KEY ENABLING TECHNOLOGIES AND FEATURES OF LTE

To meet its service and performance requirements, LTE design incorporates several important enabling radio and core network technologies. Here, we provide a brief introduction to some of the key enabling technologies used in the LTE design.

1.4.1 Orthogonal Frequency Division Multiplexing (OFDM)

One of the key differences between existing 3G systems and LTE is the use of Orthogonal Frequency Division Multiplexing (OFDM) as the underlying modulation technology. Widely deployed 3G systems such as UMTS and CDMA2000 are based on Code Division Multiple Access (CDMA) technology. CDMA works by spreading a narrow band signal over a wider bandwidth to achieve interference resistance, and performs remarkably well for low data rate communications such as voice, where a large number of users can be multiplexed to achieve high system capacity. However, for high-speed applications, CDMA becomes untenable due to the large bandwidth needed to achieve useful amounts of spreading.

OFDM has emerged as a technology of choice for achieving high data rates. It is the core technology used by a variety of systems including Wi-Fi and WiMAX.

The following advantages of OFDM led to its selection for LTE:

• **Elegant solution to multipath interference**: The critical challenge to high bit-rate transmissions in a wireless channel is intersymbol interference caused by multipath. In a multipath environment, when the time delay between the various signal paths is a significant fraction of the transmitted signal's symbol period, a transmitted symbol may arrive at the receiver during the next symbol and cause intersymbol interference (ISI). At high data rates, the symbol time is shorter; hence, it only takes a small delay to cause ISI, making it a bigger challenge for broadband wireless. OFDM is a multicarrier modulation technique that overcomes this challenge in an elegant manner. The basic idea behind multicarrier modulation is to divide a given high-bit-rate data stream into several parallel lower bit-rate streams and modulate each stream on separate carriers—often called subcarriers, or tones. Splitting the data stream into many parallel streams increases the symbol duration of each stream such that the multipath delay spread is only a small fraction of the symbol duration. OFDM is a spectrally efficient version of multicarrier modulation, where the subcarriers are selected such that they are all

orthogonal to one another over the symbol duration, thereby avoiding the need to have non-overlapping subcarrier channels to eliminate inter-carrier interference. In OFDM, any residual intersymbol interference can also be eliminated by using guard intervals between OFDM symbols that are larger than the expected multipath delay. By making the guard interval larger than the expected multipath delay spread, ISI can be completely eliminated. Adding a guard interval, however, implies power wastage and a decrease in bandwidth efficiency.

• **Reduced computational complexity**: OFDM can be easily implemented using Fast Fourier Transforms (FFT/IFFT), and the computational requirements grow only slightly faster than linearly with data rate or bandwidth. The computational complexity of OFDM can be shown to be O($BlogBT_m$) where B is the bandwidth and T_m is the delay spread. This complexity is much lower than that of a time-domain equalizer-based system—the traditional means for combating multipath interference—which has a complexity of O(B^2T_m). Reduced complexity is particularly attractive in the downlink as it simplifies receiver processing and thus reduces mobile device cost and power consumption. This is especially important given the wide transmission bandwidths of LTE coupled with multistream transmissions.

• **Graceful degradation of performance under excess delay**: The performance of an OFDM system degrades gracefully as the delay spread exceeds the value designed for. Greater coding and low constellation sizes can be used to provide fallback rates that are significantly more robust against delay spread. In other words, OFDM is well suited for adaptive modulation and coding, which allows the system to make the best of the available channel conditions. This contrasts with the abrupt degradation owing to error propagation that single-carrier systems experience as the delay spread exceeds the value for which the equalizer is designed.

• **Exploitation of frequency diversity**: OFDM facilitates coding and interleaving across subcarriers in the frequency domain, which can provide robustness against burst errors caused by portions of the transmitted spectrum undergoing deep fades. OFDM also allows for the channel bandwidth to be scalable without impacting the hardware design of the base station and the mobile station. This allows LTE to be deployed in a variety of spectrum allocations and different channel bandwidths.

• **Enables efficient multi-access scheme**: OFDM can be used as a multi-access scheme by partitioning different subcarriers among multiple users. This scheme is referred to as OFDMA and is exploited in LTE. OFDMA offers the ability to provide fine granularity in channel allocation, which can be exploited to achieve significant capacity improvements, particularly in slow time-varying channels.

• **Robust against narrowband interference**: OFDM is relatively robust against narrowband interference, since such interference affects only a fraction of the subcarriers.

• **Suitable for coherent demodulation**: It is relatively easy to do pilot-based channel estimation in OFDM systems, which renders them suitable for coherent demodulation schemes that are more power efficient.

• **Facilitates use of MIMO**: MIMO stands for multiple input multiple output and refers to a collection of signal processing techniques that use multiple antennas at both the transmitter and receiver to improve system performance. For MIMO techniques to be effective, it is required that the channel conditions are such that the multipath delays do not cause intersymbol interference—in other words, the channel has to be a flat fading channel and not a frequency selective one. At very high data rates, this is not the case and therefore MIMO techniques do not work well in traditional broadband channels. OFDM, however, converts a frequency selective broad band channel into several narrowband flat fading channels where the MIMO models and techniques work well. The ability to effectively use MIMO techniques to improve system capacity gives OFDM a significant advantage over other techniques and is one of the key reasons for its choice. MIMO and OFDM have already been combined effectively in Wi-Fi and WiMAX systems.

• **Efficient support of broadcast services**: By synchronizing base stations to timing errors well within the OFDM guard interval, it is possible to operate an OFDM network as a single frequency network (SFN). This allows broadcast signals from different cells to combine over the air to significantly enhance the received signal power, thereby enabling higher data rate broadcast transmissions for a given transmit power. LTE design leverages this OFDM capability to improve efficient broadcast services.

While all these advantages drove 3GPP to adopt OFDM as their modulation choice, it should be noted that OFDM also suffers from a few disadvantages. Chief among these is the problem associated with OFDM signals having high peak-to-average ratio (PAR), which causes non-linearities and clipping distortion when passed through an RF amplifier. Mitigating this problem requires the use of expensive and inefficient power amplifiers with high requirements on linearity, which increases the cost of the transmitter and is wasteful of power.

While the increased amplifier costs and power inefficiency of OFDM is tolerated in the downlink as part of the design, for the uplink LTE selected a variation of OFDM that has a lower peak-to-average ratio. The modulation of choice for the uplink is called Single Carrier Frequency Division Multiple Access (SC-FDMA).

1.4.2 SC-FDE and SC-FDMA

To keep the cost down and the battery life up, LTE incorporated a power efficient transmission scheme for the uplink. Single Carrier Frequency Domain Equalization (SC-FDE) is conceptually similar to OFDM but instead of transmitting the Inverse Fast Fourier Transform (IFFT) of the actual data symbols, the data symbols are sent as a sequence of QAM symbols with a cyclic prefix added; the IFFT is added at the end of the receiver. SC-FDE retains all the advantages of OFDM such as multipath resistance and low complexity, while having a low peak-to-average ratio of 4-5dB. The uplink of LTE implements a multi-user version of SC-FDE, called SC-FDMA, which allows multiple users to use parts of the frequency spectrum. SC-FDMA closely resembles OFDMA and can in fact be thought of as "DFT precoded OFDMA." SC-FDMA also preserves the PAR properties of SC-FDE but increases the complexity of the transmitter and the receiver.

1.4.3 Channel Dependent Multi-user Resource Scheduling

The OFDMA scheme used in LTE provides enormous flexibility in how channel resources are allocated. OFDMA allows for allocation in both time and frequency and it is possible to design algorithms to allocate resources in a flexible and dynamic manner to meet arbitrary throughput, delay, and other requirements. The standard supports dynamic, channel-dependent scheduling to enhance overall system capacity.

Given that each user will be experiencing uncorrelated fading channels, it is possible to allocate subcarriers among users in such a way that the overall capacity is increased. This technique, called frequency selective multiuser scheduling, calls for focusing transmission power in each user's best channel portion, thereby increasing the overall capacity. Frequency selective scheduling requires good channel tracking and is generally only viable in slow varying channels. For fast varying channels, the overhead involved in doing this negates the potential capacity gains. In OFDMA, frequency selective scheduling can be combined with multi-user time domain scheduling, which calls for scheduling users during the crests of their individual fading channels. Capacity gains are also obtained by adapting the modulation and coding to the instantaneous signal-to-noise ratio conditions for each user subcarrier.

For high-mobility users, OFDMA can be used to achieve frequency diversity. By coding and interleaving across subcarriers in the frequency domain using a uniform random distribution of subcarriers over the whole spectrum, the signal can be made

more robust against frequency selective fading or burst errors. Frequency diverse scheduling is best suited for control signaling and delay sensitive services.

1.4.4 Multi antenna Techniques

The LTE standard provides extensive support for implementing advanced multiantenna solutions to improve link robustness, system capacity, and spectral efficiency. Depending on the deployment scenario, one or more of the techniques can be used. Multiantenna techniques supported in LTE include:

• **Transmit diversity**: This is a technique to combat multipath fading in the wireless channel. The idea here is to send copies of the same signal, coded differently, over multiple transmit antennas. LTE transmit diversity is based on space-frequency block coding (SFBC) techniques complemented with frequency shift time diversity (FSTD) when four transmit antenna are used. Transmit diversity is primarily intended for common downlink channels that cannot make use of channel-dependent scheduling. It can also be applied to user transmissions such as low data rate VoIP, where the additional overhead of channel-dependent scheduling may not be justified. Transmit diversity increases system capacity and cell range.

• **Beamforming**: Multiple antennas in LTE may also be used to transmit the same signal appropriately weighted for each antenna element such that the effect is to focus the transmitted beam in the direction of the receiver and away from interference, thereby improving the received signal-to-interference ratio. Beamforming can provide significant improvements in coverage range, capacity, reliability, and battery life. It can also be useful in providing angular information for user tracking. LTE supports beamforming in the downlink.

• **Spatial multiplexing**: The idea behind spatial multiplexing is that multiple independent streams can be transmitted in parallel over multiple antennas and can be separated at the receiver using multiple receive chains through appropriate signal processing. This can be done as long as the multipath channels as seen by the different antennas are sufficiently decorrelated as would be the case in a scattering rich environment. In theory, spatial multiplexing provides data rate and capacity gains proportional to the number of antennas used. It works well under good SNR and light load conditions, and hence tends to have a more pronounced effect on peak rates rather than overall system capacity. LTE standard supports spatial multiplexing with up to four transmit antennas and four receiver antennas.

• **Multi-user MIMO**: Since spatial multiplexing requires multiple transmit chains, it is currently not supported in the uplink due to complexity and cost

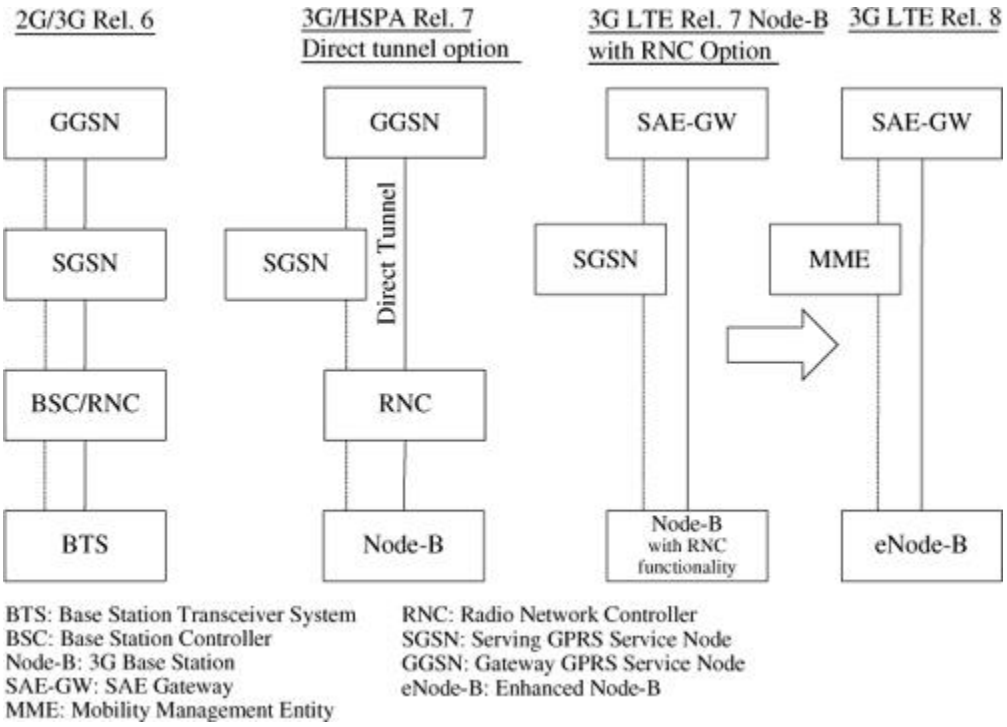
considerations. However, multi-user MIMO (MU-MIMO), which allows multiple users in the uplink, each with a single antenna, to transmit using the same frequency and time resource, is supported. The signals from the different MU-MIMO users are separated at the base station receiver using accurate channel state information of each user obtained through uplink reference signals that are orthogonal between users.

1.4.5 IP-Based Flat Network Architecture

Besides the air-interface, the other radical aspect of LTE is the flat radio and core network architecture [15]. "Flat" here implies fewer nodes and a less hierarchical structure for the network. The lower cost and lower latency requirements drove the design toward a flat architecture since fewer nodes obviously implies a lower infrastructure cost. It also means fewer interfaces and protocol-related processing, and reduced interoperability testing, which lowers the development and deployment cost. Fewer nodes also allow better optimization of radio interface, merging of some control plane protocols, and short session start-up time.

Figure 1.3 shows how the 3GPP network architecture evolved over a few releases. 3GPP Release 6 architecture, which is conceptually very similar to its predecessors, has four network elements in the data path: the base station or Node-B, radio network controller (RNC), serving GPRS service node (SGSN), and gateway GRPS service node (GGSN). Release 7 introduced a direct tunnel option from the RNC to GGSN, which eliminated SGSN from the data path. LTE on the other hand, will have only two network elements in the data path: the enhanced Node-B or eNode-B, and a System Architecture Evolution Gateway (SAE-GW). Unlike all previous cellular systems, LTE merges the base station and radio network controller functionality into a single unit. The control path includes a functional entity called the Mobility Management Entity (MME), which provides control plane functions related to subscriber, mobility, and session management. The MME and SAE-GW could be collocated in a single entity called the access gateway (a-GW). More details about the network architecture are provided in the next section.

Figure 1.3 3GPP evolution toward a flat LTE SAE architecture.



A key aspect of the LTE flat architecture is that all services, including voice, are supported on the IP packet network using IP protocols. Unlike previous systems, which had a separate circuit-switched subnetwork for supporting voice with their own Mobile Switching Centers (MSC) and transport networks, LTE envisions only a single evolved packet-switched core, the EPC, over which all services are supported, which could provide huge operational and infrastructure cost savings. It should be noted, however, that although LTE has been designed for IP services with a flat architecture, due to backwards compatibility reasons certain legacy, non-IP aspects of the 3GPP architecture such as the GPRS tunneling protocol and PDCP (packet data convergence protocol) still exists within the LTE network architecture.

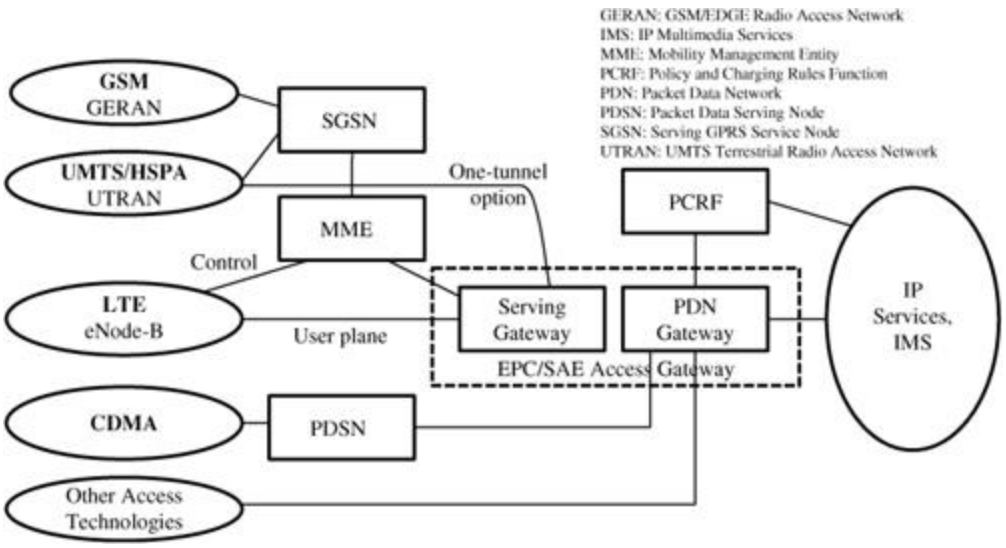
1.5 LTE NETWORK ARCHITECTURE

While the focus of this book is on the radio network aspects of LTE, a basic understanding of the overall end-to-end architecture is useful to gain an appreciation of how services are delivered over an LTE network. To that end, we provide a brief overview of the LTE network architecture in this section.

As already mentioned, the core network design presented in 3GPP Release 8 to support LTE is called the Evolved Packet Core (EPC). EPC is designed to provide a high-capacity, all IP, reduced latency, flat architecture that dramatically reduces cost and supports advanced real-time and media-rich services with enhanced quality of experience. It is designed not only to support new radio access networks such as LTE, but also provide interworking with legacy 2G GERAN and 3G UTRAN networks connected via SGSN. Functions provided by the EPC include access control, packet routing and transfer, mobility management, security, radio resource management, and network management.

The EPC includes four new elements: (1) Serving Gateway (SGW), which terminates the interface toward the 3GPP radio access networks; (2) Packet Data Network Gateway (PGW), which controls IP data services, does routing, allocates IP addresses, enforces policy, and provides access for non-3GPP access networks; (3) Mobility Management Entity (MME), which supports user equipment context and identity as well as authenticates and authorizes users; and (4) Policy and Charging Rules Function (PCRF), which manages QoS aspects. Figure 1.4 shows the end-to-end architecture including how the EPC supports LTE as well as current and legacy radio access networks.

Figure 1.4 Evolved Packet Core architecture.



A brief description of each of the four new elements is provided here:

• **Serving Gateway (SGW)**: The SGW acts as a demarcation point between the RAN and core network, and manages user plane mobility. It serves as the mobility anchor when terminals move across areas served by different eNode-B elements in

E-UTRAN, as well as across other 3GPP radio networks such as GERAN and UTRAN. SGW does downlink packet buffering and initiation of network-triggered service request procedures. Other functions include lawful interception, packet routing and forwarding, transport level packet marking in the uplink and the downlink, accounting support for per user, and inter-operator charging.

• **Packet Data Network Gateway (PGW)**: The PGW acts as the termination point of the EPC toward other Packet Data Networks (PDN) such as the Internet, private IP network, or the IMS network providing end-user services. It serves as an anchor point for sessions toward external PDN and provides functions such as user IP address allocation, policy enforcement, packet filtering, and charging support. Policy enforcement includes operator-defined rules for resource allocation to control data rate, QoS, and usage. Packet filtering functions include deep packet inspection for application detection.

• **Mobility Management Entity (MME)**: The MME performs the signaling and control functions to manage the user terminal access to network connections, assignment of network resources, and mobility management function such as idle mode location tracking, paging, roaming, and handovers. MME controls all control plane functions related to subscriber and session management. The MME provides security functions such as providing temporary identities for user terminals, interacting with Home Subscriber Server (HSS) for authentication, and negotiation of ciphering and integrity protection algorithms. It is also responsible for selecting the appropriate serving and PDN gateways, and selecting legacy gateways for hand-overs to other GERAN or UTRAN networks. Further, MME is the point at which lawful interception of signaling is made. It should be noted that an MME manages thousands of eNode-B elements, which is one of the key differences from 2G or 3G platforms using RNC and SGSN platforms.

• **Policy and Charging Rules Function (PCRF)**: The Policy and Charging Rules Function (PCRF) is a concatenation of Policy Decision Function (PDF) and Charging Rules Function (CRF). The PCRF interfaces with the PDN gateway and supports service data flow detection, policy enforcement, and flow-based charging. The PCRF was actually defined in Release 7 of 3GPP ahead of LTE. Although not much deployed with pre-LTE systems, it is mandatory for LTE. Release 8 further enhanced PCRF functionality to include support for non-3GPP access (e.g., Wi-Fi or fixed line access) to the network.

2.2 THE BROADBAND WIRELESS CHANNEL: PATH LOSS AND SHADOWING

The main goal of this chapter is to explain the fundamental factors affecting the received signal in a wireless system, and how they can be modelled using a handful of parameters. The relative values of these parameters, which are summarized in Table 2.1 and described throughout this section, make all the difference when designing a wireless communication system. In this section we will introduce the overall channel model, and discuss the *large-scale* trends that affect this model.

Table 2.1 Key Wireless Channel Parameters

Symbol	Parameter
α	Path loss exponent
v	Number of ISI taps in channel ($v+1$ is total number of taps)
σ_s	Log normal shadowing standard deviation
f_D	Doppler spread (maximum Doppler frequency), $f_D = \frac{\nu f_c}{c}$
c	Speed of light
ν	Relative speed between transmitter and receiver
f_c	Carrier frequency
λ	Carrier wavelength, $f_c\lambda = c$
T_c	Channel coherence time, $T_c \approx f_D^{-1}$
$\tau_{\max}$	Channel delay spread (maximum)
τ_{rms}	Channel delay spread (Root Mean Square)
B_c	Channel coherence bandwidth, $B_c \approx \tau^{-1}$
θ_{rms}	Angular spread (Root Mean Square)

The overall model we will use for describing the channel in discrete time is a simple tap-delay line (TDL):

(2.1)

$$h[k,t] = h_0\delta[k,t] + h_1\delta[k-1,t] + \ldots + h_v\delta[k-v,t].$$

Here, the discrete-time channel is time-varying (so changes with respect to t), and has non-negligible values over a span of $v + 1$ channel taps. Generally, we will assume that the channel is sampled at a frequency $f_s = 1/T$, where T is the symbol period,[1] and hence the duration of the channel in this case is about vT. The $v + 1$ sampled values are, in general, complex numbers.

Assuming that the channel is static over a period of $(v + 1)T$ seconds, the output of the channel can then be described as

(2.2)

$$y[k,t] = \sum_{j=-\infty}^{\infty} h[j,t]x[k-j]$$

(2.3)

$$\triangleq h[k,t] * x[k],$$

where $x[k]$ is an input sequence of data symbols with rate $1/T$, and * denotes convolution. In simpler notation, the channel can be represented as a time-varying $(v + 1) \times 1$ column vector:[2]

(2.4)

$$\mathbf{h}(t) = [h_0(t)\ h_1(t) \ldots h_v(t)]^T.$$

Although this tapped-delay line model is general and accurate, it is difficult to design a communication system for the channel without knowing some of the key attributes about $\mathbf{h}(t)$. Some likely questions one might have are

• What is the value for the total received power? In other words, what are the relative values of the h_i terms?

Answer: As we will see, a number of different effects cause the received power to vary over long (path loss), medium (shadowing), and short (fading) distances.

• How quickly does the channel change with the parameter t?

Answer: The *channel coherence time* specifies the period of time over which the channel's value is correlated. The coherence time depends on how fast the transmitter and receiver are moving relative to each other.

• What is the approximate value of the channel duration v?

Answer: This value is known as the *delay spread*, and is measured or approximated based on the propagation distance and environment.

The rest of the chapter explores these questions more deeply in an effort to characterize and explain these key wireless channel parameters, which are given in Table 2.1.

2.2.1 Path Loss

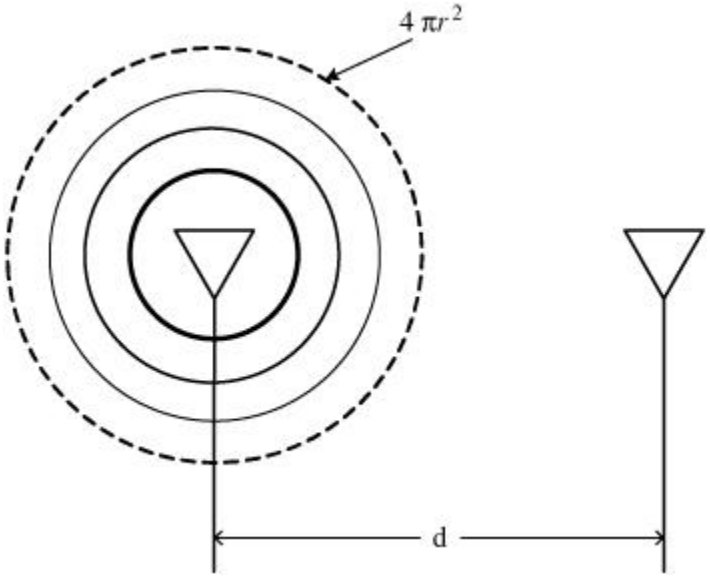
The first obvious difference between wired and wireless channels is the amount of transmitted power that actually reaches the receiver. Assuming an isotropic

antenna is used, as shown in Figure 2.2, the propagated signal energy expands over a spherical wavefront, so the energy received at an antenna a distance *d* away is inversely proportional to the sphere surface area, $4\pi d^2$. The *free-space path loss formula*, or Friis formula, is given more precisely as

(2.5)

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

Figure 2.2 Free space propagation.



where P_r and P_t are the received and transmitted powers and λ is the wavelength. In the context of the TDL model of (2.1), P_r/P_t is the average value of the channel gain; that is, $P_r/P_t = E\|\mathbf{h}\|^2$, where $E[\cdot]$ denotes the expected value, or mathematical mean. If directional antennas are used at the transmitter or receiver, a gain of G_t and/or G_r is achieved, and the received power is simply increased by the gain of these antennas.[3] An important observation from (2.5) is that since $c = f_c\lambda \Rightarrow \lambda = c/f_c$, the received power fall offs quadratically with the carrier frequency. In other words, for a given transmit power, the range is decreased when higher frequency waves are used. This has important implications for high-data rate systems, since most large bandwidths are available at higher frequencies.

RANGE VS. BANDWIDTH

As observed in (2.5), $P_r \propto \lambda^2$, which means that $P_r \propto \frac{1}{f_c^2}$. Clearly, higher frequencies suffer greater power loss than lower frequencies. As a result, lower carrier frequencies are generally more desirable, and hence very crowded. For example, compare an LTE system at 2100MHz (the new PCS frequencies) with one deployed in the newly available 700MHz spectrum. All else being equal, the 700MHz system will have nine times the received power. In fact, measurement campaigns have consistently shown that the effective path loss exponent α also increases at higher frequencies, due to increased absorption and attenuation of high frequency signals [27,32,33,55], so the difference can be greater still. From a coverage point of view, this creates a distinct advantage for 700MHz systems.

On the other hand, bandwidth at higher carrier frequencies is more plentiful, more consistently available on a global basis, and almost always less expensive. Hence, a high-rate, low-cost system would generally prefer to operate at higher frequencies, if the customer base is dense enough to overcome the coverage problems at those frequencies. Perhaps an ideal scenario from an operator standpoint would be to initially roll out LTE at low frequencies (like 700MHz) in order to serve many customers with few base stations. Then, as the subscriber base grows to a critical mass, shift to a shorter-range network operating at higher frequencies to support higher density at lower cost.

The terrestrial propagation environment is not free space. Intuitively, it seems that reflections from the Earth or other objects would actually increase the received power since more energy would reach the receiver. However, because a reflected wave often experiences a 180-degree phase shift, at relatively large distances (usually over a kilometer) the reflection serves to create destructive interference, and the common *2-ray approximation* for path loss is:

(2.6)

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

which is significantly different from free-space path loss in several respects. First, the antenna heights now assume a very important role in the propagation, as is anecdotally familiar: radio transmitters are usually placed on the highest available object. Second, the wavelength and hence carrier frequency dependence has disappeared from the formula, which is not typically observed in practice, however. Third, and crucially, the distance dependence has changed to d^{-4}, implying that energy loss is more severe with distance in a terrestrial system than in free space.

In order to more accurately describe different propagation environments, empirical models are often developed using experimental data. One of the simplest and most common is the *empirical path loss formula*:

(2.7)

$$P_r = P_t P_o \left(\frac{d_o}{d}\right)^{\alpha},$$

which groups all the various effects into two parameters, the path loss exponent α and the measured path loss P_o at a reference distance of d_o, which is often chosen as 1 meter. Although P_o should be determined from measurements, it is often well approximated (within several dB) as simply $(\lambda/4\pi)^2$ when $d_o = 1$. This simple empirical path loss formula is capable of reasonably representing most of the important path loss trends with just these two parameters, at least over some range of interest.

EXAMPLE 2.1

Consider a user in the downlink of a cellular system, where the desired base station is at a distance of 500 meters (.5 km), and there are numerous nearby interfering base stations transmitting at the same power level. If there are three interfering base stations at a distance of 1 km, three at a distance of 2 km, and ten at a distance of 4 km, use the empirical path loss formula to find the signal-to-interference ratio (SIR, i.e., the noise is neglected) when $\alpha = 3$, and then when $\alpha = 5$.

For $\alpha = 3$ and d_0 in units of kilometers, the desired received power is

(2.8)

$$P_{r,d} = P_t P_o d_o^3 (0.5)^{-3},$$

and the interference power is

(2.9)

$$P_{r,I} = P_t P_o d_o^3 \left[3(1)^{-3} + 3(2)^{-3} + 10(4)^{-3}\right].$$

The SIR expressions compute to

$$\mathrm{SIR}(\alpha = 3) = \frac{P_{r,d}}{P_{r,I}} = 2.27 \quad (3.55 \text{ dB}),$$
$$\mathrm{SIR}(\alpha = 5) = 10.32 \quad (10.32 \text{ dB}),$$

demonstrating that the overall system performance can be substantially improved when the path loss is in fact large. These calculations can be viewed as an upper bound, where the signal-to-interference plus noise ratio (SINR) is less than the SIR, due to the addition of noise. This means that as the path loss worsens, microcells grow increasingly attractive since the required signal power can be decreased down to the noise floor, and the overall performance will actually be better than in a system with lower path loss at the same transmit power level.

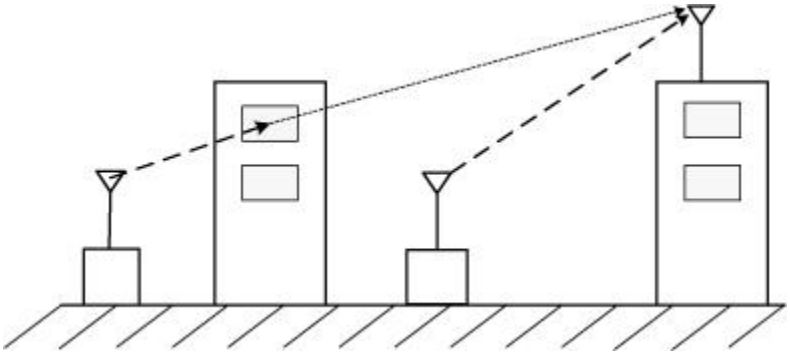
2.2.2 Shadowing

As we have seen in the last section, path loss models attempt to account for the distance-dependent relationship between transmitted and received power. However, many factors other than distance can have a large effect on the total received power. For example, as shown in Figure 2.3, obstacles such as trees and buildings may be located between the transmitter and receiver, and cause temporary degradation in received signal strength, while on the other hand a temporary line-of-sight transmission path would result in abnormally high received power. Since modelling the locations of all objects in every possible communication environment is generally impossible, the standard method of accounting for these variations in signal strength is to introduce a random effect called *shadowing*. With shadowing, the empirical path loss formula becomes

(2.10)

$$P_r = P_t P_o \chi \left(\frac{d_o}{d}\right)^{\alpha},$$

Figure 2.3 Shadowing can cause large deviations from path loss predictions.



where χ is a sample of the *shadowing* random process. Hence, the received power is now also modelled as a random process. In effect, the distance-trend in the path loss can be thought of as the mean (or expected) received power, while the χ shadowing value causes a perturbation from that expected value. It should be emphasized that since shadowing is caused by macroscopic objects, typically it has a correlation distance on the order of meters or tens of meters. Hence, shadowing is often alternatively called "large-scale fading."

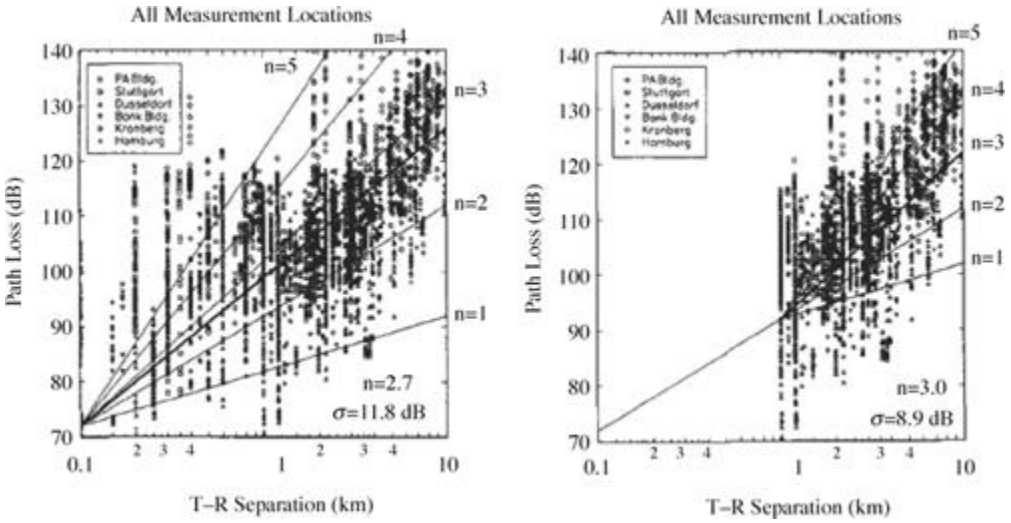
The shadowing value χ is typically modelled as a lognormal random variable, that is

(2.11)

$$\chi = 10^{x/10}, \quad \text{where } x \sim N(0, \sigma_s^2),$$

where $N(0, \sigma_s^2)$ is a Gaussian (normal) distribution with mean 0 and variance σ_s^2. With this formulation, the standard deviation σ_s is expressed in dB. Typical values for σ_s are in the 6–12 dB range. Figure 2.4 shows the very important effect of shadowing, where σ = 11.8 dB and σ = 8.9 dB, respectively.

Figure 2.4 Shadowing causes large random fluctuations about the path loss model. Figure from [42], courtesy of IEEE.



Shadowing is an important effect in wireless networks because it causes the received SINR to vary dramatically over long time scales. In some locations in a given cell, reliable high-rate communication may be nearly impossible. The system design and base station deployment must account for lognormal shadowing either through macrodiversity, variable transmit power, and/or by simply accepting that

some users will experience poor performance at a certain percentage of locations. While on occasion shadowing can be beneficial—for example, if there is an object blocking interference—it is generally detrimental to system performance because it requires a several-dB margin to be built into the system. Let's do a realistic numerical example to see how shadowing affects wireless system design.

EXAMPLE 2.2. THE EFFECT OF SHADOWING

Consider an LTE base station (BS) communicating to a subscriber, with the channel parameters $\alpha = 3$, $P_o = -40$ dB, $d_0 = 1$ m, $\sigma_s = 6$ dB. We assume a transmit power of $P_t = 1$ Watt (30 dBm),[4] a bandwidth of $B = 10$MHz and due to error correction coding, a received SNR of 14.7 dB is required for 16QAM, while just 3 dB is required for BPSK. Finally, we consider only ambient noise with a typical power spectral density of $N_o = -173$ dBm/Hz, with an additional receiver noise figure of $N_f = 5$ dB.[5]

The question is this: At a distance of 500 meters from the base station, what is the likelihood that the BS can reliably send BPSK or 16QAM?

To solve this problem, we must find an expression for received SNR, and then compute the probability that it is above the BPSK and 16QAM thresholds.

First, let's compute the received power, P_r in dB based on (2.10)

(2.12)

$$P_r(\text{dB}) = 10\log_{10} P_t + 10\log_{10} P_o - 10\log_{10} d^{\alpha} + 10\log_{10} \chi$$

(2.13)

$$= 30 \text{ dBm} - 40 \text{ dB} - 81 \text{ dB} + \chi(\text{dB}) = -91 \text{ dBm} + \chi(\text{dB})$$

Next, we can compute the total noise/interference power I_{tot} in dB similarly:

(2.14)

$$I_{tot}(\text{dB}) = N_o + N_f + 10\log_{10} B$$

(2.15)

$$= -173 \text{ dBm} + 5 \text{ dB} + 70 \text{ dB} = -98 \text{ dBm}$$

The resulting SNR $\gamma = P_r/I_{tot}$ can be readily computed in dB as

(2.16)

$$\gamma = -91 \text{ dBm} + \chi(\text{dB}) + 98 \text{ dBm} = 7 \text{ dB} + \chi(\text{dB}).$$

In this scenario, the average received SNR is 7 dB, good enough for BPSK but not good enough for 16QAM. Since we can see from Equation (2.11) that χ(dB) = x has a zero mean Gaussian distribution with standard deviation 6, the probability that we are able to achieve BPSK can be derived in terms of the complimentary cumulative distribution function (CCDF) of a normalized Gaussian random variable, which is known as the *Q function*:[6]

(2.17)

$$P[\gamma \geq 3 \text{ dB}] = P\left[\frac{\chi + 7}{\sigma} \geq \frac{3}{\sigma}\right]$$

(2.18)

$$= P\left[\frac{\chi}{6} \geq -\frac{4}{6}\right]$$

(2.19)

$$= Q\left(-\frac{4}{6}\right) = 0.75$$

And similarly for 16QAM:

(2.20)

$$P[\gamma \geq 14.7 \text{ dB}] = P\left[\frac{\chi + 7}{\sigma} \geq \frac{14.7}{\sigma}\right]$$

(2.21)

$$= Q\left(\frac{7.7}{6}\right) = .007$$

To summarize the example, while 75% of users can use BPSK modulation and hence get a raw data rate of 10MHz · 1 bit/symbol ·1/2 = 5 Mbps, less than 1% of

users can reliably use 16QAM (4 bits/symbol) for a more desirable data rate of 20 Mbps. Additionally, whereas without shadowing *all* the users could at least get low-rate BPSK through, with shadowing 25% of the users appear unable to communicate at all. Interestingly though, without shadowing 16QAM could never be sent, whereas with shadowing it can be sent a small fraction of the time. In fact, such fluctuations can be exploited using *adaptive modulation and coding*, which we will describe generally in Section 2.6.5 and specifically for LTE in Sections 9.2 and 9.6.

WHY IS THE SHADOWING LOGNORMAL?

While the primary rationale for the lognormal distribution for the shadowing value χ is accumulated evidence from channel measurement campaigns, one plausible explanation is as follows. Neglecting the path loss for a moment, if a transmission experiences N random attenuations β_i, $i = 1, 2, ..., N$ between the transmitter and receiver, the received power can be modelled as

(2.22)

$$P_r = P_t \prod_{i=1}^{N} \beta_i,$$

which can be expressed in dB as

(2.23)

$$P_r(\text{dB}) = P_t(\text{dB}) + 10 \sum_{i=1}^{N} \log_{10} \beta_i.$$

Then, using the Central Limit Theorem (CLT), it can be argued that the sum term will become Gaussian as N becomes large (and often the CLT is accurate for fairly small N), and since the expression is in dB, the shadowing is hence lognormal.

2.3 CELLULAR SYSTEMS

As explained in the previous section, due to path loss and to a lesser extent shadowing, given a maximum allowable transmit power, it is only possible to reliably communicate over some limited distance. However, path loss allows for spatial isolation of different transmitters operating on the same frequency at the same time. As a result, path loss and short-range transmissions in fact *increase* the overall capacity of the system by allowing more simultaneous transmissions to

occur. This straightforward observation is the theoretical basis for the ubiquity of modern cellular communication systems.

In this section, we briefly explore the key aspects of cellular systems, and the closely related topics of sectoring and frequency reuse. Since LTE systems are expected to be deployed primarily in a cellular architecture, the concepts presented in this section will be fundamental to understanding LTE system design and performance.

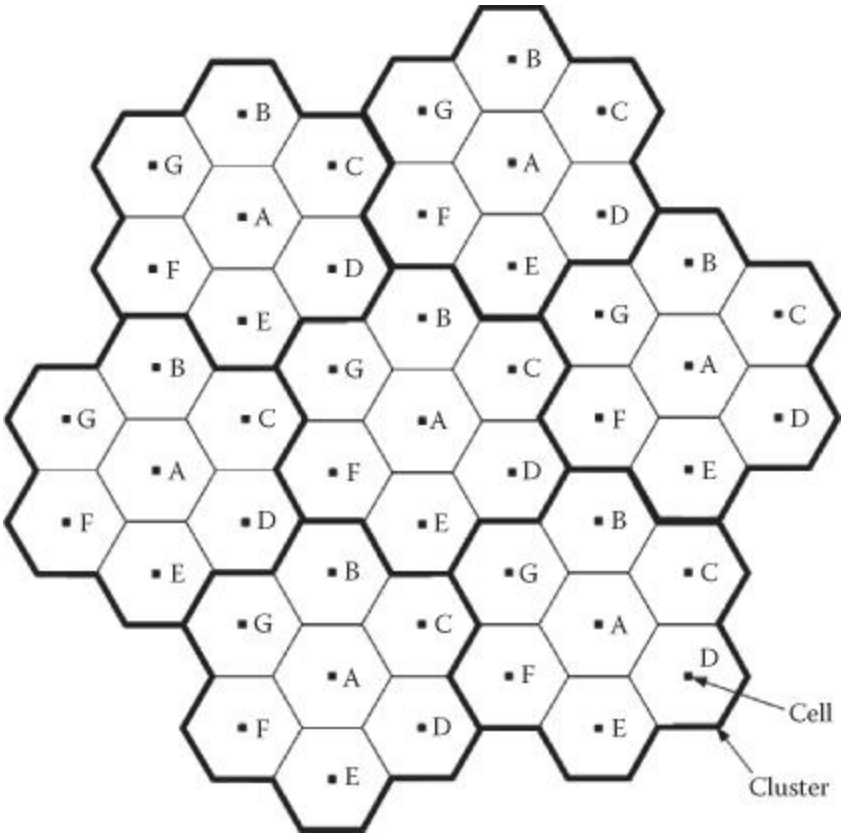
2.3.1 The Cellular Concept

In cellular systems, the service area is subdivided into smaller geographic areas called *cells* that are each served by their own base station. In order to minimize interference between cells, the transmit power level of each base station is regulated to be just enough to provide the required signal strength at the cell boundaries. Then, as we have seen, propagation path loss allows for spatial isolation of different cells operating on the same frequency channels at the same time. Therefore, the same frequency channels can be reassigned to different cells, as long as those cells are spatially isolated.

Although perfect spatial isolation of different cells cannot be achieved, from a practical point of view, the rate at which frequencies can be reused should be determined such that the interference between base stations is kept to an acceptable level. In this context, *frequency planning* is required to determine a proper frequency reuse factor and a geographic reuse pattern. The frequency reuse factor f is defined as $f \leq 1$, where $f = 1$ means that all cells reuse all the frequencies. Accordingly, $f = 1/3$ implies that a given frequency band is used by only 1 out of every 3 cells.

The reuse of the same frequency channels should be intelligently planned in order to maximize the geographic distance between the co-channel base stations. Figure 2.5 shows an example of hexagonal cellular system model with frequency reuse factor $f = 1/7$, where cells labelled with the same letter use the same frequency channels. In this model, a cluster is outlined in bold and consists of seven cells with different frequency channels. Even though the hexagonal cell shape is conceptual, it has been widely used in the analysis of a cellular system owing to its simplicity and analytical convenience.

Figure 2.5 Standard figure of a hexagonal cellular system with $f = 1/7$.



Cellular systems allow the overall system capacity to increase by simply making the cells smaller and turning down the power. In this manner, cellular systems have a very desirable scaling property—more capacity can be supplied by installing more base stations. As the cell size decreases, the transmit power of each base station also decreases correspondingly. For example, if the radius of a cell is reduced by half when the propagation path loss exponent is 4, the transmit power level of a base station is reduced by 12 dB (=10 log 16 dB).

Since cellular systems support user mobility, seamless call transfer from one cell to another should be provided. The handoff process provides a means of the seamless transfer of a connection from one base station to another. Achieving smooth handoffs is a challenging aspect of cellular system design.

Although small cells give a large capacity advantage and reduce power consumption, their primary drawbacks are the need for more base stations (and their associated hardware costs), and the need for frequent handoffs. The offered traffic in each cell also becomes more variable as the cell shrinks, resulting in inefficiency. As in most aspects of wireless systems, an appropriate tradeoff between these competing factors needs to be determined depending on the system requirements.

2.3.2 Analysis of Cellular Systems

The performance of wireless cellular systems is significantly limited by co-channel interference (CCI), which comes from other users in the same cell or from other cells. In cellular systems, other cell interference (OCI) is a decreasing function of the radius of the cell (R) and the distance to the center of the neighboring co-channel cell and an increasing function of transmit power. However, what determines performance (capacity, reliability) is the SIR, i.e., the amount of desired power to the amount of transmitted power. Therefore, if all users (or base stations) increased or decreased their power at once, the SIR and hence the performance is typically unchanged—which is known as an interference-limited system. The spatial isolation between co-channel cells can be measured by defining the parameter Z, called *co-channel reuse ratio*, as the ratio of the distance to the center of the nearest co-channel cell (D) to the radius of the cell. In a hexagonal cell structure, the co-channel reuse ratio is given by

(2.24)

$$Z = \frac{D}{R} = \sqrt{3/f}$$

where $1/f$ is the size of a cluster and the inverse of the frequency reuse factor. Obviously, a lower value of f reduces co-channel interference so that it improves the quality of the communication link and capacity. However, the overall spectral efficiency decreases with the size of a cluster, so f should be chosen just small enough to keep the received signal-to-interference-plus-noise ratio (SINR) above acceptable levels.

Since the background noise power is negligible compared to the interference power in an interference-limited environment, the received signal-to-interference ratio (SIR) can be used instead of SINR. If the number of interfering cells is N_I, the SIR for a mobile station can be given by

(2.25)

$$\frac{S}{I} = \frac{S}{\sum_{i=1}^{N_I} I_i}$$

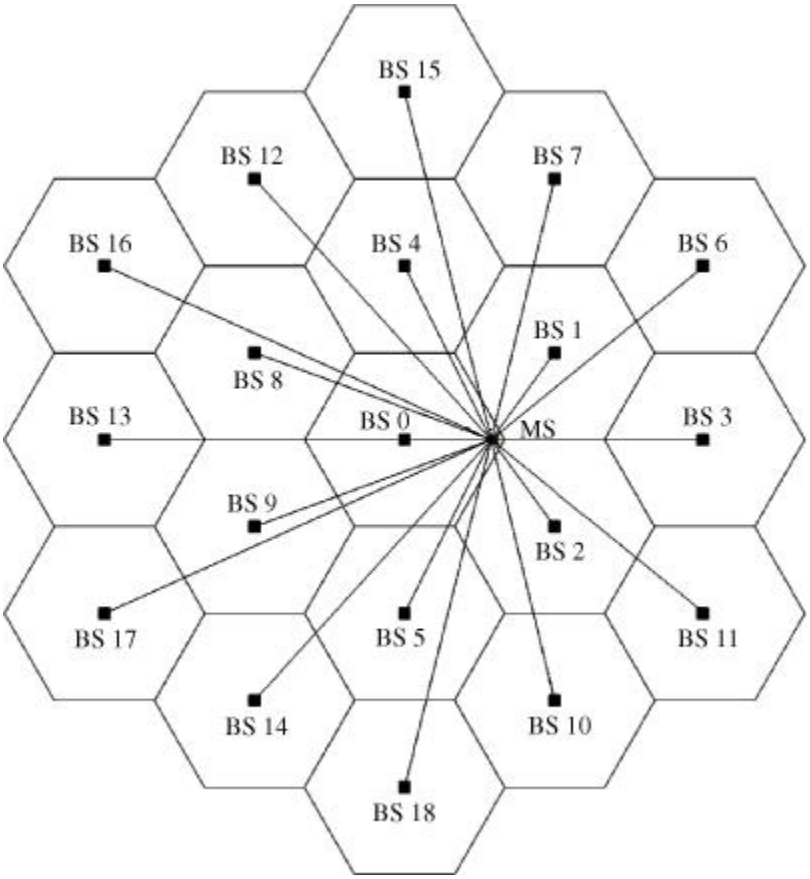
where S is the received power of the desired signal and I_i is the interference power from the ith co-channel base station. The received SIR depends on the location of each mobile station, and it should be kept above an appropriate threshold for

reliable communication. The received SIR at the cell boundaries is of great interest since this corresponds to the worst interference scenario. For example, if the empirical path loss formula given in (2.10) and universal frequency reuse are considered, the received SIR for the worst case given in Figure 2.6 is expressed as

(2.26)

$$\frac{S}{I} = \frac{\chi_0}{\chi_0 + \sum_{i=1}^{2} \chi_i + 2^{-\alpha} \sum_{i=3}^{5} \chi_i + (2.633)^{-\alpha} \sum_{i=6}^{11} \chi_i}$$

Figure 2.6 Forward link interference in a hexagonal cellular system (worst case).



where χ_i denotes the shadowing from the ith base station. Since the sum of lognormal random variables are well approximated by a lognormal random variable [15, 41], the denominator can be approximated as a lognormal random variable and then the received SIR follows a lognormal distribution [9]. Therefore, the outage probability that the received SIR falls below a threshold can be derived from the distribution. If the mean and standard deviation of the lognormal

distribution are μ and σ in dB, the outage probability is derived in the form of Q function

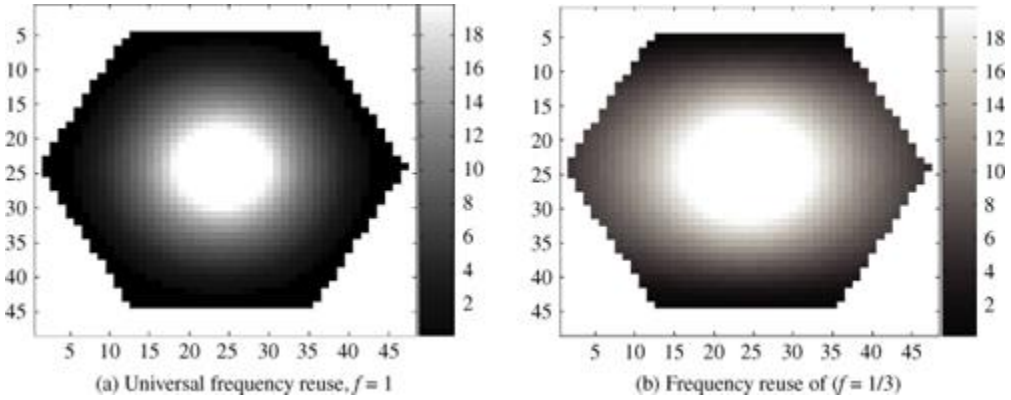
(2.27)

$$P_{\text{out}} = \mathbb{P}[\text{SIR} < \gamma] = Q\left(\frac{\gamma - \mu}{\sigma}\right)$$

where γ is the threshold SIR level in dB. Usually, the SINR at the cell boundaries is too low to achieve the outage probability design target if universal frequency reuse is adopted. Therefore, a lower frequency reuse factor is typically adopted in the system design to satisfy the target outage probability at the sacrifice of spectral efficiency.

Figure 2.7 highlights the OCI problem in a cellular system if universal frequency reuse is adopted. It shows the regions of a cell in various SIR bins of the systems with universal frequency reuse and f = 1/3 frequency reuse. The figure is based on a two-tier cellular structure and the simple empirical path loss model given in (2.7) with α = 3.5. The SIR in most parts of the cell is very low if universal frequency reuse is adopted. The OCI problem can be mitigated if higher frequency reuse is adopted as shown in Figure 2.7b. However, as previously emphasized, this improvement in the quality of communication is achieved at the sacrifice of spectral efficiency, in this case the available bandwidth is cut by a factor of 3. Frequency planning is a delicate balancing act of using the highest reuse factor possible, while still having most of the cell have at least some minimum SIR.

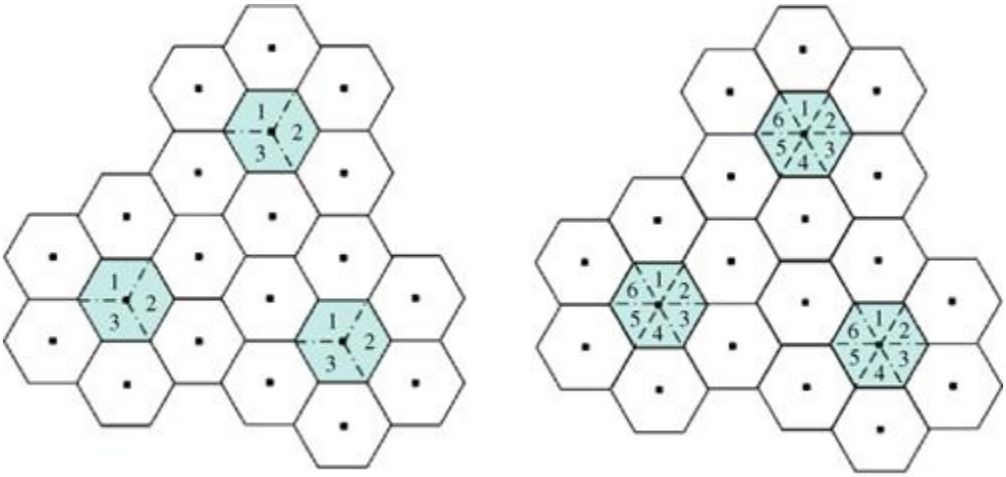
Figure 2.7 The received SIR in a cell with path-loss exponent α = 3.5. The scale on the right indicates the SIR bins, i.e., darker indicates lower SIR.



2.3.3 Sectoring

Since the SIR is so bad in most of the cell, it is desirable to find techniques to improve it without sacrificing so much bandwidth, as frequency reuse does. A popular technique is to sectorize the cells, which is effective if frequencies are reused in each cell. By using directional antennas instead of an omni-directional antenna at the base station, the co-channel interference can be significantly reduced. An illustration of sectoring is shown in Figure 2.8. Although the absolute amount of bandwidth used is 3X before (assuming 3-sector cells), the capacity increase is in fact more than 3X. No capacity is lost from sectoring because each sector can reuse time and code slots, so each sector has the same nominal capacity as an entire cell. Furthermore, the capacity in each sector is actually higher than that in a non-sectored cellular system because the interference is reduced by sectoring, since users only experience interference from the sectors at their frequency. Referring again to Figure 2.8a, if each sector 1 points the same direction in each cell, then the interference caused by neighboring cells will be dramatically reduced. An alternative way to use sectors that is not shown in Figure 2.8 is to reuse frequencies in each sector. In this case, all of the time/code/frequency slots can be reused in each sector, but there is no reduction in the experienced interference.

Figure 2.8 3-sector (120-degree) and 6-sector (60-degree) cells.

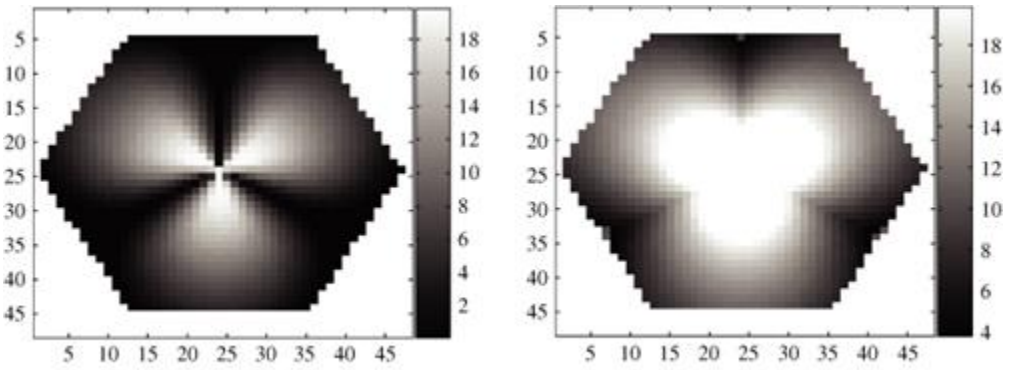


(a) 3-sector cells (b) 6-sector cells

Figure 2.9 shows the regions of a 3-sector cell in various SIR bins of the systems with universal frequency reuse and 1/3 frequency reuse. All the configurations are the same as those of Figure 2.7 except sectoring is added. Compared to Figure 2.7, sectoring improves SIR especially at the cell boundaries even when universal frequency reuse is adopted. If sectoring is adopted with frequency reuse, the

received SIR can be significantly improved as shown in Figure 2.9b where both f = 1/3 frequency reuse and 120-degree sectoring are used.

Figure 2.9 The received SIR in a sectorized cell (three sectors). The path-loss exponent = α = 3.5.



Although sectoring is an effective and practical approach to the OCI problem, it is not without cost. Sectoring increases the number of antennas at each base station and reduces trunking efficiency due to channel sectoring at the base station. Even though intersector handoff is simpler compared to intercell handoff, sectoring also increases the overhead due to the increased number of intersector handoffs. Finally, in channels with heavy scattering, desired power can be lost into other sectors, which can cause intersector interference as well as power loss.

New Approaches to Other Cell Interference

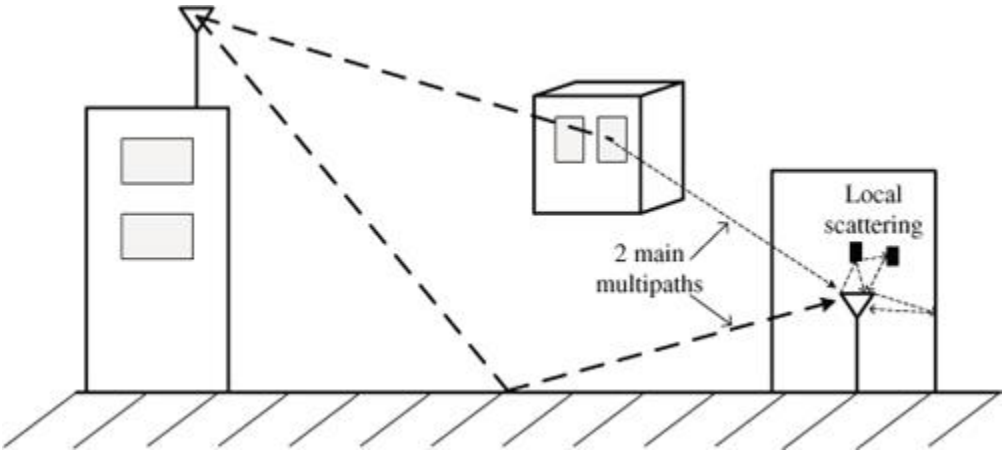
While the problem of co-channel interference has existed in cellular systems for many years, its effect on LTE is more important since LTE's requirements for high data rate and spectral efficiency—and hence high SINR—go beyond 3G systems, and LTE does not have built-in features to reduce the effects of this interference, unlike, for example, CDMA. The deployment of multiple antennas does not help, and in fact is likely to make the system even more sensitive to other cell interference [2, 5, 11]. A straightforward approach to this problem is advanced signal processing techniques at the receiver [1,11] and/or transmitter [25,43,57] as a means of reducing or cancelling the perceived interference. Although those techniques have important merits and are being actively researched and considered, they have some important shortcomings when viewed in a practical context of near-future cellular systems such as LTE. As an alternative, network-level approaches such as cooperative scheduling or encoding across base stations [8, 28, 31, 35, 48, 59], multicell power control [6, 18, 23, 38, 53, 56], and distributed antennas [7, 22, 24, 37, 58] can be considered. These network-level

approaches require relatively little channel knowledge (with the exception of cooperative encoding) and effectively reduce other cell interference through macro-diversity and efficient sharing of the spectrum. The simpler of these techniques are being actively considered for LTE and beyond.

2.4 THE BROADBAND WIRELESS CHANNEL: FADING

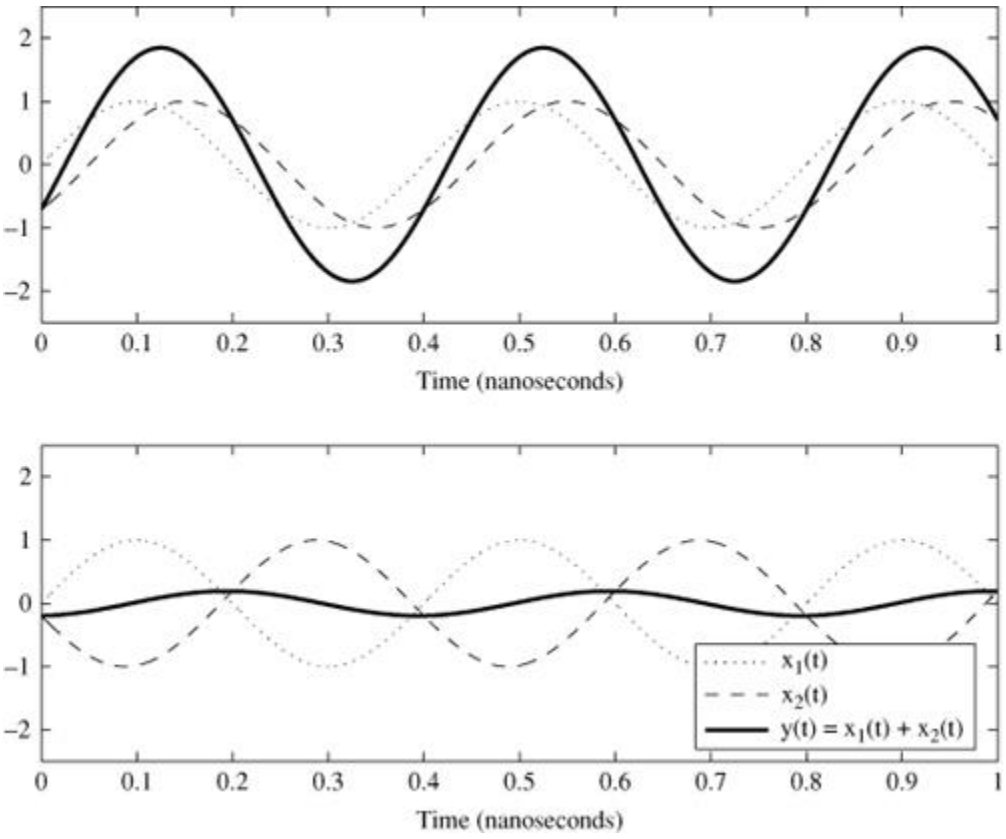
One of the more intriguing aspects of wireless channels is the fading phenomenon. Unlike path loss or shadowing, which are large-scale attenuation effects due to distance or obstacles, fading is caused by the reception of multiple versions of the same signal. The multiple received versions are caused by reflections that are referred to as *multipath*, as introduced briefly in Chapter 1. The reflections may arrive at very close to the same time—for example, if there is local scattering around the receiver—or the reflections may arrive at relatively longer intervals—for example, due to multiple different paths between the transmitter and receiver. A visualization of this is shown in Figure 2.10.

Figure 2.10 The channel may have a few major paths with quite different lengths, and then the receiver may see a number of locally scattered versions of those paths.



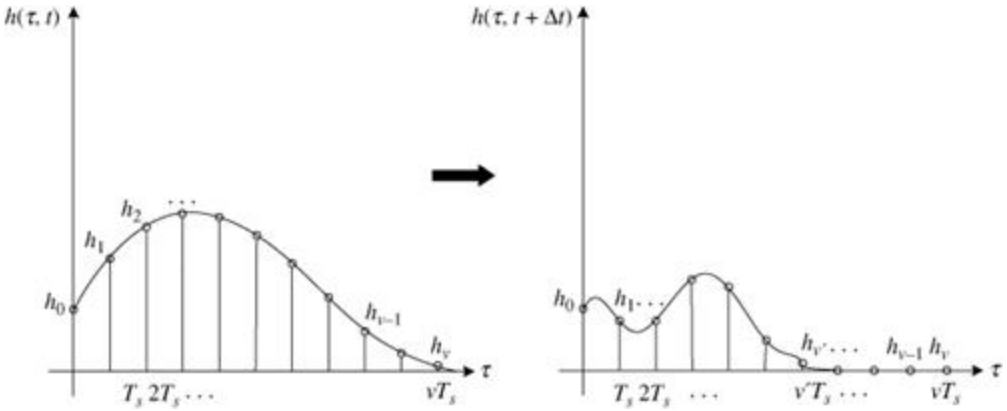
When some of the reflections arrive at nearly the same time, the combined effect of those reflections can be seen in Figure 2.11. Depending on the phase difference between the arriving signals, the interference can be either constructive or destructive, which causes a very large observed difference in the amplitude of the received signal even over very short distances. In other words, moving the transmitter or receiver even a very short distance can have a dramatic effect on the received amplitude, even though the path loss and shadowing effects may not have changed at all.

Figure 2.11 The difference between constructive interference (top) and destructive interference (bottom) at f_c = 2.5GHz is less than 0.1 nanoseconds in phase, which corresponds to about 3 cm.



To formalize this discussion, we will now return to the time-varying tapped-delay line channel model of (2.1). As either the transmitter or receiver move relative to each other, the channel response $\mathbf{h}(t)$ will change.[7] This channel response can be thought of as having two dimensions: a delay dimension τ and a time-dimension t, as shown in Figure 2.12. Since the channel changes over distance (and hence time), the values of $h_0, h_1, \ldots, h_v$ may be totally different at time t vs. time $t + \Delta t$. Because the channel is highly variant in both the τ and t dimensions, in order to be able to discuss what the channel response is, we must use statistical methods.

Figure 2.12 The delay τ corresponds to how *long* the channel impulse response lasts. The channel is time varying, so the channel impulse response is also a function of time, i.e., $h(\tau, t)$, and can be quite different at time $t + \Delta t$ than it was at time t.



The most important and fundamental function used to statistically describe broadband fading channels is the two-dimensional autocorrelation function, $A(\Delta\tau, \Delta t)$. Although this autocorrelation function is over two dimensions (and hence requires a three-dimensional plot), it can usefully be thought of as two simpler functions, $A_t(\Delta t)$ and $A_\tau(\Delta\tau)$, where $\Delta\tau$ and Δt have been set to zero, respectively. The autocorrelation function is defined as

(2.28)

$$\begin{aligned} A(\Delta\tau, \Delta t) &= E\left[h(\tau_1, t_1)h^*(\tau_2, t_2)\right] \\ &= E\left[h(\tau_1, t)h^*(\tau_2, t + \Delta t)\right] \\ &= E\left[h(\tau, t)h^*(\tau + \Delta\tau, t + \Delta t)\right] \end{aligned}$$

where in the first step we have assumed that the channel response is Wide Sense Stationary (WSS) (and hence the autocorrelation function only depends on $\Delta t = t_2 - t_1$), and in the second step we have assumed that the channel response of paths arriving at different times τ_1 and τ_2 are uncorrelated. This allows the dependence on specific times τ_1 and τ_2 to be replaced simply by $\tau = \tau_1 - \tau_2$. Channels that can be described by the autocorrelation in (2.28) are thus referred to as Wide Sense Stationary Uncorrelated Scattering (WSSUS), which is the most popular model for wideband fading channels, and relatively accurate in many practical scenarios, largely because the scale of interest for τ (usually μsecs) and t (usually msec) generally differ by a few orders of magnitude.

The next three sections will explain how many of the key wireless channel parameters can be estimated from the autocorrelation function $A(\Delta\tau, \Delta t)$, and how they are related (see Table 2.2).

Table 2.2 Summary of Broadband Fading Parameters, with Rules of Thumb

Quantity	If "Large"?	If "Small"?	LTE Design Impact
Delay Spread, τ	If $\tau \gg T$, then frequency selective	If $\tau \ll T$, then frequency flat	The larger the delay spread relative to the symbol time, the more severe the ISI.
Coherence Bandwidth, B_c	If $\frac{1}{B_c} \ll T$, then frequency flat	If $\frac{1}{B_c} \gg T$, then frequency selective	Provides a guideline to subcarrier width $B_{sc} \approx B_c/10$, and hence number of subcarriers needed in OFDM: $L \geq 10B/B_c$.
Doppler Spread, $f_D = \frac{f_c \nu}{c}$	If $f_c \nu \gg c$, then fast fading	If $f_c \nu \leq c$, then slow fading	As f_D/B_{sc} becomes non-negligible, subcarrier orthogonality is compromised.
Coherence Time, T_c	If $T_c \gg T$, then slow fading	If $T_c \leq T$, then fast fading	T_c small necessitates frequent channel estimation and limits the OFDM symbol duration, but provides greater time diversity.
Angular Spread, θ_{rms}	Non-LOS channel, lots of diversity	Effectively LOS channel, not much diversity	Multiantenna array design, beamforming vs. diversity.
Coherence Distance, D_c	Effectively LOS channel, not much diversity	Non-LOS channel, lots of diversity	Determines antenna spacing.

2.4.1 Delay Spread and Coherence Bandwidth

The delay spread is a very important property of a wireless channel, since it specifies the duration of the channel impulse response $h(\tau, t)$. Intuitively, the delay spread is the amount of time that elapses between the first arriving path and the last arriving (non-negligible) path. The delay spread can be found by inspecting $A(\Delta\tau, 0) \triangleq A_\tau(\Delta\tau)$; that is, by setting $\Delta t = 0$ in the channel autocorrelation function. $A_\tau(\Delta\tau)$ is often referred to as the *Multipath Intensity Profile*, or *power delay profile*. If $A_\tau(\Delta\tau)$ has non-negligible values from $(0,\tau_{\max})$, the maximum delay spread is $\tau_{\max}$. Intuitively, this is an important definition because it specifies how many taps v will be needed in the discrete representation of the channel impulse response, since

(2.29)

$$v \approx \frac{\tau_{\max}}{T_s},$$

where T_s is the sampling time. But, this definition is not rigorous since it is not clear what "non-negligible" means mathematically. In order to speak more quantitatively, the average and rms delay spread are often used instead of τ_{max}, and are defined as follows:

(2.30)

$$\mu_\tau = \frac{\int_0^\infty \Delta\tau A_\tau(\Delta\tau) d(\Delta\tau)}{\int_0^\infty A_\tau(\Delta\tau) d(\Delta\tau)}$$

(2.31)

$$\tau_{\text{rms}} = \sqrt{\frac{\int_0^\infty (\Delta\tau - \mu_\tau)^2 A_\tau(\Delta\tau) d(\Delta\tau)}{\int_0^\infty A_\tau(\Delta\tau) d(\Delta\tau)}}$$

Intuitively, τ_{rms} gives a measure of the "width" or "spread" of the channel response in time. A large τ_{rms} implies a highly dispersive channel in time and a long channel impulse response (i.e., large v), whereas a small τ_{rms} indicates that the channel is not very dispersive, and hence might require just a few taps to accurately characterize. A general rule of thumb is that $\tau_{max} \approx 5\tau_{rms}$.

The channel coherence bandwidth B_c is the frequency domain dual of the channel delay spread. The coherence bandwidth gives a rough measure for the maximum separation between a frequency f_1 and a frequency f_2 where the channel frequency response is correlated. That is:

$|f_1 - f_2| \leq B_c \Rightarrow H(f_1) \approx H(f_2)$
$|f_1 - f_2| > B_c \Rightarrow H(f_1)$ and $H(f_2)$ are uncorrelated

Just as τ_{max} is a "ballpark" value describing the channel duration, B_c is a ballpark value describing the range of frequencies over which the channel stays constant. Given the channel delay spread, it can be shown that

(2.32)

$$B_c \approx \frac{1}{5\tau_{\text{rms}}} \approx \frac{1}{\tau_{\max}}.$$

"Exact" relations can be found between B_c and τ_{rms} by arbitrarily defining notions of coherence, but the important and prevailing feature is that B_c and τ are inversely related.

2.4.2 Doppler Spread and Coherence Time

Whereas the power delay profile gave the statistical power distribution of the channel over time for a signal transmitted for just an instant, the *Doppler power spectrum* gives the statistical power distribution of the channel versus frequency for a signal transmitted at just one exact frequency (generally normalized as $f = 0$ for convenience). Whereas the power delay profile was caused by multipath between the transmitter and receiver, the Doppler power spectrum is caused by *motion* between the transmitter and receiver. The Doppler power spectrum is the Fourier transform of $A_t(\Delta t)$, that is:

(2.33)

$$\rho_t(\Delta f) = \int_{-\infty}^{\infty} A_t(\Delta t) e^{-\Delta f \cdot \Delta t} (d\Delta t)$$

Unlike the power delay profile, the Doppler power spectrum is non-zero strictly for $\Delta f \in (-f_D, f_D)$, where f_D is called the maximum Doppler, or Doppler spread. That is, $\rho_t(\Delta f)$ is strictly "band-limited." The Doppler spread is

(2.34)

$$f_D = \frac{v f_c}{c},$$

where v is the maximum speed between the transmitter and receiver, f_c is the carrier frequency, and c is the speed of light. As can be seen, over a large bandwidth the Doppler will change since the frequency over the entire bandwidth is *not* f_c. However, as long as the communication bandwidth $B << f_c$, the Doppler power spectrum can be treated as approximately constant. This generally is true for all but ultrawideband systems.

Due to the time-frequency uncertainty principle,[8] since $\rho_t(\Delta f)$ is strictly band-limited, its time-frequency dual $A_t(\Delta t)$ cannot be strictly time-limited. Since $A_t(\Delta t)$ gives the correlation of the channel over time, what this means is that strictly speaking, the channel exhibits non-zero correlation between any two time instants. However, in practice it is possible to define a channel coherence time T_c, which,

similarly to coherence bandwidth, gives the period of time over which the channel is significantly correlated. Mathematically:

$|t_1 - t_2| \leq T_c \Rightarrow \mathbf{h}(t_1) \approx \mathbf{h}(t_2)$
$|t_1 - t_2| > t_c \Rightarrow \mathbf{h}(t_1)$ and $\mathbf{h}(t_2)$ are uncorrelated

The coherence time and Doppler spread are also inversely related,

(2.35)

$$T_c \approx \frac{1}{f_D}.$$

This makes intuitive sense: if the transmitter and receiver are moving fast relative to each other and hence the Doppler is large, the channel will change much more quickly than if the transmitter and receiver are stationary.

Values for the Doppler spread and the associated channel coherence time are given in Table 2.3 for two plausible LTE frequency bands. This table demonstrates one of the reasons that mobility places severe constraints on the system design. At high frequency and mobility, the channel may change up to 1000 times per second, placing a large burden on overhead channels, channel estimation algorithms, and making the assumption of accurate transmitter channel knowledge questionable. Additionally, the large Doppler at high mobility and frequency can also degrade the OFDM subcarrier orthogonality, as discussed in Chapter 3.

Table 2.3 Doppler Spreads and Approximate Coherence Times for LTE at Pedestrian, Vehicular, and Maximum Speeds

f_c	Speed (km/hr)	Speed (mph)	Max. Doppler, f_D (Hz)	Coherence Time, $T_c \approx \frac{1}{f_D}$ (msec)
700MHz	2	1.2	1.3	775
700MHz	45	27	29.1	34
700MHz	350	210	226.5	4.4
2.5GHz	2	1.2	4.6	200
2.5GHz	45	27	104.2	10
2.5GHz	350	210	810	1.2

2.4.3 Angular Spread and Coherence Distance

In this chapter we have focused on how the channel response varies over time, and how to quantify its delay and correlation properties. However, channels also vary

over space. We will not attempt to rigorously treat all the aspects of spatial-temporal channels, but we will summarize a few important points.

The rms angular spread of a channel can be denoted as θ_{rms}, and refers to the statistical distribution of the angle of the arriving energy. A large θ_{rms} implies that channel energy is coming in from many directions, whereas a small θ_{rms} implies that the received channel energy is more focused. A large angular spread generally occurs when there is a lot of local scattering, and this results in more statistical diversity in the channel, whereas more focused energy results in less statistical diversity.

The dual of angular spread is coherence distance, D_c. As the angular spread increases, the coherence distance decreases, and vice versa. A coherence distance of d means that any physical positions separated by d have an essentially uncorrelated received signal amplitude and phase. An approximate rule of thumb is [13]

(2.36)

$$D_c \approx \frac{.2\lambda}{\theta_{\text{rms}}}.$$

For the case of Rayleigh fading, which assumes a uniform angular spread, the well-known relation is

(2.37)

$$D_c \approx \frac{9\lambda}{16\pi}.$$

An important trend to note from the above relations is that the coherence distance increases with the carrier wavelength λ, so higher-frequency systems have shorter coherence distances.

Angular spread and coherence distance are particularly important in multiple antenna systems. The coherence distance gives a rule of thumb for how far antennas should be spaced apart, in order to be statistically independent. If the coherence distance is very small, antenna arrays can be effectively employed to provide rich diversity. The importance of diversity will be introduced later in Section 2.6. On the other hand, if the coherence distance is large, it may not be possible due to space constraints to take advantage of spatial diversity. In this case,

it would be preferable to have the antenna array cooperate and use beamforming. The tradeoffs between beamforming and linear array processing will be discussed in Chapter 4.

2.5 MODELLING BROADBAND FADING CHANNELS

In order to design and benchmark wireless communication systems, it is important to develop channel models that incorporate their variations in time, frequency, and space. The two major classes of models are *statistical* and *empirical*. Statistical models are simpler, and are useful for analysis and simulations. The empirical models are more complicated but usually represent a specific type of channel more accurately.

A PEDAGOGY FOR DEVELOPING STATISTICAL MODELS

The methods we will describe for modelling wireless channels are broken into three steps:

1. Section 2.5.1: First consider just a single channel sample corresponding to a single principle path between the transmitter and receiver, that is:

$h(\tau, t) \rightarrow h_0\delta(\tau, t)$

Attempt to quantify: How is the value of $|h_0|$ statistically distributed?

2. Section 2.5.2: Next consider how this channel sample h_0 evolves over time, that is:

$h(\tau, t) \rightarrow h_0(t)\delta(\tau)$

Attempt to quantify: How does the value $|h_0(t)|$ change over time? That is, how is $h_0(t)$ correlated with some $h_0(t + \Delta t)$?

3. Sections 2.5.2 and 2.5.3: Finally, $h(\tau, t)$ is represented as a general time varying function. One simple approach would be to model $h(\tau, t)$ as a general multipath channel with $v + 1$ tap values. The channel sample value for each of these taps is distributed as determined in step 1 above, and evolves over time as specified by step 2.

2.5.1 Statistical Channel Models

As we have noted, the received signal in a wireless system is the superposition of numerous reflections, or multipath components. The reflections may arrive very closely spaced in time—for example, if there is local scattering around the receiver—or the reflections may arrive at relatively longer intervals. As we saw in Figure 2.11, when the reflections arrive at nearly the same time, constructive and destructive interference between the different reflection causes the envelope of the aggregate received signal $r(t)$ to vary substantially.

In this section, we will overview statistical methods that can be used to characterize the amplitude and power of $r(t)$ when all the reflections arrive at about the same time. In terms of the previous section, we will first consider the special case of the Multipath Intensity Profile where $A_r(\Delta\tau) \approx 0$ for $\Delta\tau \neq 0$. That is, we only concern ourselves with the scenario where all the received energy arrives at the receiver at the same instant—this is step 1. In practice, this is only true when the symbol time is much greater than the delay spread, i.e., $T \gg \tau_{max}$, so these models are often said to be valid for "narrowband fading channels." In addition to assuming a negligible multipath delay spread, we will first consider just a "snapshot" value of $r(t)$, and we will provide statistical models for its amplitude and power under various assumptions. In the following section, we will consider how these statistical values are correlated in time, frequency, and space—this is step 2. Then finally, we will relax all the assumptions and consider how wideband fading channels evolve in time, frequency, and space—this is step 3.

Rayleigh Fading

If the number of scatterers is large and the angles of arrival between them are uncorrelated, from the Central Limit Theorem it can be shown that the in-phase (cosine) and quadrature (sine) components of $r(t)$, denoted as $r_I(t)$ and $r_Q(t)$, follow two independent time-correlated Gaussian random processes.

Consider a snapshot value of $r(t)$ at time $t = 0$, and note that $r(0) = r_I(0) + r_Q(0)$. Since the values $r_I(0)$ and $r_Q(0)$ are Gaussian random variables, it can be shown that the distribution of the envelope amplitude $|r| = \sqrt{r_I^2 + r_Q^2}$ is Rayleigh, and the received power $|r|^2 = r_I^2 + r_Q^2$ is exponentially distributed. Formally,

(2.38)

$$f_{|r|}(x) = \frac{2x}{P_r} e^{-x^2/P_r}, \quad x \geq 0,$$

and

(2.39)

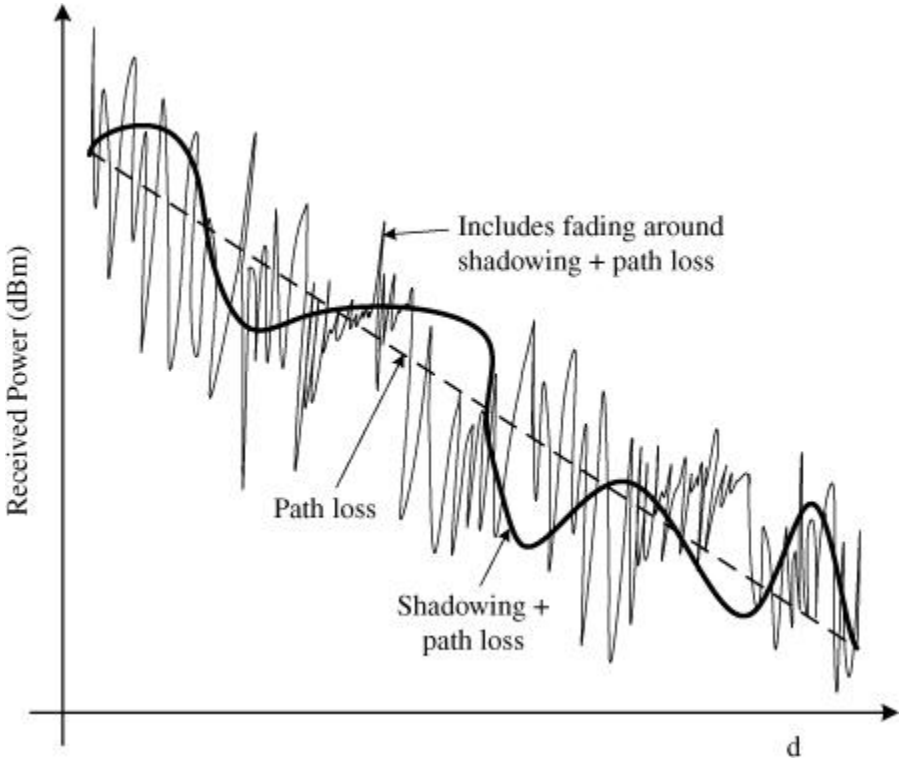
$$f_{|r|^2}(x) = \frac{1}{P_r} e^{-x/P_r}, \quad x \geq 0,$$

where P_r is the average received power due to shadowing and path loss, as described for example in Equation (2.10). The path loss and shadowing determine the mean received power (assuming they are fixed over some period of time), and the total received power fluctuates around this mean due to the fading. This is demonstrated in Figure 2.13. It can also be noted that in this setup, the Gaussian random variables r_I and r_Q each have zero mean and variance $\sigma^2 = P_r/2$. The phase of $r(t)$ is defined as

(2.40)

$$\theta_r = \tan^{-1}\left(\frac{r_Q}{r_I}\right),$$

Figure 2.13 The three major channel attenuation factors are shown in terms of their relative spatial (and hence temporal) scales.



which is uniformly distributed from 0 to 2π, or equivalently from $[-\pi, \pi]$ any other contiguous full period of the carrier signal.[9]

Line-of-Sight Channels—The Ricean Distribution

An important assumption in the Rayleigh fading model is that all the arriving reflections have a mean of zero. This will not be the case if there is a dominant path, for example, a line-of-sight (LOS) path, between the transmitter and receiver. For a LOS signal, the received envelope distribution is more accurately modelled by a Ricean [36] distribution, which is given by

(2.41)

$$f_{|r|}(x) = \frac{x}{\sigma^2} e^{-(x^2+\mu^2)/2\sigma^2} I_0\left(\frac{x\mu}{\sigma^2}\right), \quad x \geq 0,$$

where μ^2 is the power of the LOS component and I_0 is the 0th order, modified Bessel function of the first kind. Although this expression is more complicated than a Rayleigh distribution, it is really a generalization of the Rayleigh distribution. This can be confirmed by observing that $\mu = 0 \Rightarrow I_0(\frac{x\mu}{\sigma^2}) = 1$, so the Ricean distribution reduces to the Rayleigh distribution in the absence of a LOS component. Except in this special case, the Ricean phase distribution θ_r is not uniform in [0, 2π] and is not described by a straightforward expression.

Since the Ricean distribution depends on the LOS component's power μ^2, a common way to characterize the channel is by the relative strengths of the LOS and scattered paths. This factor K is quantified as

(2.42)

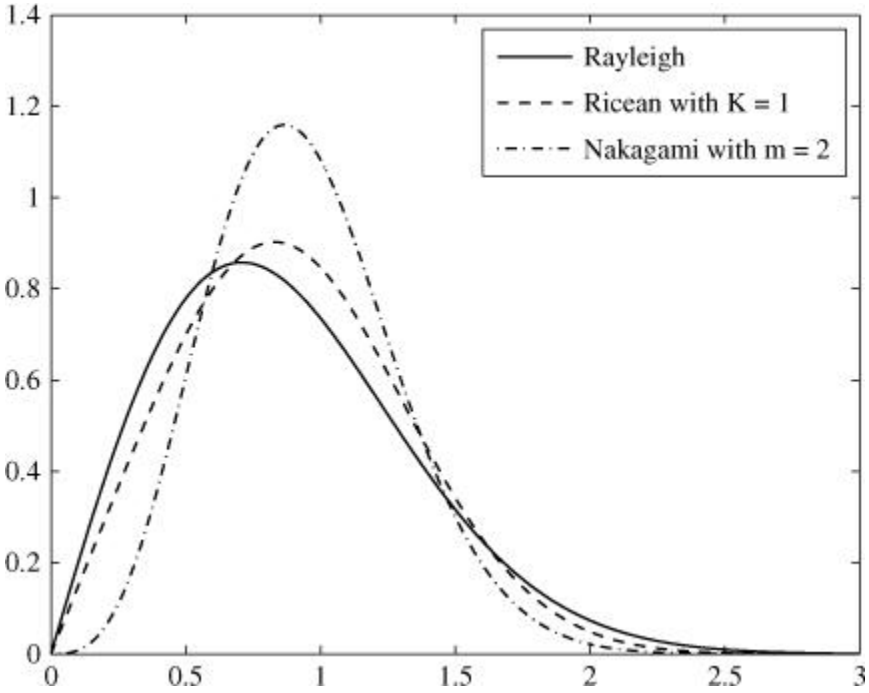
$$K = \frac{\mu^2}{2\sigma_2}$$

and is a natural description of how strong the LOS component is relative to the non-LOS (NLOS) components. For $K = 0$, again the Ricean distribution reduces to Rayleigh, and as $K \rightarrow \infty$, the physical meaning is that there is only a single LOS path and no other scattering. Mathematically, as K grows large, the Ricean distribution is quite Gaussian about its mean μ with decreasing variance, physically meaning that the received power becomes increasingly deterministic.

The average received power under Ricean fading is just the combination of the scattering power and the LOS power, i.e., $P_r = 2\sigma^2 + \mu^2$. Although it is not straightforward to directly find the Ricean power distribution $f_{|r|^2}(x)$, the Ricean envelope distribution in terms of K can be found by subbing $\mu^2 = KP_r/(K + 1)$ and $2\sigma^2 = P/(K + 1)$ into (2.41).

Although its simplicity makes the Rayleigh distribution more amenable to analysis than the Ricean distribution, the Ricean distribution is usually a more accurate depiction of wireless broadband systems, which typically have one or more dominant components. This is especially true of fixed wireless systems, which do not experience fast fading and often are deployed to maximize LOS propagation (see Figure 2.14 for a comparison of the most popular fading distributions).

Figure 2.14 Probability distributions $f_{|r|}(x)$ for Rayleigh, Ricean w/K = 1, and Nakagami with m = 2. All have average received power P_r = 1.



A More General Model: Nakagami-m Fading

The last statistical fading model that we will discuss is the Nakagami-m fading distribution [30]. The probability density function (PDF) of Nakagami fading is parameterized by m and given as

(2.43)

$$f_{|r|}(x) = \frac{2m^m x^{2m-1}}{\Gamma(m) P_r^m} e^{-mx^2/P_r}, \quad m \geq 0.5.$$

Although this expression appears to be just as ungainly as (or even more so than) the Ricean distribution, the dependence on x is simpler and hence the Nakagami distribution can in many cases be used in tractable analysis of fading channel

performance [45]. Additionally, it is more general as $m = (K + 1)^2/(2K + 1)$ gives an approximate Ricean distribution, and $m = 1$ gives a Rayleigh. As $m \rightarrow \infty$, the receive power tends to be a constant, P_r. The power distribution for Nakagami fading is

(2.44)

$$f_{|r|^2}(x) = \left(\frac{m}{P_r}\right)^m \frac{x^{m-1}}{\Gamma(m)} e^{-mx/P_r}, \quad m \geq 0.5.$$

Similarly, the power distribution is also amenable to integration.

2.5.2 Statistical Correlation of the Received Signal

The statistical methods in the last section discussed how *samples* of the received signal were statistically distributed. We considered three specific statistical models—Rayleigh, Ricean, and Nakagami-m—and provided the probability density functions (PDFs) that gave the likelihoods of the received signal envelope and power at a given time instant. What is of more interest, though, is how to link those statistical models with the channel autocorrelation function, $A_c(\Delta\tau, \Delta t)$, in order to understand how the envelope signal $r(t)$ evolves over time, or changes from one frequency or location to another.

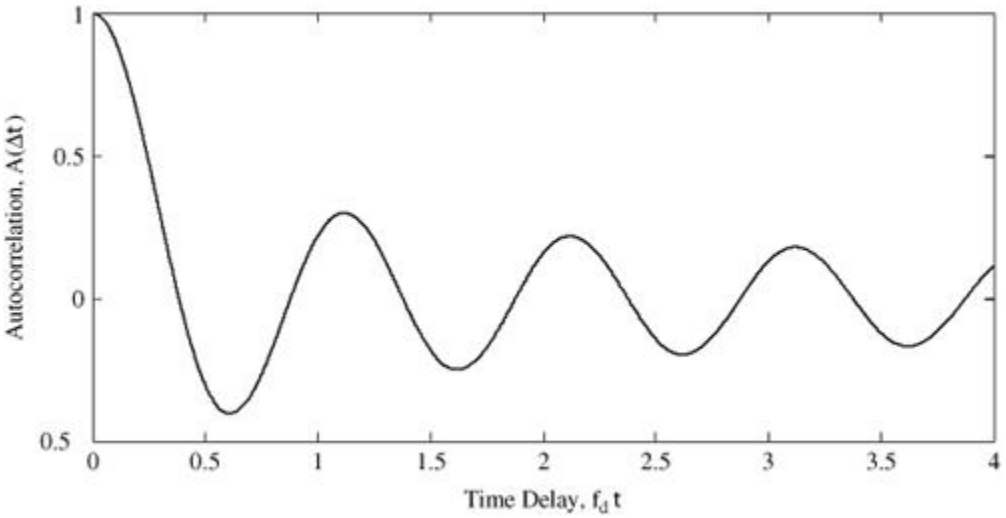
For simplicity and consistency, we will use Rayleigh fading as an example distribution during this section, but the concepts apply equally for any PDF. We will first discuss correlation in different domains separately, but will conclude with a brief discussion of how the correlations in different domains interact.

Time Correlation

In the time domain, the channel $h(\tau = 0,t)$ can intuitively be thought of as consisting of approximately one new sample from a Rayleigh distribution every T_c seconds, with the values in between interpolated. But, it will be useful to be more rigorous and accurate in our description of the fading envelope. As discussed in Section 2.4, the autocorrelation function $A_t(\Delta t)$ describes how the channel is correlated in time (see Figure 2.15). Similarly, its frequency domain Doppler power spectrum $\rho_t(\Delta f)$ provides a band-limited description of the same correlation since it is simply the Fourier transform of $A_t(\Delta t)$. In other words, the power spectral density of the channel $h(\tau = 0, t)$ should be $\rho_t(\Delta f)$. Since uncorrelated random variables have a flat power spectrum, this means that a sequence of independent complex Gaussian random numbers can be multiplied by

the desired Doppler power spectrum $\rho_t(\Delta f)$, and then by taking the Inverse Fast Fourier Transform (IFFT), a correlated narrowband sample signal $h(\tau = 0, t)$ can be generated. The signal will have a time correlation that is defined by $\rho_t(\Delta f)$, and be Rayleigh due to the Gaussian random samples in frequency.

Figure 2.15 Autocorrelation of the signal envelope in time, $A_c(\Delta t)$, which here is normalized by the Doppler f_D. For example, from this figure it can be seen that for $\Delta t \approx 0.4/f_D$, which means that after $0.4/f_D$ seconds, the fading value is uncorrelated with the value at time 0.



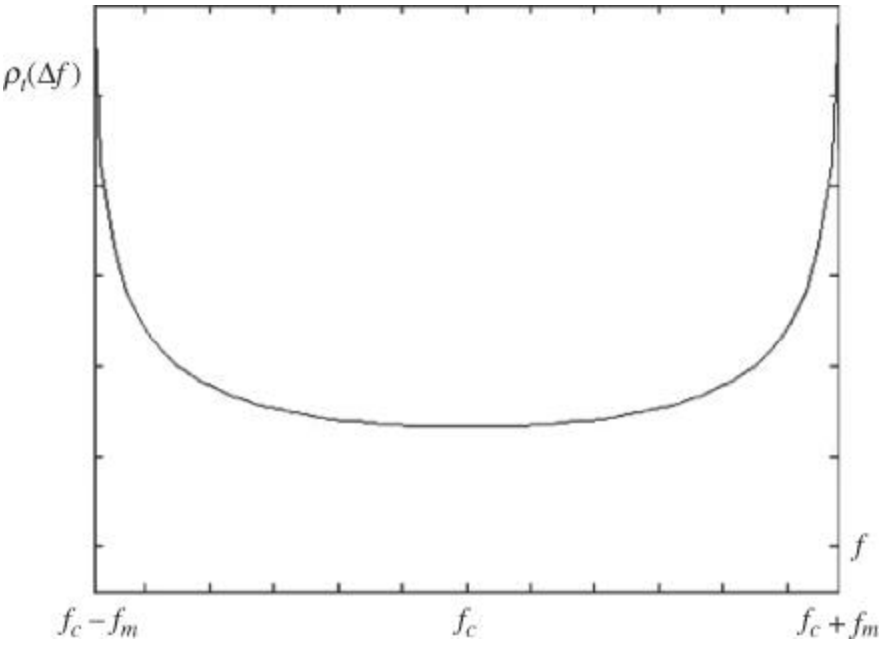
For the specific case of uniform scattering [26], it can been shown that the Doppler power spectrum becomes

(2.45)

$$\rho_t(\Delta f) = \begin{cases} \frac{P_r}{4\pi} \frac{1}{f_D \sqrt{1 - \left(\frac{\Delta f}{f_D}\right)^2}}, & |\Delta f| \leq f_D \\ 0, & \Delta_f > f_D \end{cases}$$

A plot of this realization of $\rho_t(\Delta f)$ is shown in Figure 2.16. It is well known that the Inverse Fast Fourier Transform of this function is the 0th order Bessel function of the 1st kind, which is often used to model the time autocorrelation function $A_c(\delta t)$, and hence predict the time correlation properties of narrowband fading signals.

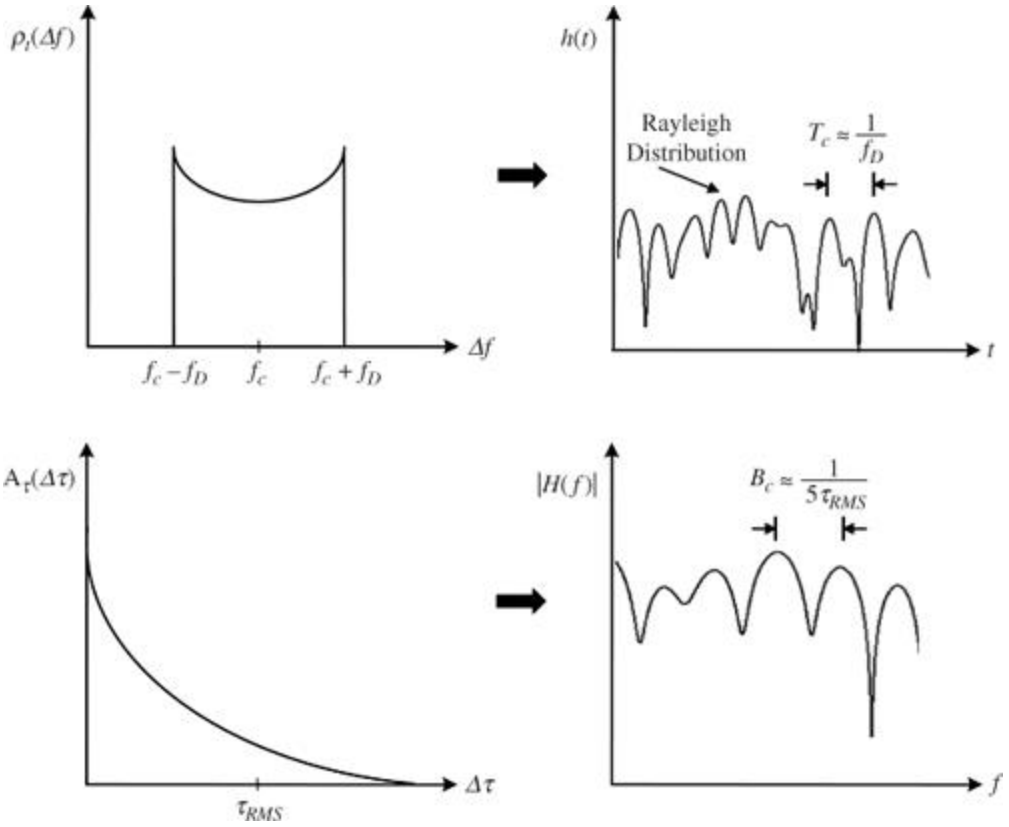
Figure 2.16 The spectral correlation due to Doppler, $\rho_t(\Delta f)$ for uniform scattering, i.e., Equation (2.45).



Frequency Correlation

Similar to time correlation, a simple intuitive notion of fading in frequency is that the channel in the frequency domain, $H(f, t = 0)$, can be thought of as consisting of approximately one new random sample every B_c Hz, with the values in between interpolated. The Rayleigh fading model assumes that the received quadrature signals in time are complex Gaussian. Similar to the development in the last section where complex Gaussian values in the frequency domain can be converted to a correlated Rayleigh envelope in the time domain, complex Gaussian values in the time domain can likewise be converted to a correlated Rayleigh frequency envelope $|H(f)|$. (See Figure 2.17).

Figure 2.17 The shape of the Doppler power spectrum $\rho_t(\Delta f)$ determines the correlation envelope of the channel in time (top). Similarly, the shape of the Multipath Intensity Profile $A_\tau(\Delta\tau)$ determines the correlation pattern of the channel frequency response (bottom).



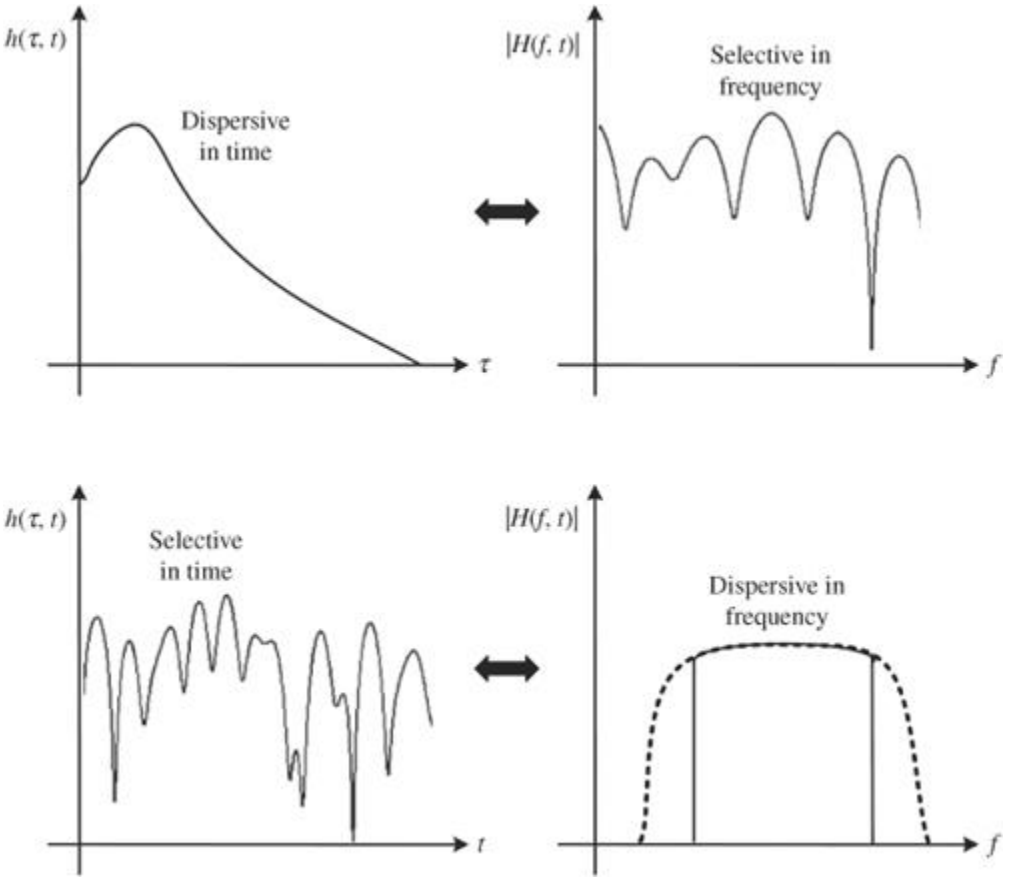
The correlation function that maps from uncorrelated time domain (τ domain) random variables to a correlated frequency response is the Multipath Intensity Profile, $A_\tau(\Delta\tau)$. This makes sense: just as $\rho_t(\Delta f)$ describes the channel time correlation in the frequency domain, $A_\tau(\Delta\tau)$ describes the channel frequency correlation in the time domain. Note that one familiar special case is the case where there is only one arriving path, in which case $A_\tau(\Delta\tau) = \delta(\Delta\tau)$. Hence, the values of $|H(f)|$ are correlated over all frequencies since the Fourier transform of $\delta(\Delta\tau)$ is a constant over all frequency. We refer to this scenario as "flat fading," and in practice whenever $A_\tau(\Delta\tau)$ is narrow (i.e., $\tau_{\max} \ll T$), the fading is approximately flat.

If the arriving quadrature components are approximately complex Gaussian, then a correlated Rayleigh distribution might be a reasonable model for the gain $|H(f)|$ on each subcarrier of a typical OFDM system. These gain values could also be generated by a suitably modified version of the provided simulation, where in particular the correlation function used changes from that in (2.45) to something like an exponential or uniform distribution, or any function that reasonably reflects the Multipath Intensity Profile $A_\tau(\Delta\tau)$,

The Dispersion-Selectivity Duality

As we have seen, there are two quite different effects from fading that we can refer to as *selectivity* and *dispersion*. By *selectivity*, we mean that the signal's received value is changed by the channel over time or frequency. By *dispersion*, we mean that the channel is dispersed, or spread out, over time or frequency. Selectivity and dispersion are time-frequency duals of each other: selectivity in time causes dispersion in frequency, and selectivity in frequency causes dispersion in time—or vice versa. This is illustrated in Figure 2.18.

Figure 2.18 The dispersion-selectivity duality: Dispersion in time causes frequency selectivity, while dispersion in frequency causes time selectivity.



For example, we have seen that the Doppler effect causes dispersion in frequency, as described by the Doppler power spectrum $\rho_t(\Delta f)$. This means that frequency components of the signal received at a specific frequency f_0 will be dispersed about f_0 in the frequency domain with a probability distribution function described by $\rho_t(\Delta f)$. As we have seen, this dispersion can be interpreted as a time varying amplitude, or selectivity, in time.

Similarly, a dispersive multipath channel that causes the paths to be received over a period of time τ_{max} causes selectivity in the frequency domain, known as frequency-selective fading. Because symbols are traditionally sent one after another in the time domain, time dispersion usually causes much more damaging interference than frequency dispersion, since adjacent symbols are smeared together.

Multidimensional Correlation

In order to present the concepts as clearly as possible, we have thus far treated time, frequency, and spatial correlations separately. In reality, signals are correlated in all three domains.

A broadband wireless data system with mobility and multiple antennas is an example of a system where all three types of fading will play a significant role. The concept of doubly selective (in time and frequency) fading channels [39] has received recent attention for OFDM. The reason that the combination of these two types of correlation is important is that in the context of OFDM, they appear to compete with each other. On one hand, a highly frequency-selective channel—resulting from a long multipath channel as in a wide area wireless broadband network—requires a large number of potentially closely spaced subcarriers to effectively combat the intersymbol interference and small coherence bandwidth. On the other hand, a highly mobile channel with a large Doppler causes the channel to fluctuate over the resulting long symbol period, which degrades the subcarrier orthogonality. In the frequency domain, the Doppler frequency shift can cause significant inter-carrier interference as the carriers become more closely spaced. Although the mobility and multipath delay spread must reach fairly severe levels before this doubly selective effect becomes significant, this is a problem facing mobile LTE systems that does not have a comparable precedent. The scalable nature of the LTE physical layer—notably variable numbers of subcarriers and guard intervals—will allow custom optimization of the system for different environments and applications.

2.5.3 Empirical Channel Models

The parametric statistical channel models discussed thus far in the chapter do not take into account specific wireless propagation environments. While exactly modelling a wireless channel requires complete knowledge of the surrounding scatterers (e.g., buildings, plants, etc.), the time and computational demands of such a methodology is unrealistic due to the near infinite number of possible

transmit-receive locations, and the fact that objects are subject to movement. Therefore, empirical and semi-empirical wireless channel models have been developed to accurately estimate the path loss, shadowing, and small-scale fast fading. Although these models are generally not analytically tractable, they are very useful for simulations and to fairly compare competing designs. Empirical models are based on extensive measurement of various propagation environments, and they specify the parameters and methods for modeling the typical propagation scenarios in different wireless systems. Compared to parametric channel models, the empirical channel models take into account realistic factors such as angle of arrival (AoA), angle of departure (AoD), antenna array fashion, angle spread (AS), and antenna array gain pattern.

Different empirical channel models exist for different wireless scenarios, such as suburban macro, urban macro, urban micro cells, and so on. For channels experienced in different wireless standards, the empirical channel models are also different. In the following sections, we briefly introduce the common physical parameters and methodologies used in several major empirical channel models. These models are also applicable to the multiple antenna systems described in Chapter 4.

LTE Channel Models for Path Loss

In this section we briefly introduce the empirical LTE channel models, which are widely used in modelling the outdoor macro- and micro-cell wireless environments. These are also referred to as “3GPP” channel models as they derive from the earlier channel models from the same standards body.

First, we need to specify the environment where an empirical channel model is used, e.g., suburban macro, urban macro, or urban micro environment. The BS to BS distance is typically larger than 3 km for a macro-cell environment and less than 1 km for an urban micro-cell environment.

The path loss can then be specified by empirical models for these different scenarios. For the 3GPP macro-cell environment, the path loss is given by the so-called COST Hata model, which is given by the following easily computable, if not immediately intuitive, formula:

(2.46)

$$PL_c[\text{dB}] = (44.9 - 6.55\log_{10}(h_b))\log_{10}(d) + 46.3 + 33.9\log_{10}(f_c) - 13.82\log_{10}(h_b) - a(h_m) + C_o,$$

where h_b is the BS antenna height in meters, f_c is the carrier frequency in MHz, d is the distance between the BS and MS in kilometers, $a(h_m)$ is a relatively negligible correction function for the mobile height defined as $a(h_m) = (1.1\log_{10}(f_c) - 0.7)h_m - 1.56\ \log_{10}(f_c) - 0.8$ where h_m is the mobile antenna height in meters. The COST Hata model is generally considered to be accurate when d is between 100 meters and 20 km and $f_c \in (1500, 2000)$MHz.

Since LTE systems are also envisioned to operate below 1500MHz, for example at 700MHz, the empirical channel model used in such scenarios is the Hata model, which is closely related to the COST Hata model, but with slightly different parameters. Several slightly different Hata models exist, depending on whether the environment is urban, suburban, or for open areas. The Hata Model for Urban Areas is:

(2.47)

$$PL_u[\text{dB}] = (44.9 - 6.55\log_{10}(h_b))\log_{10}(d) + 69.55 + 26.16\log_{10}(f_c) - 13.82\log_{10}(h_b) + C_1,$$

where C_1 is a corrective factor that further varies depending on the size of the city, but for a medium or small city is

$C_1 = 0.8 + (1.1\log_{10}(f_c) - 0.7)h_m - 1.56\ \log_{10}(f_c)$

The Hata Model for both Suburban and Open Areas derives from the Urban model. The Suburban path loss is given as

$$PL_s[\text{dB}] = PL_u - 2\left(\log_{10}\left(\frac{f_c}{28}\right)\right)^2 - 5.4$$

while the Open Area Hata Model is

$PL_o[\text{dB}] = PL_u - 4.78(\log_{10} f_c)^2 + 18.33\ \log_{10}(f_c) - 40.94$

We do not suggest that readers spend too long searching for deep meaning in these equations—they can be viewed essentially as statistical curvefits based on experimentation.

LTE Channel Models for Multipath

The 3GPP channel models also include considerations for multipath modelling and scattering. The received signal at the mobile receiver consists of *N* time-delayed versions of the transmitted signal. The *N* paths are characterized by powers and delays that are chosen according to prescribed channel generation procedures, as follows. The number of paths *N* ranges from 1 to 20 and is dependent on the specific channel models. For example, the 3GPP channel model has $N = 6$ multipath components. The power distribution normally follows the exponential profile, but other power profiles are also supported. In the next subsection, specific semi-empirical models are given for the power profiles.

Each multipath component further corresponds to a cluster of *M* subpaths, where each subpath characterizes the incoming signal from a scatterer. The *M* subpaths define a cluster of adjacent scatterers, and therefore have the same multipath delay. The *M* subpaths have random phases and subpath gains, specified by the given procedure in different stands. For 3GPP, the phases are random variables uniformly distributed from 0 to 360 degrees, and the subpath gains are given by the following equation.

The Angle of Departure (AoD) is usually within a narrow range in outdoor applications due to the lack of scatterers around the BS transmitter, and is often assumed to be uniformly distributed in indoor applications. The Angle of Arrival (AoA) is typically assumed to be uniformly distributed due to the abundance of local scattering around the mobile receiver. The final channel is created by summing up the *M* subpath components. In the 3GPP channel model, the *n*th multipath component from the *u*th transmit antenna to the *s*th receive antenna, is given as

(2.48)

$$h_{u,s,n}(t) = \sqrt{\frac{P_n \sigma_{SF}}{M}} \sum_{m=1}^{M} \begin{pmatrix} \sqrt{G_{BS}(\theta_{n,m,AoD})} \exp\left(j\left[kd_s \sin\left(\theta_{n,m,AoD} + \Phi_{n,m}\right)\right]\right) \\ \times \sqrt{G_{BS}(\theta_{n,m,AoA})} \exp\left(jkd_u \sin\left(\theta_{n,m,AoA}\right)\right) \\ \times \exp\left(jk\|\mathbf{v}\| \cos\left(\theta_{n,m,AoA} - \theta_v\right) t\right) \end{pmatrix}$$

where the following are the key parameters, shown visually in Figure 2.19:

• P_n is the power of the *n*th path, following exponential distribution.

• σ_{SF} is the lognormal shadow fading, applied as a bulk parameter to the *n* paths. The shadow fading is determined by the delay spread (DS), angle spread (AS), and

shadow fading (SF) parameters, which are correlated random variables generated with specific procedures.

• M is the number of subpaths per path.

• $\theta_{n,m,AoD}$ is the AoD for the mth subpath of the nth path.

• $\theta_{n,m,AoA}$ is the AoA for the mth subpath of the nth path.

• G_{BS} ($\theta_{n,m,AoD}$) is the BS antenna gain of each array element.

• G_{BS} ($\theta_{n,m,AoA}$) is the MS antenna gain of each array element.

• k is the wave number $\frac{2\pi}{\lambda}$ where λ is the carrier wavelength in meters.

• d_s is the distance in meters from BS antenna element s from the reference (s = 1) antenna.

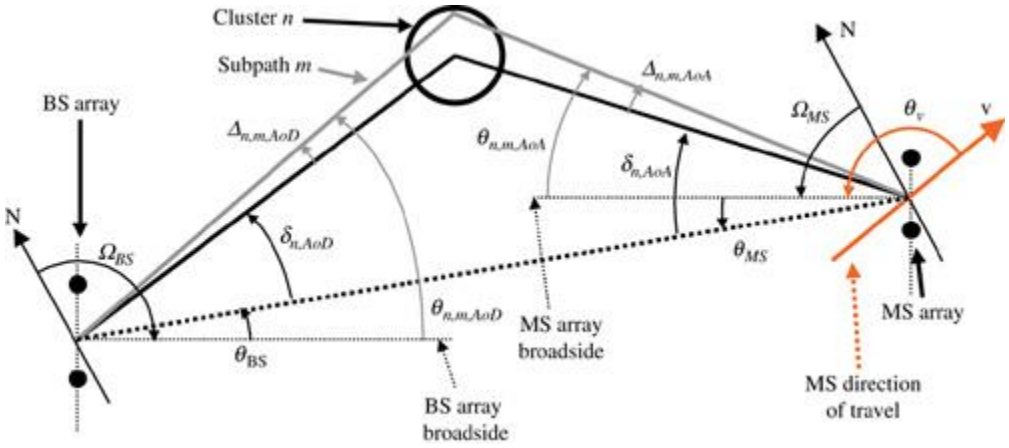
• d_u is the distance in meters from MS antenna element u from the reference (u = 1) antenna.

• $\Phi_{n,m}$ is the phase of the mth subpath of the nth path, uniformly distributed between 0 and 360 degrees.

• $\|\mathbf{v}\|$ is the magnitude of the MS velocity vector, which consists of the velocity of the MS array elements.

• θ_v is the angle of the MS velocity vector.

Figure 2.19 3GPP channel model for MIMO simulations.



LTE Semi-Empirical Channel Models

The above empirical multipath model provides a very thorough description of a number of propagation environments. However, due to the sheer number of parameters involved, constructing a fully empirical channel model is relatively time-consuming and computationally intensive. Semi-empirical channel models

are therefore alternatives, which provide the accurate inclusion of the practical parameters in a real wireless system, while maintaining the simplicity of statistical channel models.

Well-known examples of the simpler multipath channel models include the 3GPP2 Pedestrian A, Pedestrian B, Vehicular A, and Vehicular B models, suited for low-mobility pedestrian mobile users and higher mobility vehicular mobile users. The power delay profile of the channel is determined by the number of multipath taps and the power and delay of each multipath component. Each multipath component is modelled as independent Rayleigh fading with a different power level, and the correlation in the time domain is created according to a Doppler spectrum corresponding to the specified speed. The Pedestrian A is a flat fading model corresponding to a single Rayleigh fading component with a speed of 3 km/hr, while the Pedestrian B model corresponds to a power delay profile with four paths of delays [0 .11 .19 .41] microseconds, and the power profile given as [1 0.1071 0.0120 0.0052], also at 3 km/hr. For the vehicular A model, the mobile speed is specified at 30 km/hr. Four multipath components exist, each with delay profile [0 0.11 0.19 0.41] microseconds and power profile [1 0.1071 0.0120 0.0052]. For the vehicular B model, the mobile speed is 30 km/h, with six multipath components, delay profile [0 0.2 0.8 1.2 2.3 3.7] microseconds and power profile [1 0.813 0.324 0.158 0.166 0.004]. These models are often referred to as Ped A/B and Veh A/B. It is easy to see how these four models can be used to coarsely model the scenarios of low (Ped) and high (Veh) mobility, and small (A) and large (B) delay spreads.

The LTE standard has additionally defined *Extended* delay profiles with increased multipath resolution known as Extended Pedestrian A, Extended Vehicular A, and Extended Typical Urban. These profiles are given in Tables 2.4, 2.5, and 2.6.

Table 2.4 Extended Pedestrian A Model

Delay [nsec]	0	30	70	90	110	190	410
Relative Power [dB]	0	−1.0	−2.0	−3.0	−8.0	−17.2	−20.8

Table 2.5 Extended Vehicular A Model

Delay [nsec]	0	30	150	310	370	710	1090	1730	2510
Relative Power [dB]	0	−1.5	−1.4	−3.6	−0.6	−9.1	−7.0	−12.0	−16.9

Table 2.6 Extended Typical Urban Model

Delay [nsec]	0	50	120	200	230	500	1600	2300	5000
Relative Power [dB]	−1.0	−1.0	−1.0	0.0	0.0	0.0	−3.0	−5.0	−7.0

These empirical channel models follow the fundamental principles of the statistical parametric models discussed previously in this chapter, while considering empirical measurement results. As such, semi-empirical channel models are suitable for link-level simulations and performance evaluation in real broadband wireless environments.

2.6 MITIGATION OF NARROWBAND FADING

There are many different techniques used to overcome narrowband fading in modern communication systems, but most can be collectively referred to as *diversity*. Because the received signal power is random, if several (mostly) uncorrelated versions of the signal can be received, chances are good that at least one of the versions has adequate power. Without diversity, high data rate wireless communication is virtually impossible. We now briefly quantify the potential cost of narrowband fading, before describing the principal techniques used in LTE to overcome these deleterious effects.

2.6.1 The Effects of Unmitigated Fading

The probability of bit error (BER) is the principle metric of interest for the physical layer (PHY) of a communication system. For a QAM-based modulation system, the BER in an additive white Gaussian noise (AWGN, no fading) can accurately be approximated by the following bound [16]:

(2.49)

$$P_b \leq 0.2e^{-1.5\mathrm{SNR}/(M-1)}$$

where $M \geq 4$ is the MQAM alphabet size.[10] Note that the probability of error decreases very rapidly (exponentially) with the SNR, so decreasing the SNR linearly causes the BER to increase exponentially. Since the channel is constant, the BER is constant over time.

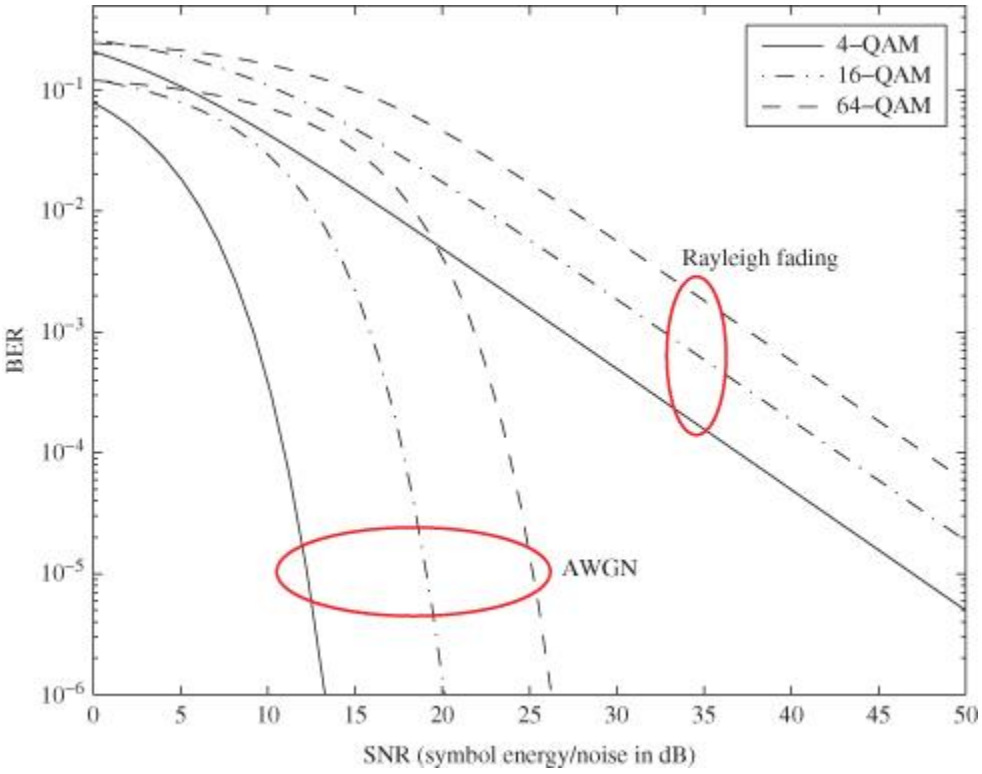
However, in a fading channel, the BER become a random variable that depends on the instantaneous channel strength, and the occasional instances when the channel is in a deep fade therefore dominate the average BER. When the required average BER is very low (say 10^{-6}), virtually all errors are made while in deep fades. The average BER varies depending on the precise constellation used, but roughly follows the relationship

(2.50)

$$\bar{P}_b \propto \frac{M}{SNR}$$

The key point being that now BER goes down very slowly with SNR, only inversely. This trend is captured plainly in Figure 2.20, where we see that at reasonable system BERs like $10^{-5} - 10^{-6}$, the required SNR is over 30 dB higher in fading! Clearly, it is not desirable, or even possible, to increase the power by over a factor of 1000 to overcome occasional deep fades. Furthermore, in an interference-limited system, increasing the power will not significantly raise the effective SINR.

Figure 2.20 Flat fading causes a loss of at least 20–30 dB at reasonable BER values.



Although BER is a more analytically convenient measure since it is directly related to the SINR, for example via (2.49), a more common and relevant measure in LTE is the Packet Error Rate (PER), or equivalently Block Error Rate (BLER) or Frame Error Rate (FER). All these measures refer to the probability that at least one bit is in error in a block of L bits. This is the more relevant measure since the detection of a single bit error in a packet by the Cyclic Redundancy Check (CRC) requires the packet to either be discarded by the receiver or retransmitted. An expression for PER is

(2.51)

$$\mathrm{PER} \leq 1 - (1 - P_b)^L$$

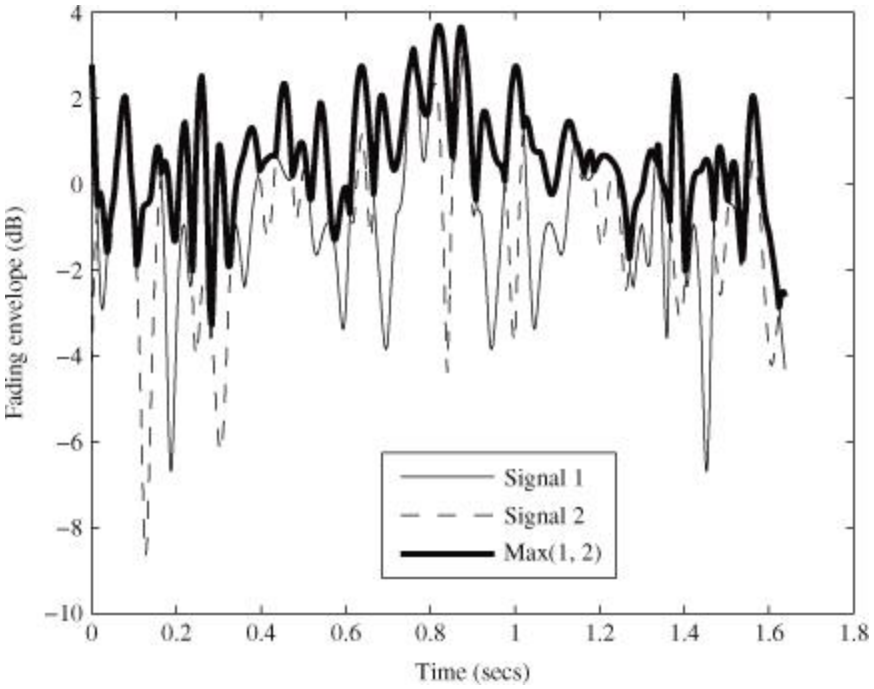
where P_b is the BER and L is the packet length. This expression is true with equality when all bits are equally likely to be in error. If the bit errors are correlated, then the PER actually improves. It is clear that PER and BER are directly related, so reducing PER and BER are roughly equivalent objectives.

Diversity is the key to overcoming the potentially devastating performance loss from fading channels, and to improving PER and BER. We now discuss the three most important types of diversity deployed in LTE systems.

2.6.2 Spatial Diversity

Spatial diversity is a powerful form of diversity, and particularly desirable since it does not necessitate redundancy in time or frequency. It usually is achieved by having two or more antennas at the receiver and/or the transmitter. The simplest form of space diversity consists of two receive antennas, where the stronger of the two signals is selected. As long as the antennas are spaced sufficiently, the two received signals will undergo approximately uncorrelated fading. This type of diversity is sensibly called *selection diversity*, and is illustrated in Figure 2.21. Even though this simple technique completely discards "half" of the received signal, most of the deep fades can be avoided and the average SNR is also increased. More sophisticated forms of spatial diversity include receive antenna arrays (two or more antennas) with maximal ratio combining, transmit diversity using space-time codes, transmit precoding, and other combinations of transmit and receive diversity. Spatial signaling techniques are so important, the ultimate success of LTE, that Chapter 4 is dedicated to this topic.

Figure 2.21 Simple two-branch selection diversity eliminates most deep fades.



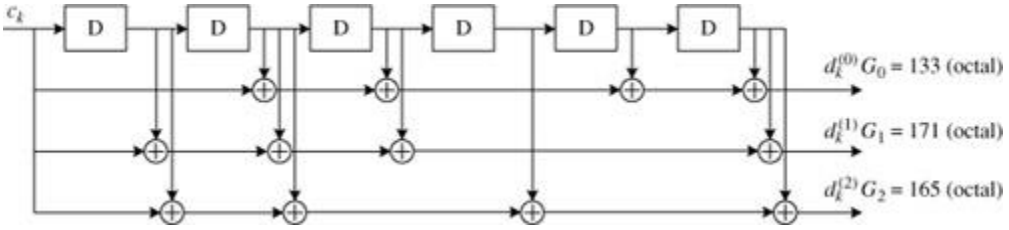
2.6.3 Coding and Interleaving

A ubiquitous form of diversity in nearly all contemporary digital communication systems is the natural pair of coding and interleaving. Traditionally thought of as a form of time diversity, in a multicarrier system they also can capture frequency diversity.

By coding, we mean the use of error correction codes (ECCs), which is also sometimes known as forward error correction. ECCs efficiently introduce redundancy at the transmitter to allow the receiver to recover the input signal even if the received signal is significantly degraded by attenuation, interference, and noise. Coding techniques can be categorized by their *coding rate* $r \leq 1$, which is the inverse of the redundancy added. For example, the output of a rate $\frac{1}{3}$ code has three times the original rate, or in other words has introduced two redundant bits for every original information bit. Equivalently, if the transmission rate is held constant, a rate $\frac{1}{3}$ code lowers the transmitted bit rate by a factor of 3, which is to say the data rate is multiplied by $\frac{1}{3}$. Somewhat counter-intuitively, though, well-designed error correction codes actually *increase* the achieved data rate even for $r < 1$ because the reliability increase they provide is so great that the number of bits per symbol that can be successfully transmitted increases by a factor greater than $1/r$, producing a net gain.

In Figure 2.22 we reproduce a figure of the convolutional encoder defined by LTE for use in the Broadcast Channel (BCH). This is clearly a rate $\frac{1}{3}$ code since there is one input bit (c_k) and 3 outputs $d_k^{(i)}$. The constraint length of this code is 7; equivalently, there are 6 delay elements or 64 possible states. The generator polynomial **G**, which consists of the generators G_i for each of the three outputs, are by a somewhat archaic convention denoted in octal notation. For example, G_0 = 133 in binary form is 1 0 1 1 0 1 1, where a 0 means the output does not include this tap and a 1 means it does. Therefore, $d_k^{(0)}$ includes modulo-2 summed contributions from the input and after delay elements 2, 3, 5, and 6. Notably, all optimal convolutional codes will include in each output the first and last taps, for maximum memory. The job of the decoder is to take degraded output symbols $\hat{d}_k$ after demodulation, and produce an estimate $\hat{c}_k$ of the original information signal c_k. If for a given packet $\hat{c}_k = c_k$, then the packet was successfully received; otherwise, it must be retransmitted. The most common decoding technique for convolutional codes is the reduced-state sliding-window maximum likelihood sequence estimator, popularly known as the Viterbi decoder [50]. A detailed discussion of how convolutional coding and decoding works is outside the scope of this book, and the interested reader is referred to the vast literature on this topic, for example [29, 52].

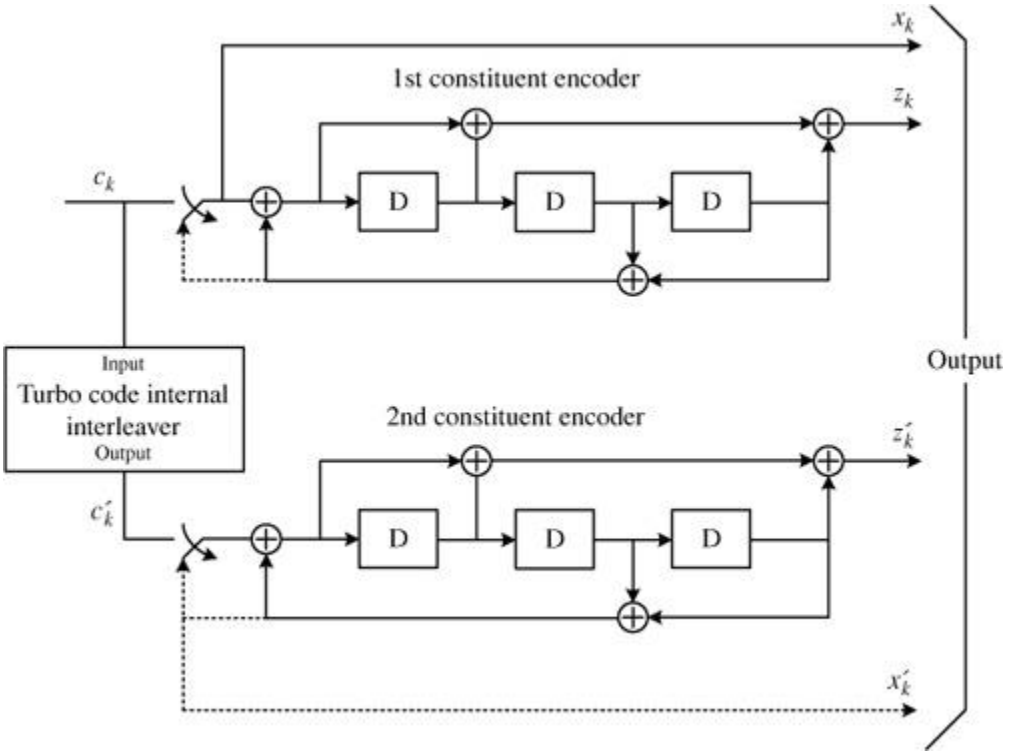
Figure 2.22 The rate $\frac{1}{3}$ convolutional encoder defined by LTE for use in the Broadcast Channel (BCH).



Turbo codes [4] build upon convolutional codes to provide increased resilience to errors through iterative decoding [3, 21]. A rate $\frac{1}{3}$ turbo code is also deployed by LTE as shown in Figure 2.23, including for the uplink and downlink shared channels. In particular, the encoder is a parallel concatenated convolutional code that comprises an 8-state rate $\frac{1}{2}$ *systematic* encoder and an 8-state rate 1 systematic encoder that operates on an interleaved input sequence, for a net coding rate of $\frac{1}{3}$. By *systematic*, we mean that one output is generated by a linear modulo-2 sum of the current encoder state that is a function of both the input bit(s) and the previous states (i.e., there is feedback in the state machine), while the other outputs are simply passed through to the output, like x_k in Figure 2.23. Note that the

convolutional encoder in Figure 2.22 was not systematic: there is no feedback in that state machine, and all outputs are linear combinations of the current state in addition to the current input.

Figure 2.23 The rate $\frac{1}{3}$ parallel concatenated turbo encoder defined by LTE for use in the uplink and downlink shared channels, among others.



Although not shown in either figure, codes in LTE can also be *punctured*, which means that some of the output coded bits are simply dropped, in order to lower the transmission rate. For example, if the output of a rate $\frac{1}{2}$ convolutional code had a puncturing factor of $\frac{1}{4}$, this means that out of every four output bits, one is dropped. Hence, the effective code rate would become $\frac{2}{3}$, since only three coded bits are transmitted for every two information bits. At the decoder, a random or fixed coded bit is inserted in the decoding process, which of course has a 50% chance of being incorrect. The decoded sequence is therefore less reliable than in the case of a lower-rate, unpunctured code, but if the received SINR was relatively good, the transmitted bit sequence can usually still be decoded correctly with the intended benefit of having a higher overall data rate in the same bandwidth, due to using a lower rate code. In general, a punctured code has worse performance than an unpunctured code of the same rate. For example, a rate $\frac{1}{2}$ code with every fourth bit punctured to achieve a rate $\frac{2}{3}$ code will usually have worse performance than a

well-designed rate $\frac{2}{3}$ code that actually produced three bits as a linear combination of every two input bits using two-state machines. The reason puncturing is used is one of simplicity and complexity: since encoders and decoders must have a different structure depending on the coding polynomial, switching between many different encoder structures places a large complexity burden particularly on the decoder. Simply puncturing the same code to achieve different coding rates allows the decoder structure to remain the same regardless of the code rate.

Interleaving is typically used in both convolutional coding and turbo coding. For use with a conventional convolutional code, the interleaver shuffles coded bits to provide robustness to burst errors that can be caused by either bursty noise and interference, or a sustained fade in time or frequency. In such noise or fading, many adjacent bits in the transmitted signal will be degraded, which without interleaving makes it very difficult for the error correcting decoder to reconstruct the intended bits, even accounting for considerable redundancy. Interleaving seeks to spread out coded bits so that the effects of a burst error, after deinterleaving, are spread roughly evenly over a frame, or block.

For turbo coding, the intuition is a bit different: here an interleaver is used between the concatenated codes in order to provide statistical independence between the two encoder outputs. At the receiver, the decoders for each encoder pass their soft outputs (conditional probabilities) back and forth via a deinterleaver that decorrelates these values. The decoder proceeds to iterate back and forth between each decoder until the symbol estimates converge, i.e., the interleaver is no longer able to decorrelate the soft outputs.

For both conventional convolutional codes and turbo codes, the interleaver block size would, from a data reliability standpoint, ideally be quite large: several coherence times or coherent bandwidths for convolutional codes in order to maximize the attained diversity, and the larger the better for turbo codes and other iterative decoders as well. However, large interleaving blocks cause long interleaving and deinterleaving delays since typically the entire block must be populated with coded bits (a serial operation) before the interleaving is performed and the output generated. For this reason, the interleaver block size is usually constrained to be at most over a single packet, and often much less than that. Deinterleaving delays have been one of the primary impediments to turbo-coding since they cause considerable latency. Nevertheless, interleaving has proven very effective in allowing ECCs designed for constant, time-invariant additive noise channels to also work well on fading, time-variant noisy channels.

Specifically to LTE, interleaving is used in several capacities. It is used in the turbo code as seen in Figure 2.23, in the rate matching functions, and finally there is an additional channel interleaver used to send control information over the uplink shared channel in order to spread it out over a wide range of subcarriers.

2.6.4 Automatic Repeat Request (ARQ)

Another technique that is often used in modern wireless communication systems including LTE is ARQ (automatic repeat request) and Hybrid-ARQ. ARQ simply is a MAC layer retransmission protocol that allows erroneous packets to be quickly retransmitted. Such a protocol works in conjunction with PHY layer ECCs and parity checks to ensure reliable links even in hostile channels. Since a single bit error causes a packet error, with ARQ the entire packet must be retransmitted even when nearly all of the bits already received were correct, which is clearly inefficient. Imagine the situation where the same packet is "dropped" twice in a row, despite the fact that 99% of its bits were received correctly. In such cases, it is likely that every bit was received correctly in one of the two packets.

Hybrid-ARQ combines the two concepts of ARQ and FEC to avoid such waste, by combining received packets. Hybrid-ARQ, therefore, is able to extract additional time diversity in a fading channel as well. In H-ARQ a channel encoder such as a convolution encoder or turbo encoder is used to generate additional redundancy to the information bits. However, instead of transmitting all the encoded bits (systematic bits + redundancy bits), only a fraction of the encoded bits are transmitted. This is achieved by puncturing some of the encoded bits to create an effective code rate greater than the native code rate of the encoder. After transmitting the encoded and punctured bits, the transmitter waits for an acknowledgment from the receiver telling it whether the receiver was able to successfully decode the information bits from the transmission. If the receiver was able to decode the information bits, then nothing else needs to be done. If, on the other hand, the receiver was unable to decode the information bits, then the transmitter can resend another copy of the encoded bits.

In type I H-ARQ—commonly referred to as *chase combining*—during a retransmission the transmitter sends a copy of the encoded bits that is identical to the first transmission and the receiver soft combines the received bits with the previous transmission. Since the encoded bits are the same in all the transmissions and the noise/interference is uncorrelated, the receiver can combine all the transmissions to increase the effective SINR. Thus, with every subsequent transmission the error probability is reduced and this process continues until the

receiver is able to decode the information without error. In type II H-ARQ—commonly referred to as *incremental redundancy*—the transmitter changes the bits that are punctured during a retransmission. This way with every retransmission the effective code rate at the receiver decreases, which reduces the error probability.

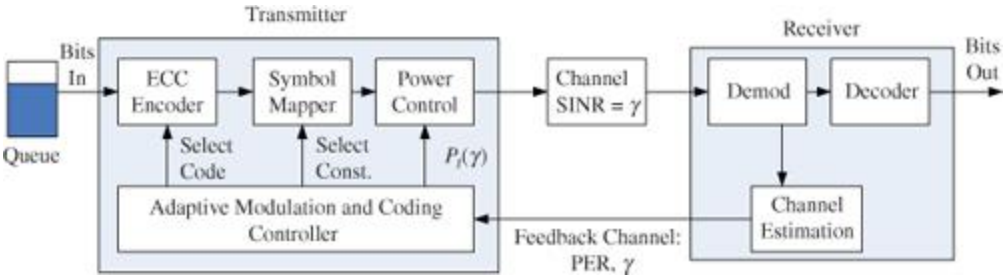
2.6.5 Adaptive Modulation and Coding (AMC)

LTE systems employ adaptive modulation and coding (AMC) in order to take advantage of fluctuations in the channel over time and frequency. The basic idea is quite simple: transmit as high a data rate as possible when and where the channel is good, and transmit at a lower rate when and where the channel is poor in order to avoid excessive dropped packets. Lower data rates are achieved by using a small constellation—such as QPSK—and low rate error correcting codes such as rate $\frac{1}{3}$ turbo codes. The higher data rates are achieved with large constellations—such as 64QAM—and less robust error correcting codes, for example, either higher rate (like $\frac{3}{4}$) codes, or in LTE's case, *punctured* turbo codes for the reduced-complexity reasons discussed in the previous section.

To perform AMC, the transmitter must have some knowledge of the instantaneous channel gain. Once it does, it can choose the modulation technique that will achieve the highest possible data rate while still meeting a BER or packet error rate (PER) requirement. An alternative objective is to pick the modulation and/or coding combination that simply maximizes the successful throughput, i.e., max $br(1 - \text{PER})$ where $b = \log_2 M$ is the number of bits per symbol (modulation type) and $r \leq 1$ is the coding rate. For example, in (2.49) it can be seen that as the constellation alphabet size M increases, the BER also increases. Since the data rate is proportional to $\log_2 M$, we would like to choose the largest alphabet size M such that either the required BER/PER is met or the throughput is maximized.

A block diagram of an AMC system is given in Figure 2.24. For simplicity, we will first consider just a single user system attempting to transmit as quickly as possible through a channel with a variable SINR, for example, due to fading. The goal of the transmitter is to transmit data from its queue as rapidly as possible, subject to the data being demodulated and decoded reliably at the receiver. Feedback is critical for adaptive modulation and coding: the transmitter needs to know the "channel SINR" γ—which is defined as the received SINR γ_r divided by the transmit power P_t (which itself is usually a function of γ). The received SINR is thus $\gamma_r = P_t\gamma$.

Figure 2.24 Adaptive modulation and coding block diagram.



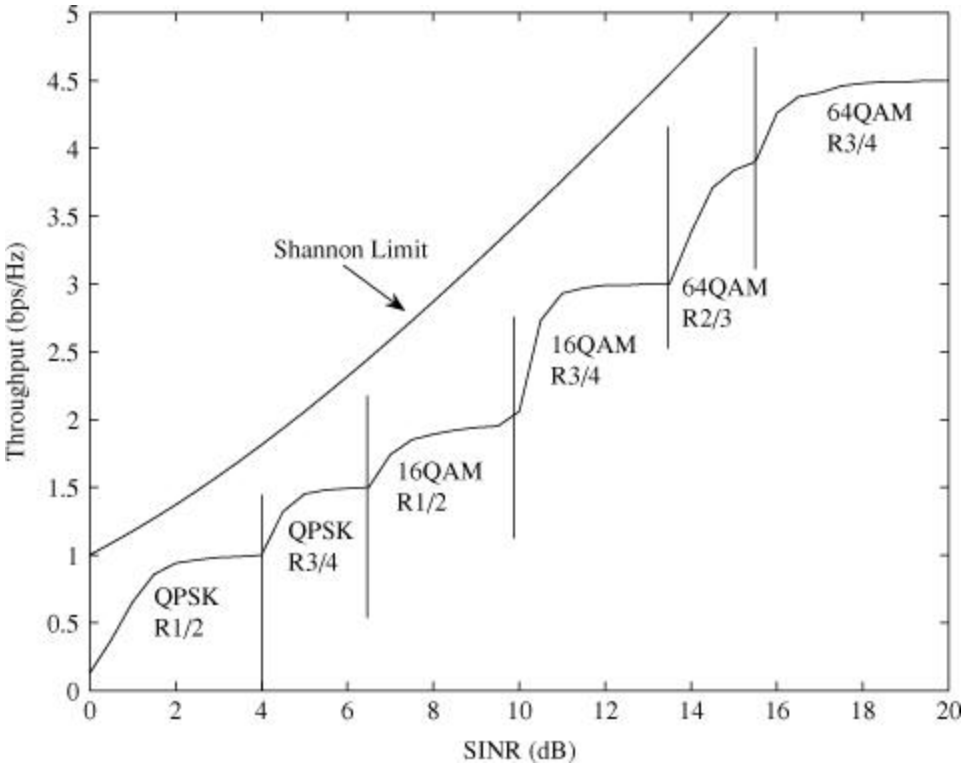
A Practical Example of AMC

Figure 2.25 shows a possible realization of AMC, using three different code rates ($\frac{1}{2}$, $\frac{2}{3}$, $\frac{3}{4}$), and three different modulation types (QPSK, 16QAM, 64QAM). We note that these are not the precise configurations used in LTE, but are used here as a demonstrative example. Using just six of the possible nine combinations, a "gap" to Shannon capacity[11] of approximately 2–4 dB can be maintained over a wide range of SINR. This allows the throughput to increase as the SINR increases. In this example, the lowest offered data rate is QPSK and rate $\frac{1}{2}$ turbo codes, while the highest data rate burst profile is with 64QAM and rate $\frac{3}{4}$ turbo codes. The achieved throughput normalized by the bandwidth is defined as

(2.52)

$$T = (1 - \text{PER}) r \log_2(M) \text{ bps/Hz}$$

Figure 2.25 Throughput vs. SINR, assuming the best available constellation and coding configuration is chosen for each SINR. Only 6 configurations are used in this figure, and the turbo decoder is a max log MAP decoder with 8 iterations of message passing.



where PER is the packet error rate, $r \leq 1$ is the coding rate, and M is the number of points in the constellation. For example, 64QAM with rate $\frac{3}{4}$ codes achieves a maximum throughput of 4.5 bps/Hz (when PER $\rightarrow 0$), while QPSK with rate $\frac{1}{2}$ codes achieves a best case throughput of 1 bps/Hz. Some combinations are never used: for example, 64QAM with rate $\frac{1}{2}$ coding has the same data rate as 16QAM with rate $\frac{3}{4}$ coding (3 bits/symbol) but worse performance over all SINR values, while the slight gain from using a rate $\frac{2}{3}$ code with QPSK or 16QAM does not justify its inclusion.

The results shown here are for the idealized case of perfect channel knowledge and do not consider retransmissions, for example, with ARQ. In practice, the feedback will incur some delay, and perhaps also be degraded due to imperfect channel estimation or errors in the feedback channel. LTE systems heavily protect the feedback channel with error correction, so usually the main source of degradation is due to mobility, which causes channel estimates to rapidly become obsolete. Empirically, with speeds greater than about 30 km/hr (on a 2GHz carrier) even the faster feedback configurations do not allow timely and accurate channel state information to be available at the transmitter.

Tuning the Adaptive Modulation and Coding Controller

A key challenge in AMC is to efficiently control three different quantities at once: transmit power, transmit rate (constellation), and the coding rate. This corresponds to developing an appropriate policy for the AMC controller shown in Figure 2.24. Although reasonable guidelines can be developed from a theoretical study of adaptive modulation, in practice the system engineer needs to develop and fine-tune the algorithm based on extensive simulations since the performance depends on many factors. We list a few considerations here, which are by no means exhaustive.

• **PER and Received SINR** In adaptive modulation theory, the transmitter needs only to know the statistics and instantaneous channel SINR. From the channel SINR, it can determine the optimum coding/modulation strategy and transmit power [10]. In practice, however, the PER should be carefully monitored as the final word on whether the data rate should be increased (if the PER is low) or decreased to a more robust setting.

• **Automatic Repeat Request (ARQ)** ARQ allows rapid retransmissions, and Hybrid-ARQ generally increases the ideal PER operating point by about a factor of 10, e.g., from 1% to 10%. For delay-tolerant applications, it may be possible to accept a PER approaching even 70%, if chase combining is used in conjunction with H-ARQ to make use of unsuccessful packets.

• **Power Control: A balancing act** Power control is a subtle system level issue, because competing objectives must be balanced. For example, in theory the best power control policy for the capacity of parallel channels (over time or frequency) is the so-called waterfilling strategy, in which more power is allocated to strong channels, and less power allocated to weak channels [19, 20]. In practice though, the opposite approach—sometimes referred to as channel inversion—is often taken, because a sensible goal is to use as little power as possible to achieve a given target rate and reliability. Excess power over this amount causes interference to the network. For example, referring to Figure 2.25, there is almost nothing gained with a 13 dB SINR vs. an 11 dB SINR: in both cases the throughput is 3 bps/Hz. Therefore, as the SINR improved from 11 dB to 13 dB, the transmitter would be well advised to lower the transmit power, in order to save power and generate less interference to neighboring cells. Power control mechanisms in LTE are discussed in Section 9.10, and additional discussion can be found in [47, 54].

• **Adaptive Modulation in OFDMA** In an OFDMA system, each user will be allocated a block of subcarriers, each of which will have a different set of SINRs. Therefore, care needs to be paid to which constellation/coding set is chosen based on the varying SINRs across the subcarriers.

2.6.6 Combining Narrowband Diversity Techniques—The Whole Is Less Than the Sum of the Parts

It should be noted that the use of diversity in one domain can decrease the gain of diversity in another domain. For example, if selection diversity is deployed on the received signals in Figure 2.21, the resulting "combines" signal will become increasingly flat in time, since at each instant, the best signal is selected. Therefore, the *gain* (in say, dB) from adaptive modulation and coding would be less than if there was no selection combining, since there would be less variation to exploit. Although the overall robustness is maximized by using all available forms of diversity, more diversity causes the effective channel to get closer to an AWGN channel (constant SNR), so additional sources of diversity achieve diminishing returns. In short, the aggregate diversity gain from multiple techniques is less than the sum of the individual diversity gains. Failure to recognize this will result in optimistic projections of system performance.

2.7 MITIGATION OF BROADBAND FADING

In Section 2.4.1 we discussed how frequency-selective fading causes dispersion in time, which causes adjacent symbols to interfere with each other unless $T \gg \tau_{\max}$. Since the data rate R is proportional to $1/T$, high data rate systems almost invariably have a substantial multipath delay spread, i.e., $T \ll \tau_{\max}$, and experience very serious *intersymbol interference* (ISI) as a result. Choosing a technique to effectively combat ISI is a central design decision for any high data rate system. Increasingly, OFDM is the most popular choice for combatting ISI in a range of high rate systems, including 802.11/WiFi, 802.16/WiMAX, and of course LTE. OFDM is the focus of the next chapter, so now let us briefly consider the other main techniques for ISI mitigation.

2.7.1 Spread Spectrum and RAKE Receivers

Somewhat counter-intuitively, actually speeding up the *transmission* rate can help combat multipath fading, assuming the *data* rate is kept the same. Since speeding up the transmission rate for a narrowband data signal results in a wideband transmission, this technique is called *spread spectrum*. Spread spectrum techniques are generally broken into two quite different categories: direct sequence and frequency hopping. Direct sequence spread spectrum, also known as Code Division Multiple Access (CDMA), is used widely in cellular voice networks and is effective at multiplexing a large number of variable rate users in a cellular

environment. Frequency hopping is used in some low-rate wireless LANs like Bluetooth, and also for its interference averaging properties in GSM cellular networks.

Some of LTE's natural competitors for wireless broadband data services have grown out of the CDMA cellular voice networks—notably 1xEVDO (from cdma2000) and HSDPA/HSUPA (from W-CDMA)—as discussed in Chapter 1. However, CDMA is not an appropriate technology for high data rates, and 1xEVDO and HSDPA are CDMA in name only.[12] For both types of spread spectrum, a large bandwidth is used to send a relatively small data rate. This is a reasonable approach for low data rate communications, like voice, where a large number of users can be statistically multiplexed to yield a high overall system capacity. For high rate data systems, it requires each user to employ several codes simultaneously, which generally results in self-interference. Although this self-interference can be corrected with an equalizer (see next section), this largely defeats the purpose of using spread spectrum to help with intersymbol interference.

In short, spread spectrum is not a natural choice for wireless broadband networks, since by definition, the data rate of a spread spectrum system is less than its bandwidth. The same trend has been observed in wireless LANs: early wireless LANs (802.11 and 802.11b) were spread spectrum[13] and had relatively low spectral efficiency; later wireless LANs (802.11a and 802.11g) used OFDM for multipath suppression and achieved much higher data rates in the same bandwidth.

2.7.2 Equalization

Equalizers are perhaps the most logical alternative for ISI-suppression to OFDM, since they don't require additional antennas or bandwidth, and have moderate complexity. Equalizers are implemented at the receiver, and attempt to reverse the distortion introduced by the channel. Generally, equalizers are broken into two classes: linear and decision-directed (nonlinear).

Linear Equalizers

A linear equalizer simply runs the received signal through a filter that roughly models the inverse of the channel. The problem with this approach is that it inverts not only the channel, but also the received noise. This noise enhancement can severely degrade the receiver performance, especially in a wireless channel with deep frequency fades. Linear receivers are relatively simple to implement, but achieve poor performance in a time-varying and severe-ISI channel.

Nonlinear Equalizers

A nonlinear equalizer uses previous symbol decisions made by the receiver to cancel out their subsequent interference, and so is often called a decision feedback equalizers (DFE). Recall that the problem with multipath is that many separate paths are received at different time offsets, so that prior symbols cause interference with later symbols. If the receiver knows the prior symbols, it can subtract their interference out. One problem with this approach is that it is common to make mistakes about what the prior symbols were (especially at low SNR), which causes "error propagation." Also, nonlinear equalizers pay for their improved performance relative to linear receivers with sophisticated training and increased computational complexity.

Maximum Likelihood Sequence Detection

Maximum Likelihood Sequence Detection (MLSD) is the optimum method of suppressing ISI, but has complexity that scales like $O(M^v)$, where as before M is the constellation size and v is the channel delay. Therefore, MLSD is generally impractical on channels with a relatively long delay spread or high data rate, but it is often used in some low–data rate outdoor systems like GSM. For a high–data rate broadband wireless channel, MLSD is not expected to be practical in the foreseeable future, although suboptimal approximations like delayed-decision feedback sequence estimation (DDFSE)—which is a hybrid of MLSE and decision feedback equalization [12] and reduced-state sequence estimation (RSSE) [14]—are reasonable suboptimal approximations for MLSE in practical scenarios [17].

2.7.3 Multicarrier Modulation: OFDM

The philosophy of multicarrier modulation is that rather than fighting the time-dispersive ISI channel, why not utilize its diversity? For this, a large number of subcarriers (L) are used in parallel, so that the symbol time for each goes from $T \rightarrow LT$. In other words, rather than sending a single signal with data rate R and bandwidth B, why not send L signals at the same time, each having bandwidth B/L and data rate R/L? In this way, if $B/L \ll B_c$, each of the signals will undergo approximately flat fading and the time dispersion for each signal will be negligible. As long as the number of subcarriers L is large enough, the condition $B/L \ll B_c$ can be met. This elegant idea is the basic principle of Orthogonal Frequency Division Multiplexing, or OFDM. In the next chapter, we will take a close look at this increasingly popular modulation technique, discussing its theoretical basis and implementation challenges.

2.7.4 Single-Carrier Modulation with Frequency Domain Equalization

A primary drawback of the OFDM approach to ISI-suppression is that the transmit signal has a high peak-to-average ratio (PAR) relative to a single carrier signal. Put simply, the dynamic range of the transmit power is too large, which results in either significant clipping and distortion, or in a requirement for highly linear power amps (which are expensive and inefficient). Needless to say, both of these outcomes are bad.

Is there a way to effectively do OFDM without generating a high PAR? The answer is yes: one can transmit a single carrier signal with a cyclic prefix, which has a low PAR, and then do all the processing at the receiver. Said processing consists of a Fast Fourier Transform (FFT) to move the signal into the frequency domain, a 1-tap frequency equalizer (just like in OFDM), and then an Inverse FFT to convert back to the time domain for decoding and detection. In addition to eliminating OFDM's PAR problem, an additional advantage of this approach for the uplink is the potential to move the FFT and IFFT operations to the base station. In LTE, however, because multiple uplink users share the frequency channel at the same time, the mobile station still must perform FFT and IFFT operations.