Chapter 1

Introduction

Digital communication systems are ubiquitous. From cell phones to Bluetooth, to WiFi, to cable modems: digital communication systems are present in everyday life. This book explores in depth at how these systems work and their fundamental limits. We begin in this Chapter with a high level understanding of digital communications to understand the tradeoffs in designing a communication system. The remainder of the book delves into the signals used in digital communications, the receivers employed, and the performance. One main challenge of all communication systems is noise which limits the data rate for which reliable communication is possible. Another limitation is the propagation mechanism that can distort and greatly attenuate the transmitted signal. These limitations can be overcome, to some extent with error control coding discussed at the end of this book.

Objectives of this Chapter

- Understand the difference between digital and analog communication systems.
- Understand the important parameters of a digital communication system including energy efficiency and bandwidth efficiency.
- Understand signal vectors, their rate and minimum distance.
- Understand the fundamental tradeoffs between energy efficiency and bandwidth efficiency.

The goal of communication systems is to transmit information from one location and *reliably* receive that information at a second location at a *high rate*. There are two important resources that are needed for reliable communication: energy and spectrum. The first resource is the *energy* used at the transmitter and the receiver. This is usually limited by the energy available from a battery for mobile applications such as cell phones, solar array power for satellites or by regulations that limit transmitted power. Part of the energy consumed at the transmitter is radiated from the transmitter in order to be able to receive a signal at the receiver. Only part of the radiated energy is actually received by the receiving antenna. Part of the total energy, apart from the radiated energy, is consumed in components in the transmitter and receiver. For example, amplifiers are not 100% efficient in converting energy from a power supply to output energy: they consume more energy than they put out. The energy received limits the rate at which information can be communicated reliably. The second resource is the *bandwidth or spectrum* used to transmit the signal. Spectrum is typically allocated by a government agency such as the Federal Communication Commission (FCC). The available spectrum limits the rate that information can be reliably communicated. The *rate* that information is communicated is usually measured in bits per second. The reliability is often measured in terms of the probability of making an error when deciding at the receiver the transmitted information. The fundamental challenge is that noise is present at the receiver. Noise in a receiver also limits the rate at which reliable communication is possible.

The design of a communication system requires careful consideration of the relation between the rate at which information can be reliably communicated, the resources used (energy and bandwidth), and the amount of noise present in the receiver. The ratio of the energy received from the transmitter per bit of information transmitted to the noise level is known as the *signal-to-noise ratio*, a unit-less quantity. The ratio of the information rate to the bandwidth is called the *bandwidth efficiency*. The bandwidth efficiency is measured in units of bits/second/Hz. There is a fundamental tradeoff between the signal-to-noise ratio and the bandwidth efficiency. A higher (received) signal-to-noise ratio allows for a higher bandwidth efficiency. At low signal-to-noise ratios reliable communication is still possible, but only if the bandwidth efficiency is low. This tradeoff is discussed in Section 1.5.

Besides noise there are various channel impairments, such as multipath fading, that can limit the rate of reliable communications. Complexity of implementation of the transmitter and receiver can also may limit the rate at which reliable communication is possible. Some of these will be discussed later in this book.

There are two basic types of communication systems: digital and analog communication systems. In a *digital* communication system during any finite time interval there is a *finite* number of possible waveforms transmitted. Typically a group of bits is mapped to a set of possible waveforms that have a certain time duration. Most modern communication systems are digital communication systems such as cellular communications, WiFi, Cable TV, Bluetooth. In an *analog* communication system, during any finite time interval there are a potentially *infinite* number of possible waveforms transmitted. Typically a continuous audio waveform is transformed into a transmitted signal by modulating the amplitude or frequency of a sinusoidal signal, known as the carrier frequency. Examples of analog communication systems include AM and FM radio, and TV signals prior to the adoption of the current TV transmission standard. In a digital communication system, because there are a finite number of possible transmitted signals, the receiver *decides*, based on the received signal, which of the finite number of transmitted signals was most likely to have caused the received signal. The receiver either decides correctly which signal was transmitted or there is an error. In an analog communication system the receiver processes the received signal to recover the analog waveform at the input of the transmitter. This may involve *estimating*, based on the received signal, what was the input of the transmitter. Exact recovery is not possible in an analog communication system because of noise. Exact recovery is possible with digital transmission.

There are various advantages of a digital communication system over an analog communication system. Because information is typically represented as bits in a digital communication system there is the possibility of regenerating the data from a noisy received signal without any errors or distortion. Another advantage is that sophisticated digital processing can be performed without significant cost due to the vast increase in number of transistors on a chip for doing calculations needed for processing. Also, because exact recover is possible a digital signal can be relayed over a number of different links without necessarily increasing the error rate. There is one significant disadvantage of digital communication relative to analog communication and that is the need for *synchronization*. In a digital communication system information is conveyed by way of waveforms representing information (e.g. bits) that have a certain duration (start time and end time). After the transmission of one waveform, transmission of another waveform representing new information can commence. Knowing the time when the waveforms changes from communicating one piece of information (e.g. a bit) to another piece of information (e.g. another bit) is necessary in a digital communication system. This is the problem of synchronization. Synchronization of this sort is not necessary in an analog communication system since the waveforms are just continuous representations of some analog source (e.g. audio). Most of present day communication systems are digital communications system which require synchronization but have the advantages mentioned above. This book is focused solely on digital communication systems.

In a digital communication system a set of information bits is mapped into a waveform for transmission. One bit can determine one of two possible waveforms. Two bits can determine one of four possible waveforms. Noise at the receiver will be part of the received signal. *The receiver decides which transmitted waveform most likely caused the received signal.* Once a decision on the transmitted waveform is made the corresponding set of bits that was mapped to that waveform at the transmitter is determined. One performance measure for *digital* communication systems is the probability of making an error in deciding which waveform was transmitted. Also of interest is the probability of an error for one bit. This is often called the bit error rate (BER). Often a sequence of bits is called a packet and the packet error probability or packet error rate (PER) is of interest. This is also called the block error probability or the frame error rate. The probability of error (bit or packet) depends on various parameters. including the power or energy consumed, the bandwidth utilized, the data rate, the delay, the complexity and the level of noise at the receiver.

There are a variety of important parameters in a digital communication system. These are listed below.

- **Power or Energy:** The more power or energy available at the receiver the higher the rate at which reliable communication is possible. However, the goal is to achieve the highest rate of reliable communication with the minimum required transmission power. Bluetooth devices have a maximum transmitted power of 100mWatts. WiFi devices have a maximum transmitted power of 1 Watt. A satellite might transmit 100's of Watts of power. Power or energy at the receiver is dependent on transmitted power or energy and propagation loss. While transmitted power is important in determining received power, performance will depend on received power.

- **Data Rate:** Generally, large data rates are desired. There are two important considerations regarding data rate. First, for a fixed amount of transmitted power, as the data rate increases the energy transmitted per bit will decrease because of decreased transmission time for each bit. That is, $E_b = P/R$ where E_b is the energy per bit (Joules/bit), P is the power (Watts=Joules/second), and R is the rate in bits/second. Second, if the date rate is large then the spectrum of the signals used will spill out of the available bandwidth so some amount of filtering will be necessary before transmission. Filtering typically introduces intersymbol interference (ISI) in which the receivers decision about one symbol will be affected by symbols transmitted previously. Large data rates for a fixed bandwidth will increase the effect of previously transmitted symbols. A wireless channel typically has an impulse response with some delay spread. That is, the received signal is delayed by different amounts on different paths. The signal corresponding to a particular bit received with the longest delay will interfere with the signal corresponding to a different bit with the shortest delay. The larger the number bits that are interfered with the more difficult it is to correct for this interference.

- **Bandwidth:** This is the amount of frequency spectrum available for use. Generally, the a government agency (e.g. FCC) allocates spectrum and provides some type of requirements for the spectrum of a transmitted signal. A specification of allowed power as a function of frequency is sometimes called a spectral mask. As an example, WiFi, circa 2000-2020, uses about 20MHz of bandwidth. Bluetooth uses about 1 MHz of bandwidth. Satellites might use 1GHz of bandwidth or more.

- **Probability of Error:** Usually the probability of bit error is the performance measure of interest and often call the bit error rate (BER). In some systems the probability of packet error or packet error rate (PER) is of interest. Different types of information may require different probabilities of error. For example in a voice call on a cellular phone whereby the voice signal is converted to bits and then transmitted, incorrect bits might distort the eventual reconstructed voice signal but not to the extent that the speech is not intelligible. However, transmission of data (e.g. bank account information) would require a much lower error probabilities than voice.

- **Delay Requirement:** Some types of communication systems, such as a phone call, require a small amount of delay or latency between the data at the transmitter and the reproduced data at the receiver. These applications are often called real-time services. For other services, such as email or file transfer, delay is not as critical. Generally, a larger delay requirement allows for better reliability.

- **Complexity:** More complexity usually implies better performance. The goal is to get the best performance for the least complexity.

- **Noise and Interference:** Noise at the receiver limits the performance (data rate or error probability). Cooling the receiver is an option to reduce the noise level in limited circumstances. Interference is an impairment caused by other transmitters either intentional or unintentional in the same frequency band or in an adjacent frequency band. Sometimes a system is designed to allow multiple transmitters to transmit simultaneously in the same frequency band. The different transmitters use signals that are sufficiently distinguishable so that a receiver can recover each of the transmitted signals. The Global Positioning System (GPS) is an example of this. However, multiple simultaneous transmissions usual requires more bandwidth than a single user transmitting in a frequency band. This technique is called spread-spectrum or code-division multiple-access (CDMA).

The overall design of a communication system depends on the relative importance of different parameters (energy, delay, error probability, data rate, bandwidth). The goal of this book is to understand the tradeoffs possible in designing a communication system between these parameters.

1.1 Communication System Coat of Arms

The overall design of a digital communication system is quite complex. To simplify the design most communication systems are broken into parts or subsystems, each with a clearly defined goal. One standard representation of subsystems is shown in Figure 1.1. The subsystems are often in pairs with one part in the transmitter and one part in the receiver. The components of this diagram are described as follows.

- **Source/Sink** The source of the channel is the subsystem that generates the data that is desired to be communicated to the sink or destination.

- **Source Encoder/Decoder:** The source encoder removes redundancy from the source data such that the output of the source encoder is a sequence of symbols from a finite alphabet. For example, in English the letter “q” is virtually always followed by the letter “u”. So the input to the source coder might be the word “question” but the source encoder could produce the output “qestion””. So one less symbol would need to be transmitted. The source decoder reverses the operation of the source encoder to determine the most probable sequence that could have caused the output. For the example above, since every appearance of the letter q is always followed by the letter u, if the input of the source decoder was the symbols “qestion” then the source decoder would add

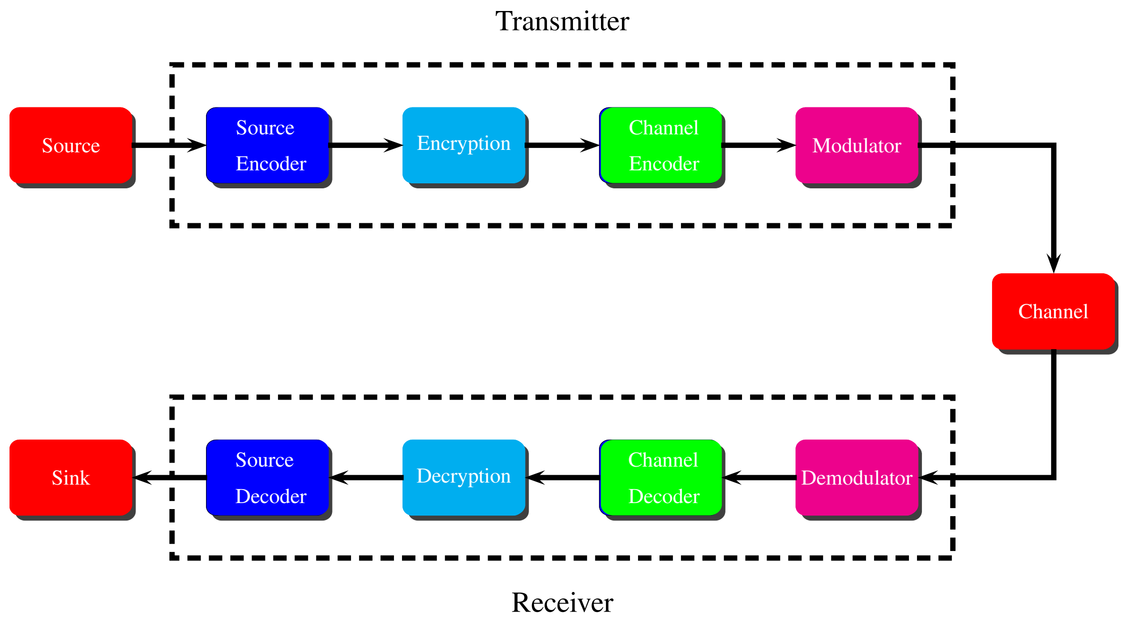


Figure 1.1: Communication System Coat of Arms

a u after every q and reproduce the original input of the source encoder. In English then, the letter u after a letter q is a redundant symbol. We can reduce the number of symbols (or bits) that need to be transmitted by removing redundant symbols. If the source produces symbols from an infinite alphabet (e.g. real numbers) then some distortion must be incurred in representing the source with a finite alphabet. For example the real number π can be represented by 3.14 and the decimal digits 3,1,4 could be transmitted. So the input to the source encoder would be some real number (say between 0 and 10) and the source encoder would round the number to two decimal digits of accuracy. The source decoder would output the quantized numbers 3,1,4 and (assuming knowledge by the source decoder of where the decimal point would go) would output 3.14 as the reproduction at the source decoder output of the source encoder input π. So symbols with an infinite alphabet, when mapped to a finite number of source encoder outputs will not be reproduced without distortion. There is a fundamental tradeoff between the number of bits used to represent a source output and the distortion incurred when reproducing the source output. This is known as the rate-distortion function.

- **Encryption/Decryption:** An encryption system will transform an input sequence $\{W_k\}$ into an output sequence $\{Z_n\}$ such that knowledge of $\{Z_n\}$ alone (without a key) makes calculation of $\{W_l\}$ extremely difficult (e.g. many years of CPU time on a fast computer). The decryption system, with the aid of a secret key, reverses the operation of the encryption device. With private key cryptography the key is only known to the encryption and decryption devices. In private key encryption the method of encryption is typically easily invertible to obtain the decryption algorithm. In another method of encryption, called public key cryptography, there is a encryption key which is made public. This key allows anyone to encrypt a message. However, even knowing this key it is not possible, at least not easily, to reverse this operation and recover the message from the encrypted message. There are some special properties of the encryption algorithm known only to the decryption device which, knowledge of, makes this operation easy. This is known as a trap door. Since the encryption key need not be kept secret for the message to be kept secret this is called public key cryptography. One example of public key cryptography uses the fact that factoring large numbers is difficult. For example, a 200 digit number would take a large number of operations to factor. This 200 digit number could be a public encryption key. What is private is the factorization of this number. The decoder uses the factors in order to reverse the encryption step. So the 200 digit number is public but what is private is the factorization. If the number could be factored by an eavesdropper, then decryption would be possible by the eavesdropper. But since factorization is hard, the assumption is that an eavesdropper would not be able to do the factorization in the time for which the information has value.

- **Channel Encoder/Decoder:** The channel encoder introduces redundancy into data such that if there are some errors made over the channel they can be corrected. This is often called channel coding or forward error control (FEC). The channel decoder reverses the operation of the channel encoder in the absence of any channel noise. When the channel causes some errors to be made in the estimates of the transmitted messages the decoder attempts to correct these errors to the extent possible. For example if every information bit is repeated three times by the channel encoder then at the receiver a majority vote count of the bits received will be able to recover the original bit even if one of the three bits is received in error. Sometimes an error detecting code is used to prevent incorrect information from being delivered to the destination. If a decoder for an error detecting code detects an error a retransmission can be requested.

  Note: The source encoder removes *unstructured* redundancy from the source data and may cause distortion or errors in a *controlled* fashion. The channel encoder adds redundancy in a structured fashion so that the channel decoder can correct some errors caused by the channel.

- **Modulator/Demodulator:** The modulator maps a finite number of messages (e.g. some number of bits) into a set of distinguishable signals so that at the channel output it is possible to determine which signal in the set was transmitted. The demodulator processes the channel output and produces an estimate of the message that caused the output. For example a modulator could transmit a single bit (i.e. 0 or 1) by mapping the bit to a sinusoidal signal with either phase 0 or π depending on if the bit was 0 or 1. The demodulator would decide if the received signal was more likely to be caused by a 0 at the input to the modulator or by a 1 at the input to the modulator.

- **Channel:** Medium by which signal propagates from transmitter to receiver

  Examples of communication channels:

  - Noiseless channel (very good, but not reality).
  - Additive white Gaussian noise (AWGN) channel. This is the classical channel but also is a good model for the deep space channel.
  - A telephone line channel or wireless channel. These channels often produce intersymbol interference channel.
  - Fading channel (mobile communication system when transmitters are behind buildings, Satellite systems when there is rain on the earth).
  - Channels with multiple-access interference. This occurs when several users access the same frequency at the same time.
  - Hostile interference channel (jamming signals).

- Semiconductor memories (RAM's, errors due to alpha particle decay in packaging).
- Magnetic and optical disks (Compact digital disks for audio and for read only memories, errors due to scratches and dust).
- Quick Response (QR) codes used to link a user to a web page. There can be unintentional errors due to marks or intentional errors due to embedding a logo in the image.

Often the modulator-channel-demodulator part of the block diagram are thought of as a *super channel* with a finite number of inputs and a finite or infinite number of outputs.

1.2 Basic Concepts of Modulation and Demodulation: Signals

In this section we discuss the modulator, channel and demodulator part of digital communication systems. In a digital communication system the modulator has an input that is a set of one or more bits. The modulator, based on those bits, generates one of a finite number of signals or waveforms to be sent. The waveforms typically have a finite time duration. For example, a single bit (i.e. either 0 or 1) into a modulator will produce one of two signals depending on the value of the bit. A modulator transmitting one of four signals can communicate two bits of information.

The channel in a digital communication system characterizes the relation between the transmitted signal and the received signal. The signal at the receiver will differ from the transmitted signal for several reasons. One reason is that there is *propagation loss*. That is, the signal power at the receiver is typically much smaller than the power at the transmitter. Second, there is generally noise in the receiver. This noise is added to the signal received due to the transmitted signal. So the total received signal often consists of the sum of two parts: the transmitted signal after propagation loss and the noise.

The demodulator must determine which signal was transmitted and the associated bits. For the case of a single bit mapped to one of two signals, the demodulator determines which of the two possible transmitted signals was most likely to have caused the received signal. *In general, the demodulator needs to decide, based on the received signal, which of the finite number of transmitted signals most likely could have caused the received signal.* Generally, the demodulator will decide a particular transmitted signal if the received signal is "closer" or more similar to that signal than all the other signals. An error is made if one of the transmitted signals is closer or more similar to the received signal than the actual transmitted signal.

If there is more energy or power available for transmitting then more signals can be transmitted and yet have the signals remain distinguishable. Thus a larger rate (bits per second, bps) can be communicated. However, if there are too many possible transmitted signals, for a given energy, the demodulator will be unable to adequately distinguish the possible transmitted signals. In this case the probability of error may be higher than desired.

In this section, for simplicity and to illustrate the concepts, we represent the signals as vectors. Representing signals as points in the two-dimensional plane helps understand the concept of distinguishability of signals in a simplified, but yet accurate, sense. As shown in Chapter 2, the distinguishability of signal waveforms, which are functions of time, and signal vectors are identical parameters. Often these signal vectors are just two-dimensional vectors in which case the signal vectors are represented as points in a plane. When the signals are represented by two dimensions (a point in the plane) the first component is called the in-phase component, or just I component for short, and the second component is called the quadrature-phase component, or Q component for short. These points are plotted on an I-Q plane. For example, Figure 1.2 shows a set of points in the plane representing four signals. This signal set or constellation is called quadrature phase shift keying (QPSK). The phase angle of a signal point is the angle between the line connecting the signal point to the origin and the horizontal axis. In QPSK there are four different phases for the four different signal points. Thus the information about which signal point is transmitted is in the phase of the signal, as opposed to the amplitude. How the signal vectors are mapped into waveforms is discussed in Chapter 2.

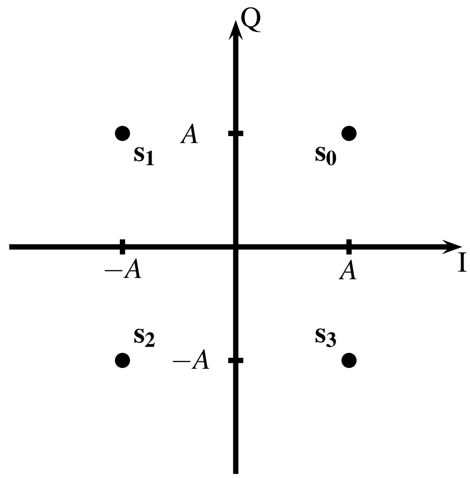


Figure 1.2: Quadrature phase shift keying (QPSK) constellation represented in I/Q plane. Four signals representing 2 bits.

Signals can be represented by vectors of length N. The length of the vector N is also called the dimension. A set of M vectors each of length N is called a ***constellation*** or a ***signal set***. We denote the M vectors by $\mathbf{s_m} = (s_{m,0}, s_{m,1}, ..., s_{m,N-1}), m = 0, 1, ..., M-1$. A constellation or signal set (M vectors of length N) has a number of parameters associated with it. These parameters include the rate of the signal set, the average signal energy per bit and the minimum squared Euclidean distance.

The ***rate of a signal set*** in units of bits per dimension, denoted by r, is the number of bits represented by the signals per dimension. If a signal set has M signals in N dimensions then transmitting one of the M signals can communicate $\log_2(M)$ bits of information. The rate is

$$r = \log_2(M)/N \quad \text{bits/dimension.}$$

For example, a set of 32 signals in two dimensions has rate 5/2 bits/dimension because 5 bits can be communicated with 32 signals in two dimensions. A closely related concept to the rate of a signal set is the rate of communication. The ***rate of communication***, denoted by R, is the number of bits communicated per second. There is a simple relation between the rate of a signal set, r, the rate of communication, R and the bandwidth W. Consider the sampling theorem which says that signals with bandwidth W can be reconstructed from samples of the signal if the sampling rate is at least twice the highest frequency. So if the highest frequency is W (i.e. frequency range from $-W$ to $+W$ then samples at a rate greater than $2W$ can recover the signal. So in T seconds the number of samples would be $2WT$. So from any set of $2WT$ samples (a vector of length or dimension $2WT$) we can construct a signal of bandwidth W lasting T seconds. In later chapters we will show in an alternate way that signals that occupy an approximate bandwidth of W Hz and that last for T seconds that can be represented with vectors of length $2WT$; that is, the waveforms can be represented using vectors of length $2WT$ or $2WT$ dimensions. For example, a signal set with signals that occupy a bandwidth of 20MHz with each signal lasting 1 μ second can be represented by vectors of length 40. With $2WT$ dimensions in T seconds there are available $2W$ dimensions per second to represent signals. So to convert from bits/dimension to bits/second we need to multiply the number of dimensions per second which is $2W$. The relation between the number bits per dimensions and the number of bits per second is $R = r(2W)$. Often the rate of communication R is normalized to the bandwidth W being used by the signals. This is called the bandwidth efficiency R/W and has units bits/second/Hz. The conversion from bits/dimension to bits/second/Hz is just a factor of 2. That is,

$$R/W = 2r. \tag{1.1}$$

Another important parameter of a signal set is the energy. The ***energy*** of a signal that is represented by a vector is the squared distance of the point from the origin or the squared length of the signal represented. The energy of signal $\mathbf{s_m} = (s_{m,0}, s_{m,1}, ..., s_{m,n-1})$ is

$$E_m = \sum_{n=0}^{N-1} |s_{m,n}|^2.$$

The average energy E of a signal set is found by averaging the energy of all the signals in the set. The average energy per bit is the average energy of the set of signals divided by the number of bits communicated. Assuming the signals are equally likely

$$E = \frac{1}{M}\sum_{m=0}^{M-1} E_m, \tag{1.2}$$

$$E_b = E/\log_2(M). \tag{1.3}$$

For example a set of 32 signals with average energy E would have energy per bit of $E_b = E/5$. The power of a signal is related to the energy of the signal and the time duration of the associated waveform. The time duration of a bit T is the inverse of the data rate $R = 1/T$. The average power P of a signal set that has average energy per bit of E_b and rate R is $P = E_b R$ where R is the rate in units of bits/second.

The distinguishability of signals in the presence of noise is related to the distances between signals. The squared Euclidean distance between two signals, represented as vectors, is

$$d_E^2(\mathbf{s_m}, \mathbf{s_l}) = \sum_{n=0}^{N-1} |s_{m,n} - s_{l,n}|^2.$$

The ***minimum squared Euclidean distance*** is the smallest squared distance between any two distinct signals represented as vectors.

$$d^2_{E,\min} = \min_{\mathbf{s_m} \neq \mathbf{s_l}} d_E^2(\mathbf{s_m}, \mathbf{s_l})$$

The minimum squared Euclidean distance is a measure of the distinguishability of signals in the presence of noise.

Example 1: Figure 1.2 shows an example of signals represented as points in the plane. In this example there are four signals representing two bits of information. Only one of the four signals is transmitted at any time depending on the current two bits of information. The four signal vectors and the mapping of information bits to signals is as follows.

$$\begin{aligned}
00 \to \mathbf{s_0} &= (s_{0,0}, s_{0,1}) = A(+1,+1) = (+A,+A) \\
01 \to \mathbf{s_1} &= (s_{1,0}, s_{1,1}) = A(-1,+1) = (-A,+A) \\
10 \to \mathbf{s_2} &= (s_{2,0}, s_{2,1}) = A(-1,-1) = (-A,-A) \\
11 \to \mathbf{s_3} &= (s_{3,0}, s_{3,1}) = A(+1,-1) = (+A,-A)
\end{aligned}$$

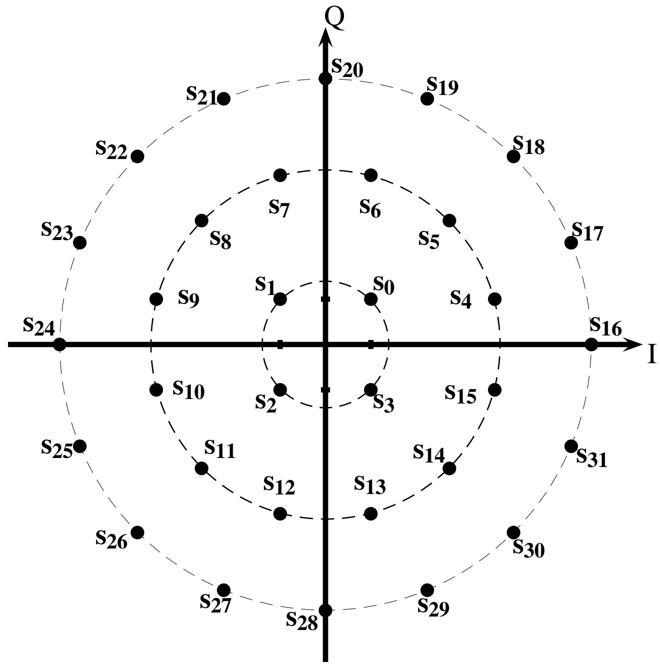


Figure 1.3: 32 signals representing 5 bits.

where A represents the amplitude in each dimension. There are $M = 4$ vectors of length $N = 2$ that communicate 2 bits of information. Each signal has energy $2A^2$. The average energy of the signal set is $E = 2A^2$. The energy per bit is $E_b = E/2 = A^2$. The rate of the signal set is 1 bit/dimension. The Euclidean distance between signals depends on the particular pair of signals. The squared Euclidean distance between s_0 and s_1 is calculated as follows.

$$\begin{aligned} d_E^2(\mathbf{s_0},\mathbf{s_1}) &= ||\mathbf{s_0}-\mathbf{s_1}||^2 \\ &= (s_{0,0}-s_{1,0})^2+(s_{0,1}-s_{1,1})^2 \\ &= (A-(-A))^2+(A-A)^2 \\ &= 4A^2. \end{aligned}$$

Similarly

$$\begin{aligned} d_E^2(\mathbf{s_0},\mathbf{s_2}) &= ||\mathbf{s_0}-\mathbf{s_2}||^2 \\ &= (s_{0,0}-s_{2,0})^2+(s_{0,1}-s_{2,1})^2 \\ &= (A-(-A))^2+(A-(A))^2 \\ &= 8A^2. \end{aligned}$$

The squared distance between $\mathbf{s_0}$ and $\mathbf{s_3}$ is the same as that between $\mathbf{s_0}$ and $\mathbf{s_1}$. The squared distance between $\mathbf{s_0}$ and $\mathbf{s_2}$ is $8A^2$. The minimum squared Euclidean distance of this signal set is $4A^2$. The rate of the signal set is $r = 2$ bits/ 2 dimensions =1 bits/dimension. The bandwidth efficiency of the signals is $R/W = 2r = 2$ bits/second/Hz.

Example 2: Another signal set is shown in Figure 1.3. This is called 32 amplitude phase shift keying (32 APSK). In this signal set there are 32 signals that can be used to communicate $\log_2(32) = 5$ bits of information. Some of these signals are shown in Table 1.1. The other signals can be determined

Bits	Signal	I Component	Q Component	Energy
00000	$\mathbf{s_0}$	+1.000	+1.000	2.000
00001	$\mathbf{s_1}$	-1.000	+1.000	2.000
00010	$\mathbf{s_2}$	-1.000	-1.000	2.000
00011	$\mathbf{s_3}$	+1.000	-1.000	2.000
00100	$\mathbf{s_4}$	+3.732	+1.000	14.928
00101	$\mathbf{s_5}$	+2.732	+2.732	14.928
00110	$\mathbf{s_6}$	+1.000	+3.732	14.928
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
10000	$\mathbf{s_{16}}$	+5.864	+0.000	34.383
10001	$\mathbf{s_{17}}$	+5.417	+2.244	34.383
10010	$\mathbf{s_{18}}$	+4.146	+4.146	34.383
10011	$\mathbf{s_{19}}$	+2.244	+5.417	34.383
10100	$\mathbf{s_{20}}$	+0.000	+5.864	34.383
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
11111	$\mathbf{s_{31}}$	+5.417	-2.244	34.383

Table 1.1: Signal points for Example 2

based on the symmetry of the constellation. The inner four points lie on a circle of radius $r_1 = \sqrt{2} = 1.414$. The signals $\mathbf{s_4}, \mathbf{s_5}, ..., \mathbf{s_{15}}$ lie on a circle with radius $r_2 = \sqrt{2}/\sqrt{1-\cos(\pi/6)} = 3.864$. The signals $\mathbf{s_{16}}, \mathbf{s_{17}}, ..., \mathbf{s_{31}}$ lie on a circle of radius $r_3 = r_2 + 2 = 5.864$. This modulation is called 32-ary amplitude and phase shift keying or 32-APSK for short since the information is contained in both the amplitude and the phase of the signal.

The signal constellation has three groups of signals with each group having a different energy. The second group with 12 signals ($\mathbf{s_4}$,...,$\mathbf{s_{15}}$) has energy that is about 7.46 times as large as the first group ($\mathbf{s_0}, \mathbf{s_1}, \mathbf{s_2}, \mathbf{s_3}$). The last group with 16 signals has energy that is about 17.19 times larger than the first group. The average energy and the average energy per bit of information then is

$$\begin{aligned} \bar{E} &= (4(2) + 12(14.93) + 16(34.39))/32 \\ &= 23.04 \\ E_b &= \bar{E}/5 = 4.61. \end{aligned}$$

The squared distance between $\mathbf{s_0}$ and $\mathbf{s_1}$ is 4. The squared distance between $\mathbf{s_4}$ and $\mathbf{s_5}$ is also 4. The squared distance between $\mathbf{s_5}$ and $\mathbf{s_{18}}$ is also 4. So the radius of the middle circle is chosen to guarantee a separation between points in the middle circle of 4. The radius of the outer circle is chosen to guarantee that the separation between points in the middle circle and points in the outer circle is also 4. The squared distance between several pairs of signals for this example are shown in Table 1.2. The minimum squared Euclidean distance for this signal set is 4. The average energy per bit (with signals used equally likely) is 4.608 and the rate of the signal set is $r = 5$ bits/2 dimensions. The bandwidth efficiency is $R/W = 5$ bits/second/Hz.

Example 3: As another example consider the signals shown in (1.4). In this example there are eight

First Signal	Second Signal	Squared Distance
$\mathbf{s_0}$	$\mathbf{s_1}$	4.00
$\mathbf{s_0}$	$\mathbf{s_4}$	7.46
$\mathbf{s_0}$	$\mathbf{s_5}$	6.00
$\mathbf{s_4}$	$\mathbf{s_5}$	4.00
$\mathbf{s_5}$	$\mathbf{s_{18}}$	4.00
$\mathbf{s_{16}}$	$\mathbf{s_{17}}$	5.23

Table 1.2: Distance between pairs of signals for Example 2

signals in eight dimensions. The eight signals described by the vectors below can be used to communicate 3 bits of information using eight dimensions. The rate of the signal set is $r = 3$ bits/8 dimensions=3/8 bits/dimension which would corresponds to a bandwidth efficiency of $R/W = 3/4$ bits/second/Hz. The energy used by each signal is $E = 8A^2$. The energy per bit is $E_b = 8A^2/3$.

$$\begin{aligned}
000 &\to \mathbf{s_0} = A(+1,+1,+1,+1,+1,+1,+1,+1) \\
001 &\to \mathbf{s_1} = A(+1,-1,+1,-1,+1,-1,+1,-1) \\
010 &\to \mathbf{s_2} = A(+1,+1,-1,-1,+1,+1,-1,-1) \\
011 &\to \mathbf{s_3} = A(+1,-1,-1,+1,+1,-1,-1,+1) \\
100 &\to \mathbf{s_4} = A(+1,+1,+1,+1,-1,-1,-1,-1) \\
100 &\to \mathbf{s_5} = A(+1,-1,+1,-1,-1,+1,-1,+1) \\
110 &\to \mathbf{s_6} = A(+1,+1,-1,-1,-1,-1,+1,+1) \\
111 &\to \mathbf{s_7} = A(+1,-1,-1,+1,-1,+1,+1,-1)
\end{aligned} \quad . \tag{1.4}$$

The squared distance between any two distinct signals is

$$d_E^2(\mathbf{s_i},\mathbf{s_l}) = 4(4A^2) = 16A^2, i \neq l.$$

The distance between distinct pairs of signals is the same because any two distinct signals differ in exactly 4 positions and in each position the squared distance is $4A^2$.

We can compare various signal sets in terms of the minimum squared Euclidean distance normalized by the energy per bit. That is, for the same amount of energy per bit, which signal set has signals that are more distinguishable. Equivalently, for the same Euclidean distance (reliability) which signal set requires less energy. The eight signals in Example 3 are farther apart than the 32 signals in Example 2. The minimum squared Euclidean distance divided by the energy per bit for Example 3 is $16A^2/(8A^2/3) = 6$. The minimum squared Euclidean distance divided by the energy per bit for Example 2 is $4/(4.608) = .868$. Thus, for the same average energy per bit the constellation of Example 3 with 8 signals has about about 6.9 times larger squared distance than the signals of Example 2. Thus the signals of the constellation (set of signals) of Example 3 will be easier to distinguish from each other than the constellation of Example 2. However, the rate of the signal set for Example 3 (.375 bits/per dimension) is about 6.67 times smaller than the rate of the signal set of Example 2 (2.5 bits per dimension). To summarize, different signal sets will offer different rates and different minimum squared Euclidean distance normalized to the energy per bit. Table 1.3 compares the rate and normalized distance for the three example signal sets. A communication system might use a signal set with a large rate and small normalized distance if the propagation loss is small (large received energy) and might use a signal set with a small rate and large normalized distance if the propagation loss is large (small received energy).

Example	r =rate (bits/dimension)	$d^2_{E,min}/E_b$
1	1	4
2	2.5	0.868
3	0.375	6

Table 1.3: Rate and normalized distance for different signal sets

There is one more parameter of a signal set that is important in many applications and that is the peak-to-average power ratio. Roughly the peak power is related to the signal with the largest energy while the average power is similarly related to the average energy. The ratio of the two is the peak-to-average power ratio (PAPR). A large PAPR is more difficult to amplify for transmission in an energy efficient manner without cause distortion. However, generally a signal set (like the 32 APSK example) can provide higher bandwidth efficiency (bits/second/Hz).

1.3 Noise

In a communication system noise is typically due, at least in part, to thermal motion of electrons in a receiver and is sometimes called thermal noise. The most widely used model for a communication system is that the received signal has two additive parts; the desired signal and the noise. The received signal is the desired signal to which noise is added. Generally, the noise fluctuates in such a way that it can contain power at all frequencies (up to some large limit). which is known as white noise. The noise is usually Gaussian meaning that the probability density function of the noise at any time is Gaussian (bell shaped density). The combination is called white Gaussian noise (WGN) and a communication system transmitting a signal to which noise is added is often called the ***additive white Gaussian noise channel*** or AWGN channel.

The amount of thermal noise is usually quantified by something called the "two-sided" power spectral density. The power spectral density describes how the noise power is distributed with frequency. The value at a particular frequency of the two-sided power spectral density value is the amount of power in the noise at that frequency per unit bandwidth. For white noise the two-sided noise power spectral density is-Boltzmann's $N_0/2$ Watts/Hz where $N_0 = kT_0$ and k is Boltzmann's constant (1.38×10^{-23} Joules/Kelvin) and T_0 is the temperature in Kelvin. That is, the noise power is the same for all frequencies. Note that the units for N_0 is Joules or Watts/Hz. It is called two-sided because it consists of power at both positive frequencies and negative frequencies. Consider a frequency band centered at a carrier frequency of f_c Hz with one-sided bandwidth of W Hz from $f_c - W/2$ to $f_c + W/2$. Such a bandwidth (for a real signal) also encompasses the frequency band $-f_c - W/2$ to $-f_c + W/2$. The combination is the "two-sided bandwidth" of $2W$ Hz. Then the noise power in this bandwidth is $N_0/2(2W) = N_0W$ which is measured in units of Watts. For example, if $T_0 = 290$ Kelvin (about 62 degrees Fahrenheit) and since Boltzmann's constant is $k = 1.38\times10^{-23}$ (Joule/degree Kelvin) then $N_0 = 4\times10^{-21}$ Watts/Hz. Often this is converted to a dB scale. To convert to dBW/Hz (which is referenced to 1W) it is necessary to take a base 10 logarithm and multiply by 10. For $T_0 = 290$ Kelvin the one-sided power spectral density is N_0 (dBW/Hz)= -204dBW/Hz = -174dBm/Hz where the later is referenced to a power of 1mWatt.

Just as signals can be represented by points in either a two-dimensional or higher dimension space, noise can be represented by points or vectors in space. Typically noise is represented by infinite length vectors where each component is a random variable. However, if the signal is represented by vectors of length 2 then typically only two of the noise components are relevant. For white Gaussian noise each

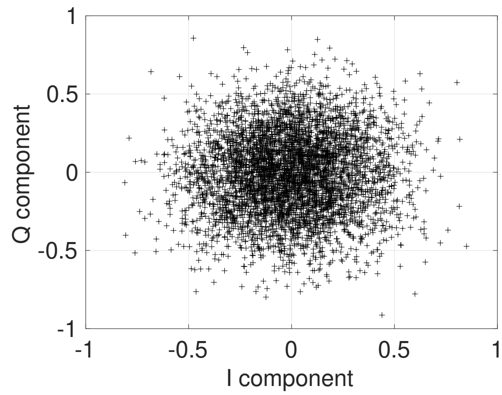


Figure 1.4: Noise samples

random variable in the noise vector has a Gaussian (or normal or bell-shaped) distribution with mean is 0 and variance $N_0/2$. Furthermore the noise in each dimension is independent of the noise in other dimensions. Concepts related to random variables will be reviewed in Chapter 3. In Figure 1.4 we show a collection of noise vectors of length 2 where each component is an independent Gaussian random variable. As can be seen it is more probable that the noise is small (close to the origin) than large. The noise is also circularly symmetric. That is, the noise in any direction has the same variance. A histogram of either the I component or the Q component with enough points generated should match a Gaussian density (bell shape).

1.4 Signals and Noise

Because noise is present in the receiver the received signal is different from the transmitted signal. The received signal is the desired signal plus noise. For example if the signal is a vector of length 2 then the relevant noise is of length 2 and the relevant received signal can be expressed as vector of length 2. Suppose that signal $\mathbf{s_i} = (s_{i,0}, s_{i,1})$ is transmitted. Then the received vector is

$$\begin{aligned} r_0 &= s_{i,0} + \eta_0 \\ r_1 &= s_{i,1} + \eta_1 \end{aligned}$$

where η_0, η_1 is the noise in each of the components of the received signal. In Figures 1.5 and 1.6 the received signal points for different noise levels for QPSK and 32APSK signal constellations are shown. Generally the receiver will try to find which signal point (without noise) is closest to the received point. An error will be made in deciding which signal is sent when the noise causes the received signal point to be closer to one of the signals that was not transmitted. Because the noise is random there is a probability that an error is made. As can be seen as the noise level increases (or the signal-to-noise ratio decreases) the received signal is often as close to an incorrect signal as it is to a correct signal. As such the probability that the receiver will make an error will become large as the noise level increases. Reducing the rate of transmission (with the same energy) will increase the distinguishability of signals and lower the error probability.

In these examples we see that larger distinguishability is possible if the rate of transmission is lower. However, it is possible with the same energy to increase the distinguishability of signals without reducing the rate of transmission. This is accomplished using many dimensions for the signal rather than just two dimensions or eight dimensions.

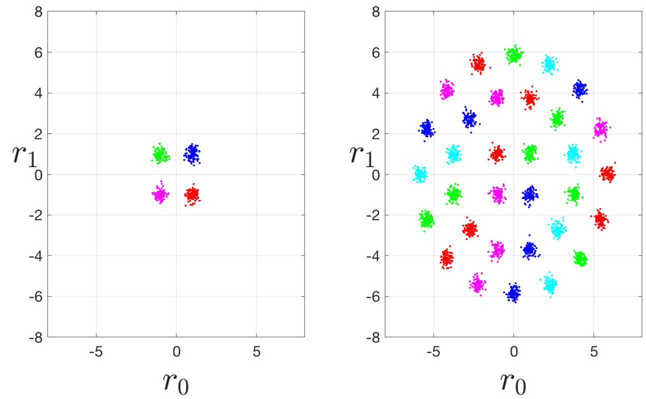


Figure 1.5: Received signals with small noise level (or large signal-to-noise ratio)

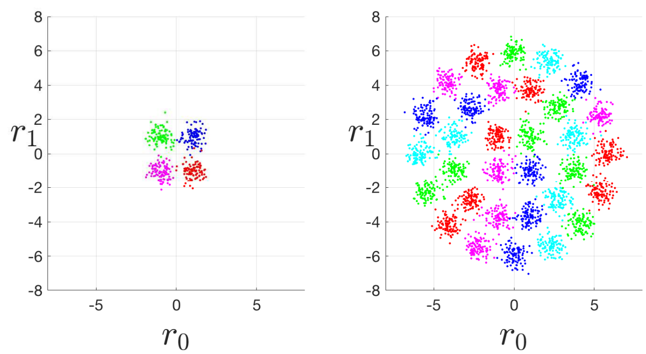


Figure 1.6: Received signals with large noise level (or small signal-to-noise ratio)

1.5 Fundamental Tradeoffs in Communications

There are two fundamental problems in a communication system. The first problem involves determining how large of a rate (bits/second) can information be reliably communicated over a noisy channel with a given amount of bandwidth and energy per bit of information. The information here is a sequence of independent bits which are equally likely to be 0 or 1. Generally, the bandwidth and amount of received energy limit the data rate. If the energy and/or the bandwidth is increased then the data rate can be increased. The second problem involves how accurately can a source of information, such as a voice signal or an image, be represented by a sequence of bits and then reproduced from those bits. As more and more bits are used to represent the source of information the distortion incurred when the source is reproduced from those bits should decrease or equivalently there will be higher fidelity reproduction. These problems translate into tradeoffs. The first problem yields a tradeoff between the rate of transmission and the energy per bit used for transmission. There is some largest data rate that can be achieved with a given amount of energy and a given amount of bandwidth. Similarly, for the second problem, there is some minimum distortion that can be achieved with a given bit rate of data that is used to represent a source of information. Below we quantify these tradeoffs.

1.5.1 First Fundamental Tradeoff: Bandwidth Efficiency-Energy Efficiency Tradeoff

The first fundamental tradeoff is between energy efficiency and bandwidth efficiency. The fundamental tradeoff derived in this section is also a fundamental limit. That is, it describes the best possible performance that can be achieved for a given rate or for a given energy used per bit of information. As such, it is a way to determine how well a particular communication systems performs. Comparing the energy used per bit and the rate of a particularly designed communication system to the fundamental limit determines the potential for further improved designs.

A communication system can be designed with larger bandwidth efficiency (bits/second/Hz) at the expense of an increase in the amount of energy needed per bit of information (Joules per bit). Figure 1.7 shows the system under consideration. A source generates bits at a certain rate R bits per second. Some type of processing id done to generate a waveform with bandwidth W and power P. The channel adds noise. The received signal is processed attempting to reproduce the original bits. Arbitrary processing is allowed between the source and the channel and between the channel and the sink.

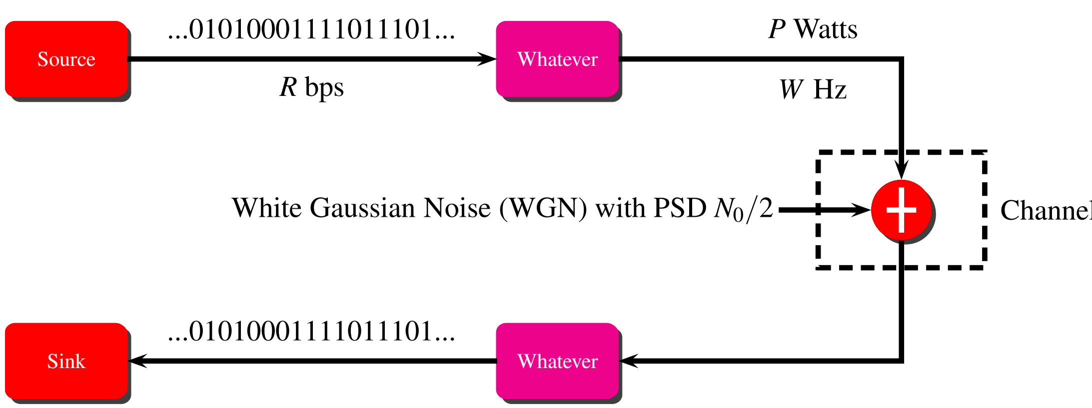


Figure 1.7: One fundamental tradeoff

We now provide the fundamental relationship between the data rate that can be reliably communicated, the received power and the bandwidth. Consider a source that produces equally likely data bits (0s and 1s) at rate R bits/second. A signal (waveform) is transmitted such that the *received* power is P (Watts). The transmitted signal has (absolute) bandwidth W (Hz). Noise is added to the transmitted signal. The noise is white (power at all frequencies of interest), Gaussian and has two-sided power spectral density $N_0/2$ Watts/Hz. The power of the noise in a bandwidth $f_c - W/2$ to $f_c + W/2$ Hz can be determined as follows. The two-sided bandwidth (including positive and negative frequencies) is $2W$ Hz. The two-sided noise power spectral density, $N_0/2$, multiplied by the two-sided bandwidth, $2W$ results in $N_0 W$ Watts of noise power. The channel that just adds Gaussian noise is called an additive white Gaussian noise (AWGN) channel.

In 1948 Claude Shannon published a paper in which he determined the tradeoff between data rate, bandwidth, signal power and noise power for reliable communications for an additive white Gaussian noise (AWGN) channel. Let W be the bandwidth (in Hz), R be the data rate (in bits per second), P be the *received* signal power (in Watts), $N_0/2$ the noise power spectral density (in Watts/Hz). Shannon showed that reliable communication is possible if and only if

$$R < C = W\log_2(1+\frac{P}{N_0 W}). \tag{1.5}$$

The maximum rate C is called the channel capacity.

The result of Shannon (not proven here) states that there exists a communication system (a method of transforming bits into waveforms and processing the received signal to recover the original bits) with arbitrarily small error probability provided the rate R is less than the capacity C. The book by Soni and Goodman puts into perspective the contribution by Shannon [1].

> Before Shannon's 1948 "A Mathematical Theory of Communication," a century of common sense and engineering trial and error said that noise–the physical world's tax on our messages–had to be lived with. Shannon proved that noise could be defeated, that information sent from Point A could be received with perfection at Point B, not just often, but essentially always. He gave engineers the conceptual tools to digitize information and send it flawlessly, a result considered hopelessly utopian up until the moment Shannon proved it was not.

For large values of W the maximum rate (capacity) approaches

$$\lim_{W\to\infty} W\log_2(1+\frac{P}{N_0 W}) = \frac{P}{N_0 \ln(2)} = 1.4426(P/N_0). \tag{1.6}$$

Thus for large W (much larger than P/N_0) the achievable rate grows linearly with P/N_0. In Figure 1.8 the capacity in bits/second is plotted against the bandwidth for different values of P/N_0. For W about 30 times P/N_0 the asymptotic limit is fairly accurate (e.g. within 2 percent).

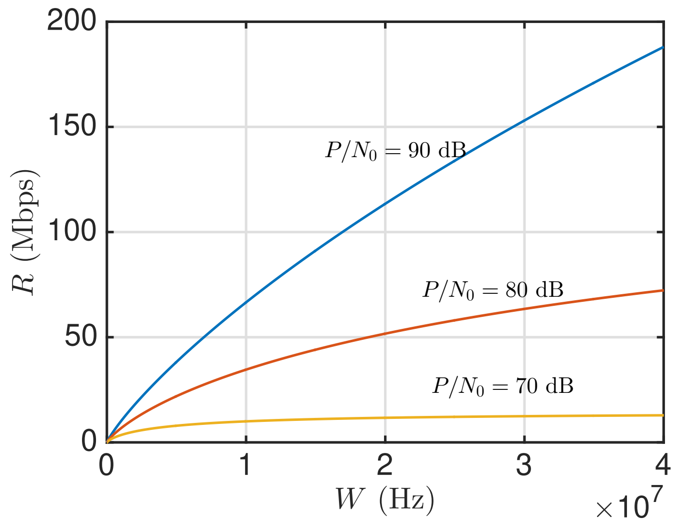


Figure 1.8: Capacity as a function of bandwidth W

It is instructive to rewrite the capacity in terms of the energy per bit rather than the power. Let E_b be the energy received per bit of information. The energy per bit, E_b, is the product of the received power, P, and the duration of a bit, T. The duration of a bit is the inverse of the date rate: $T = 1/R$. Thus

$$E_b = P/R \quad \text{or} \quad P = E_b R.$$

Using this relation we can express the capacity formula as

$$R/W < \log_2(1 + \frac{E_b}{N_0}\frac{R}{W}).$$

Solving this for E_b/N_0 we obtain the required signal-to-noise ratio for reliable communications in terms of either the bandwidth efficiency R/W or the rate in bits/dimension r:

$$E_b/N_0 > \frac{2^{R/W}-1}{R/W} = \frac{2^{2r}-1}{2r}. \tag{1.7}$$

$$\tag{1.8}$$

In (1.7) R/W has units of bits/second/Hz and is the bandwidth efficiency of the communication system. It is the rate per unit bandwidth that information is communicated. The interpretation is that reliable communication is possible with *bandwidth efficiency* R/W provided that the *signal-to-noise ratio*, E_b/N_0, is larger than the right hand side of (1.7). As can be seen from (1.7), the larger the bandwidth efficiency, the larger required signal-to-noise ratio.

For small values of R/W the smallest value of E_b/N_0 where reliable communication is possible is $\ln(2) = 0.693$. That is,

$$\min_R \frac{2^{R/W}-1}{R/W} = \lim_{R/W \to 0} \frac{2^{R/W}-1}{R/W} = \ln(2).$$

This also follows from (1.6).

When the range of values for energy or power are vast we usually employ a dB scale. The conversion when dealing with energy or power (as opposed to current or voltage) is

$$E_b/N_0(\text{dB}) = 10\log_{10}(E_b/N_0).$$

The smallest signal-to-noise ratio for reliable communication (at low rates) is

$$E_b/N_0 > \log(2) = 0.693$$
$$E_b/N_0(\text{dB}) > 10\log_{10}(0.693) = -1.59\text{dB}.$$

There are several ways to graphically depict the tradeoff described by the capacity formula. The capacity formula describes an achievable region. That is, the capacity formula describes a set of values for E_b/N_0 and R/W for which there exists a communication system. Another result of Shannon is that if the rate is larger than the capacity then every communication system has error probability approaching 1 when the number of dimensions of the signal becomes large. So the complement of the achievable region is a region where the error probability will approach 1 for large dimensional signals. We will just label this region as "not achievable."

Figure 1.9 shows the capacity as a function of the signal-to-noise ratio E_b/N_0. For each E_b/N_0 (dB) there is a maximum bandwidth efficiency R/W that is possible. Communication system exist with R/W and E_b/N_0 (dB) below or to the right of the curve. Figure 1.10 shows the minimum signal-to-noise ration E_b/N_0 (dB) as a function of the R/W. For each R/W there is a minimum signal-to-noise ratio E_b/N_0 (dB) for which reliable communication systems exist. Communication systems exist with R/W and E_b/N_0 (dB) above or to the left of the curve. Table 1.4 shows for various R/W the threshold signal-to-noise ratio, E_b/N_0 (dB), where the error probability can be made arbitrarily small. When the signal-to-noise ratio (E_b/N_0) is above the threshold there exists a communication system of a given rate which can have arbitrarily small error probability. So for a communication system using a bandwidth

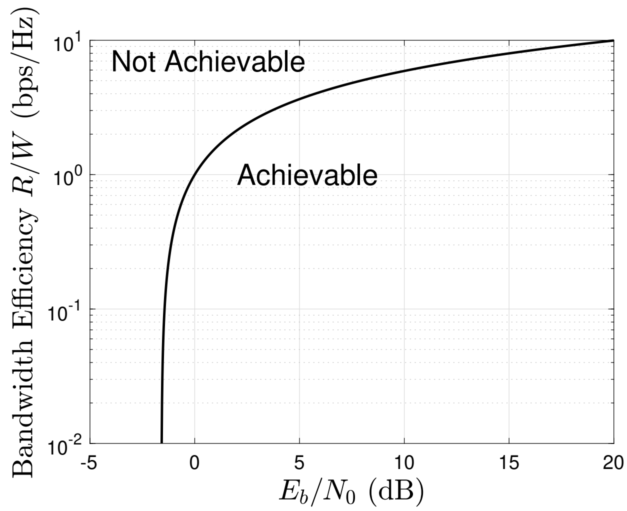


Figure 1.9: Maximum bandwidth efficiency (R/W) versus E_b/N_0 (dB).

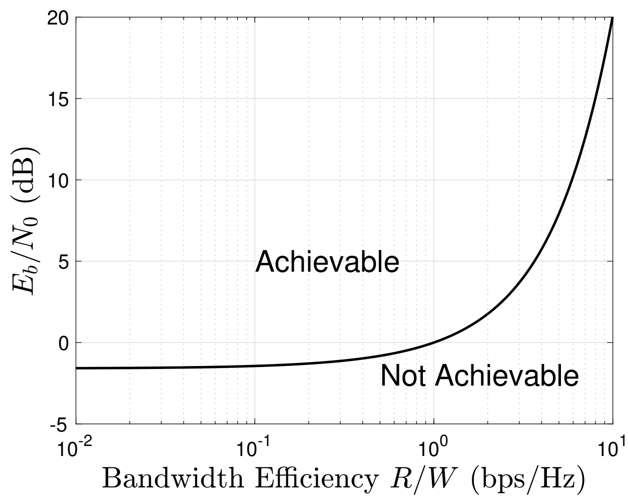


Figure 1.10: Minimum signal-to-noise ratio E_b/N_0 (dB) versus R/W

efficiency of 1 bps/Hz a minimum of 0dB for E_b/N_0 (dB) is required to achieve arbitrarily small error probability.

The capacity formula only provides a tradeoff between energy efficiency and bandwidth efficiency. Complexity is essentially infinite, as is delay. The model of the channel is rather benign in that no signal fading is assumed to occur. The capacity theorem says that we can communicate with error probability near zero at rates below the capacity or equivalently at values of E_b/N_0 above a threshold that depends on the rate.

Example 4: A WiFi router with a single transit antenna and receive antenna uses 20MHz of bandwidth. The transmit power is 250mWatts (24dBm) but the received power is only -85dBm or 3.16 pW because of propagation loss. The noise power spectral density level is $N_0 = -174$dBm/Hz$= -204$ dBW/Hz. Determine the largest data rate for which it is possible to have reliable communications.

$$\begin{aligned} N_0 W &= 10^{(-204/10)}(20 \times 10^6) = 7.962 \times 10^{-14} \\ P/(N_0 W) &= 10^{(-115/10)}/(7.962 \times 10^{-15}) = 39.72. \end{aligned}$$

R/W	Minimum E_b/N_0 (dB)	R/W	Minimum E_b/N_0 (dB)
0	-1.59	1.0	0
0.1	-1.44	2.0	1.76
0.2	-1.26	3.0	3.68
0.3	-1.13	4.0	5.74
0.4	-0.98	5.0	7.92
0.5	-0.82	6.0	10.21
0.6	-0.66	7.0	12.58
0.7	-0.50	8.0	15.03
0.8	-0.33	9.0	17.54
0.9	-0.17	10.0	20.10

Table 1.4: Threshold signal-to-noise ratio

Alternatively we can do the calculations in dB.

$$
\begin{aligned}
[N_0 W]_{\text{dBW}} &= -204+73=-131 \text{ dBW} \\
[P]_{\text{dBW}} &= -115 \text{ dBW} \\
[P/(N_0 W)]_{\text{dB}} &= -115+131=16 \text{ dB}=39.72.
\end{aligned}
$$

The maximum rate is then calculated as follows.

$$
\begin{aligned}
R &< W\log_2\left(1+\frac{P}{N_0 W}\right) \\
R &< 20\times 10^6 \log_2(1+39.72) \\
&= 107 \text{ Mbps} \\
R/W &= 5.35\text{bps/Hz}.
\end{aligned}
$$

While it is beyond the scope of this book to prove the capacity result of Shannon, it is interesting to understand the methodology used in the proof. First Shannon made what is known as a random coding argument which is the following. First, consider every possible communication system and then determine the average of the error probability when averaged over all possible communication systems. Then claim that at least one of the communication systems would perform at least as well as the average performance. While it is difficult to compute the performance of every possible communication system it is not so difficult to compute the average performance over all communication systems. Shannon considered a set of M signals in N dimensions and determined the average performance over all possible sets of signals with a receiver that is optimal. Second, Shannon let the number of dimensions N become large as well as the number of signals M in a way such that the rate of communications ($r=\log_2(M)/N$) in bits/dimension approaches a constant r. If the rate is smaller than the capacity then as N and M become large the average error probability goes to zero. As such at least one of the communication systems has error probability approaching zero; they cannot all have error probabilities bounded away from 0 while the average goes to 0. In Section 6.6 we will demonstrate a similar analysis that shows existing of signals with error probability approaching 0 at rates less than something smaller than the capacity. So while Shannon showed the existence of a communication system with arbitrarily small error probability provided the rate is less than the capacity, he did not identify any communication system that could achieve what he showed was possible. It has been the job of communication engineers since the discovery of Shannon to find communication systems that are close to achieving the capacity of the channel but are not so complicated that they are not practical. Communication engineers have come within a few tenths of a dB in finding practical ways to achieve what Shannon said was possible but it took more that 50 years for this to occur. This accomplishment was the result of better signal design and the significant advance in digital processing chips.

1.5.2 Second Fundamental Tradeoff: Rate-Distortion Tradeoff

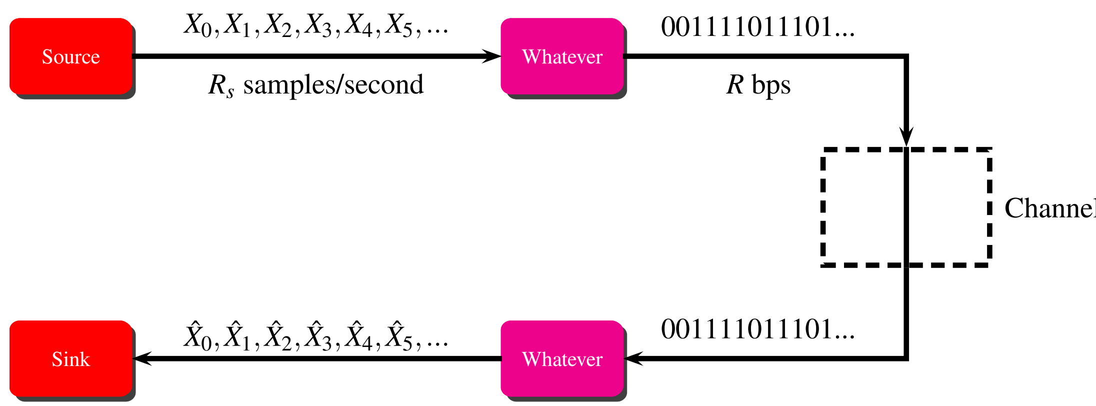


Figure 1.11: Source Coding Framework

Next we consider the tradeoff between representing a source of information accurately and the rate at which bits are used to represent the source as shown in Figure 1.11. In the system shown a source of information (either bits or real numbers or waveforms) is to be communicated to a sink by representing the source by bits. And then at the sink recovering the original information from the bits. The processing at the source side is done by a *source encoder*. The processing at the sink side is done by a *source decoder*. The basic idea is to represent a source of information by encoded bits produced by a source encoder. The source encoded bits are either stored or transmitted. These bits are then either retrieved from memory or received from the transmitter without error. We assume that the solution of the first problem has resulted in a method of communicating the information perfectly. After retrieving/receiving these source encoded bits, the source decoder attempts to reconstruct the original source of information. The larger the number of source encoded bits that are used to represent the source output, the greater the accuracy in reproducing the original source output at the source decoder output. Arbitrary processing is allowed in generating the bits from the source and arbitrary processing is allowed in attempting to reproduce the original source information in the decoder. In this section we consider two examples and demonstrate the trade-off between the rate at which the source (number of source encoded bits per source sample) is represented and the accuracy at which the source decoder can recreate the original source sequence.

Gaussian Source

Suppose a source produces samples at a rate of R_s samples/second. Each sample is a Gaussian distributed random variable (mean 0, variance σ^2). The samples are independent. For example,

$$X_0 = -0.4326,\ X_1 = -1.6656,\ X_2 = +0.1253,$$
$$X_3 = +0.2877,\ X_4 = -1.1465,\ X_5 = +1.1909, \ldots$$

We want to represent these samples with a sequence of bits at rate R bps. So there would be R/R_s bits per sample in the representation. So if $R/R_S = 3.5$ then we represent each sample with 3.5 bits (e.g. 2 samples represented with 7 bits). The goal is to as accurately as possible (minimal distortion) reproduce the source samples with the fewest possible bits per sample or the lowest bit rate R. After the bits have been retrieved perfectly at a destination an estimate $\hat{X}_i$ of the original source sample X_i is created. The distortion is the mean-squared-error between the original source sample and the reproduction of that sample; that is, $D = E[(X_i - \hat{X}_i)^2]$. Shannon showed that reproduction of a Gaussian source with mean-squared distortion D is possible provided

$$R/R_s > \begin{cases} \frac{1}{2}\log_2(\frac{\sigma^2}{D}), & D < \sigma^2 \\ 0, & D > \sigma^2. \end{cases}$$

The distortion is often measured on a log scale and expressed in deciBels. $D_{dB} = 10\log_{10} D$. If R/R_s increases by 1, that is, there is one extra bit per sample, then the minimal possible distortion will decrease by a factor of 4 (6dB). Figure 1.12 shown the tradeoff between the rate and distortion for a Gaussian source and a mean-squared error distortion criteria.

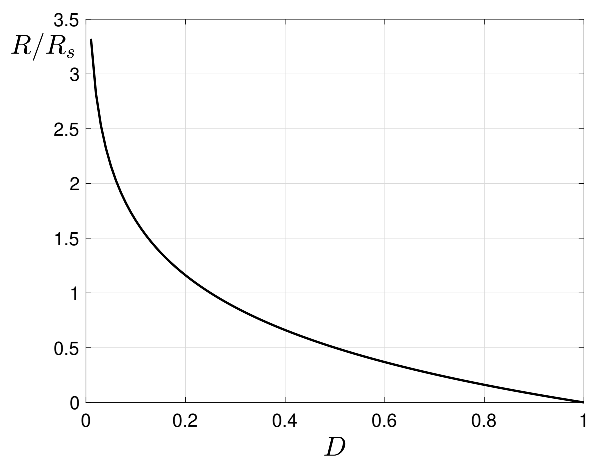


Figure 1.12: Rate-distortion function

Binary Source

As a second example, suppose the source samples X_i are bits (0 or 1) with $P\{X_i = 1\} = p_s$, $P\{X_i = 0\} = 1 - p_s$. Is it possible to encode or represent (store or transmit) the output of the source using less then 1 encoded bit per source output? If we want to represent the source output accurately (without any errors) then we need to store or transmit

$$R = H_2(p_s) = -p_s\log_2(p_s) - (1-p_s)\log_2(1-p_s)$$

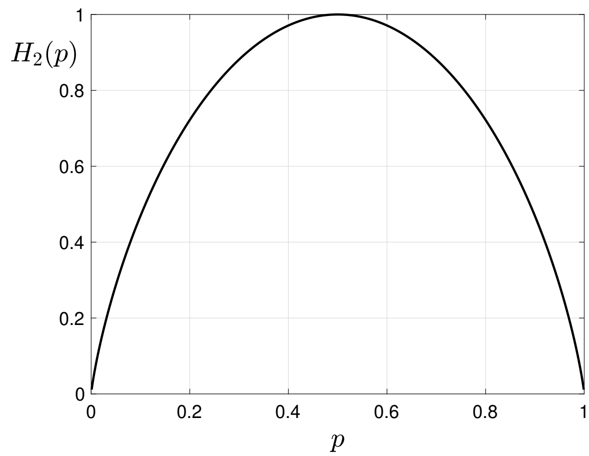


Figure 1.13: Binary entropy function

bits for every source sample (bit). If $p_s = 1/2$ then $H_2(p_s) = 1$ and 1 bit is required for each source sample. If $p_s = 0$ then the source always produces zeros and we don't need to transmit any bits because we know in advance that the bits are always 0. Similarly if $p_s = 1$ we don't need to transmit any bits ($R = 0$) because $H_2(1) = H_2(0) = 0$. The entropy function is shown in Figure 1.13.

Now consider that we only want a certain reliability, that is, if we only want the source bits that were encoded and then stored or transmitted to be reconstructed correctly with at least a certain probability $1 - p$ or we want the error probability to be less than p. Then could we store or transmit fewer encoded bits? If so, how many bits are required to store if we only want a probability of correct to be, say 99%? Clearly if we only store or transmit 98 out of every 100 source bits and guess the two bits that are not stored or transmitted we would have a probability of correct of 99/100 because on average we would guess one of the two bits we did not store correctly. Is this the best that can be done? The answer is no. We can do better. Suppose we require a probability bit correctly recovering a sample (source bit) from the encoded bits of $(1 - p)$ so the probability of incorrectly recovering a sample (source bit) is p. If the source produces samples (bits) with probability p_s of a 1 and $1 - p_s$ of a 0, then the rate R/R_s per source sample at which we need to represent those source samples (bits) must satisfy

$$R/R_s > H_2(p_s) - H_2(p)$$

where $H_2(x) = -x\log_2(x) - (1-x)\log_2(1-x)$. If the source has equally likely bits (that is, $p_s = 1/2$) then $H_2(p_s) = 1$ and the required rate is

$$R/R_s > 1 - H_2(p)$$

We focus the rest of the discussion on this case of $p_s = 1/2$. If the required error probability p is only 1/2 (not a practical case) and every source bit is equally likely ($p_s = 1/2$) then we can just guess every bit and so we don't need to store or transmit any encoded bits. If the required error probability is larger than 1/2 then we still need to encoded some bits. This is because, recovering the bits with error probability p greater than 1/2 is the same as recovering the bits with error probability $1 - p$ since we can easily flip all the bits. That is, switch every 0 to 1 and every 1 to 0. When we do this we get error probability $1 - p$. So the rate at which we need to represent the source is the same whether we require error probability 1% or 99%. In Figure 1.14 the rate at which data can be communicated with reliability p is shown. For example, if the desired reliability for reconstruction was 99% (1% error rate) this shows that we need only to store/transmit at a rate of about 0.92 encoded bit/source bit. A 98% reliability (2% error rate)

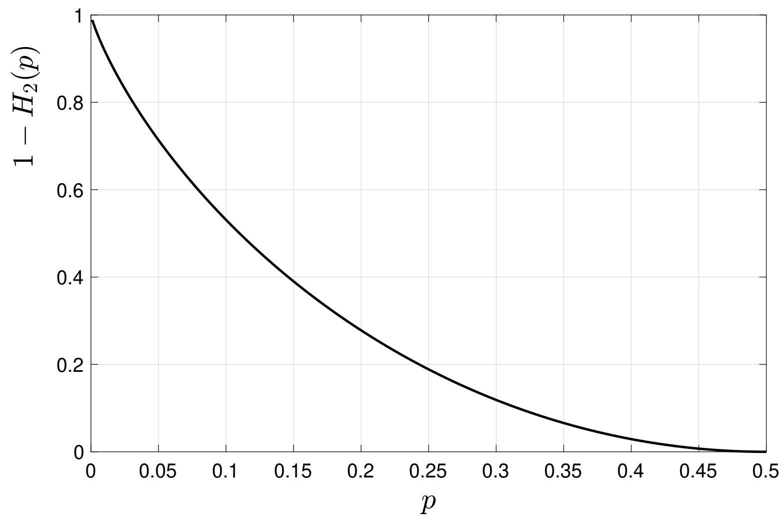


Figure 1.14: Largest rate which a binary source can be compressed

requires a rate of about 0.86 encoded bit/source bit. A reliability of 89% (11% error rate) can be obtained with transmitting at a rate of only 0.50 encoded bit/source source bit.

1.5.3 Joint Source-Channel Coding Problem

These two fundamental tradeoffs can be combined into a single relation indicating the possibility of transmission with a certain reliability given a certain amount of resources. Suppose for a communication channel we have a certain amount of (received) power, a given amount of bandwidth and a certain noise level. We wish to communicate a source of information over this channel with as low of a distortion as possible. The bandwidth, power and noise level limit how many bits/second we can communicate reliably. So the representation of the source of information with bits should not exceed this limit. The limit on how fast we can transmit information reliably will limit the accuracy that can be achieved in reconstructing the source at the destination.

Gaussian source-Gaussian channel

Suppose we have a Gaussian source with (independent) samples at rate R_s samples/second, with variance σ^2. Suppose we have an additive white Gaussian noise channel with power spectral density $N_0/2$. Suppose we have signals with power P and bandwidth W. What distortion is possible?

Consider first compressing the source to a rate of

$$R_1 = \frac{R_s}{2}\log_2(\frac{\sigma^2}{D})$$

and then transmitting these over a channel with capacity (maximum achievable rate) of

$$C_1 = W\log_2(1+\frac{P}{N_0 W}).$$

As long as the rate R_1 of the bits generated after compression is smaller than the capacity C_1 of the channel it is possible to communicate the source to the sink with the desired distortion. That is, the possible distortion is determined by comparing the above two quantities.

$$\begin{aligned}
\frac{R_s}{2}\log_2(\frac{\sigma^2}{D}) &< W\log_2(1+\frac{P}{N_0W}) \\
D &> \frac{\sigma^2}{(1+\frac{P}{N_0W})^{2W/R_s}} \\
D &> \frac{\sigma^2}{(1+\frac{E_s}{N_0}\frac{R_s}{W})^{2W/R_s}}
\end{aligned}$$

where $E_s = P/R_s$ is the energy used per source sample. The transmission of the bits generated by the source encoder is reliable. So the only distortion comes from the source encoder. The channel limits the rate at which reliable communication is possible. This puts a constraint on the source encoder in terms of the maximum rate that the source can compress the information.

Binary source-Gaussian channel

Another example is a binary source transmitted over a Gaussian channel. This scenario is important when determining the fundamental limits for data transmission including the reliability. What are the fundamental limits of the rate at which a binary source can be communicated over an additive white Gaussian channel if the error probability is not required to be zero? In this case we must combine the theory for source coding with that of channel coding/modulation. If we start with a source producing data at a rate of R_s bits/second with probability 1/2 of the data being 0 and probability of 1/2 of the data being 1 then we can compress that data source into a source with a smaller rate by allowing some errors in the reproduction. If we require the error probability to be $P_{e,b}$ or less then we can compress the data to the rate

$$\begin{aligned}
R_2 &= R_s(1-H_2(P_{e,b})) \\
&= R_s(1+P_{e,b}\log_2(P_{e,b})+(1-P_{e,b})\log_2(1-P_{e,b})).
\end{aligned}$$

These compressed bits can be reliably communicated over a channel with bandwidth W Hz using power P watts provided

$$R_2 < C_2 = W\log_2(1+\frac{P}{N_0W}).$$

Thus error probability $P_{e,b}$ is possible provided

$$R_s(1+P_{e,b}\log_2(P_{e,b})+(1-P_{e,b})\log_2(1-P_{e,b})) < W\log_2(1+\frac{P}{N_0W}).$$

We can express this in terms of energy per source information bit by writing $E_b = P/R_s$. Using this relation the condition for achieving error probability $P_{e,b}$ is

$$1+P_{e,b}\log_2(P_{e,b})+(1-P_{e,b})\log_2(1-P_{e,b}) < \frac{W}{R_s}\log_2(1+\frac{E_b}{N_0}\frac{R_s}{W}).$$

We can express this in terms of a required signal-to-noise ratio as a function of the desired bit error probability as

$$E_b/N_0 > \frac{2^{R_s/W(1-H_2(P_{e,b}))}-1}{R_s/W} \tag{1.9}$$

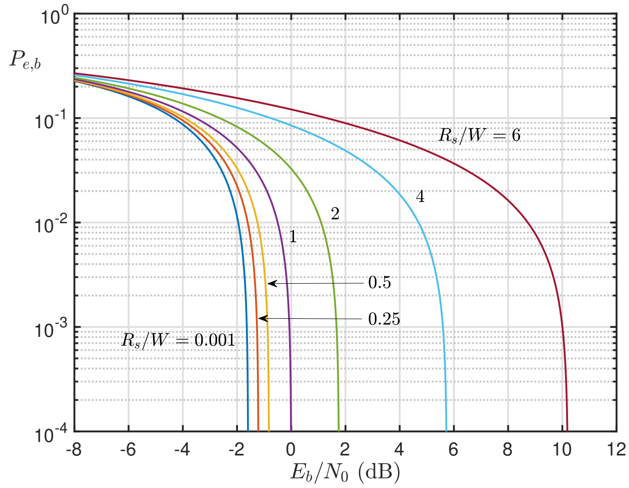


Figure 1.15: Fundamental limit on error probability and signal-to-noise ratio for different bandwidth efficiencies

where

$$H_2(P_{e,b}) = -P_{e,b}\log_2(P_{e,b}) - (1-P_{e,b})\log_2(1-P_{e,b})$$

and is called the binary entropy function.

The error probability, $P_{e,b}$, possible as a function of the signal-to-noise ratio E_b/N_0 is shown in Figure 1.15. This is shown for different values of R_s/W. There is a minimum signal-to-noise ratio at very low values of R_s/W where reliable communication is possible. This minimum is $\ln(2) = -1.59$dB.

Now we compare the fundamental limits derived above with some practical modulation techniques. Considering again the examples of 4 constellation points transmitting 2 bits of information and 32 constellation points transmitting 5 bits of information. The error probability as a function of signal-to-noise ratio (E_b/N_0) is shown in Figure 1.16 and 1.17. The rate of communications is 2 bits/second/Hz and 5 bits/second/Hz, respectively. The smallest possible error probability for a system transmitting 2 and 5 bits/second per Hertz ($R_s/W = 2,5$) is also shown in Figures 1.16 and 1.17. One point to note is that for error probabilities below about 10^{-3} the smallest value of E_b/N_0 is determined by just the capacity limit. For example, for $R_/W = 2$ the smallest E_b/N_0 from the capacity $((2^{R/W}-1)/(R/W)$ is 1.76dB and matches the plot in Figure 1.16.

There is a 7.83dB gap in required signal-to-noise ratio for a error probability around 10^{-5} between what is achieved with the 4 signal points and what is theoretically possible. There is a 10.21dB gap in required signal-to-noise ratio for a error probability around 10^{-8} between what is achieved with the 4 signal points and what is theoretically possible. There is a 7.92dB gap in required signal-to-noise ratio for a error probability around 10^{-5} between what is achieved with the 32 signal points and what is theoretically possible. There is a 10.39dB gap in required signal-to-noise ratio for a error probability around 10^{-8} between what is achieved with the 32 signal points and what is theoretically possible. The goal of part of this book is to show how this performance gap can be decreased. That is, how can we obtain a given error probability at the same rate with smaller signal-to-noise ratio.

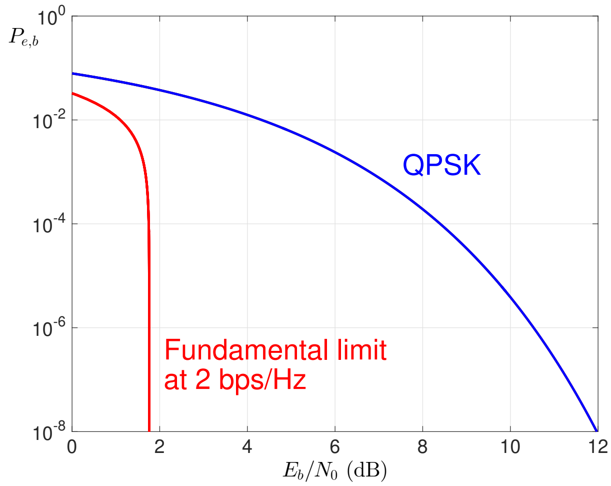


Figure 1.16: Fundamental limit on error probability and signal-to-noise ratio for bandwidth efficiency 2 bps/Hz compared to QPSK

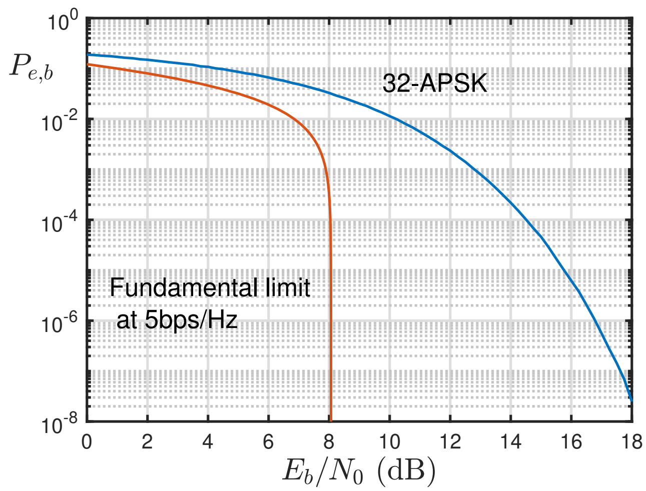


Figure 1.17: Fundamental limit on error probability and signal-to-noise ratio for bandwidth efficiency 5 bps/Hz compared to 32-APSK

1.6 Conversion to decibels (dB) units

Because of the large range of power levels (transmitted versus received) power is often expressed using a logarithmic scale. This is called the decibel scale. The decibel scale expresses the ratio of two powers (or energies). An absolute power level can be represented by a ratio to a known power level. For example, a power of 100 Watts is expressed as a ratio of 100/1 where the 1 is a one Watt reference. The conversion of a ratio of power levels to dB units is done by taking the base 10 logarithm of the power level and then multiplying by 10. So a ratio of 100 Watts/1Watt would correspond to a 20 dB power ratio. To express an absolute power level, we just need to know the reference level. There are two common reference levels; 1 Watt and 1 mWatt. When the reference is 1 Watt the power ratio is written as dBW. When the reference is 1 mWatt the power ratio is written as dBm. So a 0dBm power level corresponds to 1mWatt. A 30dBm power level is the same as a 0 dBW power level, namely 1 Watt. A -40dBW is the same as a -10dBm power level and corresponds to 0.1mWatt.

Summary of Chapter 1 Concepts:

- In a digital communication system a finite number of possible waveforms are sent during a finite time interval representing a finite number of bits.
- Transmission of information using digital communication has significant advantages over analog communication because of the processing potential of digital circuits and the regeneration potential of digital signals.
- Digital communication requires synchronization not required in analog communications system to obtain timing of the start/end of data symbols or bits.
- In digital communications transmitting one of M signals can communication $\log_2(M)$ bits of information.
- In a digital communication system M signals are often represented by M vectors of length N.
- The Euclidean distance between the vectors represent the distinguishability of the signals.
- The squared length of the vectors represents the energy used to transmit the signal.
- The rate of communication in bits per dimension is $r = \log_2(M)/N$; that is, the base 2 logarithm of the number of vectors divided by the length of the vectors represents the rate of communications.
- There are $2W$ dimensions per second available to represent signals in a bandwidth of W Hz. That is, signals of duration T seconds occupying a bandwidth of W Hz can be represented with vectors with length $2WT$.
- The rate of reliable communications is limited by either the bandwidth available or the received signal-to-noise ratio (i.e. the ratio of energy per bit to noise power spectral density).
- It is possible to have arbitrarily reliable communication provide the rate of transmission in bits/second is smaller than the capacity of the channel. Equivalently, reliable communication is possible provided the ratio of received energy per bit to the noise power spectral density (E_b/N_0) is larger than $(2^{R/W} - 1)/(R/W)$ where R/W is the ratio of data rate in bits/second to bandwidth W and is known as the bandwidth efficiency.
- There is a fundamental tradeoff between energy needed for reliable communications and the bandwidth efficiency in bits/second/Hz.
- There is a fundamental tradeoff between the rate in bits/second at which a source is represented by a stream of bits and the distortion incurred when reproducing the source from the bit stream.
- In the overall problem of communicating a source to a sink in a given bandwidth and with a given power there is a minimum distortion that can be achieved no matter what type of processing is used to communicate the source (origin of information) over the channel to the sink (destination of the information).

Key Notation and Relations

- Number of signals: M.
- Number of information bits communicated using one of M signals is $\log_2(M)$.
- Number of dimensions: N.
- Energy per bit (Joules per bit): E_b.
- Power (Watts=Joules/second): P
- Bandwidth (Hz): W
- Time duration of signals (s): T
- Rate of a signal set (bits/dimension): $r = \log_2(M)/N$.
- Rate of communications (bits/second): $R = \log_2(M)/T$.
- Relation between bandwidth (Hz=1/second), time (seconds) and number of dimensions (unitless): $N = 2WT$.
- Number of dimensions per second $N/T = 2W$.
- Relation between rate of communications R (bits/second), rate of a signal set r (bits/dimension) and bandwidth W: $R/W = 2r$,
- Relation between power P (Joules/second), energy per bit E_b (Joules/bit) and rate R (bits/second): $E_b = P/R$, $P = E_b R$.
- Noise power spectral density (Watts/Hz or Joules): $N_0/2 = kT_0/2$, k is Boltzmann's constant $=1.38 \times 10^{-23}$ (Joule/degree Kelvin), T_0 is temperature in Kelvin.
- Signal-to-noise ratio E_b/N_0 (unitless).
- Capacity (bits per second): $R < C = W \log_2(1 + \frac{P}{N_0 W})$
- Capacity (bits per dimension): $r < c = \frac{1}{2}\log_2(1 + (\frac{2E}{N_0}))$ where E is the energy per dimension so that $E = E_b r$ and E_b is energy per bit.
- Capacity Theorem: Reliable communication is possible if and only if $R < C$ or $r < c$ or

$$
\begin{aligned}
E_b/N_0 &> \frac{2^{R/W} - 1}{R/W} \\
E_b/N_0 &> \frac{2^{2r} - 1}{2r}.
\end{aligned}
$$

1.7 Problems

1. A communication system transmits one of 8 equally likely signals. The signals (waveforms) are represented by the vectors shown below.

   (a) Determine how many information bits can be sent using these signals.

   (b) Determine the energy of each of the signals and the average energy per information bit.

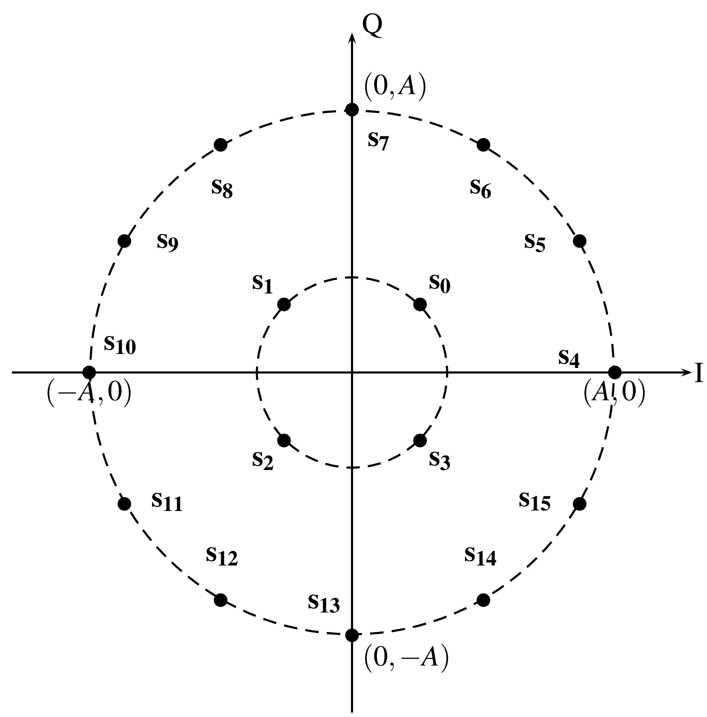
   (c) Determine the squared Euclidean distance between signals s_0 and all the other signals.

   (d) Determine the rate of communication in bits/dimension for these signals.

$$\begin{aligned}
s_0 &= (-1,-1,-1,-1,-1)\\
s_1 &= (-1,-1,+3,-3,+3)\\
s_2 &= (-1,+3,-3,+3,-1)\\
s_3 &= (-1,+3,+1,+1,+3)\\
s_4 &= (+3,-3,+3,-1,-1)\\
s_5 &= (+3,-3,-1,-3,+3)\\
s_6 &= (+3,+1,+1,+3,-1)\\
s_7 &= (+3,+1,-3,+1,+3)
\end{aligned}$$

2. A communication system transmits one of 16 equally likely signals. The signals (waveforms) are represented by the vectors shown below.

$$\begin{aligned}
\mathbf{s_0} &= (+1,+1)\\
\mathbf{s_1} &= (-1,+1)\\
\mathbf{s_2} &= (-1,-1)\\
\mathbf{s_3} &= (+1,-1)\\
\mathbf{s_4} &= (A,0)\\
\mathbf{s_5} &= (A\cos(2\pi/12), A\sin(2\pi/12))\\
\mathbf{s_6} &= (A\cos(2\pi2/12), A\sin(2\pi2/12))\\
\mathbf{s_7} &= (0,A)\\
\mathbf{s_8} &= (A\cos(2\pi4/12), A\sin(2\pi4/12))\\
\mathbf{s_9} &= (A\cos(2\pi5/12), A\sin(2\pi5/12))\\
\mathbf{s_{10}} &= (-A,0)\\
\mathbf{s_{11}} &= (A\cos(2\pi7/12), A\sin(2\pi7/12))\\
\mathbf{s_{12}} &= (A\cos(2\pi8/12), A\sin(2\pi8/12))\\
\mathbf{s_{13}} &= (0,-A)\\
\mathbf{s_{14}} &= (A\cos(2\pi10/12), A\sin(2\pi10/12))\\
\mathbf{s_{15}} &= (A\cos(2\pi10/12), A\sin(2\pi10/12))
\end{aligned}$$



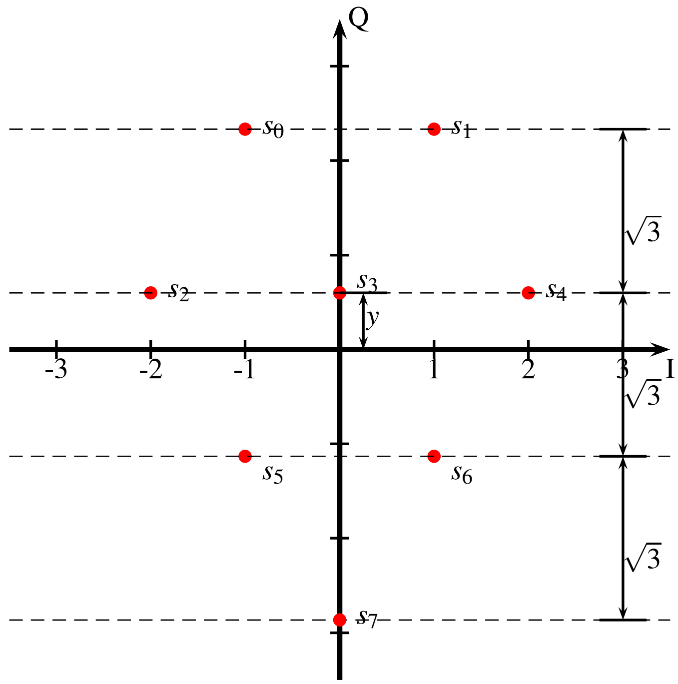
(a) Determine how many information bits can be sent using these signals and determine the rate of communications in bits/dimension.

(b) Determine the smallest value of A so that the minimum squared Euclidean distance between signals is 4.

3. A modulator transmits 3 bits of information using 8 equally likely signals in two dimensions. The signal vectors are given as

$$\begin{aligned} s_0 &= A(-1, y+\sqrt{3}) \\ s_1 &= A(1, y+\sqrt{3}) \\ s_2 &= A(-2, y) \\ s_3 &= A(0, y) \\ s_4 &= A(2, y) \\ s_5 &= A(-1, y-\sqrt{3}) \\ s_6 &= A(1, y-\sqrt{3}) \\ s_7 &= A(0, y-2\sqrt{3}) \end{aligned}$$

(a) Determine the optimum value of the parameter y to minimize the average signal energy transmitted.



(b) Determine the minimum squared Euclidean distance between any two signals.

(c) Determine the rate of communication in bits/dimension.

4. A first signal set with $M = 16$ signals in two dimensions that can transmit 4 bits of information has

the following signals.

$$
\begin{aligned}
s_0 &= A(-3,-3), & s_8 &= A(+1,-3)\\
s_1 &= A(-3,-1), & s_9 &= A(+1,-1)\\
s_2 &= A(-3,+1), & s_{10} &= A(+1,+1)\\
s_3 &= A(-3,+3), & s_{11} &= A(+1,+3)\\
s_4 &= A(-1,-3), & s_{12} &= A(+3,-3)\\
s_5 &= A(-1,-1), & s_{13} &= A(+3,-1)\\
s_6 &= A(-1,+1), & s_{14} &= A(+3,+1)\\
s_7 &= A(-1,+3), & s_{15} &= A(+3,+3)
\end{aligned}
$$

(a) Determine the average energy of this set of signals assuming each signal is used equally likely in terms of A.

(b) Determine the average energy per information bits E_b of this set of signals in terms of A.

(c) Determine the minimum squared Euclidean distance d_E^2 between any distinct pair of signals.

(d) Determine the ratio of minimum squared Euclidean distance to energy per bit d_E^2/E_b.

(e) Determine the rate in terms of bits per dimension.

A second signal set with $M = 16$ signals in 16 dimensions that can transmit 4 bits of information has the following signals.

$$
\begin{aligned}
s_0 &= A(+1,+1,+1,+1,+1,+1,+1,+1,+1,+1,+1,+1,+1,+1,+1,+1)\\
s_1 &= A(+1,-1,+1,-1,+1,-1,+1,-1,+1,-1,+1,-1,+1,-1,+1,-1)\\
s_2 &= A(+1,+1,-1,-1,+1,+1,-1,-1,+1,+1,-1,-1,+1,+1,-1,-1)\\
s_3 &= A(+1,-1,-1,+1,+1,-1,-1,+1,+1,-1,-1,+1,+1,-1,-1,+1)\\
s_4 &= A(+1,+1,+1,+1,-1,-1,-1,-1,+1,+1,+1,+1,-1,-1,-1,-1)\\
s_5 &= A(+1,-1,+1,-1,-1,+1,-1,+1,+1,-1,+1,-1,-1,+1,-1,+1)\\
s_6 &= A(+1,+1,-1,-1,-1,-1,+1,+1,+1,+1,-1,-1,-1,-1,+1,+1)\\
s_7 &= A(+1,-1,-1,+1,-1,+1,+1,-1,+1,-1,-1,+1,-1,+1,+1,-1)\\
s_8 &= A(+1,+1,+1,+1,+1,+1,+1,+1,-1,-1,-1,-1,-1,-1,-1,-1)\\
s_9 &= A(+1,-1,+1,-1,+1,-1,+1,-1,-1,+1,-1,+1,-1,+1,-1,+1)\\
s_{10} &= A(+1,+1,-1,-1,+1,+1,-1,-1,-1,-1,+1,+1,-1,-1,+1,+1)\\
s_{11} &= A(+1,-1,-1,+1,+1,-1,-1,+1,-1,+1,+1,-1,-1,+1,+1,-1)\\
s_{12} &= A(+1,+1,+1,+1,-1,-1,-1,-1,-1,-1,-1,-1,+1,+1,+1,+1)\\
s_{13} &= A(+1,-1,+1,-1,-1,+1,-1,+1,-1,+1,-1,+1,+1,-1,+1,-1)\\
s_{14} &= A(+1,+1,-1,-1,-1,-1,+1,+1,-1,-1,+1,+1,+1,+1,-1,-1)\\
s_{15} &= A(+1,-1,-1,+1,-1,+1,+1,-1,-1,+1,+1,-1,+1,-1,-1,+1)
\end{aligned}
$$

These signals are written into a Matlab file on Canvas. So you can download this file and you need not type in all the signals.

(f) Determine the average energy of this set of signals assuming each signal is used equally likely in terms of A.

(g) Determine the average energy per information bits E_b of this set of signals in terms of A.

(h) Determine the minimum squared Euclidean distance d_E^2 between any distinct pair of signals.

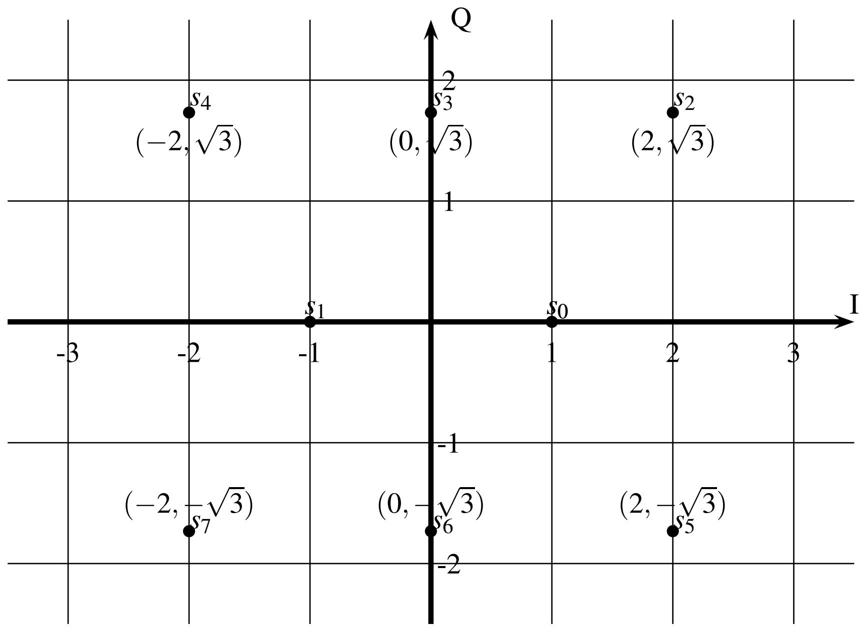
(i) Determine the ratio of minimum squared Euclidean distance to energy per bit d_E^2/E_b. (This should be independent of A.)

(j) Determine the rate in terms of bits per dimension.

(k) Which of the two signal sets is better from a bandwidth efficiency point of view and which is better from an energy efficiency point of view?

5. The eight constellation points for an equal probable signal set are shown below.

$$s_0 = (1,0), \quad s_1 = (-1,0), \quad s_2 = (2,\sqrt{3}), \quad s_3 = (0,\sqrt{3})$$
$$s_4 = (-2,\sqrt{3}), \quad s_5 = (2,-\sqrt{3}), \quad s_6 = (0,-\sqrt{3}), \quad s_7 = (-2,-\sqrt{3})$$



(a) Determine the average energy of this signal set.

(b) Determine the average energy per bit.

(c) Determine the distance between signal s_0 and every other signal.

(d) Determine the rate of communication in bits/dimension.

6. One of the two XM satellites (Rythym or Blues) transmits a signal with a power of 1350 Watts =31.3dBW. The received power is $P_r = -117$dBW $=2\times10^{-12}$ Watts. The bandwidth used is 12.5MHz and the noise level is $N_0 = -204$ dBW/Hz $= 4\times10^{-21}$ Watts/Hz.

(a) Determine the maximum data rate that can be communicated reliably using one of these satellites.

(b) Determine the corresponding E_b/N_0 (in dB) for a system operating at capacity (rate equal to the capacity) and the bandwidth efficiency R/W.

7. Consider the UWB channel which goes from 3.1GHz to 10.6 GHz. Suppose the noise power spectral density is $N_0 = kT = (1.38\times10^{-23})(290) = 4\times10^{-21}$ Watts/Hz. Here k is Boltzman's constant and T is the temperature in Kelvin. A temperature of 290 K corresponds to 62 degree Fahrenheit.

The allowed transmitted power *density* is -41.3dBm/MHz =-71.3dB/MHz. (Note 0dBm=1mW, 30dBm=1W, -30dBm=1 μW).

(a) For the given frequency band determine the total power that can be transmitted.

Suppose the received power is related to the transmitted power by

$$P_r = P_t h_t^2 h_r^2 / d^4$$

where the d is the distance in meters (independent of frequency), h_t is the height of the transmitting antenna (in meters) and h_r is the height of the receiving antenna (in meters).

(b) Compute the largest possible data rate that can be communicated reliably with both antennas at a height of 1m at a distance of 100 m and 1000 m.

8. (a) A communication system is to be designed. The allocated (absolute) bandwidth is 100kHz. It is desired to communicate 300kbits/sec very reliably (error probability close to zero). What is the smallest value of E_b/N_0 (in dB) for which this is possible?

   (b) A channel with absolute bandwidth $W = 100$kHz, power $P = 5$ watts= 5 Joules/sec and two sided noise power spectral density $N_0/2 = 1.778 \times 10^{-3}$ Watts/ Hz is used. The source is an i.i.d. Gaussian source with mean 0 and variance 1. What is the minimum possible distortion (mean square error) if the source is sampled at rate 4000 samples/sec.

9. Consider a Wi-Fi (802.11) system circa 2010. The bandwidth is 20MHz, and the data rates possible are 3, 6, 9, 12, 18, 24, 36, 48, 54 Mbps. Determine the minimum required signal-to-noise ratio E_b/N_0 in dB for each of these data rates and the bandwidth specified for reliable communications.

Bibliography

[1] J. Soni and R. Goodman, *A mind at play: How Claude Shannon invented the information age*, Simon and Schuster, 2017.

[2] A. Leon-Garcia, *Probability, statistics, and random processes for electrical engineering*, Pearson Education, 2017.

[3] B. Hajek, *Random processes for engineers*, Cambridge University Press, 2015.

[4] M. B. Pursley, *Random processes in linear systems*, Prentice Hall, 2002.

[5] F. Amoroso, “The bandwidth of digital data signals,” *IEEE Communications magazine*, vol. 18, no. 6, pp. 13–24, 1980.

[6] E. Agrell, J. Lassing, E. G. Strom, and T. Ottosson, “On the optimality of the binary reflected gray code,” *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3170–3182, 2004.

[7] J. Lassing, E. G. Strom, E. Agrell, and T. Ottosson, “Computation of the exact bit-error rate of coherent m-ary psk with gray code bit mapping,” *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1758–1760, 2003.

[8] P. K. Sinha, P. Biswas, and A. Sinha, “Design and implementation of high speed digital multi-programme system based on itu-t j. 83 annexures a and c,” in *2015 IEEE International Conference on Signal Processing, Informatics, Communication and Energy Systems (SPICES)*, pp. 1–5. IEEE, 2015.

[9] Y. Okumura, “Field strength and its variability in vhf and uhf land-mobile radio service,” *Rev. Electr. Commun. Lab.*, vol. 16, pp. 825–873, 1968.

[10] M. Hata, “Empirical formula for propagation loss in land mobile radio services, ieee transa-ctions on vehicular technology,” *Vol. VT-29*, no. 3, , 1980.

[11] H. Hashemi, “The indoor radio propagation channel,” *Proceedings of the IEEE*, vol. 81, no. 7, pp. 943–968, 1993.

[12] N. Zhang and S. W. Golomb, “Sixty-phase generalized Barker sequences,” *IEEE Transactions on Information theory*, vol. 35, no. 4, pp. 911–912, 1989.

[13] M. Pursley, “Performance evaluation for phase-coded spread-spectrum multiple-access communication-part i: System analysis,” *IEEE Transactions on communications*, vol. 25, no. 8, pp. 795–799, 1977.

[14] *Global Positioning System Standard Positioning Service Signal Specification*, 1995.

[15] D. Chu, "Polyphase codes with good periodic correlation properties (corresp.)," *IEEE Transactions on information theory*, vol. 18, no. 4, pp. 531–532, 1972.

[16] S. W. Golomb et al., *Shift register sequences*, Aegean Park Press, 1967.

[17] R. J. McEliece, *Finite fields for computer scientists and engineers*, volume 23, Springer Science & Business Media, 2012.

[18] P. Fan and M. Darnell, *Sequence design for communications applications*, volume 1, Research Studies Press, 1996.

[19] D. V. Sarwate and M. B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences," *Proceedings of the IEEE*, vol. 68, no. 5, pp. 593–619, 1980.