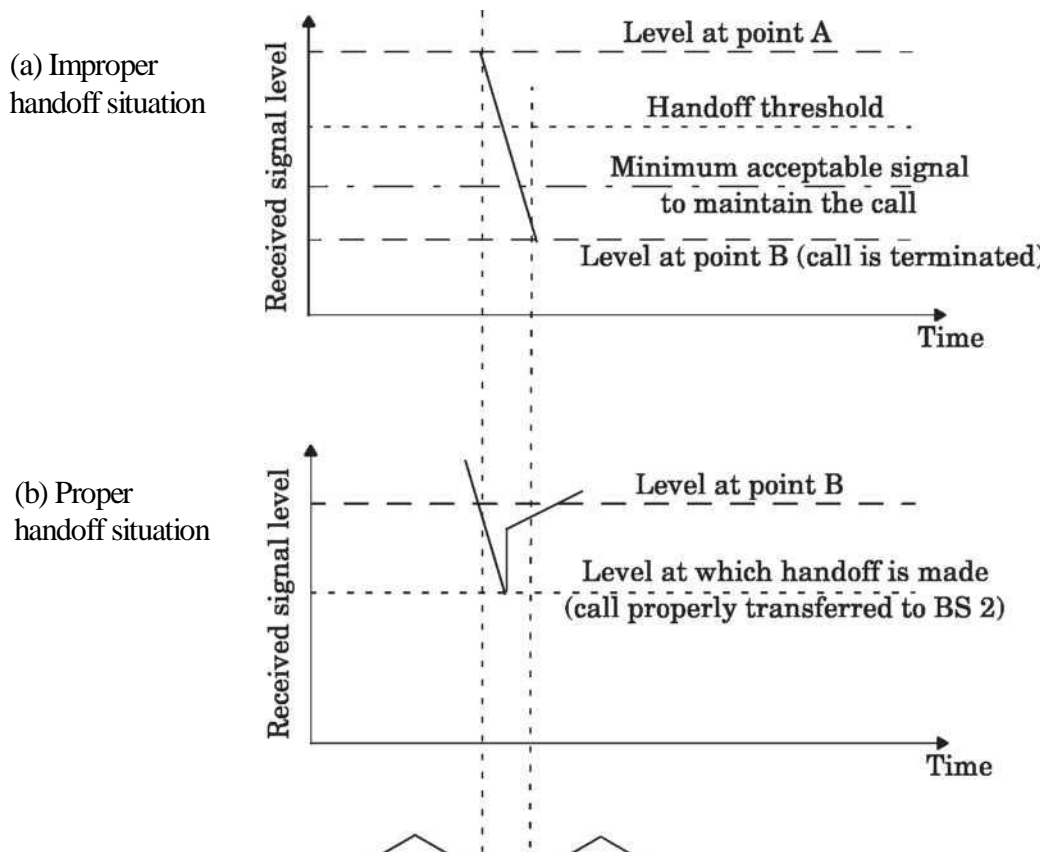


Figure 3 Illustration of a handoff scenario at cell boundary.

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.

Prioritizing Handoffs

One method for giving priority to handoffs is called the *guard channel concept*, whereby a fraction of the total available channels in a cell is reserved exclusively for handoff requests from ongoing calls which may be handed off into the cell. This method has the disadvantage of reducing the total carried traffic, as fewer channels are allocated to originating calls. Guard channels, however, offer efficient spectrum utilization when dynamic channel assignment strategies, which minimize the number of required guard channels by efficient demand-based allocation, are used.

Queuing of handoff requests is another method to decrease the probability of forced termination of a call due to lack of available channels. There is a tradeoff between the decrease in probability of forced termination and total carried traffic. Queuing of handoffs is possible due to the fact that there is a finite time interval between the time the received signal level drops below the handoff threshold and the time the call is terminated due to insufficient signal level. The delay time and size of the queue is determined from the traffic pattern of the particular service area. It should be noted that queuing does not guarantee a zero probability of forced termination, since large delays will cause the received signal level to drop below the minimum required level to maintain communication and hence lead to forced termination.

Practical Handoff Considerations

In practical cellular systems, several problems arise when attempting to design for a wide range of mobile velocities. High speed vehicles pass through the coverage region of a cell within a matter of seconds, whereas pedestrian users may never need a handoff during a call. Particularly with the addition of microcells to provide capacity, the MSC can quickly become burdened if high speed users are constantly being passed between very small cells. Several schemes have been devised to handle the simultaneous traffic of high speed and low speed users while minimizing the handoff intervention from the MSC. Another practical limitation is

the ability to obtain new cell sites.

Although the cellular concept clearly provides additional capacity through the addition of cell sites, in practice it is difficult for cellular service providers to obtain new physical cell site locations in urban areas. Zoning laws, ordinances, and other nontechnical barriers often make it more attractive for a cellular provider to install additional channels and base stations at the same physical location of an existing cell, rather than find new site locations. By using different antenna heights (often on the same building or tower) and different power levels, it is possible to provide “large” and “small” cells which are co-located at a single location. This technique is called the *umbrella cell* approach and is used to provide large area coverage to high speed users while providing small area coverage to users traveling at low speeds. Figure 4 illustrates an umbrella cell which is colocated with some smaller microcells. The umbrella cell approach ensures that the number of handoffs is minimized for high speed users and provides additional microcell channels for pedestrian users. The speed of each user may be estimated by the base station or MSC by evaluating how rapidly the short-term average signal strength on the RVC changes over time, or more sophisticated algorithms may be used to evaluate and partition users. If a high speed user in the large umbrella cell is approaching the base station, and its velocity is rapidly decreasing, the base station may decide to hand the user into the co-located microcell, without MSC intervention.

Another practical handoff problem in microcell systems is known as *cell dragging*. Cell dragging results from pedestrian users that provide a very strong signal to the base station. Such a situation occurs in an urban environment when there is a line-of-sight (LOS) radio path between the subscriber and the base station. As the user travels away from the base station at a very slow speed, the average signal strength does not decay rapidly. Even when the user has traveled well beyond the designed range of the cell, the received signal at the base station may be above the handoff threshold, thus a handoff may not be made. This creates a potential interference and traffic management problem, since the user has meanwhile traveled deep within a neighboring cell. To solve the cell dragging problem, handoff thresholds and radio coverage parameters must be adjusted carefully.

In first generation analog cellular systems, the typical time to make a handoff, once the signal level is deemed to be below the handoff threshold, is about 10 seconds. This requires that the value for Δ be on the order of 6 dB to 12 dB. In digital cellular systems such as GSM, the mobile assists with the handoff procedure by determining the best handoff candidates, and the handoff, once the decision is made, typically requires only 1 or 2 seconds.

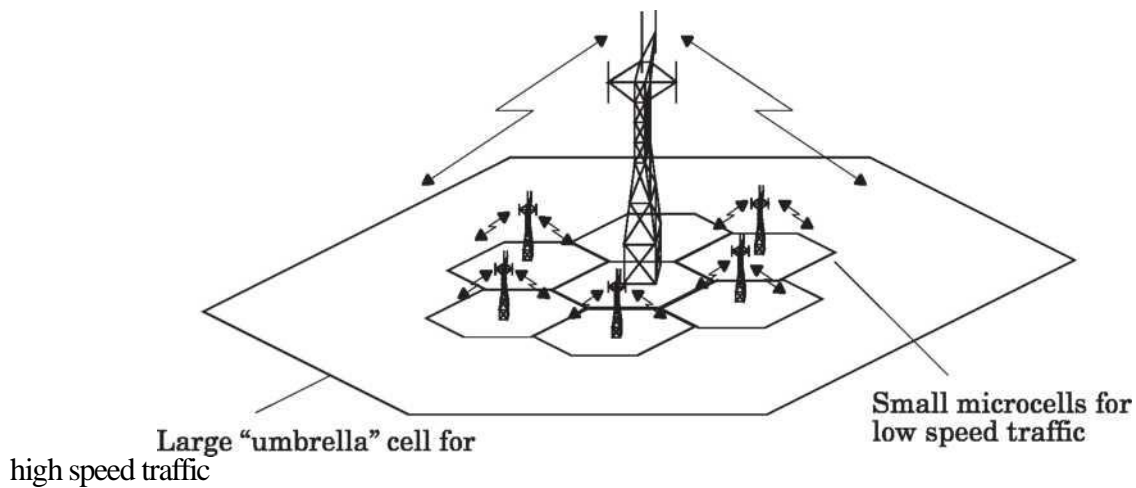


Figure 3.4 The umbrella cell approach.

Consequently, Δ is usually between 0 dB and 6 dB in modern cellular systems. The faster handoff process supports a much greater range of options for handling high speed and low speed users and provides the MSC with substantial time to “rescue” a call that is in need of handoff.

Another feature of newer cellular systems is the ability to make handoff decisions based on a wide range of metrics other than signal strength. The **co-channel and adjacent channel interference** levels may be measured at the base station or the mobile, and this information may be used with conventional signal strength data to provide a multi-dimensional algorithm for determining when a handoff is needed.

Unlike channelized wireless systems that assign different radio channels during a handoff (called a **hard handoff**), spread spectrum mobiles share the same channel in every cell. Thus, the term *handoff* does not mean a physical change in the assigned channel, but rather that a different base station handles the radio communication task. By simultaneously evaluating the received signals from a single subscriber at several neighboring base stations, the MSC may actually decide which version of the user’s signal is best at any moment in time. This technique exploits macroscopic space diversity provided by the different physical locations of the base stations and allows the MSC to make a “soft” decision as to which version of the user’s signal to pass along to the PSTN at any instance. The ability to select between the instantaneous received signals from a variety of base stations is called **soft handoff**.

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 noncellular 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.

1. 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 (see Table 1). For a hexagonal geometry,

$$Q = D/R = \sqrt{3N} \quad (4)$$

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.

Let i_0 be the number of co-channel interfering cells. Then, the signal-to-interference ratio (S/I or SIR) for a mobile receiver which monitors a forward channel can be expressed as

$$S/I = S / \sum_{i=1}^{i_0} I_i \quad (5)$$

where, S is the desired signal power from the desired base station and I is the interference power caused by the i^{th} interfering co-channel cell base station. If the signal levels of co-channel cells are known, then the S/I ratio for the forward link can be found using Equation (5).

Table 1 Co-channel Reuse Ratio for Some Values of N

	Cluster Size (N)	Co-channel Reuse Ratio (Q)
$i = 1, j = 1$	3	3
$i = 1, j = 2$	7	4.58
$i = 0, j = 3$	9	5.20
$i = 2, j = 2$	12	6

Propagation measurements in a mobile radio channel show that the average received signal strength at any point decays as a power law of the distance of separation between a transmitter and receiver. The average received power P_r at a distance d from the transmitting antenna is approximated by

$$P_r = P_0 (d/d_0)^{-n} \quad (6)$$

Or

$$P_r(\text{dBm}) = P_0(\text{dBm}) - 10 n \log (d/d_0) \quad (7)$$

where, P_0 is the power received at a close-in reference point in the far field region of the antenna at a small distance d_0 from the transmitting antenna and n is the path loss exponent. Now consider the forward link where the desired signal is the serving base station and where the interference is due to co-channel base stations. If D_i is the distance of the i th interferer from the mobile, the received power at a given mobile due to the i th interfering cell will be proportional to $(D_i)^{-n}$. The path loss exponent typically ranges between two and four in urban cellular systems.

When the transmit power of each base station is equal and the path loss exponent is the same throughout the coverage area, S/I for a mobile can be approximated as

$$\frac{S}{I} = \frac{R^{-n}}{\sum_{i=1}^{i_0} (D_i)^{-n}} \quad (8)$$

Considering only the first layer of interfering cells, if all the interfering base stations are equidistant from the desired base station and if this distance is equal to the distance D between cell centers, then Equation (8) simplifies to

$$\frac{S}{I} = \frac{(D/R)^n}{i_0} = \frac{(\sqrt{3N})^n}{i_0} \quad (9)$$

Equation (9) relates S/I to the cluster size N , which in turn determines the overall capacity of the system from Equation (2). Using an exact cell geometry layout, it can be shown for a seven-cell cluster, with the mobile unit at the cell boundary, the mobile is approximately $D - R$ from the two nearest cochannel interfering cells and approximately $D + R/2$, D , $D - R/2$, and $D + R$ from the other interfering cells in the first tier, as shown rigorously in . Using the approximate geometry shown in Figure 5, Equation (8), and assuming $n = 4$, the signal-to-interference ratio for the worst case can be closely approximated as

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

Equation (10) can be rewritten in terms of the co-channel reuse ratio Q , as

$$\frac{S}{I} = \frac{1}{2(Q - 1)^{-4} + 2(Q + 1)^{-4} + 2Q^{-4}} \quad (10)$$

(11)

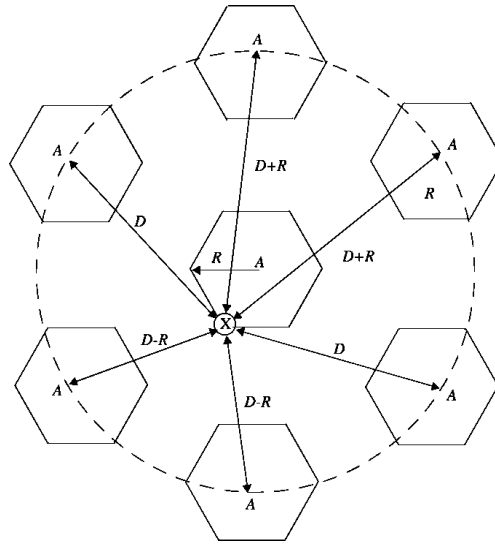


Figure 5 Illustration of the first tier of co-channel cells for a cluster size of $N = 7$. An approximation of the exact geometry is shown here. When the mobile is at the cell boundary (point X), it experiences worst case co-channel interference on the forward channel. The marked distances between the mobile and different co-channel cells are based on approximations made for easy analysis.

2. 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 "bleedover" 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 pass band, the signal-to-interference ratio at the base station for the weak mobile (before receiver filtering) is approximately.

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

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

(12)

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.

Trunking and Grade of Service

Cellular radio systems rely on *trunking* to accommodate a large number of users in a limited radio spectrum. The concept of trunking allows a large number of users to share the relatively small number of channels in a cell by providing access to each user, on demand, from a pool of available channels. In a trunked radio system, each user is allocated a channel on a per call basis, and upon termination of the call, the previously occupied channel is immediately returned to the pool of available channels.

Trunking exploits the statistical behavior of users so that a fixed number of channels or circuits may accommodate a large, random user community. The telephone company uses trunking theory to determine the

number of telephone circuits that need to be allocated for office buildings with hundreds of telephones, and this same principle is used in designing cellular radio systems. There is a trade-off between the number of available telephone circuits and the likelihood of a particular user finding that no circuits are available during the peak calling time. As the number of phone lines decreases, it becomes more likely that all circuits will be busy for a particular user. In a trunked mobile radio system, when a particular user requests service and all of the radio channels are already in use, the user is blocked, or denied access to the system. In some systems, a queue may be used to hold the requesting users until a channel becomes available.

To design trunked radio systems that can handle a specific capacity at a specific “grade of service,” it is essential to understand trunking theory and queuing theory. The fundamentals of trunking theory were developed by Erlang. Today, the measure of traffic intensity bears his name. One Erlang represents the amount of traffic intensity carried by a channel that is completely occupied (i.e. one call-hour per hour or one call-minute per minute). For example, a radio channel that is occupied for thirty minutes during an hour carries 0.5 Erlangs of traffic.

The ***grade of service (GOS)*** is a measure of the ability of a user to access a trunked system during the busiest hour. The busy hour is based upon customer demand at the busiest hour during a week, month, or year. The grade of service is a benchmark used to define the desired performance of a particular trunked system by specifying a desired likelihood of a user obtaining channel access given a specific number of channels available in the system. It is the wireless designer’s job to estimate the maximum required capacity and to allocate the proper number of channels in order to meet the *GOS*. *GOS* is typically given as the likelihood that a call is blocked, or the likelihood of a call experiencing a delay greater than a certain queuing time.

Definitions of Common Terms Used in Trunking Theory

Set-up Time: The time required to allocate a trunked radio channel to a requesting user.

Blocked Call: Call which cannot be completed at time of request, due to congestion. Also referred to as a *lost call*.

Holding Time: Average duration of a typical call. Denoted by H (in seconds).

Traffic Intensity: Measure of channel time utilization, which is the average channel occupancy

measured in Erlangs. This is a dimensionless quantity and may be used to measure the time utilization of single or multiple channels. Denoted by A .

Load: Traffic intensity across the entire trunked radio system, measured in Erlangs.

Grade of Service (GOS): A measure of congestion which is specified as the probability of a call being blocked (for Erlang B), or the probability of a call being delayed beyond a certain amount of time (for Erlang C).

Request Rate: The average number of call requests per unit time. Denoted by X seconds⁻¹.

The traffic intensity offered by each user is equal to the call request rate multiplied by the holding time. That is, each user generates a traffic intensity of A_u Erlangs given by

$$A_u = \lambda H \quad (13)$$

where H is the average duration of a call and X is the average number of call requests per unit time for each user. For a system containing U users and an unspecified number of channels, the total offered traffic intensity A , is given as

$$A = U A_u \quad (14)$$

Furthermore, in a C channel trunked system, if the traffic is equally distributed among the channels, then the traffic intensity per channel, A_c , is given as

$$A_c = U A_u / C \quad (15)$$

Note that the offered traffic is not necessarily the traffic which is *carried* by the trunked system, only that which is *offered* to the trunked system. When the offered traffic exceeds the maximum capacity of the system, the carried traffic becomes limited due to the limited capacity (i.e. limited number of channels). The maximum possible carried traffic is the total number of channels, C , in Erlangs.

There are **two types of trunked systems** which are commonly used. The first type offers no queuing for call requests. That is, for every user who requests service, it is assumed there is no setup time and the user is given immediate access to a channel if one is available. If no channels are available, the requesting user is blocked without access and is free to try again later. This type of trunking is called ***blocked calls cleared*** and assumes that calls arrive as determined by a Poisson distribution. Furthermore, it is assumed that there are an infinite

number of users as well as the following: (a) there are memoryless arrivals of requests, implying that all users, including blocked users, may request a channel at any time; (b) the probability of a user occupying a channel is exponentially distributed, so that longer calls are less likely to occur as described by an exponential distribution; and (c) there are a finite number of channels available in the trunking pool. This is known as an M/M/m/m queue, and leads to the derivation of the Erlang B formula (also known as the *blocked calls cleared* formula). The Erlang B formula determines the probability that a call is blocked and is a measure of the GOS for a trunked system which provides no queuing for blocked calls. The Erlang B formula is given by

$$Pr[blocking] = \frac{\frac{A^C}{C!}}{\sum_{k=0}^C \frac{A^k}{k!}} = GOS \quad (16)$$

where, C is the number of trunked channels offered by a trunked radio system and A is the total offered traffic. While it is possible to model trunked systems with finite users, the resulting expressions are much more complicated than the Erlang B result, and the added complexity is not warranted for typical trunked systems which have users that outnumber available channels by orders of magnitude. Furthermore, the Erlang B formula provides a conservative estimate of the GOS, as the finite user results always predict a smaller likelihood of blocking.

The second kind of trunked system is one in which a queue is provided to hold calls which are blocked. If a channel is not available immediately, the call request may be delayed until a channel becomes available. This type of trunking is called ***Blocked Calls Delayed***, and its measure of GOS is defined as the probability that a call is blocked after waiting a specific length of time in the queue. To find the GOS, it is first necessary to find the likelihood that a call is initially denied access to the system. The likelihood of a call not having immediate access to a channel is determined by the Erlang C formula derived is

$$Pr[delay > 0] = \frac{A^C}{A^C + C! \left(1 - \frac{A}{C}\right) \sum_{k=0}^{C-1} \frac{A^k}{k!}} \quad (17)$$

If no channels are immediately available the call is delayed, and the probability that the delayed call is forced to wait more than t seconds is given by the probability that a call is delayed, multiplied by the conditional probability that the delay is greater than t seconds. The GOS of a trunked system where blocked calls are delayed is hence given by

$$Pr[delay > t] = Pr[delay > 0] Pr[delay > t | delay > 0] \quad (18)$$

$$= Pr[delay > 0] \exp(-(C-A)t/H)$$

The average delay D for all calls in a queued system is given by

$$D = Pr[delay > 0] \frac{H}{C - A} \quad (19)$$

where, the average delay for those calls which are queued is given by $H/(C - A)$.

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 sectorized cells, and enable the base station to oversee all handoff chores related to the microcells, thus reducing the computational load at the MSC. These popular capacity improvement techniques will be explained in detail.

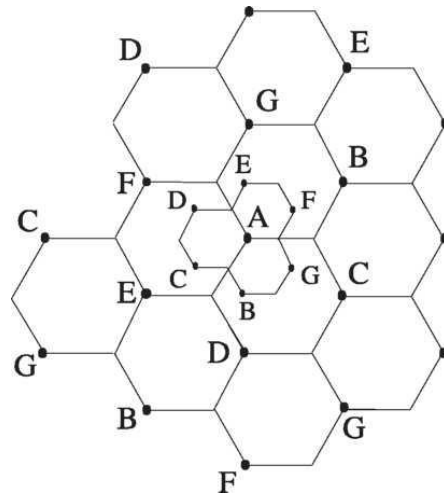
1. Cell Splitting

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 Figure 1 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 (see Equation (4)) between co-channel cells.

An example of cell splitting is shown in Figure 8. In Figure 8, 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. In the example shown in Figure 8, the smaller cells were added in such a way as to preserve the frequency reuse plan of the system. For example, the microcell base station labeled G was placed half way between two larger stations utilizing the same channel set G. This is also the case for the other microcells in the figure. As can be seen from Figure 8, cell splitting merely scales the geometry of the cluster. In this case, the radius of each new microcell is half that of the original cell.



For the new
of these cells must be

cells to be smaller in size, the transmit power
reduced. The transmit power of the new cells

with radius half that of the original cells can be found by examining the received power P_r at the new and old cell boundaries and setting them equal to each other. This is necessary to ensure that the frequency reuse plan for the new microcells behaves exactly as for the original cells. For Figure 8.

Figure 8 Illustration of cell splitting.

$$P_r[\text{at old cell boundary}] = P_{t1}R^{-n} \quad (20)$$

$$P_r[\text{at new cell boundary}] = P_{t2}(R/2)^{-n} \quad (21)$$

where, P_{t1} and P_{t2} are the transmit powers of the larger and smaller cell base stations, respectively, and n is the path loss exponent. If we take $n = 4$ and set the received powers equal to each other, then

$$P_{t2} = P_{t1}/16 \quad (22)$$

In other words, the transmit power must be reduced by 12 dB in order to fill in the original coverage area with microcells, while maintaining the S/I requirement.

In practice, not all cells are split at the same time. It is often difficult for service providers to find real estate that is perfectly situated for cell splitting. Therefore, different cell sizes will exist simultaneously. In such situations, special care needs to be taken to keep the distance between cochannel cells at the required minimum, and hence channel assignments become more complicated. Also, handoff issues must be addressed so that high speed and low speed traffic can be simultaneously accommodated. When there are two cell sizes in the same region as shown in Figure 8, Equation (22) shows that one cannot simply use the original transmit power for all new cells or the new transmit power for all the original cells. If the larger transmit power is used for all cells, some channels used by the smaller cells would not be sufficiently separated from co-channel cells. On the other hand, if the smaller transmit power is used for all the cells, there would be parts of the larger cells left unserved. For this reason, channels in the old cell must be broken down into two channel groups, one that corresponds to the smaller cell reuse requirements and the other that corresponds to the larger cell reuse requirements. The larger cell is usually dedicated to high speed traffic so that handoffs occur less frequently.

The two channel group sizes depend on the stage of the splitting process. At the beginning of the cell splitting process, there will be fewer channels in the small power groups. However, as demand grows, more channels will be required, and thus the smaller groups will require more channels. This splitting process continues until all the channels in an area are used in the lower power group, at which point cell splitting is complete within the region, and the entire system is rescaled to have a smaller radius per cell. *Antenna downtilting*, which deliberately focuses radiated energy from the base station toward the ground (rather than toward the horizon), is often used to limit the radio coverage of newly formed microcells.

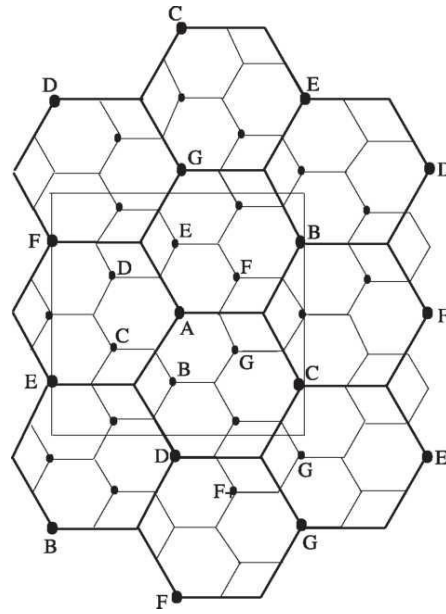


Figure 3.9 Illustration of cell splitting within a 3 km by 3 km square centered around base station A.

2. Sectoring

Cell splitting achieves capacity improvement by essentially rescaling the system. By decreasing the cell radius R and keeping the co-channel reuse ratio DIR unchanged, cell splitting increases the number of channels per unit area.

However, another way to increase capacity is to keep the cell radius unchanged and seek methods to decrease the DIR ratio. As we now show, *sectoring* increases SIR so that the cluster size may be reduced. In this approach, first the SIR is improved using directional antennas, then capacity improvement is achieved by reducing the number of cells in a cluster, thus increasing the frequency reuse. However, in order to do this successfully, it is necessary to reduce the relative interference without decreasing the transmit power.

The co-channel interference in a cellular system may be decreased by replacing a single omnidirectional antenna at the base station by several directional antennas, each radiating within a specified sector. By using directional antennas, a given cell will receive interference and transmit with only a fraction of the available co-channel cells. The technique for decreasing co-channel interference and thus increasing system performance by using directional antennas is called *sectoring*. The factor by which the

co-channel interference is reduced depends on the amount of sectoring used. A cell is normally partitioned into three 120° sectors or six 60° sectors as shown in Figure 10 (a) and (b).

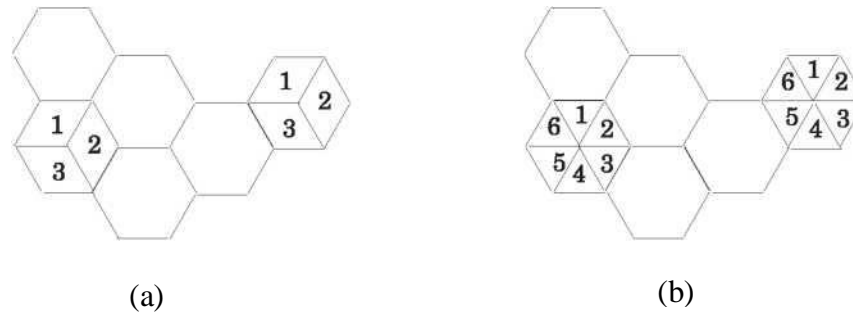


Figure 3.10 (a) 120 sectoring; (b) 60 sectoring.

When sectoring is employed, the channels used in a particular cell are broken down into sectorized groups and are used only within a particular sector, as illustrated in Figure 10(a) and (b). Assuming seven-cell reuse, for the case of 120° sectors, the number of interferers in the first tier is reduced from six to two. This is because only two of the six co-channel cells receive interference with a particular sectorized channel group. Referring to Figure 11, consider the interference experienced by a mobile located in the right-most sector in the center cell labeled “5”. There are three cochannel cell sectors labeled “5” to the right of the center cell, and three to the left of the center cell. Out of these six co-channel cells, only two cells have sectors with antenna patterns which radiate into the center cell, and hence a mobile in the center cell will experience interference on the forward link from only these two sectors. This S/I improvement allows the wireless engineer to then decrease the cluster size N in order to improve the frequency reuse, and thus the system capacity. In practical systems, further improvement in S/I is achieved by down-tilting the sector antennas such that the radiation pattern in the vertical (elevation) plane has a notch at the nearest co-channel cell distance.

Thus, sectoring reduces interference, which amounts to an increase in capacity. In practice, the reduction in interference offered by sectoring enable planners to reduce the cluster size N , and provides an additional degree of freedom in assigning channels. The penalty for improved S/I and the resulting capacity improvement from the shrinking cluster size is an increased number of antennas at each base station, and a decrease in trunking efficiency due to channel sectoring at the base station. Since sectoring reduces the coverage area of a particular group of channels, the number of handoffs increases, as well. Fortunately, many modern base stations support sector-ization and allow mobiles to be handed off from sector to sector within the same cell

without intervention from the MSC, so the handoff problem is often not a major concern.

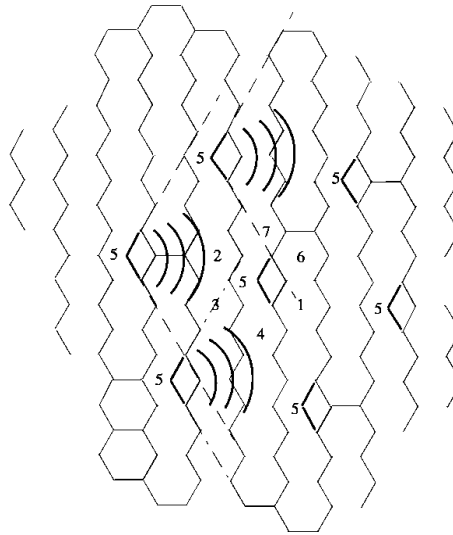


Figure 11 Illustration of how 120° sectoring reduces interference from co-channel cells. Out of the 6 co-channel cells in the first tier, only two of them interfere with the center cell. If omnidirectional antennas were used at each base station, all six co-channel cells would interfere with the center cell.

It is the loss of traffic due to decreased trunking efficiency that causes some operators to shy away from the sectoring approach, particularly in dense urban areas where the directional antenna patterns are somewhat ineffective in controlling radio propagation. Because sectoring uses more than one antenna per base station, the available channels in the cell must be subdivided and dedicated to a specific antenna. This breaks up the available trunked channel pool into several smaller pools, and decreases trunking efficiency.

3. Microcell Zone Concept

The increased number of handoffs required when sectoring is employed results in an increased load on the switching and control link elements of the mobile system. A solution to this problem was presented by Lee . This proposal is based on a microcell concept for seven cell reuse, as illustrated in Figure 13. In this scheme, each of the three (or possibly more) zone sites (represented as Tx/Rx in Figure 13) are connected to a single base station and share the same radio equipment. The zones are connected by coaxial cable, fiber optic cable, or microwave link to the base station. Multiple zones and a single base station make up a cell. As a mobile travels within the cell, it is served by the zone with the strongest signal. This approach is superior to sectoring since antennas are placed at the outer edges of the cell, and any base station channel may be assigned to any zone by the base station.

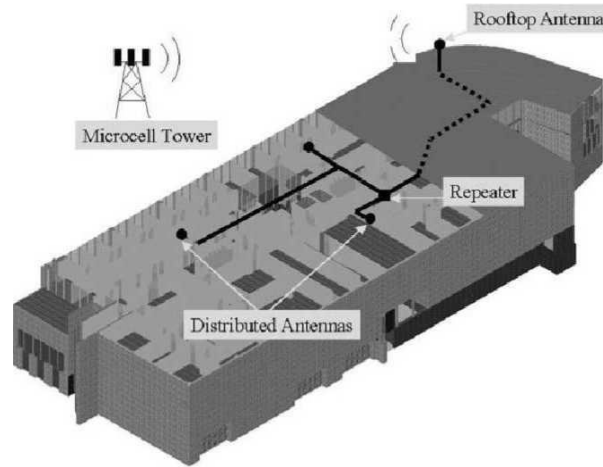


Figure 3.12 Illustration of how a distributed antenna system (DAS) may be used inside a building.

As a mobile travels from one zone to another within the cell, it retains the same channel. Thus, unlike in sectoring, a handoff is not required at the MSC when the mobile travels between zones within the cell. The base station simply switches the channel to a different zone site. In this way, a given channel is active only in the particular zone in which the mobile is traveling, and hence the base station radiation is localized and interference is reduced. The channels are distributed in time and space by all three zones and are also reused in co-channel cells in the normal fashion. This technique is particularly useful along highways or along urban traffic corridors.

The advantage of the zone cell technique is that while the cell maintains a particular coverage radius, the co-channel interference in the cellular system is reduced since a large central base station is replaced by several lower powered transmitters (zone transmitters) on the edges of the cell. Decreased co-channel interference improves the signal quality and also leads to an increase in capacity without the degradation in trunking efficiency caused by sectoring. With respect to the zone microcell system, since transmission at any instant is confined to a particular zone, this implies that a D_z/R_z of 4.6 (where D_z is the minimum distance between active co-channel zones and R_z is the zone radius) can achieve the required link performance.

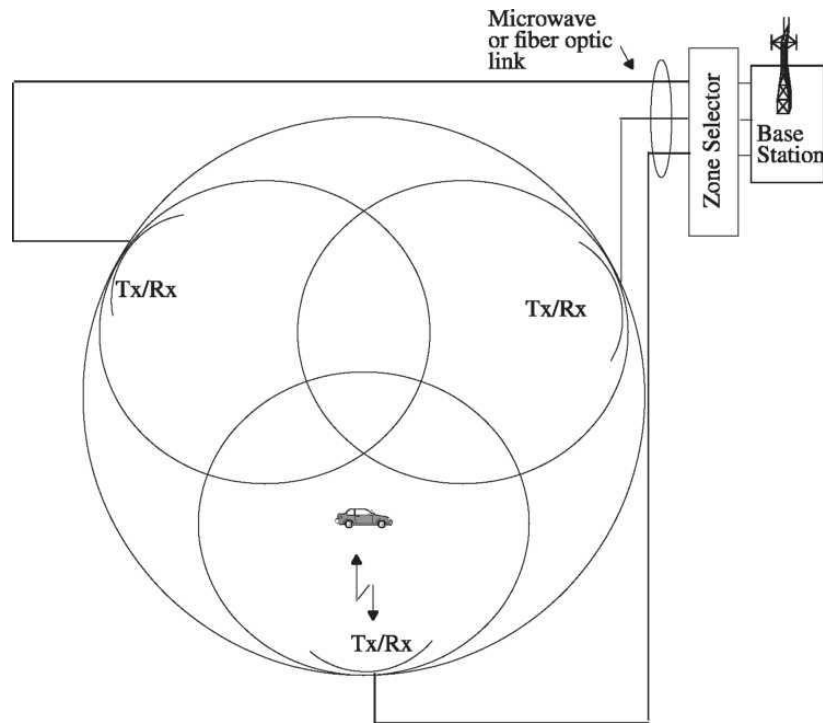


Figure 3.13 The microcell concept

In Figure 14, let each individual hexagon represents a zone, while each group of three hexagons represents a cell. The zone radius R_z is approximately equal to one hexagon radius. Now, the capacity of the zone microcell system is directly related to the distance between co-channel cells, and not zones. This distance is represented as D in Figure 14. For a D_z/R_z value of 4.6, it can be seen from the geometry of Figure 14 that the value of co-channel reuse ratio, D/R , is equal to three, where R is the radius of the cell and is equal to twice the length of the hexagon radius. Using Equation (4), $D/R = 3$ corresponds to a cluster size of $N = 3$. This reduction in the cluster size from $N = 7$ to $N = 3$ amounts to a 2.33 times increase in capacity for a system completely based on the zone microcell concept. No loss in trunking efficiency is experienced. Zone cell architectures are being adopted in many cellular and personal communication systems.

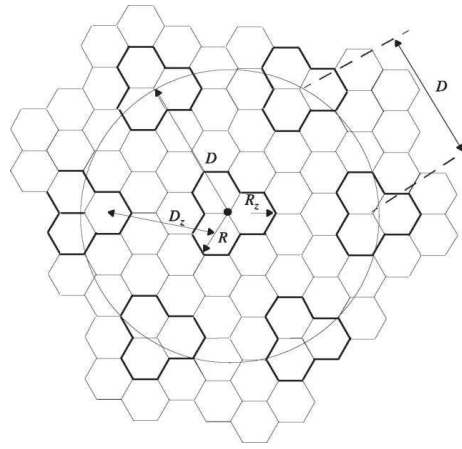
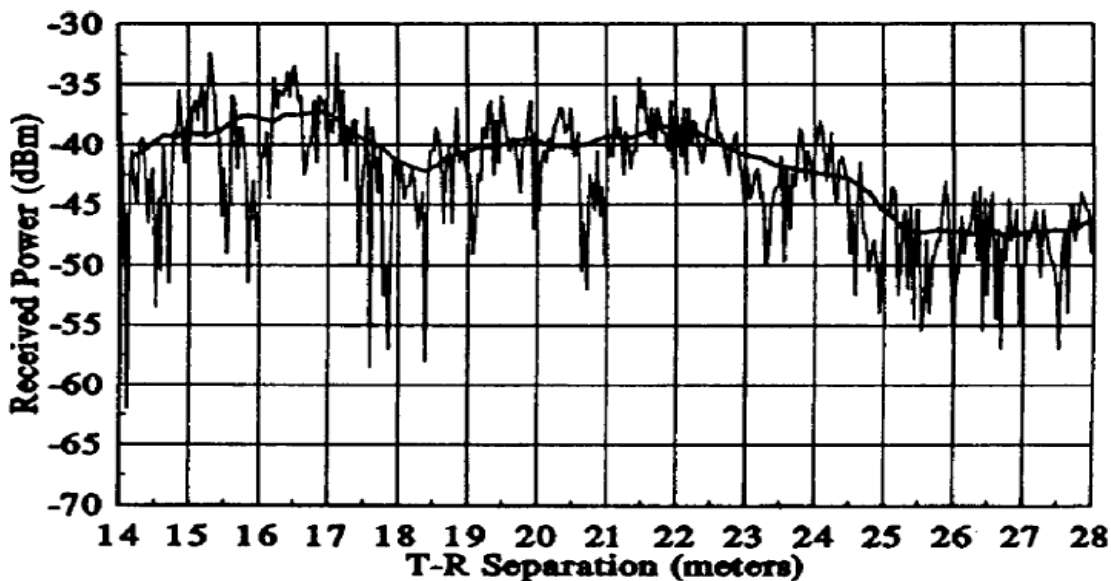


Figure 14 Define D , D_z , R , and R_z for a microcell architecture with $N=7$. The smaller hexagons form zones and three hexagons (outlined in bold) together form a cell. Six nearest co-channel cells are shown.

Wireless Propagation Mechanism

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

Propagation models have traditionally focused on predicting the average received signal strength at a given distance from the transmitter, as well as the variability of the signal strength in close spatial proximity to a particular location. Propagation models that predict the mean signal strength for an arbitrary transmitter-receiver (T-R) separation distance are useful in estimating the radio coverage area of a transmitter and are called *large-scale propagation models*, since they characterize signal strength over large T -R separation distances (several hundreds or thousands of meters). On the other hand, propagation models that characterize the rapid fluctuations of the received signal strength over very short travel distances (a few wavelengths) or short time durations (on the order of seconds) are called *small-scale* or *fading* models. As a mobile moves over very small distances, the instantaneous received signal strength may fluctuate rapidly giving rise to small-scale fading. The reason for this is that the received signal is a sum of many contributions coming from different directions. Figure below illustrates small-scale fading and the more gradual large-scale variations for an indoor radio communication system. Notice in the figure that the signal fades rapidly (smallscale fading) as the receiver moves, but the local average signal changes much more gradually with distance.



Small-scale and large-scale fading

Basic Methods of Propagation

Reflection, diffraction and scattering are the three fundamental phenomena that cause signal propagation in a mobile communication system, apart from LoS communication. The most important parameter, predicted by propagation models based on above three phenomena, is the

received power. The physics of the above phenomena may also be used to describe small scale fading and multipath propagation. The following subsections give an outline of these phenomena.

Reflection

Reflection occurs when an electromagnetic wave falls on an object, which has very large dimensions as compared to the wavelength of the propagating wave. For example, such objects can be the earth, buildings and walls. When a radio wave falls on another medium having different electrical properties, a part of it is transmitted into it, while some energy is reflected back. Let us see some special cases. If the medium on which the e.m. wave is incident is a dielectric, some energy is reflected back and some energy is transmitted. If the medium is a perfect conductor, all energy is reflected back to the first medium. The amount of energy that is reflected back depends on the polarization of the e.m. wave. Another particular case of interest arises in parallel polarization, when no reflected occurs in the medium of origin. This would occur, when the incident angle would be such that the reflected Coefficient is equal to zero. This angle is the Brewster's angle. By applying laws of electro-magnetics, it is found to be

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

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

$$\epsilon_i = \epsilon_r$$

and

$$E_i = E_r \quad \text{for vertical polarization, and}$$

$$E_i = -E_r \quad \text{for horizontal polarization.}$$

Diffraction

Diffraction is the phenomenon due to which an EM wave can propagate beyond the horizon, around the curved earth's surface and obstructions like tall buildings.

As the user moves deeper into the shadowed region, the received field strength decreases. But the diffraction field still exists and it has enough strength to yield a good signal. This phenomenon can be explained by the Huygens principle, according to which, every point on a wave front acts as point sources for the production of secondary wavelets, and they combine to produce a new wave front in the direction of propagation. The propagation of secondary wavelets in the shadowed

region results in diffraction. The field in the shadowed region is the vector sum of the electric field components of all the secondary wavelets that are received by the receiver.

Scattering

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

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

Free Space Propagation Model

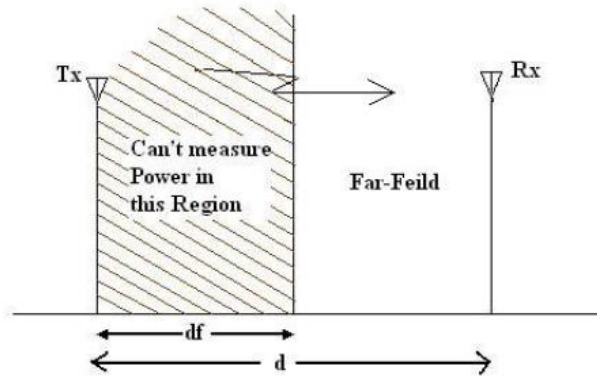


Figure 1.1: Free space propagation model, showing the near and far fields

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

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 d^2 L}$$

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

$$G = 4\pi A_e / \lambda^2.$$

The effective aperture A_e is related to the physical size of the antenna, and λ is related to the carrier frequency by

$$\lambda = \frac{c}{f} = \frac{2\pi c}{\omega_c}$$

where f is the carrier frequency in Hertz, ω_c is the carrier frequency in radians per second, and c is the speed of light given in meters/s. The values for P_t and P_r must be expressed in the same units, and G_t and G_r are dimensionless quantities. The miscellaneous losses L ($L \geq 1$) are usually due to transmission line attenuation, filter losses, and antenna losses in the communication system. A value of $L = 1$ indicates no loss in the system hardware. The Friis free space equation shows that the received power falls off as the square of the T-R separation distance. This implies that the received power decays with distance at a rate of 20 dB/decade.

An *isotropic* radiator is an ideal antenna which radiates power with unit gain uniformly in all directions, and is often used to reference antenna gains in wireless systems. The *effective isotropic radiated power* (*EIRP*) is defined as

$$EIRP = P_t G_t$$

and represents the maximum radiated power available from a transmitter in the direction of maximum antenna gain, as compared to an isotropic radiator.

In practice, *effective radiated power* (*ERP*) is used instead of EIRP to denote the maximum radiated power as compared to a half-wave dipole antenna (instead of an isotropic antenna). In practice, antenna gains are given in units of dBi (dB gain with respect to an isotropic antenna) or dBd (dB gain with respect to a half-wave dipole).

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

$$PL(dB) = 10 \log P_t / P_r.$$

The fields of an antenna can broadly be classified in two regions, the far field and the near field. It is in the far field that the propagating waves act as plane waves and the power decays inversely

with distance. The far field region is also termed as Fraunhofer region and the Friis equation holds in this region. Hence, the Friis equation is used only beyond the far field distance, d_f , which is dependent upon the largest dimension of the antenna as:

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

Also we can see that the Friis equation is not defined for $d=0$. For this reason, we use a close in distance, d_0 , as a reference point. The power received, $P_r(d)$, is then given by:

$$P_r(d) = P_r(d_0)(d_0/d)^2.$$

$$P_r(d) \text{ dBm} = 10\log\left[\frac{P_r(d_0)}{0.001 \text{ W}}\right] + 20\log\left(\frac{d_0}{d}\right)$$

where $P_r(d_0)$ is in units of watts.

Ground (Two Ray) Reflection Model

In a mobile radio channel, a single direct path between the base station and a mobile is seldom the only physical means for propagation, and hence the free space propagation model by Friis Equation is in most cases inaccurate when used alone. The two-ray ground reflection model shown in Figure 1.2 is a useful propagation model that is based on geometric optics, and considers both the direct path and a ground reflected propagation path between transmitter and receiver. This model has been found to be reasonably accurate for predicting the large-scale signal strength over distances of several kilometers for mobile radio systems that use tall towers, as well as for line-of-sight microcell channels in urban environments.

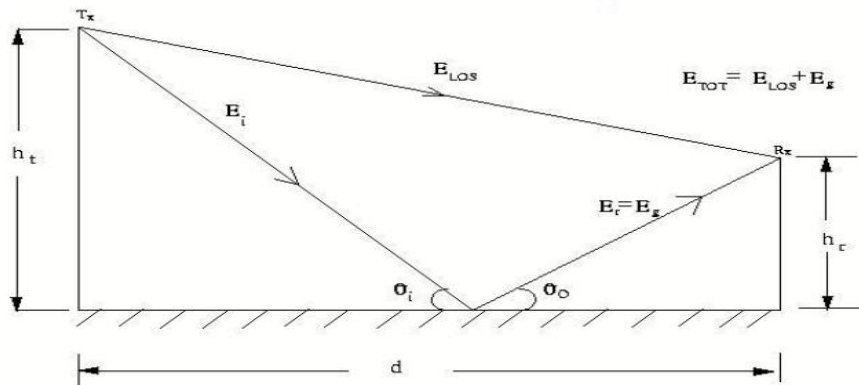


Fig 1.2 Two-ray reflection model

In most mobile communication systems, the maximum T-R separation distance is at most only a few tens of kilometers, and the earth may be assumed to be flat. The total received Efield, E_{TOT} , is then a result of the direct line-of-sight component, E_{LOS} , and the ground reflected component, E_g , distance between the Tx and the Rx. From Figure 1.2, the total transmitted and received electric fields are

$$E_T^{TOT} = E_i + E_{LOS},$$

$$E_R^{TOT} = E_g + E_{LOS}.$$

Referring to Figure , ht is the height of the transmitter and hr is the height of the receiver. If E_0 is the free space E-field (in units of V/m) at a reference distance d_0 from the transmitter, then for $d > d_0$, the free space propagating E-field is given by:

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

where $|E(d, t)| = E_0 d_0 / d$ represents the envelope of the E-field at d meters from the transmitter. Two propagating waves arrive at the receiver: the direct wave that travels a distance d' ; and the reflected wave that travels a distance d'' . The E-field due to the line-of-sight component at the receiver can be expressed as

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

and the E-field for the ground reflected wave, which has a propagation distance of d'' , can be expressed as

$$E_g(d'', t) = \Gamma \frac{E_0 d_0}{d''} \cos\left(\omega_c\left(t - \frac{d''}{c}\right)\right)$$

According to laws of reflection in dielectrics

$$\theta_i = \theta_0$$

$$E_g = \Gamma E_i$$

$$E_t = (1 + \Gamma)E_i$$

where Γ is the reflection coefficient for ground. For small values of θ_i (i.e., grazing incidence), the reflected wave is equal in magnitude and 180° out of phase with the incident wave.

The resultant E-field, assuming perfect horizontal E-field polarization and ground reflection (i.e., $\Gamma = -1$ and $E_t = 0$), is the vector sum of E_{LOS} and E_g , and the resultant total E-field envelope is given by

$$|E_{TOT}| = |E_{LOS} + E_g|$$

The electric field $E_{TOT}(d, t)$ can be expressed as

$$E_{TOT}(d, t) = \frac{E_0 d_0}{d} \cos\left(\omega_c\left(t - \frac{d}{c}\right)\right) + (-1) \frac{E_0 d_0}{d} \cos\left(\omega_c\left(t - \frac{d}{c}\right)\right)$$

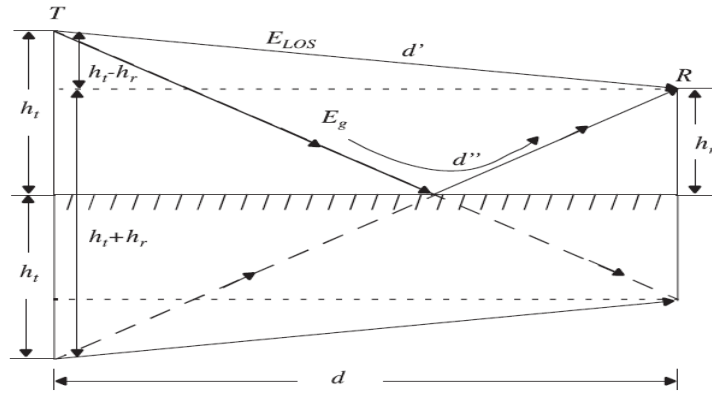


Fig 1.3 The method of images is used to find the path difference between the line-of-sight and the ground reflected paths.

Using the *method of images*, which is demonstrated by the geometry of Figure 1.3, the path difference, Δ , between the line-of-sight and the ground reflected paths can be expressed as

$$\Delta = d - d' = \sqrt{(h_t + h_r)^2 + d^2} - \sqrt{(h_t - h_r)^2 + d^2}$$

When the T-R separation distance d is very large compared to $h_t + h_r$, it can be simplified using a Taylor series approximation

$$\Delta = d - d' \approx \frac{2h_t h_r}{d}$$

Once the path difference is known, the phase difference $\phi \Delta$ between the two E-field components and the time delay τ_d between the arrival of the two components can be easily computed using the following relations. Let E_0 be the free space electric field (in V/m) at a reference distance d_0 .

$$\theta_{\Delta} = \frac{2\pi\Delta}{\lambda} = \frac{\Delta\omega_c}{c}$$

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

It should be noted that as d becomes large, the difference between the distances d' and d'' becomes very small, and the amplitudes of E_{LOS} and E_g are virtually identical and differ only in phase. That is

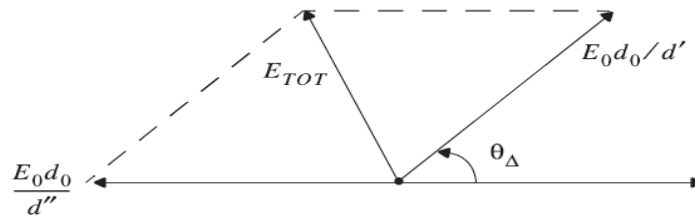
$$\left| \frac{E_0 d_0}{d} \right| \approx \left| \frac{E_0 d_0}{d'} \right| \approx \left| \frac{E_0 d_0}{d''} \right|$$

If the received E-field is evaluated at some time, say at $t = d''/c$,

$$\begin{aligned} E_{TOT}\left(d, t = \frac{d''}{c}\right) &= \frac{E_0 d_0}{d'} \cos\left(\omega_c\left(\frac{d'' - d'}{c}\right)\right) - \frac{E_0 d_0}{d''} \cos 0^\circ \\ &= \frac{E_0 d_0}{d'} \angle \theta_{\Delta} - \frac{E_0 d_0}{d''} \\ &\approx \frac{E_0 d_0}{d} [\angle \theta_{\Delta} - 1] \end{aligned}$$

where d is the distance over a flat earth between the bases of the transmitter and receiver antennas. Referring to the phasor diagram of Figure below which shows how the direct and ground reflected rays combine, the electric field (at the receiver) at a distance d from the transmitter can be written as

$$|E_{TOT}(d)| = \sqrt{\left(\frac{E_0 d_0}{d}\right)^2 (\cos \theta_{\Delta} - 1)^2 + \left(\frac{E_0 d_0}{d}\right)^2 \sin^2 \theta_{\Delta}}$$



Phasor diagram showing the electric field components of the line-of-sight, ground reflected, and total received E-fields

$$|E_{TOT}(d)| = \frac{E_0 d_0}{d} \sqrt{2 - 2 \cos \theta_{\Delta}}$$

$$|E_{TOT}(d)| = 2 \frac{E_0 d_0}{d} \sin\left(\frac{\theta_\Delta}{2}\right)$$

Above Equation is an important expression, as it provides the exact received E-field for the two-ray ground reflection model. One notes that for increasing distance from the transmitter, $E_{TOT}(d)$ decays in an oscillatory fashion.

$$\sin(\theta_\Delta/2) \approx \theta_\Delta/2$$

$$\frac{\theta_\Delta}{2} \approx \frac{2\pi h_t h_r}{\lambda d} < 0.3 \text{ rad}$$

$$E_{TOT}(d) \approx \frac{2E_0 d_0}{d} \frac{2\pi h_t h_r}{\lambda d} \approx \frac{k}{d^2} \text{ V/m}$$

where k is a constant related to E_0 , the antenna heights, and the wavelength.

The received power at a distance d from the transmitter for the two-ray ground bounce model can be expressed as

$$P_r = P_t G_t G_r \frac{h_t^2 h_r^2}{d^4}$$

As seen from Equation, at large distances, the received power falls off with distance raised to the fourth power, or at a rate of 40 dB/decade. This is a much more rapid path loss than is experienced in free space. Note also that at large values of d , the received power and path loss become independent of frequency. The path loss for the two-ray model (with antenna gains) can be expressed in dB as

$$PL(\text{dB}) = 40 \log d - (10 \log G_t + 10 \log G_r + 20 \log h_t + 20 \log h_r)$$

Multipath Wave Propagation and Fading

Multipath Propagation

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

Multipath & Small-Scale Fading

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

Fading

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

Multipath Fading Effects

In principle, the following are the main multipath effects:

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

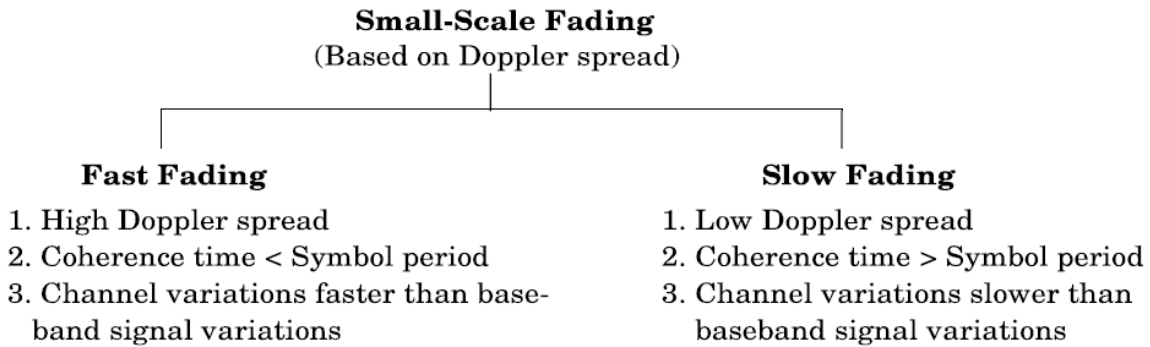
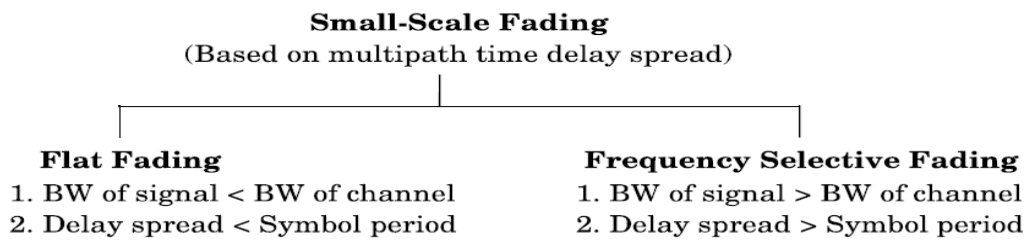
Factors Influencing Fading

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

- (1) **Multipath propagation** - Multipath is the propagation phenomenon that results in radio signals reaching the receiving antenna by two or more paths. The effects of multipath include constructive and destructive interference, and phase shifting of the signal.
- (2) **Speed of the mobile** - The relative motion between the base station and the mobile results in random frequency modulation due to different doppler shifts on each of the multipath components.
- (3) **Speed of surrounding objects** - If objects in the radio channel are in motion, they induce a time varying Doppler shift on multipath components. If the surrounding objects move at a greater rate than the mobile, then this effect dominates fading.
- (4) **Transmission Bandwidth of the signal** - If the transmitted radio signal bandwidth is greater than the “bandwidth” of the multipath channel (quantified by coherence bandwidth), the received signal will be distorted.

Types of Small-Scale Fading

Depending on the relation between the signal parameters (such as bandwidth, symbol period, etc.) and the channel parameters (such as rms delay spread and Doppler spread), different transmitted signals will undergo different types of fading. The time dispersion and frequency dispersion mechanisms in a mobile radio channel lead to four possible distinct effects, which are manifested depending on the nature of the transmitted signal, the channel, and the velocity. While multipath delay spread leads to *time dispersion* and *frequency selective fading*, Doppler spread leads to *frequency dispersion* and *time selective fading*. The two propagation mechanisms are independent of one another. Figure shows a tree of the four different types of fading.



Fading Effects Due to Multi path Time Delay Spread

Flat fading

If the mobile radio channel has a constant gain and linear phase response over a bandwidth which is greater than the bandwidth of the transmitted signal, then the received signal will undergo *flat fading*. This type of fading is historically the most common type of fading described in the technical literature. In flat fading, the multipath structure of the channel is such that the spectral characteristics of the transmitted signal are preserved at the receiver. However the strength of the received signal changes with time, due to fluctuations in the gain of the channel caused by multipath. The characteristics of a flat fading channel are illustrated in Figure. It can be seen from Figure that if the channel gain changes over time, a change of amplitude occurs in the received signal. Over time, the received signal $r(t)$ varies in gain, but the spectrum of the transmission is preserved. In a flat fading channel, the reciprocal bandwidth of the transmitted signal is much larger than the multipath time delay spread of the channel, and $h_b(t; \tau)$ can be approximated as having no excess delay (i.e., a single delta function with $\tau = 0$). Flat fading channels are also known as *amplitude varying channels* and are sometimes referred to as *narrowband channels*, since the bandwidth of the applied signal is *narrow* as compared to the channel flat fading bandwidth.

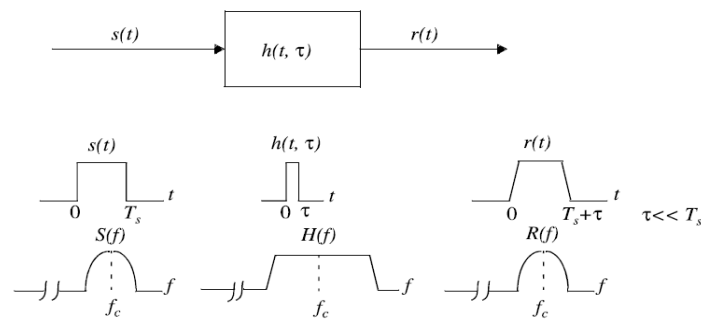
To summarize, a signal undergoes flat fading if

$$B_S \ll B_C$$

and

$$T_S \gg \sigma_\tau$$

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



Flat fading channel characteristics

Frequency Selective Fading

If the channel possesses a constant-gain and linear phase response over a bandwidth that is smaller than the bandwidth of transmitted signal, then the channel creates *frequency selective fading* on the received signal. Under such conditions, the channel impulse response has a multipath delay spread which is greater than the reciprocal bandwidth of the transmitted message waveform. When this occurs, the received signal includes multiple versions of the transmitted waveform which are attenuated (faded) and delayed in time, and hence the received signal is distorted.

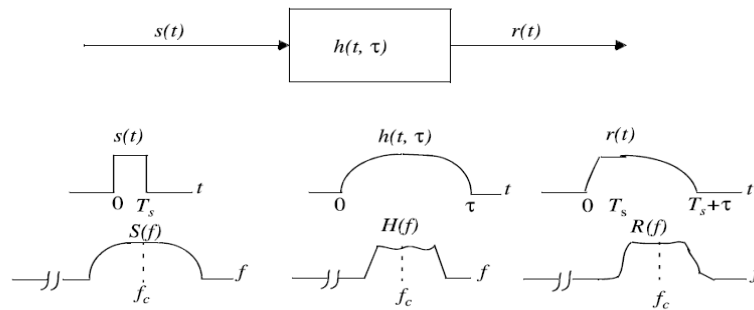
Frequency selective fading is due to time dispersion of the transmitted symbols within the channel. Thus the channel induces *inter symbol interference* (ISI). Viewed in the frequency domain, certain frequency components in the received signal spectrum have greater gains than others.

Frequency selective fading channels are much more difficult to model than flat fading channels since each multipath signal must be modeled and the channel must be considered to be a linear filter. It is for this reason that wideband multipath measurements are made, and models are developed from these measurements. Figure below illustrates the characteristics of a frequency selective fading channel.

For frequency selective fading, the spectrum $S(f)$ of the transmitted signal has a bandwidth which is greater than the coherence bandwidth B_c of the channel. Viewed in the frequency domain, the channel becomes frequency selective, where the gain is different for different frequency components. Frequency selective fading is caused by multipath delays which approach or exceed the symbol period of the transmitted symbol. Frequency selective fading channels are also known as *wideband channels* since the bandwidth of the signal $s(t)$ is wider than the bandwidth of the channel impulse response. As time varies, the channel varies in gain and phase across the spectrum of $s(t)$, resulting in time varying distortion in the received signal $r(t)$. To summarize, a signal undergoes frequency selective fading if,

$$B_s > B_c \quad \text{and}$$

$$T_s < \sigma_\tau$$



Frequency selective fading channel characteristics

Fading Effects Due to Doppler Spread

Fast Fading

Depending on how rapidly the transmitted baseband signal changes as compared to the rate of change of the channel, a channel may be classified either as a *fast fading* or *slow fading* channel. In a *fast fading channel*, the channel impulse response changes rapidly within the symbol duration. That is, the coherence time of the channel is smaller than the symbol period of the transmitted

signal. This causes frequency dispersion (also called time selective fading) due to Doppler spreading, which leads to signal distortion. Viewed in the frequency domain, signal distortion due to fast fading increases with increasing Doppler spread relative to the bandwidth of the transmitted signal. Therefore, a signal undergoes fast fading if

$$T_S > T_C$$

and

$$B_S < B_D$$

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

Slow Fading

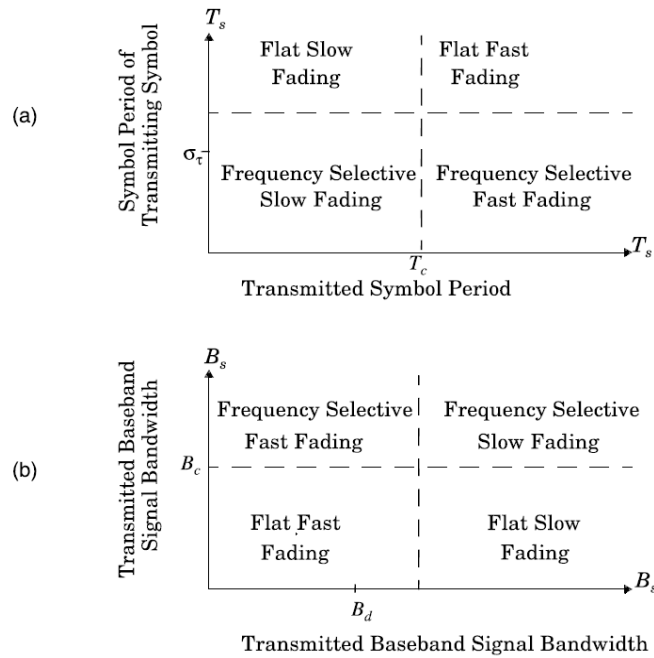
In a *slow fading channel*, the channel impulse response changes at a rate much slower than the transmitted baseband signal $s(t)$. In this case, the channel may be assumed to be static over one or several reciprocal bandwidth intervals. In the frequency domain, this implies that the Doppler spread of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

$$T_S \ll T_C$$

and

$$B_S \gg B_D$$

It should be emphasized that fast and slow fading deal with the relationship between the time rate of change in the channel and the transmitted signal, and not with propagation path loss models.



Matrix illustrating type of fading experienced by a signal as a function of: (a) symbol period; and (b) baseband signal bandwidth

Diversity Techniques

Diversity is a powerful communication receiver technique that provides wireless link improvement at relatively low cost. Diversity exploits the random nature of radio propagation by finding independent (or at least highly uncorrelated) signal paths for communication. In virtually all applications, diversity decisions are made by the receiver, and are unknown to the transmitter.

The diversity concept can be explained simply. If one radio path undergoes a deep fade, another independent path may have a strong signal. By having more than one path to select from, both the instantaneous and average SNRs at the receiver may be improved.

In order to prevent deep fades from occurring, *microscopic diversity techniques* can exploit the rapidly changing signal. For example, the small-scale fading reveals that if two antennas are separated by a fraction of a meter, one may receive a null while the other receives a strong signal. By selecting the best signal at all times, a receiver can mitigate small-scale fading effects ; this is called ***antenna diversity or space diversity***.

Large-scale fading is caused by shadowing due to variations in both the terrain profile and the nature of the surroundings. In deeply shadowed conditions, the received signal strength at a mobile can drop well below that of free space. By selecting a base station which is not shadowed when others are, the mobile can improve substantially the average signal-to-noise ratio on the forward link. This is called *macroscopic diversity*, since the mobile is taking advantage of large separations (the macrosystem differences) between the serving base stations.

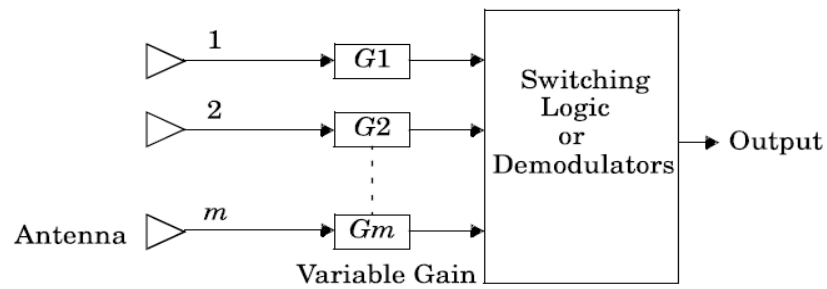
Macroscopic diversity is also useful at the base station receiver. By using base station antennas that are sufficiently separated in space, the base station is able to improve the reverse link by selecting the antenna with the strongest signal from the mobile.

Space diversity reception methods can be classified into *four categories*:

1. Selection diversity
2. Feedback diversity
3. Maximal ratio combining
4. Equal gain diversity

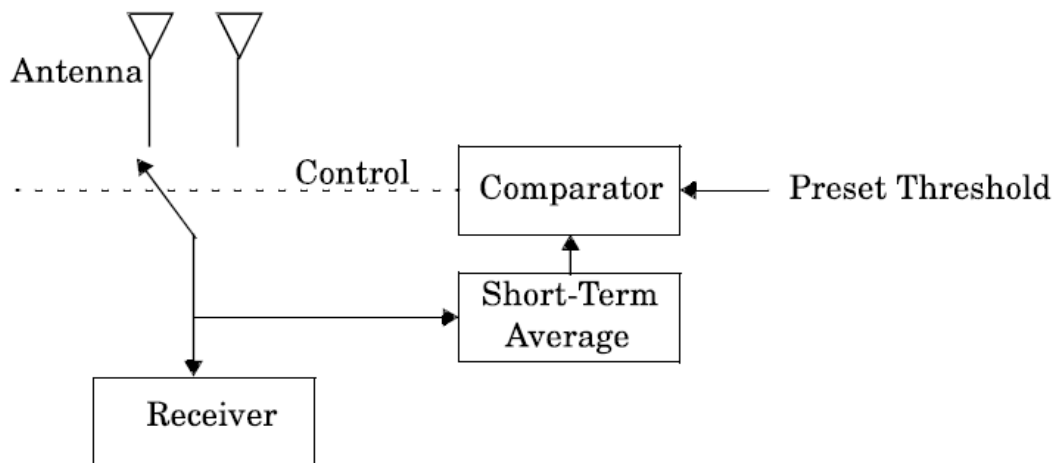
Selection Diversity

A block diagram of this method is similar to that shown in Figure, where m demodulators are used to provide m diversity branches whose gains are adjusted to provide the same average SNR for each branch. The receiver branch having the highest instantaneous SNR is connected to the demodulator. The antenna signals themselves could be sampled and the best one sent to a single demodulator. In practice, the branch with the largest $(S + N)/N$ is used. A practical selection diversity system cannot function on a truly instantaneous basis, but must be designed so that the internal time constants of the selection circuitry are shorter than the reciprocal of the signal fading rate.

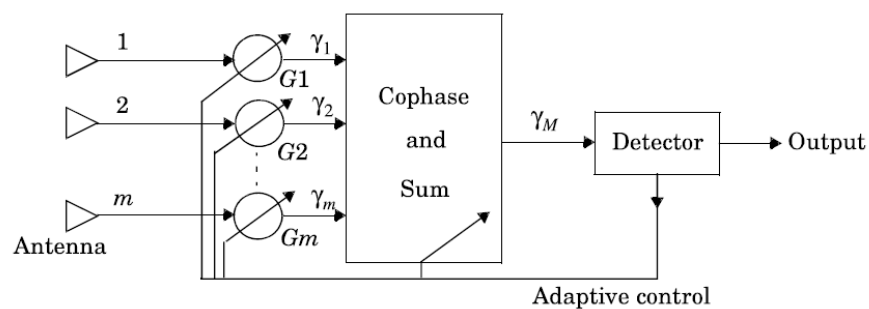


Feedback or Scanning Diversity

Scanning diversity is very similar to selection diversity except that instead of always using the best of M signals, the M signals are scanned in a fixed sequence until one is found to be above a predetermined threshold. This signal is then received until it falls below threshold and the scanning process is again initiated. The resulting fading statistics are somewhat inferior to those obtained by the other methods, but the advantage with this method is that it is very simple to implement-only one receiver is required. A block diagram of this method is shown in Figure.



Maximal Ratio Combining



In this method, the signals from all of the M branches are weighted according to their individual signal voltage to noise power ratios and then summed. Figure shows a block diagram of the technique. Here, the individual signals must be cophased before being summed (unlike selection diversity) which generally requires an individual receiver and phasing circuit for each antenna element. Maximal ratio combining produces an output SNR equal to the sum of the individual SNRs. Thus, it has the advantage of producing an output *with an acceptable SNR even when none of the individual signals are themselves acceptable*. This technique gives the best statistical reduction of fading of any known linear diversity combiner.

Equal Gain Combining

In certain cases, it is not convenient to provide for the variable weighting capability required for true maximal ratio combining. In such cases, the branch weights are all set to unity, but the signals from each branch are co-phased to provide *equal gain combining* diversity. This allows the receiver to exploit signals that are simultaneously received on each branch.

Polarization Diversity

The comparatively high cost of using space diversity at the base station prompts the consideration of using orthogonal polarization to exploit polarization diversity. Most subscribers are no longer using vertical polarization due to hand-tilting when the portable cellular phone is used. This recent phenomenon has sparked interest in polarization diversity at the base station.

Multiple versions of a signal are transmitted and received via antennas with different polarization. A diversity combining technique is applied on the receiver side combines pairs of antennas with orthogonal polarizations. Reflected signals can undergo polarization changes depending on the medium through which they are travelling. A polarisation difference of 90° will result in an attenuation factor of up to 34dB in signal strength. By pairing two complementary polarizations, this scheme can immunize a system from polarization mismatches that would otherwise cause signal fade.

Line-of-sight microwave link use polarization diversity to support two simultaneous users on the same radio channel. It is assumed that the signal is transmitted from a mobile with vertical (or

horizontal) polarization. It is received at the base station by a polarization diversity antenna with two branches. A polarization diversity antenna is composed of two antenna elements which are orthogonal to each other. Some of the vertically polarized signals transmitted are converted to the horizontal polarized signal because of multipath propagation and viz.

Frequency Diversity

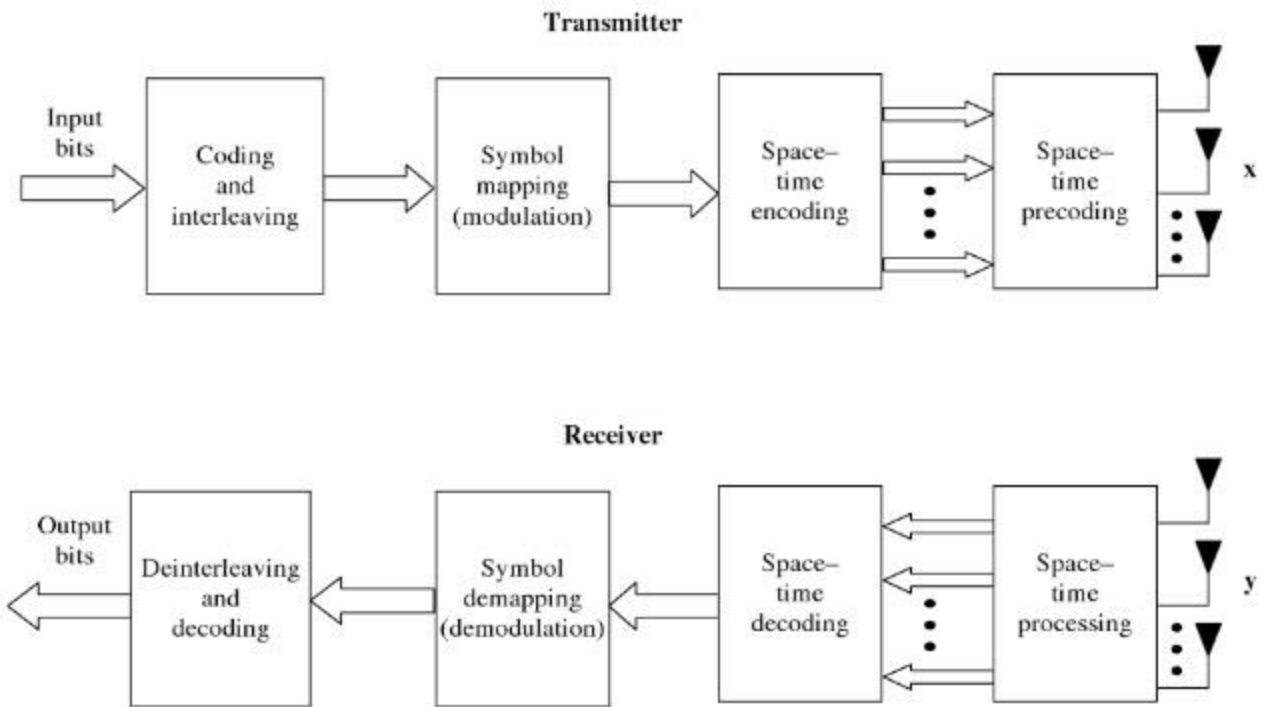
Frequency diversity is implemented by transmitting information on more than one carrier frequency. The rationale behind this technique is that frequencies separated by more than the coherence bandwidth of the channel will be uncorrelated and will thus not experience the same fades. If a particular frequency undergoes a fade, the composite signal will still be demodulated.

Time Diversity

Time diversity repeatedly transmits information at time spacings that exceed the coherence time of the channel, so that multiple repetitions of the signal will be received with independent fading conditions, thereby providing for diversity. Each replica experiences a particular multipath delay, so that a better estimate of the original signal may be formed at the receiver.

Basic Block Diagram of MIMO System

Below figure shows the basic block diagram of MIMO systems. The information bits to be transmitted are encoded by using a conventional encoder. And that's to be interleaved. The interleaved codeword is mapped to data symbols (Quadrature amplitude modulation symbols) by using a symbol mapper.



Basic Block Diagram of MIMO System

These data symbols are input to a space-time encoder that outputs, one or more spatial data streams. The spatial data streams are mapped to the transmit antennas by space-time pre-coding block. The signals launched from the transmit antennas propagate through the channel and arrive at the receive antenna array. The receiver collects the signals at the output of each receive antenna element and reverses the transmitter operations in order to decode the data: receive space-time processing, followed by space-time decoding, symbol demapping, deinterleaving, and decoding.