

Chapter 2

Introduction to Wireless Communications

Dinesh Rajan

Abstract This chapter provides an introduction to the basic concepts of a wireless communication system. First, a simple point-to-point wireless link is considered and the key techniques at the transmitter and receiver that are required for a successful communication are discussed. The wireless medium being inherently a shared medium, multiple access methods to support numerous users in a given area is critical. This chapter also discusses several multiuser communication schemes. Specifically, two popular multiuser transmission strategies - CDMA and OFDM, are described and their merits and challenges discussed. We also provide a brief introduction to Shannon's theoretic capacity concept and discuss simple coding methods used in wireless channels.

2.1 Introduction

Digital communication systems have advanced significantly in the past few decades. Coupled with the progress in the semiconductor industry which obeys Moore's law, inexpensive and power efficient mobile communication devices have become widely prevalent. Cellular networks have become ubiquitous in most of the developed world and continues to see exponential growth in developing countries. Wireless local area networks (WLANs) based on the 802.11 standard have opened new venues for high speed data access in homes, offices and hot-spots. Bluetooth based commercial short range personal area networks have enabled several new commercial technologies such as hands-free cellular headsets and remote controllers for video game consoles. Further, the data rates for these technologies are constantly increasing and enabling several new applications as evident from the hugely popular *app stores* from Apple and Android.

Dinesh Rajan

Department of Electrical Engineering, Bobby B. Lyle School of Engineering, Southern Methodist University, Dallas, TX 75275-0340, USA, e-mail: rajand@smu.edu

The goal of this chapter is to introduce the basic components of a wireless communications system. To understand the functioning of digital communication devices, we first introduce the layering principle that is commonly used in the design of these systems.

2.1.1 Layered Architecture

The task of transmitting information from a transmitter device to a receiver device is a complex process that requires careful design, analysis, and implementation. This challenging task is usually accomplished with a layered architecture such as the open systems interconnection (OSI) model proposed by the international standards organization (ISO), which is shown in [Figure 2.1](#). An OSI model, typically consists of 7 layers which function together to accomplish the required data transmission task. The 7 layers and a brief description of their functionality are:

- *Physical Layer*: This layer performs the basic task of converting digital bits into signals that may be sent and recovered over the transmission medium.
- *Data Link Layer*: This layer provides a certain amount of data reliability over a particular link and also assists in separating the signals from the various users that are transmitted over the same channel.
- *Network Layer*: The network layer provides functionality such as routing to enable data packets to be efficiently sent over multiple hops in a network.
- *Transport Layer*: This layer provides a mechanism to ensure end-to-end reliability of data. The transport layer is also tasked with creating congestion control mechanisms that reduces packet losses in a network occurring due to buffer overflows.
- *Session Layer*: The session layer provides services such as directory assistance to enable efficient and reliable data delivery.
- *Presentation Layer*: This layer typically provides data encryption and compression functionality.
- *Application Layer*: This layer includes all other tasks involved in the data transmission process as well as preparing the data for user or machine consumption.

The advantages of a layered architecture is that the design and implementation of each layer is simplified and can be done independent of the implementation of the other layers, as long as the interfaces with the layer above and below it are standardized. This freedom in implementation allows many of these layers to be reused in many different systems. For instance, the transmission control protocol (TCP) functionality (which forms the backbone for the Internet) could be the same irrespective of what physical layer transmission scheme is used. Also, changes in the implementation of one layer are transparent to the other layers. One of the disadvantages of a layered architecture is that the resulting end-to-end system performance may not be optimized. For instance, the use of TCP functionality which was originally

developed for wired networks, in the wireless domain results in severe throughput degradation due to the time varying nature of the wireless channel [4].

Most systems however, use the simplified Internet layering scheme which was proposed in a Request for Comments (RC) document of the Internet Engineering Task Force (IETF) [11]. This scheme broadly consists of 4 layers: link layer, Internet layer, transport layer, and application layer [11]. Note that several variations of these layering methods exist and different authors interpret the RFCs in a slightly different manner [42].

Next, we look at the main elements of a digital communication system.

2.1.2 Digital Communication System

The basic functional blocks in a digital communication system are shown in Figure 2.2. The digital signal to be transmitted is first processed through a source encoder [37]. Most common digital data files have a significant amount of redundancy.

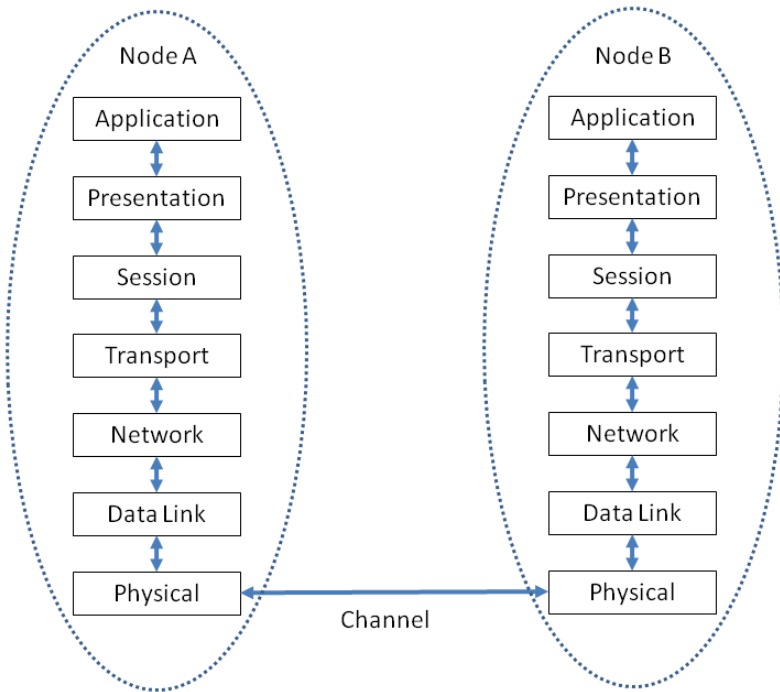


Fig. 2.1 Layered architecture

The source encoder leverages this redundancy to compress the data to the fewest number of bits required to meet the desired fidelity. For instance, text and document files could be subject to a lossless compression using a utility such as WinZip; image files could be subject to a lossy compression such as JPEG. If the signal to be transmitted is an analog signal such as voice, this signal is first processed through an analog-to-digital conversion using a sampling and quantization device before feeding it to the source encoder. In certain systems, the output of the source encoder is passed through an encryption mechanism to ensure data integrity and security [44].

The compressed (and possibly encrypted) data is then passed through a channel encoder. The primary goal of a channel encoder is to add redundancy to the data to protect it against errors experienced when transmitted over the channel [26]. Channel coding may also be used to detect errors in data packets using simple parity check mechanisms. Error detection allows the higher layers of the system to request a retransmission of the erroneous packets. Channel coding schemes are further discussed in Section 2.4.

The output of the channel encoder is passed through a modulator which effectively converts the digital bit stream to an analog signal that is appropriate for transmission over the channel. Examples of popular modulation schemes and their performance are discussed in Section 2.2. The modulated signal is then transmitted

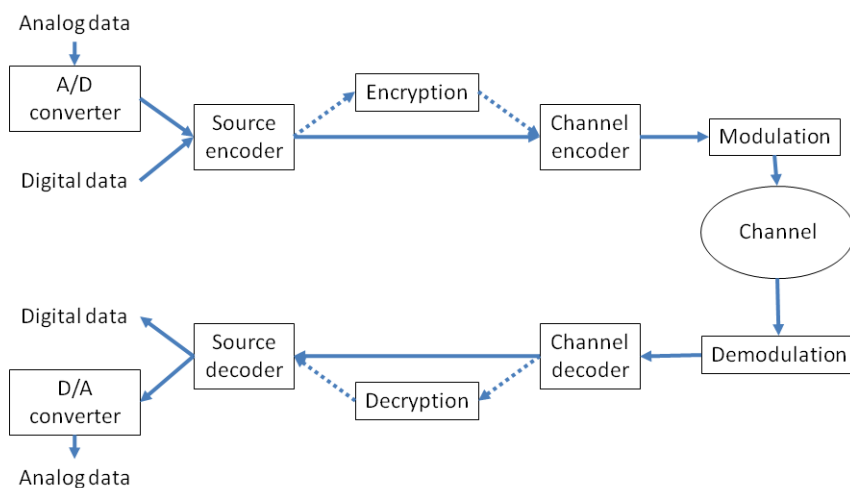


Fig. 2.2 Block diagram of digital communication system

over the desired medium. For instance, in a wireless channel this signal is radiated out from an antenna.

At the receiver, these blocks are processed in the exact reverse order. First, the received signal is demodulated, *i.e.* the noisy analog received signal is converted to a sequence of bits. These bits are then passed through a channel decoder, which leverages the redundant bits to reduce the effective errors in the data bits. If channel encryption is used at the transmitter, then the signal is passed through an equivalent decryptor. Finally, the output bits from the channel decoder (or decryptor) is processed in a source decoder which converts it back into the original signal that is desired to be transmitted. While in legacy systems, *hard* decisions about the detected bits are sent to the channel decoder, the use of *soft* information about the bits (e.g. bit probabilities) was later shown to improve performance significantly [33]. With the increased capabilities of processing elements, sophisticated receiver techniques have also been proposed. For instance, the use of iterative detection and decoding using soft information is proposed in [50]. In such iterative processing, the output of the decoder is fed back to the bit detection block to reduce the overall bit error rate.

The separation of the source and channel encoding as indicated in Figure 2.2 has certain advantages. The design of the source encoder can be optimized to the characteristics of the source, independent of the channel over which it is transmitted. Similarly, the channel encoder can be optimized for the channel of interest, independent of the source that is transmitted. One potential disadvantage of this separation principle could be the loss in optimality of performance as compared to a system with a single joint source-channel encoder. Fortunately, one of the results in Shannon's pioneering work [39] is that the separation of the source and channel encoding does not entail any loss in asymptotic performance. As a result, the source coding and channel coding research developed in parallel for many years without many significant interactions. Note, however, that Shannon's separation principle only proves optimality under certain assumptions including processing over very large data lengths, which could significantly increase the delay, memory, and computational complexity requirements [46]. For a system with finite resources (memory and processing power) it turns out that joint source-channel coders do outperform separable source and channel coders [20, 35]. Moreover, Shannon's separation principle does not hold in multiuser settings [22]. Further details on source coding and joint source-channel coding is available in [37, 40].

Since the focus of this chapter is on wireless systems, we discuss the basics of digital modulation in Section 2.2. We introduce the Shannon theoretic concept of channel capacity in Section 2.3. The channel capacity is defined as the maximum rate at which data can be *reliably* transmitted over the channel. The channel capacity concept however does not provide practical coding schemes that achieve this capacity. Practical code design and analysis is covered in Section 2.4. We then focus attention on multiuser communication systems in Section 2.5. Specifically, in Section 2.5.1 we consider random access methods enabled by algorithms at the data link layer. Common wideband multiuser methods at the physical layer, including direct sequence (DS) code division multiple access (CDMA), frequency hop-

ping (FH) CDMA, and orthogonal frequency division multiplexing (OFDM), are covered in Sections 2.5.2.1, 2.5.2.2, and 2.5.2.3, respectively. Finally, we discuss several advanced wireless transceiver techniques in Section 2.6.

Due to the wide range of topics involved in the design of wireless systems, in this chapter, we discuss only the key results and the underlying intuition. Detailed mathematical proofs and derivations of the main results are omitted due to space constraints; relevant references are provided as appropriate. To stay in focus we only consider the transmission of digital data. As and when possible, we provide performance metrics that can be easily used by network designers to ensure that desired quality of service guarantees are provided by their designs.

2.2 Digital Modulation in Single User Point-to-point Communication

Consider the transmission of digital data at rate R bits per second. Let T_s denote the symbol period in which M bits are transmitted. Then, $R = M/T_s$. The information sequence to be transmitted is denoted by a sequence of real symbols $\{I_n\}$. This sequence could be the output of the channel coding block shown in Figure 2.2. A traditional modulation scheme consists of transmitting one of 2^M different signals every T_s seconds. In orthogonal modulation the signal $s(t)$ is given as

$$s(t) = \sum_n I_n q_n(t), \quad (2.1)$$

where $q_n(t)$ form a set of orthonormal pulses, *i.e.* $\int q_i(t)q_j(t)dt = \delta_{i-j}$, where δ_m is the Kronecker delta function which is defined as follows: $\delta_0 = 1$, and $\delta_m = 0, m \neq 0$.

In many traditional modulation schemes these orthonormal waveforms are generated as time-shifts of a basic pulse $p(t)$. Examples of $p(t)$ include the raised cosine and Gaussian pulse shapes. For example, in pulse-amplitude modulation (PAM), the transmit signal $s(t)$ equals

$$s(t) = \sum_n I_n p(t - nT_s). \quad (2.2)$$

In (2.2) for a 2-PAM system, $I_n \in \{-A, +A\}$, where the amplitude A is selected to meet average and peak power constraints at the transmitter. Similarly, for a 4-PAM system $I_n \in \{-3A, -A, +A, +3A\}$.

Consider the transmission of data through an ideal additive white Gaussian noise (AWGN) channel, in which the received signal, $r(t)$ is given by,

$$r(t) = s(t) + n(t), \quad (2.3)$$

where input signal $s(t)$ has an average power constraint P and the noise $n(t)$ has a constant power spectral density of $N_0/2$. The additive noise channel is one of the simplest and most widely analyzed channel. In practical systems, the noise at the

receiver is a combination of several independent effects. Hence, using the central limit theorem (CLT) [31], the noise distribution is well approximated by a Gaussian density. Realistic wireless channel models are more sophisticated and are discussed in Chapter 3 and also in several references [29, 36].

At the receiver, the signal $r(t)$ given in (2.3) is first passed through a correlator to obtain

$$y_n = I_n \int_{T_s} p(t - nT_s) p(t - mT_s) dt + \int_{T_s} n(t) \cdot p(t - mT_s) dt. \quad (2.4)$$

It is known that the output of the correlator is a sufficient statistic for bit detection, *i.e.* all information about I_n that is contained in $y(t)$ is preserved in y_n . The orthonormality condition requires that the choice of the pulse shape should satisfy:

$$\int p(t - nT_s) p(t - mT_s) dt = \delta_{n-m}. \quad (2.5)$$

Under this condition, the output $y_n = I_n + z_n$, where it can be easily shown that z_n is a zero mean, Gaussian random variable with variance $N_0/2$. Under this condition, the output also does not have any inter-symbol interference.

The detection process involves making a decision on which symbol was transmitted given y_n is received. The detector design depends on the performance metric that one desires to optimize. For instance, the detector that minimizes the probability of error in the detection process, is the maximum a posteriori probability (MAP) detector. The MAP detector selects the \hat{I}_n that maximizes $p(\hat{I}_n/y_n) = \frac{p(y_n/\hat{I}_n)p(\hat{I}_n)}{p(y_n)}$. If all possible symbols I_n are *a priori* equally likely, then the MAP detector is the same as the maximum likelihood (ML) receiver. The ML receiver maximizes $p(y_n/\hat{I}_n)$. Since the effective noise z_n is Gaussian, the MAP receiver also reduces to the minimum Euclidean receiver, *i.e.*, the constellation point that is the closest to the received signal is selected as an estimate for the transmitted signal. The decision region corresponding to this Euclidean distance based receiver is simple to derive and is shown in Figure 2.3 for two different modulation schemes.

The probability of error in this case can be computed as the tail probability of the Gaussian noise variable. For instance, the conditional probability of symbol error given symbol “10” is transmitted can be evaluated as the area under the tails of the Gaussian probability density function (pdf) as shown in Figure 2.3. The width of the Gaussian pdf equals the variance of the noise in the system. The overall probability of symbol error can be calculated by averaging over all possible transmit symbols. This probability of symbol error for M -ary PAM is given by

$$P_e \approx 2Q \left(\sqrt{\frac{6 \log_2(M) SNR_{bit}}{M^2 - 1}} \right), \quad (2.6)$$

where, $SNR_{bit} = \frac{E_b}{N_0}$ is the SNR per bit and Q is the tail probability of the standard normal distribution [53]. The average bit energy $E_b = P_{avg} T_s$ where P_{avg} is the average power of the transmit symbols. A plot of this error probability as a function

of SNR is given in Figure 2.4 for different values of M . Clearly, we can see that for a given SNR the error rate is higher for larger M . Also, for a given M , the error probability decreases with increasing SNR. In the presence of multiple users and other interference sources, the probability of error would depend on the signal to interference ratio (SIR) instead of just the SNR. The error performance can be further decreased at a given SNR using error control codes as discussed in Section 2.4. The probability of bit error can be evaluated from the probability of symbol error as follows. Typically, bits are assigned to symbols using a process known as *gray coding* in which neighboring symbols differ in only 1 bit as shown in Figure 2.3. With such a bit assignment process, a symbol error will cause only a single bit error with high probability. Hence the probability of bit error $P_{bit-error} \approx \frac{1}{\log_2(M)} P_e$.

The transmitted signal in the above case is actually a signal in baseband. Typically, this signal is modulated up to a pass-band frequency, f_c before transmission. This pass band signal could be obtained as $s(t)e^{-j2\pi f_c t}$ which is however, a complex signal. To make the transmitted signal real, we can add its complex conjugate to obtain $s_p(t) = s(t)e^{-j2\pi f_c t} + s(t)e^{j2\pi f_c t}$.

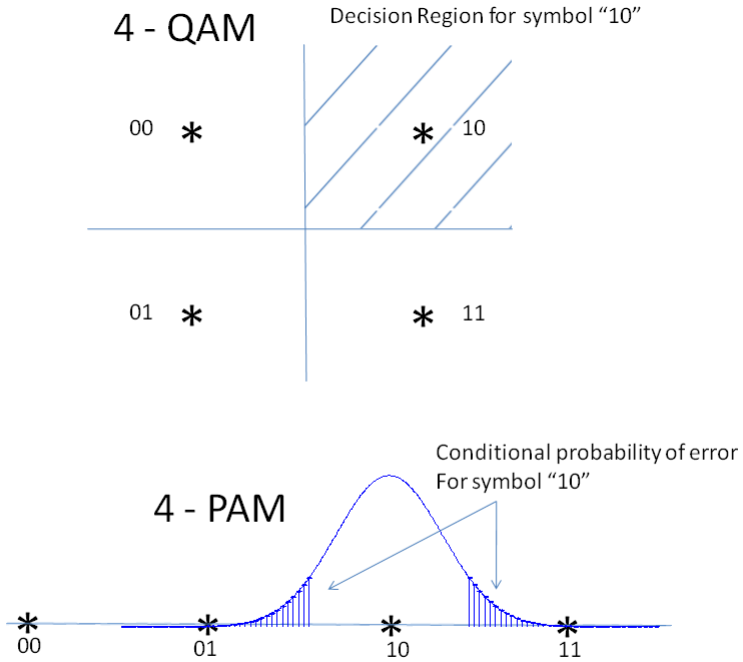


Fig. 2.3 Euclidean distance based detection and error calculation for various modulation schemes. The figure also demonstrates the concept of gray coding, in which neighboring symbols only differ in 1 bit.

The bandwidth of the signal is defined as the range of positive frequencies that are transmitted. It can easily be seen that the bandwidth of $s_p(t)$ equals twice the bandwidth of $s(t)$; this process is referred to as double sideband modulation. Clearly, this process is wasteful in bandwidth.

One way to use the bandwidth more efficiently is to use a complex signal constellation $u(t)$ instead of a real $s(t)$. Thus the transmitted passband signal is now $u(t)e^{-j2\pi f_c t} + u^*(t)e^{j2\pi f_c t}$ where $u^*(t)$ is the complex conjugate of $u(t)$.

Examples of such complex constellation are the quadrature amplitude modulation (QAM) signals given in Figure 2.5. The error performance of QAM signals can be derived in a manner similar to that for PAM signals and is given by [34]

$$P_e \leq 4Q \left(\sqrt{\frac{3 \log_2(M) SNR_{bit}}{M-1}} \right). \quad (2.7)$$

It turns out that for high SNR and large values of M this upper bound is quite tight. The probability of error versus SNR for QAM is given in Figure 2.4.

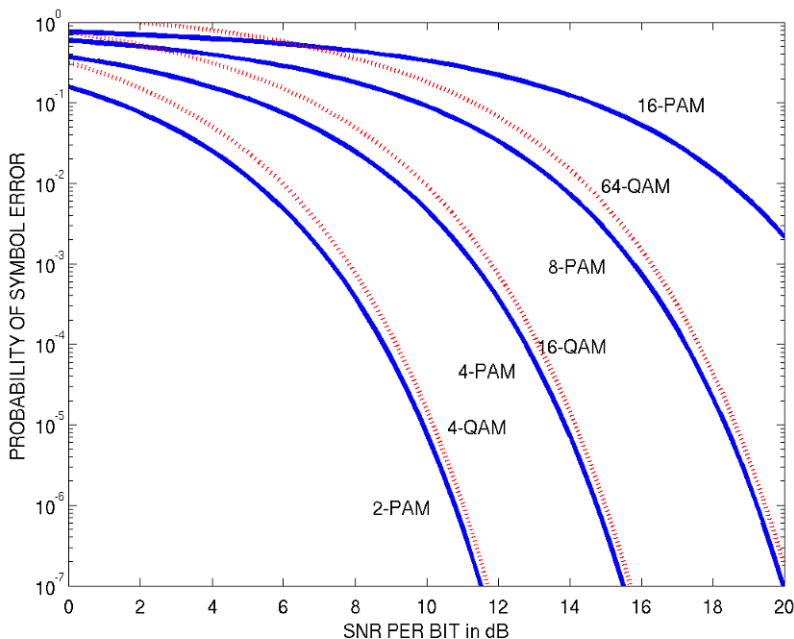


Fig. 2.4 Probability of symbol error with specific PAM and QAM schemes.

2.3 Information Theoretic Capacity and Coding

Consider the discrete time additive Gaussian noise channel given by

$$y_n = x_n + z_n \quad (2.8)$$

where, transmit signal x_n has power constraint P and Gaussian noise z_n has 0 mean and variance of σ^2 . A fundamental question that arises is to find the maximum rate at which data can be reliably transmitted over this channel. This question is answered in an emphatic manner by Shannon's landmark paper [39]. For this ideal discrete time real AWGN channel the Shannon capacity [39] is given by

$$C_{ideal} = 0.5 \log(1 + SNR) \text{ bits per channel use} \quad (2.9)$$

The interpretation of this capacity is that for transmission at any rate $R > C_{ideal}$, the probability of error in decoding that message at the receiver is necessarily bounded away from 0. Shannon also showed that information can be transmitted at any rate $R < C_{ideal}$ with arbitrarily small probability of error.

This capacity can be achieved by using a *random* Gaussian codebook as follows. To transmit R bits per use of a channel, a codebook is constructed containing $M = 2^{nR}$ codewords that each spans n symbols. Each symbol in the codebook is selected as an independent Gaussian random variable with zero mean and variance N . This codebook is revealed to both the transmitter and receiver. The codeword to be transmitted is selected at random (with uniform density) based on nR information bits. The n symbols of the selected codeword are transmitted in n uses of the channel. An example of such a codebook is shown in [Figure 2.6](#).

At the receiver, decoding is done using a typical set decoder [14, 54]. A typical set for a sequence of independent realizations of a random variable is the set of all sequences for which the empirical entropy is close to the true entropy of the random variable. In this decoding method, the receiver compares the received sequence with all possible 2^{nR} codewords to see which pairs of the received sequence and codewords belong to the *jointly typical set*. If the transmitted codeword is the one and only one codeword that is jointly typical with the received sequence, then the decoder can “correctly” decode what was sent. If the transmitted codeword sequence is not jointly typical with the received sequence or if more than one codeword is jointly typical with the received sequence, the decoder makes an error. It can be proved that for any desired upper bound on error probability ϵ , there exists a sequence of codes that achieve error below ϵ for asymptotically large n .

The typical set decoding process has exponential complexity which makes it impractical to implement in real systems. Real systems use some low complexity decoders based typically on the maximum likelihood (ML) criterion or maximum a posteriori probability (MAP) criterion as discussed in Section 2.4. Further, rather than using *random* codes, practical systems use codes with structure. The fundamental challenge in code design is to find codes with *just enough* structure to enable

design of a low complexity decoder and *just enough* randomness to ensure good performance.

A plot of the capacity versus SNR is given in Figure 2.7. As expected, the capacity shows a linear increase with SNR for very small SNR values and a logarithmic increase for higher SNR's. Also shown in the figure is the achievable rate for various PAM schemes. This achievable rate is calculated as the mutual information between an equiprobable PAM constellation input to the channel and the output of the channel [16]. As expected the rate for the M -PAM is upper bounded by $\log_2(M)$.

In addition to the complexity issue mentioned before, another challenge in using the random coding formulation in practical systems is that it does not give any indication of the size of codebook required to achieve a desired error probability. To provide some theoretical guidance on the performance of the best coding scheme for a given length, the theory of error exponents has been developed [18]. For a given code of length n , the error exponent is the logarithm of the error probability. Specifically, for an ensemble of random codes, it can be shown that the average error probability is upper bounded as,

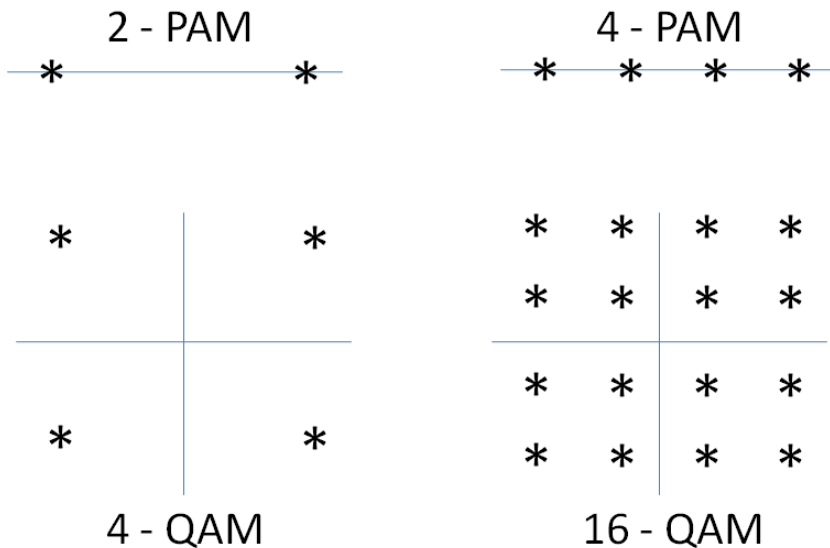


Fig. 2.5 Different digital modulation schemes

$$P_{\text{error}} \leq e^{-nE_{\text{random}}(R)} \quad (2.10)$$

where, $E_{\text{random}}(R)$ is the random coding exponent and R is the rate of transmission. This random coding exponent can be calculated based on the channel characteristics. For the Gaussian noise channel,

$$E_{\text{random}}(R) = 1 - \beta + \frac{SNR}{2} + 0.5 \log\left(\beta - \frac{SNR}{2}\right) + 0.5 \log(\beta) - R, \quad (2.11)$$

where $\beta = 0.5\left[1 + \frac{SNR}{2} + \sqrt{1 + \frac{SNR^2}{4}}\right]$.

This expression for the error exponent is valid for transmission rates $R < 0.5 \log\left[0.5 + \frac{SNR}{4} + 0.5 \sqrt{1 + \frac{SNR^2}{4}}\right]$. For rates above this threshold and below the capacity, the expression for the error exponent is slightly different and is given in [18]. Further, a lower bound on the decoding error probability for any code over a particular channel is given by the sphere packing error exponent, which can be computed in a manner similar to the random coding error exponent [18].

	Symbol 1	Symbol 2	...	Symbol n
Codeword 1	C ₁₁	C ₁₂	...	C _{1n}
⋮		⋮		
Codeword k	C _{k1}	C _{k2}	...	C _{kn}
⋮		⋮		
Codeword M	C _{M1}	C _{M2}	...	C _{Mn}

Fig. 2.6 Illustration of a random codebook used to prove achievability of Shannon's channel capacity.

Similar to the discrete time channel, for a continuous time additive white Gaussian noise channel with bandwidth W the capacity can be calculated as

$$C_{AWGN} = W \log(1 + P/N_0W) \text{ bits/sec} \quad (2.12)$$

Further, the capacity of wireless fading channels has been well investigated under several conditions [10]. The concept of channel capacity has also been extended to multiuser channels. We illustrate the idea in 2 different settings.

1. Uplink Channel: Consider a simple uplink channel in which multiple users are communicating with a single base station. Again, consider the simple Gaussian noise channel in which the received signal $y_{bs}(n)$ at the base station is given by

$$y_{bs}(n) = \sum_{i=1}^K x_i(n) + z_{bs}(n) \quad (2.13)$$

where $x_i(n)$ is the signal transmitted by user i and z_{bs} is the effective noise at the base station receiver, which is modeled as having a Gaussian pdf. Similar to the capacity of a single user channel, in this case, the capacity region is defined as the set of all K -tuples of achievable rates for the various users.

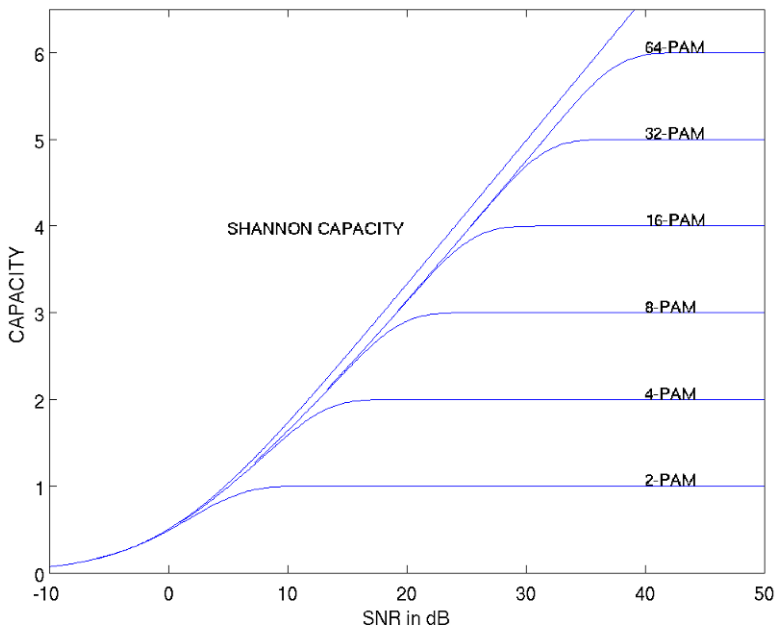


Fig. 2.7 Capacity of AWGN channel and maximum possible transmission rates with various modulation schemes.

For instance, for the $K = 2$ user channel, this capacity is given by the convex hull of rate pairs (R_1, R_2) satisfying the following inequalities:

$$\begin{aligned} R_1 &\leq 0.5 \log \left(1 + \frac{P_1}{N} \right) \\ R_2 &\leq 0.5 \log \left(1 + \frac{P_2}{N} \right) \\ R_1 + R_2 &\leq 0.5 \log \left(1 + \frac{P_1 + P_2}{N} \right) \end{aligned} \quad (2.14)$$

where P_i is the average power constraint at transmitter i and N is the variance of the noise at the receiver. This capacity region is illustrated in Figure 2.8. The interpretation of this capacity is that any rate pair within the boundary of the pentagon capacity region can be simultaneously achieved for both users with any desired error probability. Further, any rate pair outside the boundary cannot be simultaneously supported for both users and still ensure low probability of error. Also shown in the figure is the achievable rate with *naive* TDMA, in which user i exclusively access the channel for a fraction α_i of the time, where $\sum_i \alpha_i = 1$. The achievable rates with FDMA is also shown in the figure. It can be seen that FDMA achieves the boundary of the capacity region at one point; this point is the case when the whole bandwidth is allocated to the different users in proportion to their powers. Recognize that in a TDMA scheme, user i is only active for a portion α_i of the time, where $\sum_{k=1}^K \alpha_k = 1$. Hence, if we adjust the transmission power of user i to $\frac{P_i}{\alpha_i}$ its performance will improve. The performance of this variation of TDMA is identical to the FDMA scheme shown in Figure 2.8.

2. Downlink Channel: Now, consider a downlink channel in which a base station is transmitting to multiple users. Let $x_{bs}(n)$ be the signal transmitted by the base station. The received signal $y_i(n)$ at the i^{th} user is given by

$$y_i(n) = x_{bs}(n) + z_i(n) \quad (2.15)$$

where $z_i(n)$ is the noise at the receiver and is modeled as being Gaussian with 0 mean and variance of N_i . Without loss in generality, consider the case that $N_i < N_j \forall i < j$. Such a channel is denoted as a degraded broadcast channel [14]. One simple interpretation of the degraded Gaussian broadcast channel is that any signal that can be decoded at user j can also be decoded at user i , since user i has lower noise on the average than user j . For simplicity, consider the 2-user channel. The capacity in this case is given by the convex hull of rate pairs (R_1, R_2) satisfying the following inequalities:

$$\begin{aligned} R_1 &\leq 0.5 \log \left(1 + \frac{\alpha P}{N_1} \right) \\ R_2 &\leq 0.5 \log \left(1 + \frac{(1 - \alpha)P}{N_2 + \alpha P} \right) \end{aligned} \quad (2.16)$$

where P is the power constraint at the base station. This capacity region is illustrated in [Figure 2.9](#). Again, the achievable rate with FDMA is also shown in the figure.

These theoretical capacity bounds are very useful for network designers in planning and resource allocation issues.

2.4 Channel Coding

Channel codes, also referred to as forward error correction (FEC) codes, are used to reduce the effective bit error rate when data is sent over a lossy medium such as a wireless channel. There are two main categories of FEC codes - block codes and convolutional codes. We now briefly discuss the properties of these codes. For simplicity, we only discuss binary codes in this chapter. Codes over other alphabets are discussed in several references such as [26, 51].

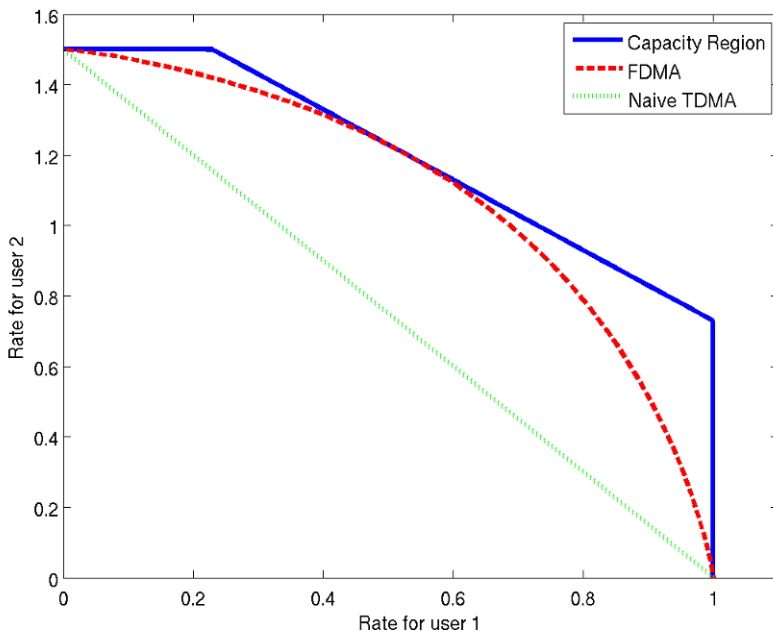


Fig. 2.8 Illustration of capacity of an uplink channel

2.4.1 Block Codes

In a simple block code, the information data stream is parsed into non-overlapping groups of a fixed size, say k bits. Each group of k bits is then mapped to a group of $n > k$ bits independent of the bits in the other groups. This mapping into a higher dimensional space essentially defines the code and is responsible for the error correction and detection capabilities of the code. Since each group of n bits are essentially independent of other groups, this type of coding is sometimes referred to as memoryless. The rate, r , of the channel code is defined as $r = \frac{k}{n}$. The random coding described in Section 2.3 and also depicted in Figure 2.6 is a block code (over a non-binary alphabet) with $k = \log_2(M)$ and n coded symbols.

The channel encoding can be achieved using a generator matrix \mathbf{G} of size $k \times n$ which uniquely defines the code. Let \mathbf{x} and \mathbf{y} denote, respectively, the $1 \times k$ and $1 \times n$ vectors of the information bits and coded bits. Then $\mathbf{y} = \mathbf{x}\mathbf{G}$. For each code, there is also a unique parity check matrix \mathbf{H} of size $n - k \times n$ such that $\mathbf{y}\mathbf{H}^T = 0$, for all valid codewords \mathbf{y} , which also implies that $\mathbf{G}\mathbf{H}^T = 0$.

There are several decoding algorithms for block codes depending on the type of code and the channel over which the code is used. These algorithms offer the system designer several options to trade-off the complexity of the decoding process

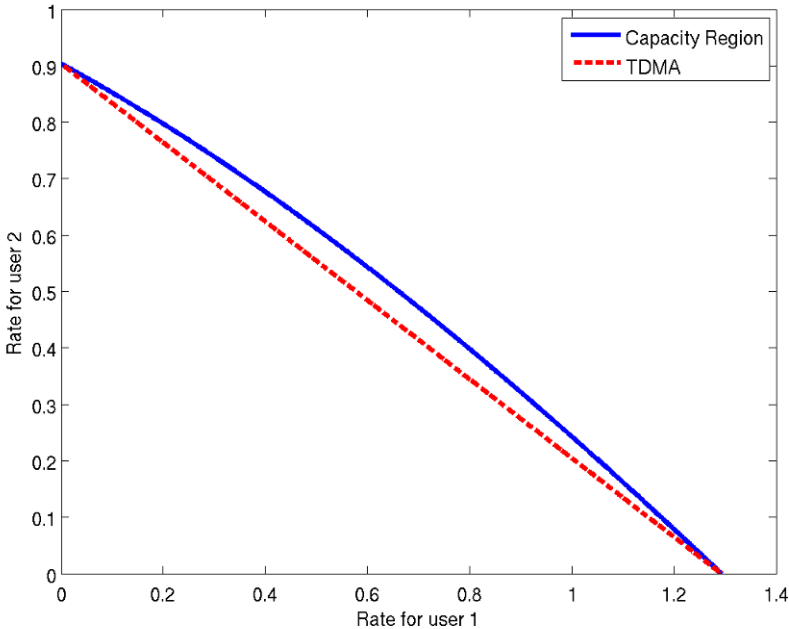


Fig. 2.9 Illustration of capacity of a downlink channel

with the performance achieved. For example, consider a simple binary symmetric channel (BSC) which randomly flips each of the coded bits with a certain probability. Hence, the received vector $\mathbf{b} = \mathbf{y} + \mathbf{e}$, where \mathbf{e} is a binary error vector in which a value of 1 indicates that the corresponding bit is flipped.

The maximum likelihood decoding in this case selects the codeword that is closest to \mathbf{b} in terms of the Hamming distance,

$$\hat{\mathbf{y}} = \arg \min d_H(\mathbf{c}, \mathbf{b}) \quad (2.17)$$

This ML decoding can be implemented using a standard array decoder [27]. The standard array decoder is similar to a look up table based decoder, with a separate entry for each possible received vector \mathbf{b} . The standard array gives the ML estimate of the error vector for each of the 2^n possible received vectors. Thus, the decoder simply looks for the error corresponding to the vector \mathbf{b} and then subtracts the error vector from \mathbf{b} to obtain the ML estimate of the transmitted codeword.

A major problem with this type of decoding is that the memory requirements for storing the standard array becomes prohibitively large for even reasonable size codebooks. Further, the search procedure to find the correct element in the array is also computationally intensive. To reduce these complexities, another decoding process called syndrome decoding is sometimes used. The syndrome, \mathbf{s} of vector \mathbf{b} is defined as $\mathbf{s} = \mathbf{b}\mathbf{H}^T = (\mathbf{y} + \mathbf{e})\mathbf{H}^T$. Since $\mathbf{y}\mathbf{H}^T = 0$ it turns out that the value of the syndrome only depends on the error vector. Thus, in syndrome decoding a table of all 2^{n-k} possible syndromes and the error vectors associated with them are stored. Decoding is achieved by simply computing the syndrome for the vector \mathbf{b} and then looking up the associated error vector from the table. It turns out that in many cases the complexity and memory requirements of syndrome decoding are still prohibitively high. In a typical minimum distance decoder, which is often used over Gaussian wireless channels, the decoder calculates the distance of the received signal to each of the 2^k possible transmit sequences and selects the one with the minimum distance. Decoders with lower memory and complexity requirements have been derived which effectively exploit the structure in the codes [23].

The error correction and detection performance of a channel code depends on the *distance spectrum* of the code. Essentially, the distance spectrum is an indication of the separation between the 2^k codewords of length n . The calculation of this distance spectrum is, however, not an easy task in many scenarios. Fortunately, the performance of the code can be approximated using a single quantity referred to as minimum distance, typically denoted by d . This minimum distance d equals the minimum of the Hamming distance between all pairs of codewords. The Hamming distance between two binary strings is the number of positions in which the two strings differ. For the special case of linear block codes, the minimum distance of the code also equals the minimum (non-zero) weight of the code.

The weight distribution of codes is conveniently represented in polynomial form as $B(z) = \sum_{i=0}^n B_i z^i$, where B_i is the number of codewords with Hamming weight i . For example, the weight distribution of Hamming codes is given by

$$B(z) = \frac{1}{n+1} \left[(1+z)^n + n(1-z)(1-z^2)^{(n-1)/2} \right] \quad (2.18)$$

The calculation of this minimum distance is a non-trivial task in many cases. There are several options available to a system designer in selecting the appropriate code.

1. Use a code from an existing family of codes for which the minimum distance has already been calculated. Examples of such codes are the Hamming codes which have $k = 2^m - m - 1$, $n = 2^m - 1$ for any $m \geq 3$. It is well known that the minimum distance for all Hamming codes is $d = 3$. The generator matrix \mathbf{G} for a Hamming code, has a specific structure. The columns of \mathbf{G} are the $2^m - 1$ binary codewords of length k excluding the all zero codeword. The order of these columns does not affect the properties of the code.
2. Use bounds to approximate the Hamming distance. For instance, for a (n, k, d) code, the Hamming bound gives the minimum amount of redundancy $(n - k)$ required as

$$n - k \geq \log_2 V(n, \lfloor \frac{d-1}{2} \rfloor) \quad (2.19)$$

where $V(n, t) = \sum_{j=0}^t C_j^n$ is the number of binary vectors of length n that are at most Hamming distance t from a given vector.

3. Evaluate the minimum distance using numerical simulations.

There is a basic trade-off between code rate r and minimum distance d . The smaller the code rate, the larger is the minimum distance and vice-versa. Given a code with a certain minimum distance d , a simple bound on the probability of decoding error for a bounded-distance based decoder is given by

$$P_{\text{decoding-error}} \leq \sum_{j=t}^n \binom{n}{j} \alpha^j (1 - \alpha)^{n-j} \quad (2.20)$$

where α is the crossover probability of the BSC and $t = \lfloor \frac{d-1}{2} \rfloor$. The probability of bit error can be easily lower and upper bounded as $\frac{1}{k} P_{\text{decoding-error}} \leq P_{\text{bit-error}} \leq P_{\text{decoding-error}}$.

2.4.2 Convolutional Codes

In a convolutional code, similar to a block code, the information bits are parsed into non-overlapping groups of k bits and each group is mapped to $n > k$ bits. However, unlike block codes, the mapping is dependent on the bits in the prior groups which introduces memory in the system. The n coded bits thus depends both on the input k bits as well as the state of the system. The constraint length, K , of the code is an indicator of the effective memory in the system.

One advantage of a convolutional code is that the encoder can be implemented as a digital filter, using shift registers and adders. An example of such an encoder is shown in Figure 2.10. While there are several analytical representations for convolutional codes, for simplicity we only present a couple of graphical representations.

Consider for instance the rate $1/2$ code shown in Figure 2.10. In this case, for each input bit x_i , the encoder produces two output bits, $y_i(1) = x_i + x_{i-2}$ and $y_i(2) = x_i + x_{i-1}$. The state diagram corresponding to this code is also shown in Figure 2.10. The advantage of this approach is that it clearly represents the relationship between the current state and next state as a function of the input and the output bits. One disadvantage of the state diagram approach is that the temporal sequence of coded bits is difficult to represent with this approach. For this reason, the equivalent trellis is used and is shown in Figure 2.11.

There are two main types of decoders for convolutional codes. The first type of decoder is the maximum likelihood sequence decoder which is typically implemented using a Viterbi algorithm [48]. The second decoder is the maximum a posteriori probability decoder which is implemented using the BCJR algorithm [3]. The Viterbi algorithm has much lower complexity than the BCJR algorithm whereas the performance of both are nearly similar. The description of these algorithms are fairly complex and hence relegated to the appropriate references [33]. It turns out that the trellis description shown in Figure 2.11 is useful in describing the Viterbi decoding algorithm. The enumeration of the performance of convolutional codes has received significant attention in the literature [9, 27] and is not duplicated here for conciseness.

The performance of convolutional codes over an AWGN channel is shown in Figure 2.12 for three different codes with rates $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$. The constraint length of these codes were set to $K = 3$. Clearly, we can see that the effective bit error rate reduces as the rate of the code decreases. Also, in this case, the performance of the soft decoder is about 2dB better than the hard decoder.

Convolutional codes are also the basic elements of the hugely popular Turbo codes [5]. The original Turbo codes are essentially a concatenation of two convolutional codes acting on different pseudo-random permutations of the input bits. The performance of Turbo codes is close to the best possible theoretical limits.

The choice of whether to use block codes or convolutional codes in a specific scenario has been a subject of extensive debate between code designers. Some coding theorists believe that convolutional codes offer superior performance over block codes for the same complexity. Many real systems use more sophisticated coding methods such as trellis coded modulation [45] and concatenated codes [15].

2.5 Multiuser Communication

One of the features of the wireless medium is its broadcast nature which makes simultaneous use of the spectrum by multiple users in a certain region challenging. This multiple access is accomplished using techniques at physical and/or data link

layers. Two of the simplest and intuitive methods of multiple access is to assign the different users non-overlapping time-slots and/or frequency bands. Such schemes are referred to as time-division multiple access (TDMA) and frequency division multiple access (FDMA). The popular GSM standard uses a combination of time and frequency division multiplexing to support multiple users. For instance, the basic GSM standard specifies different groups of carrier frequencies for each base station. Within each base station, each carrier frequency is shared among 7 different users in a time-division manner.

2.5.1 Data Link Layer Random Access Methods

Traditional data networks such as Ethernet which primarily support packet data traffic (which is bursty in nature) tend to enable random access using a data link layer protocol. The basic scenario being considered in the design of these protocols is as

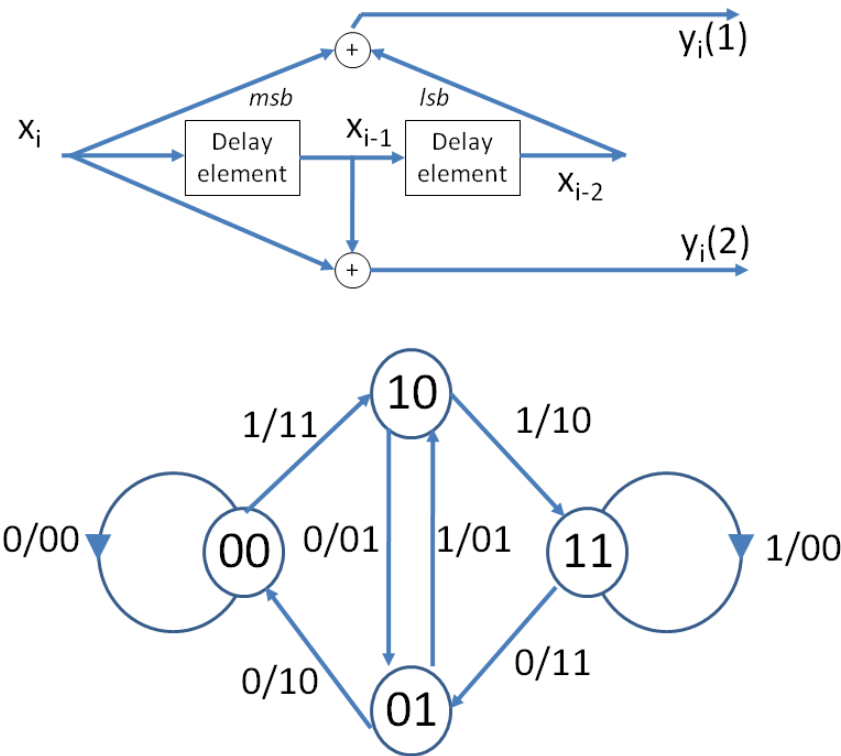


Fig. 2.10 Digital filter implementation of convolutional encoding and state diagram representation of convolutional codes

follows. Let there be N distributed nodes that are sharing the wireless medium. Traffic arrivals at each node is modeled as a random process, such as a Poisson point process. Some basic assumptions are typically made to enable simple analysis of these protocols. For instance, if two or more nodes transmit a packet at the same time, a *collision* is said to occur and the receiver will be unable to accurately decode either of these packets.¹ It is also assumed that some form of instantaneous feedback mechanism exists from the receiver to the transmitter indicating whether a packet is received successfully or not. This feedback in turns enables the transmit node to retransmit packets that are not received correctly.

The earliest and simplest form of multiple access is the well known Aloha method [1]. In an Aloha system, users simply transmit their packets whenever they have data to be sent. Clearly, such a scheme might cause significant amount of collisions if multiple packets are sent simultaneously. Further, if a collision is detected, the nodes which are involved in the collision, delay their subsequent retransmissions by random amounts. This randomization will allow the multiple competing nodes to reduce their probability of collision in their retransmissions. It turns out that the maximum throughput of Aloha is severely limited. Specifically, if the wireless chan-

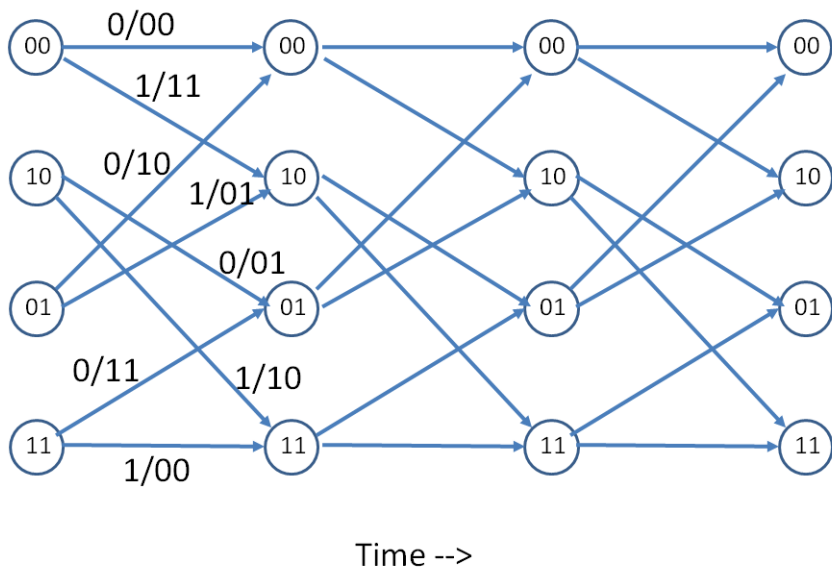


Fig. 2.11 Depiction of convolutional code using a trellis.

¹ This assumption has been modified in recent years to address the issue of *capture*, which is discussed later.

nel can support a peak data rate of R , then with random multiple access using the Aloha protocol, the maximum effective data rate is only $\frac{1}{2e}R$, or approximately 18% of data rate. However, for lightly loaded systems Aloha is a simple strategy that offers reasonable performance. An excellent historical perspective of Aloha written by the original inventor including recent application areas is given in [2].

Several variations of Aloha including slotted-Aloha have also been proposed which improve the performance of the basic scheme [7]. In a slotted Aloha system, the entire time is discretized into time-slots of fixed duration and nodes are allowed to transmit packets only at the beginning of a time-slot. Further, each node that has a packet to transmit will only transmit it with a finite probability q at the beginning of the next time-slot. These changes to Aloha protocol, reduces the number of collisions and increases the overall throughput. The maximum data rate of a slotted Aloha system can be shown to equal $\frac{1}{e}R$.

Many recent multiple access (MA) strategies are based on the idea of carrier sense multiple access (CSMA) [7, 25]. Several modifications of CSMA have been proposed in the literature. In a CSMA scheme, each node listens to the medium to see if the medium is free or if there exists any other signal in the medium. If the wireless medium is believed to be free then the node transmits its data after waiting for a random amount of time. If the medium is sensed as being busy, then

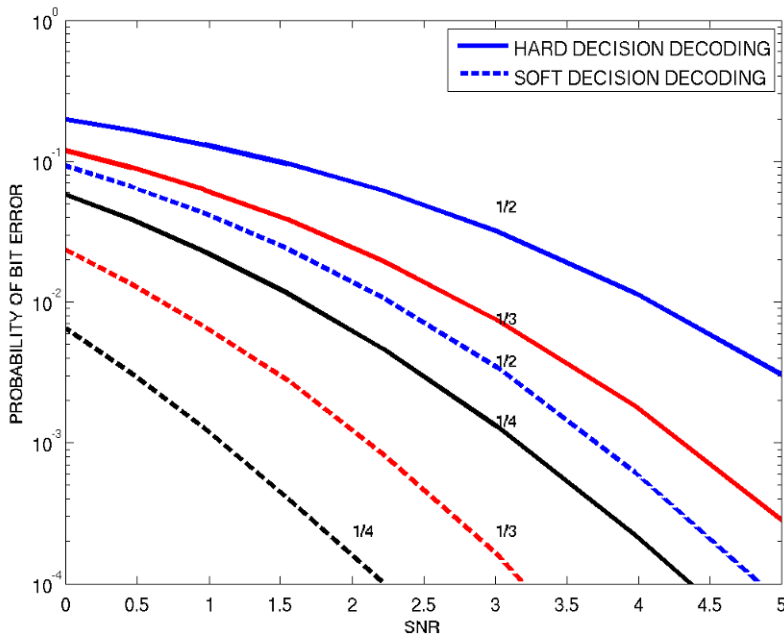


Fig. 2.12 Performance of hard decision and soft decision decoding of convolutional codes over an AWGN channel. The rate of the codes are $\frac{1}{2}$, $\frac{1}{3}$, and $\frac{1}{4}$.

the node backs off from transmission for a random amount of time and then senses the channel again. However, CSMA still suffers from collisions if multiple nodes are constantly monitoring a busy medium waiting for it to become free. As soon as the medium becomes free all these nodes might begin transmission simultaneously resulting in a packet collision.

While CSMA offers significant improvement in performance over Aloha type protocols, it suffers from the hidden node and exposed node problems. Consider the scenario shown in [Figure 2.13](#), where nodes *B* and *C* are trying to communicate with node *A*. If nodes *B* and *C* are not within range of each other, they cannot detect the presence of the other node's signal and might try to simultaneously communicate with *A*. These transmissions would likely result in a collision and node *A* may not be able to decode either of these signals. This problem is commonly referred to as the hidden node problem.

Now consider, the same scenario as shown in [Figure 2.13](#). If node *A* is trying to send data to node *B*, then node *C* is prevented from transmitting due to carrier sense aspect. However, it is possible for node *C* to simultaneously transmit data to a different node since its signal is not received at node *B* and hence will not cause any interference. This problem is called the exposed node problem.

A related problem that sometimes occurs in wireless networks is the issue of *capture*. For example, if signals from two different transmitters are received at a particular node at significantly different power levels, then the transmitter with the stronger signal will *capture* the medium. This capture will prevent the weaker transmitter from sending data and lead to severe unfairness in the resource allocation.

Virtual carrier sense is proposed as a solution to overcome the hidden node problem [19]. In virtual carrier sense, each node that wants to transmit a packet first sends a short request-to-send (RTS) packet to the intended destination. If the destination node receives this packet successfully and determines that there are no other active signals in its vicinity, it replies with a short clear-to-send (CTS) packet. These RTS and CTS packets also include a field indicating the amount of time that the medium is expected to be busy. All nodes that receive either the RTS or CTS packet will not attempt transmission during the period of time specified in these packets. The use of RTS/CTS mechanism can also solve the exposed node problem under certain scenarios; for instance, using the MACAW protocol [8]. The fairness issues associated with capture can also be solved using more sophisticated back-off mechanisms.

While the use of RTS/CTS coupled with a CSMA variant scheme reduces collisions significantly, it can add significant overhead especially if the packet size is small. The other problem is that in some scenarios nodes might back-off unnecessarily thereby reducing the number of simultaneous transmissions in the network and effectively reducing the network throughput. The popular 802.11 based standard uses a combination of CSMA with the RTS/CTS mechanism for multiple access.

2.5.2 Physical layer Multiuser Signaling Schemes

Traditional cellular networks which primarily support voice tend to also build multiple access capabilities in the physical layer signaling. Next generation wideband networks tend to use a variation of CDMA or OFDM, which are described next.

2.5.2.1 Direct Sequence Code Division Multiple Access DS-CDMA

In a direct sequence CDMA system, all users use the same frequency and time to communicate. User separation is achieved by modulating the bits of each user with a distinct spreading sequence. By a careful selection of these spreading sequences, it is possible to recover the bits of each user from a linear combination of their transmit sequences. In many systems, these spreading codes are generated to form an orthonormal set. Although we only discuss uplink communications in this section, CDMA can also be used in downlink channels. The CDMA2000 standard specified by the International Telecommunication Union (ITU) is based on DS-CDMA technology.

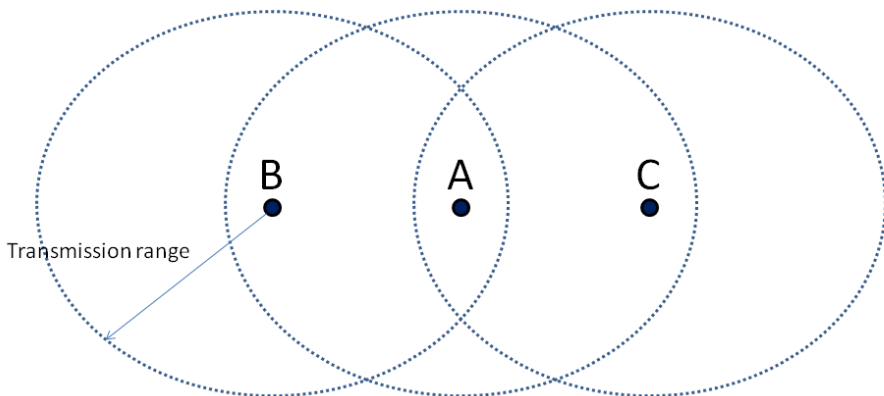


Fig. 2.13 Topology showing the hidden node problem. If both nodes *B* and *C* transmit at the same time, their signals collide at node *A*.

Consider an uplink CDMA system with K users. A simplified block diagram of the entire uplink CDMA system is shown in Figure 2.14. The transmit signal $x_k(t)$ for user k at time t is given by,

$$x_k(t) = \sqrt{E_k} \sum_{m=0}^{\infty} b_k^{(m)} \sum_{l=0}^{N_c-1} c_{k,l} p(t - mT_s - lT_c), \quad (2.21)$$

where, $b_k^{(m)}$ is the m^{th} bit of user k , E_k is the energy, $c_{k,l}$ is the spreading sequence and $p(t)$ is the normalized pulse of duration T_c . The chip period and symbol period are represented by T_c and T_s , respectively. The ratio $\frac{T_s}{T_c}$ is referred to as the spreading gain of the DS-CDMA system. The received signal, $y(t)$, at the base station equals,

$$y(t) = \sum_k h_k x_k(t) + n(t), \quad (2.22)$$

where, h_k is the channel between user k and the base-station and $n(t)$ is the additive noise at the base-station receiver. For simplicity, we have not shown the temporal or spectral variations of the channel h_k . Detailed models for h_k are discussed in Chapter 3.

This received signal, $y(t)$, is passed through a bank of parallel filters that are matched to the spreading codes of the various users. The output, w_k of the k^{th}

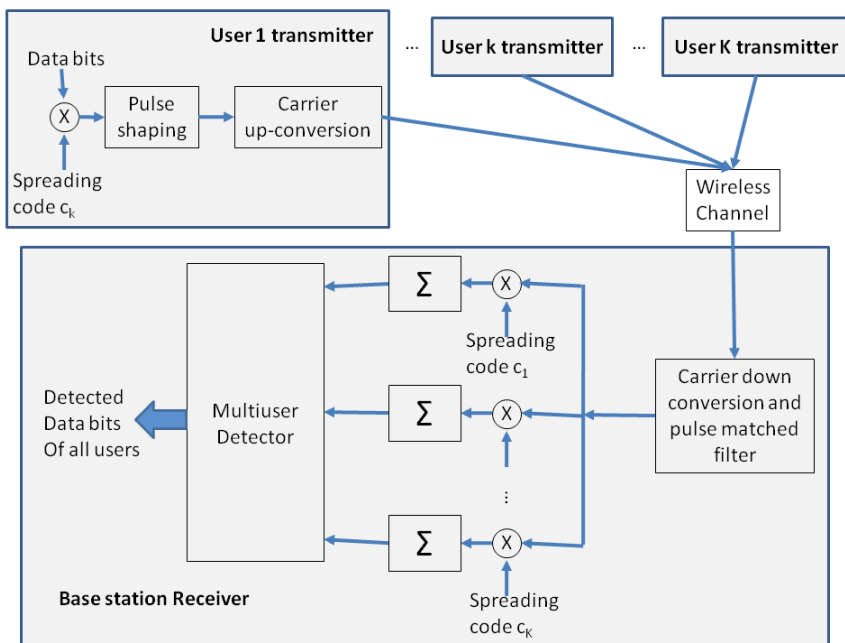


Fig. 2.14 Basic block diagram of an uplink DS-CDMA system.

matched filter for the m^{th} time period equals

$$w_k^{(m)} = \sqrt{E_k} b_k^{(m)} h_k \langle c_k, c_k \rangle + \sqrt{E_k} b_k^{(m)} \sum_{j \neq k} h_j \langle c_k, c_j \rangle + n_{eff} \quad (2.23)$$

where $\langle ., . \rangle$ represents the inner product of the two vectors. In the special case that the spreading codes for the various users are selected to belong to an orthonormal set (*i.e.* $\langle c_k, c_j \rangle = 0$ and $\langle c_k, c_k \rangle = 1$, for $k \neq j$), then (2.23) reduces to the form

$$w_k^{(m)} = \sqrt{E_k} h_k b_k^{(m)} + n_{eff} \quad (2.24)$$

Effectively, the output of the matched filter in (2.24) equals the scaled signal of user k corrupted by additive white Gaussian noise. The effect of the multiuser interference is completely eliminated due to the orthogonality of the codes selected.

In the general case of non-orthogonal codes, an equivalent discrete time model for the output of the matched filters for bit m of all users can be written in matrix form as

$$\mathbf{w}^{(k)} = \mathbf{R} \mathbf{A} \mathbf{b}^{(k)} + \mathbf{n} \quad (2.25)$$

where $\mathbf{w}^{(k)} = [w_1^{(k)}, w_2^{(k)}, \dots, w_K^{(k)}]$, $\mathbf{b} = [b_1^{(k)}, b_2^{(k)}, \dots, b_K^{(k)}]$ and $\mathbf{n} = [n_1, n_2, \dots, n_K]$. Matrix \mathbf{R} is the cross-correlation between the various spreading sequences and matrix \mathbf{A} is a diagonal matrix of the channel gains $\{h_k\}$.

Figure 2.15 illustrate the various signals in a DS-CDMA system. The figure shows the first 5 bits in a simple 2-user CDMA system. The bit sequences are modulated by spreading sequences of length 8. The received signal is the summation of the transmit signals of the 2 users plus additive noise. For simplicity, in this figure the attenuation in the channel is not considered and all users are assumed to have the same transmit power. The signals recovered at the output of the matched filter are also shown in the figure. In this example, the spreading sequences of the two users are orthogonal and since the users are also synchronized, there is no multiuser interference and the bits are only corrupted by noise. The detected bit sequences are also shown in the figure and clearly demonstrates effectiveness of the DS-CDMA approach.

The simple analysis above assumes a perfect time synchronization among the various users connected to the same base station receiver. In practical systems, such synchronization is difficult to achieve. Analysis of CDMA with asynchronous users is provided in [47]. The equivalent received signal can be written in a manner similar to (2.25) but with the matrices redefined to include the output of a vector of bits for each user.

In the general case of nonorthogonal codes and asynchronous user transmissions, computing the optimal receiver and quantifying its performance is a challenging task. The general principle behind multiuser detection is to simultaneously detect the signal of all the users from $\{w_1, w_2, \dots, w_K\}$. There are several well known multiuser detectors that offer a trade-off in performance and implementation complexity [47]. Typically, the performance of a MUD depends on the signal-to-interference (SIR) ratio before bit detection.

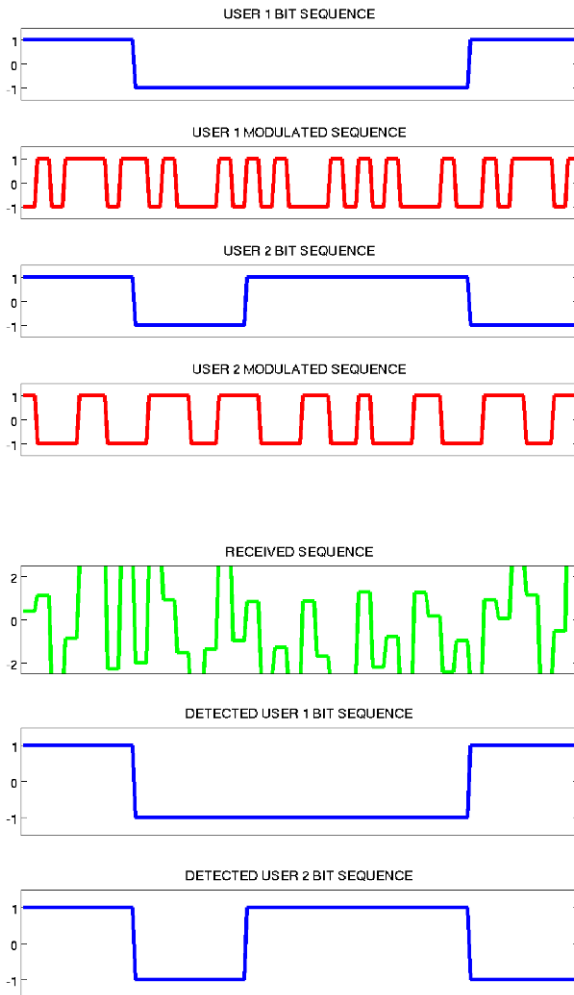


Fig. 2.15 Simple illustration of DS-CDMA signals. The bit sequences for users 1 and 2 are $[1 \ -1 \ -1 \ -1 \ 1]$ and $[1 \ -1 \ 1 \ 1 \ -1]$, respectively. The corresponding spreading codes are $[-1 \ 1 \ 1 \ -1 \ 1 \ 1 \ 1 \ -1 \ 1]$ and $[1 \ -1 \ -1 \ -1 \ 1 \ 1 \ -1 \ -1]$. In this case, the two spreading codes are orthogonal and hence the multiuser interference is completely eliminated at the receiver and all bits for both users are detected correctly.

In summary, the main advantages of CDMA are: i) efficient utilization of frequency spectrum with a reuse factor of 1 in cellular systems, which leads to increased system throughput [49], ii) efficient allocation of resources (spreading

codes) leads to better utilization in bursty traffic scenarios, and iii) better immunity against narrow band interference.

One of the main challenges with DS-CDMA is the near-far problem, *i.e.*, the received power for various users can vary by orders of magnitude depending on the relative distances of the users from the base station and the fading conditions. This large difference in received power will cause the signal of certain users to completely swamp out the signals of other users. The solution proposed to overcome this problem is to carefully control the transmit power of the users using a power control algorithm [52].

2.5.2.2 Frequency Hopping Code Division Multiple Access (FH-CDMA)

The basic idea in a frequency hopping CDMA system is to temporally vary the frequency used by each user in a pseudo random manner as illustrated in [Figure 2.16](#). Since each user selects the pattern in a pseudo random manner, the probability of more than one user selecting the same frequency in a given time-slot is low. Thus packet collisions are reduced. Depending on the rate at which users change the frequency, the hopping is termed as slow-hopping or fast hopping. In addition to providing multiple access capability, FH-CDMA also has the advantage that it provides frequency diversity and robustness to narrow band interference. For instance, if another system is using a narrow frequency band, then the user employing FH-CDMA will only be affected for a small fraction of the time when its hop-frequency matches the interference band; the transmission during the rest of the time-slots are not affected by the interference. Similarly, in a wideband wireless channel (formally defined in Chapter 3) the channel attenuation is a function of frequency. So, even if the channel encounters severe attenuation due to fading at a particular frequency, the other frequencies may be less attenuated. Since the FH-CDMA transmits over multiple time-slots and frequencies, the overall probability of successful transmission is increased.

One of the challenges in a FH-CDMA system is to ensure that the transmitter and intended receiver hop using the same pattern. The protocol required to ensure that these sequences match is fairly sophisticated. The popular Bluetooth standard [12] is based on FH-CDMA technology. The standard specifies an elaborate sequence of steps that nodes need to follow in order to discover and share a frequency hopping pattern based on the unique Bluetooth identity of each node and their system clocks. The standard also employs an adaptive frequency hopping protocol. In adaptive frequency hopping, the hopping pattern that each node uses can be varied with time depending on the interference encountered [21, 32].

2.5.2.3 Orthogonal Frequency Division Multiplexing (OFDM)

OFDM is essentially a combined method for modulating and multiplexing multiple signals onto a wideband channel. The primary objective of OFDM is an effi-

cient system to transmit information over parallel channels. Consider for instance a simple frequency division multiplexing (FDM) system in which the available bandwidth B Hz is divided into $m = B/b$ non-overlapping segments of width b Hz. A traditional modulation pulse and carrier can be used in each of these segments. However, in an OFDM system, the carrier frequency on each of these smaller segments (typically referred to as subcarriers) are selected in such a manner that they are orthogonal with each other as shown in Figure 2.17. As can be seen from the figure, at the center frequency of each subcarrier, the contributions from the other subcarriers is 0. This orthogonality is the key to the superior spectral efficiency of OFDM since it allows numerous subcarriers to be used within the given band without causing any interference between subcarriers.

There are several variations of OFDM [24] and it has become widely used in next generation wireless standards such as WiMAX. OFDM based signaling is also used in wire-line systems where it is frequently referred to as discrete multitone modulation (DMT). In this chapter, we only focus on a very basic OFDM system and provide the key insights into its performance.

Consider the basic blocks in an OFDM system as shown in Figure 2.18. The input data stream is first parsed into groups of N bits. Without loss in generality consider the transmission of one input vector \mathbf{X} of length N . This input is first passed through

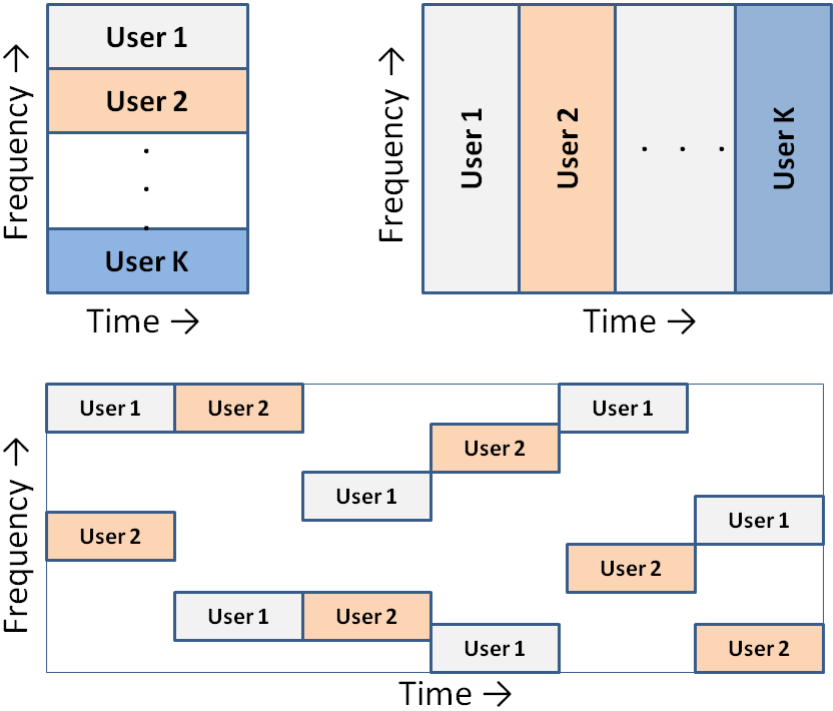


Fig. 2.16 Simple illustration of FDMA, TDMA and frequency hopping CDMA

a length N Inverse Discrete Fourier Transform (IDFT) resulting in output \mathbf{x} . A cyclic prefix which is a replica of the first c symbols is appended to the end of the symbol stream. This sequence is then transmitted over the channel.

Let the channel be represented by a linear, time-invariant (LTI) system with impulse response g_n . Consequently, the received signal y_n is given by the convolution of the input signal and the channel response, as

$$y_n = \sum_k g_k x_{n-k} + z_n \quad (2.26)$$

where z_n is the additive Gaussian noise. At the receiver the portion corresponding to the cyclic prefix is first removed. Then the resulting signal is passed through a Discrete Fourier Transform (DFT). The output, Y of the DFT block is given by

$$\begin{aligned} Y = DFT(y_n) &= DFT(g_n) \cdot DFT(y_n) + DFT(z_n) \\ &= h_n \cdot X_n + Z_n \end{aligned} \quad (2.27)$$

Effectively, the signal Y can be treated as the output of a set of parallel channels with gains h_n and noise Z_n . The properties of the effective discrete wideband channel (represented by g_n and h_n) are discussed in Chapter 3 of this book. As long as the length of the cyclic prefix c is greater than the spread of the channel, represented by the size of g_n , there is no intersymbol interference or interchannel interference. The use of the cyclic prefix effectively converts the linear convolution of the channel into a cyclic convolution.

It turns out that an OFDM system also offers the flexibility to alter the rate and power on each subcarrier separately. The advantage of such adaptation is that it helps achieve capacity of the wideband channel. One of the characteristics of a wideband channel is that the channel gain is different on each subcarrier, *i.e.* the h_n varies with n . It is well known that the optimal power allocation across subcarriers is based on what is commonly referred to as water-filling [14]. In a variation of OFDM, called orthogonal frequency division multiple access (OFDMA), multiple users are assigned non-overlapping groups of subcarriers within the entire band. The allocation of the subcarriers can be fixed or adaptively vary based on channel measurements. Such adaptive subcarrier allocation plays a critical role in achieving multiuser diversity in next generation wireless systems such as WiMAX.

In summary, the main advantages of an OFDM system are: i) low complexity implementation because of the use of Fast Fourier Transform (FFT) to compute the DFT [30, 41], ii) high spectral efficiency since multiple signals are superimposed in a given amount of bandwidth, and iii) effective equalization of the fading channel characteristics.

A major challenge with an OFDM system is that it is sensitive to timing errors and frequency offsets between transmitter and receiver. OFDM systems also have a high peak-to-average ratio which makes implementation a challenge. The power amplifier used before transmission typically has a narrow operating range where its performance is linear. Thus, the high peak-to-average ratio poses a challenge in ensuring efficient operation of the power amplifier.

2.5.3 Power Control

Determining the transmission power to use in a wireless system is a crucial task which affects the overall performance. The transmission power is controlled to achieve a variety of desired objectives, such as: i) provide performance guarantees even in the presence of a time varying channel attenuation, ii) conserve battery power in mobile devices by regulating the transmission power to the bare minimum required to provide desired level of service, or iii) reduce the amount of interference caused to other users in a multiuser system.

The variety and complexity of power control problems that arise in wireless systems are too numerous to cover in this chapter. For conciseness, we provide two examples of power control problems.

Cellular Uplink: We will consider the uplink of a cellular network in which the mobile users attempt to minimize their transmission powers to meet their desired performance level. To keep the discussion general we will consider both orthogonal and non-orthogonal signaling schemes used by multiple users.

Let set M and set B represent, respectively, the set of mobile users and base stations in the network. Mobile user $m \in M$ is connected to base station $b_m \in B$. Let set $K_m \subset M$ denote the set of mobile users whose transmissions interfere with

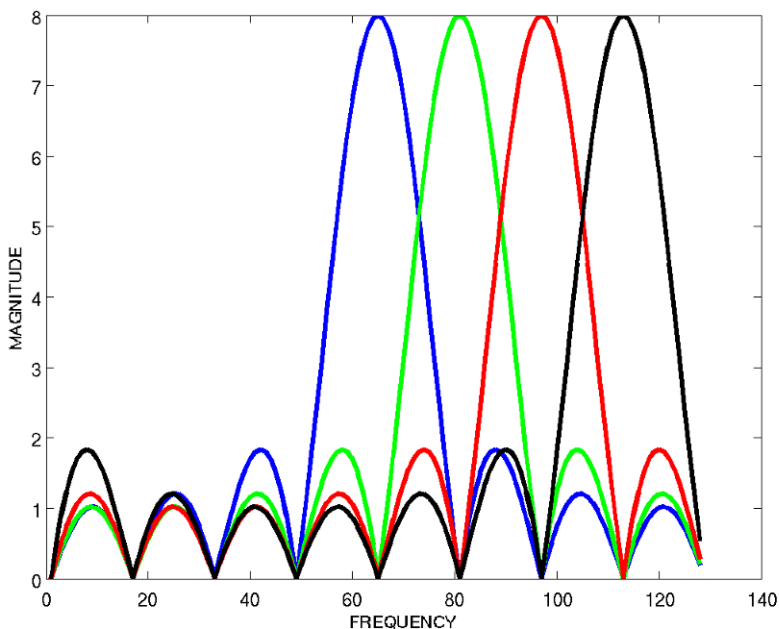


Fig. 2.17 Simple illustration orthogonality in an OFDM system

the transmission of user m . For instance, in a CDMA network using non-orthogonal codes and a frequency reuse of 1, $K_m = M - m \forall m \in M$. In a GSM system, K_m will only include mobiles from cells other than b_m and which use the same frequency as m .

The average channel attenuation between mobile $m \in M$ and base station $b \in B$ is given by h_{mb} . Let the transmit power of mobile m equal P_m and the receiver noise at base station b be denoted by N_b . Thus, the received signal-to-interference ratio SIR_{mb} for mobile m at base station b is given by,

$$SIR_{mb} = \frac{P_m |h_{mb}|^2}{\sum_{i \in K_m} P_i |h_{ib}|^2 + N_b} \quad (2.28)$$

As noted before, the obtained average SIR is proportional to a variety of performance metrics including throughput, and bit error rate. Thus, a typical performance requirement might be to ensure a desired minimum level of signal-to-interference ratio, \bar{SIR} is achieved for each user. There is clearly a conflicting requirement for each user. For instance, increasing P_m results in an increase in the ratio SIR_{mb} but also decreases the ratio SIR_{jk} for all users j such that $m \in K_j$. Thus, if each user employed a greedy approach of increasing their power when their received SIR is

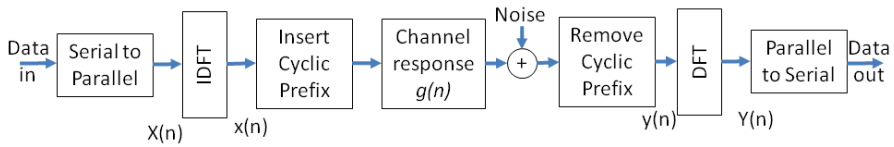


Fig. 2.18 Basic block diagram of an OFDM system.

lower than \bar{SIR} and not changing their power when their SIR is greater than \bar{SIR} all user powers would increase to infinity and the system would become unstable.

However, it can be proved that in this case, there exists an optimum transmit power level for each user that results in a stable operating point. This optimum operating point can be obtained as the solution to the following optimization problem:

$$\min \sum_{m \in M} P_m \quad (2.29)$$

$$s.t. SIR_{mb} \geq \bar{SIR} \quad (2.30)$$

As the keen reader might recognize, (2.29) is a centralized problem and requires global knowledge of all parameters to solve. However, it turns out that there exists a simple distributed and iterative solution to this problem. This solution only relies on each link observing its own SIR during iteration n and modifying its transmission power in iteration $n + 1$, as follows:

$$P_m(n+1) = \frac{\bar{SIR}(m)}{SIR_{mb}(n)} P_m(n) \quad (2.31)$$

One of the features of this solution is that the updates for each user can occur independent of the other users and in a completely asynchronous manner. In most cases, the number of iterations required to achieve convergence is small. It is typically assumed that all system parameters including channel gains, network topology variations due to mobility, variation in the number of users occur on a time-scale that is larger than what is required to obtain convergence.

Several variations of this problem and resulting solutions have been studied in the literature. For instance, problems that make the solution robust, opportunistic, and couple power control with beam-forming and scheduling have been investigated [13].

Ad-hoc Networks: The second example of power control is from ad-hoc networks. An ad-hoc network is a collection of nodes that communicate with each other in a multihop manner without the support of a backbone network. In an ad-hoc network, power control is also used for network topology control (for instance to achieve a fully connected network), or to improve the efficiency of the routing protocol or the MAC contention process. Further, the diverse nature of the nodes might cause a link to not have bidirectional connectivity.

Let c_{ij}^1 be the indicator function for first order or direct connectivity between nodes i and j . That is, $c_{ij}^1 = 1$ if node i can directly communicate with node j . Set M denotes the set of all nodes. In the case of asymmetric links, $c_{ij}^1 \neq c_{ji}^1$. Let c_{ij}^k indicate whether a k -hop connection exists between nodes i and j . Then,

$$c_{ij}^k = I[\{\max_{0 \leq n \leq k, l \in M, l \neq i, j} (c_{il}^n + c_{lj}^{k-n})\} > 2] \quad (2.32)$$

where $I[x]$ is the indicator function that equals 1 if the condition x is true, and equals 0, otherwise. Let the power of node i be denoted by P_i . In a more general

setting, a node may choose to use a different power level to communicate with each of its neighboring nodes. At the other extreme, the protocol design may restrict all nodes in the network to use the same transmission power. Recognize that c_{ij}^1 depends on the transmission power used by node i . Computing the minimum network power that ensures a fully connected network can be formally posed as the following optimization problem:

$$\min \sum_{i \in M} P_i \quad (2.33)$$

$$s.t. \quad P_i < P_i^{max} \quad (2.34)$$

$$c_{ij}^1(P_i) = c_{ji}^1(P_j) \quad (2.35)$$

$$\exists \quad k : c_{ij}^k = 1 \forall i, j \in M, i \neq j \quad (2.36)$$

Recognize that (2.35) imposes the condition that each link is bidirectional and (2.36) ensures that the network is fully connected.

Several protocols have been developed to solve problems of this nature in a distributed manner. We now describe a popular protocol, COMPOW [28], which selects a common transmit power for all nodes, thereby ensuring each link is bidirectional ($P_i = P_j, \forall i, j$).

COMPOW computes and maintains multiple routing tables at each node, one for each possible transmit power level. A routing table is essentially a look up table that a node uses to determine the next hop to forward the packets for each possible destination. COMPOW then determines the smallest power level at each node such that the list of nodes that can be reached (using a finite number of hops) from that particular node using the selected power, equals the list of nodes that can be reached from that node using the highest transmission power. In other words, network connectivity is not sacrificed in reducing the transmit power levels.

The key features of COMPOW are: i) the authors show that there is no significant loss in asymptotic (large number of nodes) network capacity by using a common transmit power for all nodes, ii) existence of low power routes using a low common transmit power for all nodes, and iii) the contention at the data link layer is also minimized.

2.6 Advanced Transceiver Methods

In this section we discuss selected advanced techniques used in state of the art wireless systems. These techniques are also an area of active research and form a critical component of future wireless networks.

2.6.1 Multiple Antenna Systems

One of the most promising methods to improve the performance of wireless systems is the use of multiple antennas at the transmitter and receiver. Such systems are broadly classified as multi-input-multi-output (MIMO) systems. The capacity of a MIMO system with n_t transmit and n_r receive antennas is given by [17, 43]

$$C_{MIMO} = \log \det(\mathbf{I}_{n_r} + \mathbf{H}\mathbf{P}\mathbf{H}^*), \quad (2.37)$$

where, \mathbf{H} is the $n_r \times n_t$ matrix of the channel gains between the various users, \mathbf{I}_{n_r} is the identity matrix of size n_r , \mathbf{P} is the diagonal matrix of size n_t with the diagonal elements being the power transmitted on the various transmit antennas and \mathbf{H}^* denotes the complex conjugate of the matrix \mathbf{H} . For simplicity, the noise at the receiver is normalized to 1 in (2.37). For channels with certain types of fading characteristics, this capacity grows asymptotically with SNR as

$$C_{MIMO} \approx \min(n_t, n_r) \log(SNR). \quad (2.38)$$

Thus, the capacity grows linearly with the number of antennas, if both the number of transmit and receive antennas increases.

The multiple antennas can also be used for improving the diversity of the system. The diversity of the system is defined as the exponent of the SNR in the expression for probability of error. For instance, the maximum diversity of a Gaussian noise channel is infinite. This diversity can be derived from the expression for probability of error, similar to (2.6) or (2.7). By approximating the Q -function with an exponential function, the asymptotic fall off of error rate with SNR can be computed. In contrast, the diversity order in fading channels is finite. For the MIMO system with an independent Rayleigh fading distribution between each pair of transmit and receiver antennas, the maximum diversity equals $d_{max} = n_t n_r$.

While it is desirable to use multiple antennas to increase the throughput as well as reduce the error probability, there is a fundamental trade-off between these two objectives. This trade-off is quantified by the diversity-multiplexing trade-off [55]. Specifically, the maximum diversity $d(r)$ that can be obtained when the throughput equals $r \log(SNR)$ is given by the piecewise linear boundary specified by

$$d(r) = (n_r - r)(n_t - r), \forall 0 \leq r \leq \min\{n_t, n_r\}, r \in \mathbb{Z}^+ \quad (2.39)$$

The multiple antennas can also be used for beam-forming. In digital beam-forming, the transmit node (for instance a cellular base station) uses a group of antennas to transmit and receive signals. The signal for each user is premultiplied by a different complex number at each antenna in order to shape the beam towards the direction of that user. This beam-forming is useful in reducing the interference experienced by other users in the system. Recent advances in antenna technology and signal processing methods have enabled adaptive beam-forming methods to be implemented in which the beams are adaptively designed in real-time depending on user locations and the desired performance metric to optimize. A similar adaptive

weighting of the signals received from multiple antennas can be used to enhance the signal of one user and minimize the interference from the other users.

2.6.2 Scheduling, Rate, and Power Control

One of the promising areas of recent research is in the area of cross layer optimization [6, 38]. The essential idea in cross layer optimization is to modify the design of the protocol/algorithm in one layer based on input from a different layer. Specifically, the area of cross layer scheduling, rate and power control has emerged as a useful tool to improve performance significantly.

Consider a multiuser wireless network in which each of the users encounters a different time varying wireless channel gain. Further, each user supports a different type of traffic and consequently has different rate and delay requirements. The task of the scheduler is to optimally allocate the resources (power, modulation, coding rate, time-slots, frequency bands) to the various users to optimize a desired global network level metric while also satisfying some fairness constraints. These fairness constraints are required to ensure that no single user is starved of network resources. These scheduling issues are covered in detail in Chapter 14 for general wireless systems and in Chapter 13 for WiMAX systems.

2.7 Conclusions

In this chapter, we considered the basics of a digital wireless communication system. The focus is on introducing the fundamental concepts of modulation and coding in a manner that is broadly accessible. Basic multiple access methods including popular CDMA and OFDM are also introduced.

We did not consider the various models for wireless channels. As noted before, several detailed and extensive models for the wireless channel are available and are briefly discussed in Chapter 3.

Acknowledgements The work of the author was supported in part by the Office of Naval Research through a research grant.

References

1. Abramson, N.: The ALOHA system: another alternative for computer communications. In: Proceedings of Fall Joint Computer Conference (ACM), pp. 281–285. Houston, TX (1970)
2. Abramson, N.: The Alohanet - surfing for wireless data. IEEE Comm. Mag. **47**(12), 21–25 (2009)

3. Bahl, L., Cocke, J., Jelinek, F., Raviv, J.: Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Transactions on Information Theory* **20**(2), 284–287 (1974)
4. Balakrishnan, H., Padmanabhan, V., Seshan, S., Katz, R.H.: A comparison of mechanisms for improving tcp performance over wireless links. *IEEE/ACM Transactions on Networking* **5**(6), 756–769 (1997)
5. Berrou, C., Glavieux, A., Thitimajshima, P.: Near Shannon limit error-correcting coding and decoding: Turbo-codes. In: *Proceedings of ICC*, pp. 1064–1070. Geneva, Switzerland (1993)
6. Berry, R., Yeh, E.: Cross layer wireless resource allocation - fundamental performance limits. *IEEE Signal Processing Magazine* **21**(5), 59–68 (2004)
7. Bertsekas, D.P., Gallager, R.: *Data Networks*. Prentice Hall (1992)
8. Bhargavan, V., Demers, A., Shenker, S., Zhang, L.: MACAW: A media access protocols for wireless LANs. In: *Proceedings of ACM Sigcomm*. London, England (1994)
9. Biglieri, E.: *Coding for Wireless Channels*. Springer, New York, NY (2005)
10. Biglieri, E., Proakis, J.G., Shamai, S.: Fading channels: Information-theoretic and communications aspects. *IEEE Transactions on Information Theory* **44**(6), 2619–2692 (1998)
11. Braden, R.: RFC 1122 Requirements for Internet Hosts – Communication Layers. In: *Internet Engineering Task Force, Request for Comments (RFC) document*. Network Working Group (1989)
12. Bray, J., Sturman, C.F.: *Bluetooth 1.1: Connect Without Cables*. 2nd edn. Prentice Hall, Upper Saddle River, NJ (2001)
13. Chiang, M., Hande, P., Lan, T., Tan, C.W.: Power control in wireless cellular networks. *Foundations and Trends in Networking* **2**, 381–533 (2008)
14. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*. 2nd edn. Wiley, New York (2006)
15. Forney, D.G.: *Concatenated Codes*. The MIT Press (1966)
16. Forney, D.G., Ungerboeck, G.: Modulation and coding for linear Gaussian channels. *IEEE Transactions on Information Theory* **44**(6), 2384–2415 (1998)
17. Foschini, G.J., Gans, M.J.: On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Communication* **6**, 311–355 (1998)
18. Gallager, R.G.: *Information Theory and Reliable Communication*. John Wiley and Sons, Inc. (1968)
19. Gast, M.: *802.11 Wireless Networks: The Definitive Guide*. 2nd edn. O'Reilly Media, Sebastopol, CA (2005)
20. Gastpar, M., Rimoldi, B., Vetterli, M.: To code, or not to code: Lossy source-channel communication revisited. *IEEE Transactions on Information Theory* **49**(5), 1147–1158 (2003)
21. Golmie, N., Rebala, O., Chevrollier, N.: Bluetooth adaptive frequency hopping and scheduling. In: *Proceedings of MILCOM*, pp. 1138–1142. Boston, MA (2003)
22. Gunduz, D., Erkip, E., Goldsmith, A., Poor, H.V.: Source and channel coding for correlated sources over multiuser channels. *IEEE Transactions on Information Theory* **55**(5), 3927–3944 (2009)
23. Guruswami, V., Sudan, M.: Improved decoding of Reed-Solomon codes and algebraic geometry codes. *IEEE Transactions on Information Theory* **45**(6), 1757–1767 (1999)
24. Hanzo, L., Munster, M., Choi, B.J., Keller, T.: *OFDM and MC-CDMA for Broadband Multi-User Communications, WLANs and Broadcasting*. Wiley-IEEE Press, West Sussex, England (2003)
25. Keshav, S.: *An Engineering Approach to Computer Networks*. Addison-Wesley Longman, Inc. (1997)
26. Lin, S., Costello, D.J.: *Error Control Coding: Fundamentals and Applications*. Prentice Hall, second edition (2004)
27. Moon, T.: *Error Correction Coding: Mathematical Methods and Algorithms*. Wiley-Interscience, Hoboken, NJ (2005)
28. Narayanaswamy, S., Kawadia, V., Sreenivas, R.S., Kumar, P.R.: Power control in ad-hoc networks: Theory, architecture, algorithm and implementation of the COMPOW protocol. In: *Proceedings of European Wireless 2002. Next Generation Wireless Networks: Technologies, Protocols, Services and Applications*, pp. 156–162. Florence, Italy (2002)

29. Oestges, C., Clerckx, B.: MIMO wireless communications: from real-world propagation to spacetime code design. Academic Press: Oxford (2007)
30. Oppenheim, A.V., Schaffer, R.W.: Discrete Time Signal Processing. Prentice Hall, Englewood Cliffs, NJ (1989)
31. Papoulis, A., Pillai, U.S.: Probability, Random Variables and Stochastic Processes. Mc-Graw Hill Inc., New York, NY (2002)
32. Popovski, P., Yomo, H., Prasad, R.: Dynamic adaptive frequency hopping for mutually interfering wireless personal area networks. In: Proceedings of ACM Mobihoc, vol. 4, pp. 199–209. Tokyo, Japan (2004)
33. Proakis, J.G.: Digital Communications, Fourth Edition. Mc-Graw Hill Inc., New York, NY (2001)
34. Proakis, J.G., Salehi, M.: Fundamentals of Communication Systems. Prentice Hall, Upper Saddle River, NJ (2004)
35. Ramstad, T.: Insights into mobile multimedia communications, chap. Combined Source Coding and Modulation for Mobile Multimedia Communication, pp. 415–430. Academic Press Signal Processing And Its Applications Series. Academic Press Professional, Inc., San Diego, CA (1998)
36. Rappaport, T.: Wireless Communications: Principles and Practice. 2nd edn. Prentice Hall, Upper Saddle River, NJ (2001)
37. Sayood, K.: Introduction to Data Compression. Morgan Kauffman, San Francisco, CA (2005)
38. Shakkottai, S., Rappaport, T.S., Karlsson, P.C.: Cross-layer design for wireless networks. IEEE Communications Magazine **41**(10), 74–80 (2003)
39. Shannon, C.E.: A mathematical theory of communication. Bell Systems Technical Journal **27**, 379–423 (1948)
40. Sherwood, P.G., Zeger, K.: Progressive image coding on noisy channels. In: Proceedings of Data Compression Conference, pp. 72–81. Snowbird, UT (1997)
41. Soliman, S., Srinath, M.D.: Continuous and Discrete Signals and Systems. Prentice Hall, Upper Saddle River, NJ (1998)
42. Tanenbaum, A.S.: Computer Networks. Prentice Hall, Upper Saddle River, NJ (2002)
43. Telatar, I.E.: Capacity of multi-antenna Gaussian channels. European Transactions on Telecommunications **10**(6), 585–596 (1999)
44. Trape, W., Washington, L.C.: Introduction to Cryptography with Coding Theory. Prentice Hall (2005)
45. Ungerboeck, G.: Channel coding with multilevel/phase signals. IEEE Transactions on Information Theory **25**(1), 55–66 (1982)
46. Vembu, S., Verdu, S., Steinberg, Y.: The source-channel separation theorem revisited. IEEE Transactions on Information Theory **41**(1), 44–54 (1995)
47. Verdu, S.: Multiuser Detection. Cambridge University Press, Cambridge, UK (1998)
48. Viterbi, A.J.: Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Transactions on Information Theory **13**(2), 260–269 (1967)
49. Viterbi, A.J.: CDMA : Principles of Spread Spectrum Communication. Addison-Wesley Publishing Company, Reading, MA (1995)
50. Wang, X., Poor, V.: Iterative (Turbo) soft interference cancellation and decoding for coded CDMA. IEEE Transactions on Communications **47**(7), 1046–1061 (1999)
51. Wicker, S.B.: Error Control Systems for Digital Communication and Storage. Prentice Hall, Upper Saddle River, NJ (1995)
52. Yates, R.: A framework for uplink power control in cellular radio systems. IEEE Journal on Selected Areas in Communications **13**(7), 1341–1348 (1995)
53. Yates, R., Goodman, D.J.: Probability and Stochastic Processes: A Friendly Introduction for Electrical and Computer Engineers. John Wiley and Sons Inc., Hoboken, NJ (2004)
54. Yeung, R.W.: A First Course in Information Theory. Kluwer Academic/Plenum publishers, New York, NY (2002)
55. Zheng, L., Tse, D.: Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. IEEE Transactions on Information Theory **49**(5), 1073–1096 (2003)

Wireless Network Design
Optimization Models and Solution Procedures
Kennington, J.; Olinick, E.; Rajan, D. (Eds.)
2011, XIX, 373 p.,
ISBN: 978-1-4419-6111-2