

Part I

Signal Processing and Detection

Contents

I	Signal Processing and Detection	1
1	Fundamentals of Discrete Data Transmission	3
1.1	Data Modulation and Demodulation	4
1.1.1	Waveform Representation by Vectors	6
1.1.2	Synthesis of the Modulated Waveform	7
1.1.3	Vector-Space Interpretation of the Modulated Waveforms	12
1.1.4	Demodulation	14
1.2	Discrete Data Detection	15
1.2.1	The Vector Channel Model	15
1.2.2	Optimum Data Detection	16
1.2.3	Decision Regions	19
1.2.4	Irrelevant Components of the Channel Output	19
1.3	The Additive White Gaussian Noise (AWGN) Channel	22
1.3.1	Conversion from the Continuous AWGN to a Vector Channel	23
1.3.2	Optimum Detection with the AWGN Channel	25
1.3.3	Signal-to-Noise Ratio (SNR) Maximization with a Matched Filter	27
1.4	Error Probability for the AWGN Channel	29
1.4.1	Invariance to Rotation and Translation	30
1.4.2	Union Bounding	32
1.4.3	The Nearest Neighbor Union Bound	38
1.4.4	Alternative Performance Measures	39
1.4.5	Block Error Measures	41
1.5	General Classes of Constellations and Modulation	42
1.5.1	Cubic Constellations	45
1.5.2	Orthogonal Constellations	48
1.5.3	Circular Constellations - M -ary Phase Shift Keying	54
1.6	Rectangular (and Hexagonal) Signal Constellations	54
1.6.1	Pulse Amplitude Modulation (PAM)	55
1.6.2	Quadrature Amplitude Modulation (QAM)	58
1.6.3	Constellation Performance Measures	66
1.6.4	Hexagonal Signal Constellations in 2 Dimensions	69
1.7	Additive Self-Correlated Noise	69
1.7.1	The Filtered (One-Shot) AWGN Channel	70
1.7.2	Optimum Detection in the Presence of Self-Correlated Noise	71
1.7.3	The Vector Self-Correlated Gaussian Noise Channel	74
1.7.4	Performance of Suboptimal Detection with Self-Correlated Noise	76
	Chapter 1 Exercises	77
A	Gram-Schmidt Orthonormalization Procedure	97
B	The Q Function	98

Chapter 1

Fundamentals of Discrete Data Transmission

Figure 1.1 illustrates **discrete data transmission**, which is the transmission of one message from a finite set of messages through a communication channel. A message sender at the transmitter communicates with a message receiver. The sender selects one message from the finite set, and the transmitter sends a corresponding signal (or “waveform”) that represents this message through the communication channel. The receiver decides the message sent by observing the channel output. Successive transmission of discrete data messages is known as **digital communication**. Based on the noisy received signal at the channel output, the receiver uses a procedure known as **detection** to decide which message, or sequence of messages, was sent. Optimum detection minimizes the probability of an erroneous receiver decision on which message was transmitted.

This chapter characterizes and analyzes optimum detection for a single message transmission through the channel. Dependencies between message transmissions can be important also, but the study of such inter-message dependency is deferred to later chapters.

The messages are usually digital sequences of bits, which are usually not compatible with transmission of physical analog signals through a communication channel. Thus the messages are converted into analog signals that can be sent through the channel. Section 1.1 introduces both **encoding** and **modulation** to characterize such conversion of messages into analog signals by a transmitter. Encoding is the process of converting the messages from their innate form (typically bits) into vectors of real numbers that represent the messages. Modulation is a procedure for converting the encoder-output real-number vectors into analog signals for transmission through a physical channel.

Section 1.2 studies the theory of optimal detection, which depends on a probabilistic model for the communication channel. The channel distorts the transmitted signals both deterministically and with

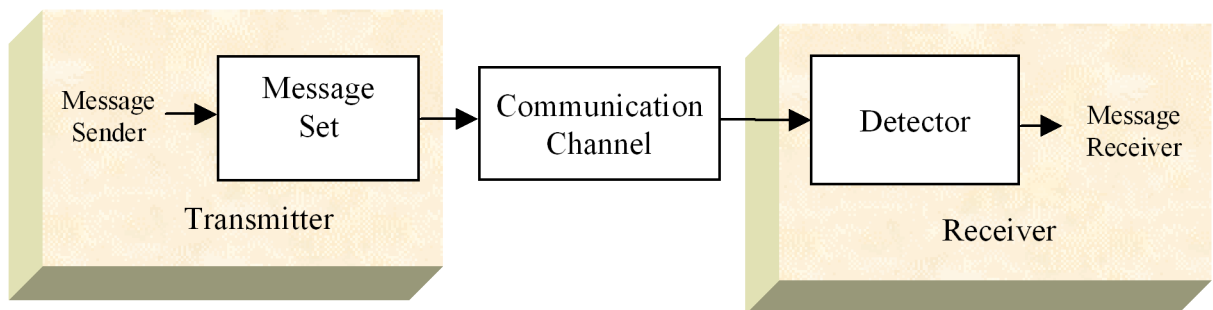


Figure 1.1: Discrete data transmission.

random noise. The noisy channel output will usually not equal the channel input and will be described only in terms of conditional probabilities of various channel-output signals. The channel-input signals have probabilities equal to the probabilities of the messages that they represent. The optimum detector will depend only on the probabilistic model for the channel and the probability distribution of the messages at the channel input. The general optimum detector specializes in many important practical cases of interest.

This chapter develops a theory of modulation and detection that uses a discrete vector representation for any set of continuous-time signals. This “vector-channel” approach was pioneered for educational purposes by Wozencraft and Jacobs in their classic text [1] (Chapter 4). In fact, the first four sections of this chapter closely parallel their development (with some updating and rearrangement), before diverging in Sections 1.5 – 1.7 and in the remainder of this text.

The general model for modulation and demodulation leads to a discussion of the relationship between continuous signals and their vector-channel representation, essentially allowing easier analysis of vectors to replace the more difficult analysis of continuous signals. Section 1.2 solves the general detection problem for the discrete vector channel. Section 1.3 shows that the most common case of a continuous Gaussian-noise channel maps easily into the discrete vector model without loss of generality. Section 1.3 then finds the corresponding optimum detector with Gaussian noise. Given the optimum detector, Section 1.4 shows methods to calculate and estimate average probability of error, P_e , for a vector channel with Additive White Gaussian Noise (AWGN). Sections 1.5 and 1.6 discuss several popular modulation schemes and determine bounds for their probability of error with AWGN. Section 1.6 focuses in particular on signals derived from rectangular lattices, a popular signal transmission format. Section 1.7 then generalizes results for the case of self-correlated Gaussian noise.

1.1 Data Modulation and Demodulation

Figure 1.2 adds more detail to the basic discrete data transmission system of Figure 1.1. The messages emanate from a message source. A **vector encoder** converts each message into a **symbol**, which is a real vector \mathbf{x} that represents the message. Each possible message corresponds to a distinct value of the symbol vector \mathbf{x} . The words “symbol” and “message” are often used interchangeably, with the tacit understanding that the symbol actually represents the message via the action of the encoder. A message from the set of M possible messages m_i $i = 0, \dots, M - 1$ is sent every T seconds, where T is the **symbol period** for the discrete data transmission system. Thus, messages are sent at the **symbol rate** of $1/T$ messages per second. The number of messages that can be sent is often measured in bits so that $b = \log_2(M)$ bits are sent every symbol period. Thus, the **data rate** is $R = b/T$ bits per second. The message is often considered to be a real integer equal to the index i , in which case the message is abbreviated m with possible values $0, \dots, M - 1$.

The **modulator** converts the symbol vector \mathbf{x} that represents the selected message into a continuous time (analog) waveform that the transmitter outputs into the channel. There is a set of possible M signal waveforms $\{x_i(t)\}$ that is in direct one-to-one correspondence with the set of M messages. The demodulator converts continuous-time channel output signals back into a channel output vector \mathbf{y} , from which the detector tries to estimate \mathbf{x} and thus also the message sent. The messages then are provided by the receiver to the message “sink”.

In any data transmission system, the physically transmitted signals are necessarily analog and continuous time. In general, the conversion of a discrete data signal into a continuous time analog signal is called **modulation**. The inverse process of converting the modulated signal back into its original discrete form is called **demodulation**. Thus, the combination of encoding and modulation in the transmitter leads to the mapping:

$$\text{discrete message } m_i \rightarrow x_i(t) \text{ continuous waveform .}$$

Conversely, the combination of demodulation and detection in the receiver leads to the mapping:

$$\text{continuous waveform } y(t) \rightarrow \hat{m} \text{ discrete message .}$$

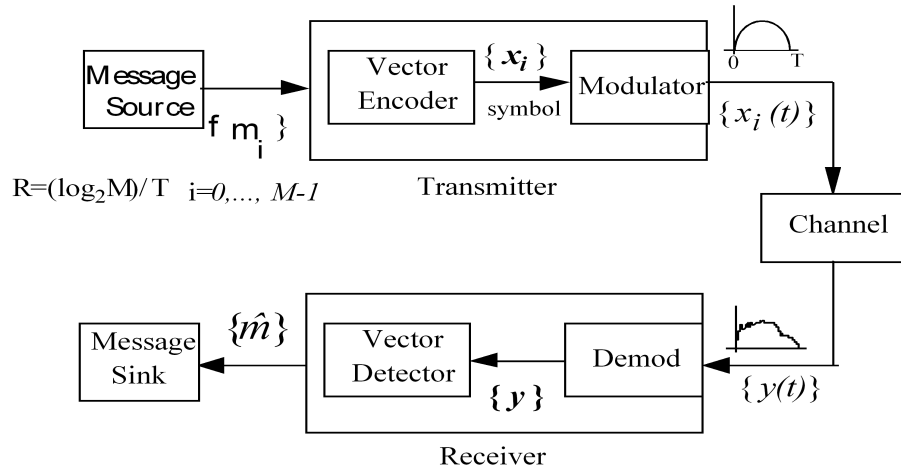


Figure 1.2: Discrete data transmission with greater detail.

When the receiver output message is not equal to the transmitter input message, an **error** occurs. An optimum receiver minimizes the probability of such errors for a given communications channel and set of message waveforms.

EXAMPLE 1.1.1 (binary phase-shift keying) Figure 1.3 repeats Figure 1.1 with a specific linear time-invariant channel that has the Fourier transform indicated. This channel essentially passes signals between 100 Hz and 200 Hz with 150 Hz having the largest gain. Binary logic familiar to most electrical engineers transmits some positive voltage level (say perhaps 1 volt) for a 1 and another voltage level (say 0 volts) for a 0 inside integrated circuits. Clearly such 1/0 transmission on this channel would not pass through the channel, leaving 0 always at the output and making a receiver detection of the correct message difficult if not impossible. Instead the two modulated signals $x_0(t) = +\cos(2\pi t)$ and $x_1(t) = -\cos(2\pi t)$ will easily pass through this channel and be readily distinguishable at the channel output. This latter type of transmission is known as BPSK for **binary phase-shift keying**. If the symbol period is 1 second and if successive transmission is used, the data rate would be 1 bit per second (1 bps).¹

In more detail, the engineer could recognize the trivial vector encoder that converts the message bit of 0 or 1 into the real one-dimensional vectors $\mathbf{x}_0 = +1$ and $\mathbf{x}_1 = -1$. The modulator simply multiplies this \mathbf{x}_i value by the function $\cos(2\pi t)$.

A variety of modulation methods are applied in digital communication systems. To develop a separate analysis for each of these formats would be an enormous task. Instead, this text uses a general vector representation for modulated signals. This vector representation leads to a single method for the analysis of the performance of the data transmission (or storage) system. This section describes the discrete vector representation of any finite or countably infinite set of continuous-time signals and the conversion between the vectors and the signals.

The analysis of the detection process will simplify for an additive white Gaussian noise (AWGN) channel through the symbol-vector approach, which was pioneered by Wozencraft and Jacobs. This approach, indicated in Figure 1.2 by the real-valued vector symbols \mathbf{x}_i and \mathbf{y} , decouples the probability-of-error analysis from the specific modulation method. Each modulation method uses a set of basis

¹However, this chapter is mainly concerned with a single transmission. Each of such successive transmissions could be treated independently because by ignoring transients at the beginning or end of any message transmission as they would be negligible in time extent on such a channel.

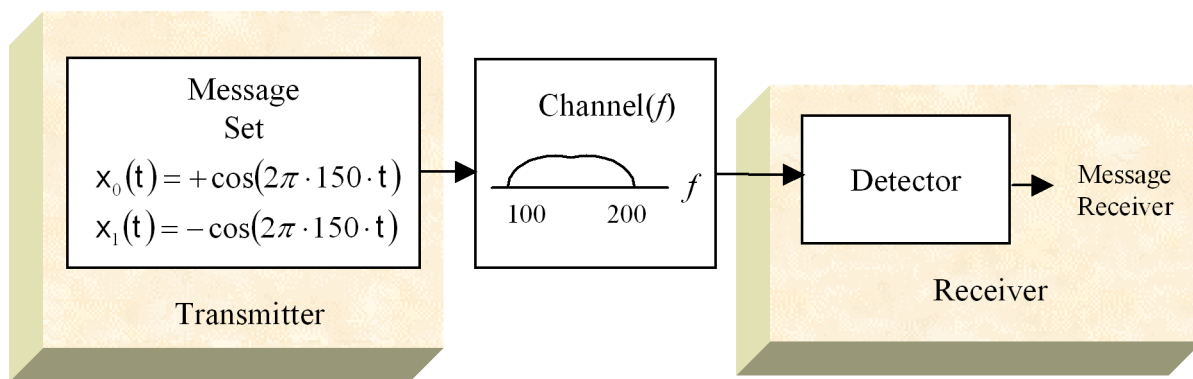


Figure 1.3: Example of channel for which 1 volt and 0 volt binary transmission is inappropriate.

functions that link the vector \mathbf{x}_i with the continuous waveform $x_i(t)$. The choice of modulation basis functions usually depends upon their spectral properties. This chapter investigates and enumerates a number of different basis functions in later sections.

1.1.1 Waveform Representation by Vectors

The reader should be familiar with the infinite-series decomposition of continuous-time signals from the basic electrical-engineering study of Fourier series in signals and systems. For the transmission and detection of a message during a finite time interval, this text considers the set of real-valued functions $\{f(t)\}$ such that $\int_0^T f^2(t)dt < \infty$ (technically known as the Hilbert space of continuous time functions and abbreviated as $L_2[0, T]$). This infinite dimensional vector space has an inner product, which permits the measure of distances and angles between two different functions $f(t)$ and $g(t)$,

$$\langle f(t), g(t) \rangle = \int_0^T f(t) \cdot g(t) dt \quad .$$

Any “well-behaved” continuous time function $x(t)$ defined on the interval $[0, T]$ decomposes according to some set of N orthonormal basis functions $\{\varphi_i(t)\}$ as

$$x(t) = \sum_{n=1}^N x_n \cdot \varphi_n(t)$$

where $\varphi_n(t)$ satisfy $\langle \varphi_n(t), \varphi_m(t) \rangle = 1$ for $n = m$ and 0 otherwise. The continuous function $x(t)$ describes the continuous-time waveform that carries the information through the communication channel. The number of basis functions that represent all the waveforms $\{x_i(t)\}$ for a particular communication system may be infinite, i.e. N may equal ∞ . Using the set of basis functions, the function $x(t)$ maps to a set of N real numbers $\{x_i\}$; these real-valued scalar coefficients assemble into an N -dimensional real-valued vector

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix} \quad .$$

Thus, the function $x(t)$ corresponds to an N -dimensional point \mathbf{x} in a vector space with axes defined by $\{\varphi_i(t)\}$ as illustrated for a three-dimensional point in Figure 1.4.

Similarly a set of continuous time functions $\{x_i(t)\}$ corresponds to a set of discrete N -dimensional points $\{\mathbf{x}_i\}$ known as a **signal constellation**. Such a geometric viewpoint advantageously enables the visualization of the distance between continuous-time functions using distances between the associated signal points in \mathcal{R}^N , the space of N -dimensional real vectors. In fact, later developments show

$$\langle x_1(t), x_2(t) \rangle = \langle \mathbf{x}_1, \mathbf{x}_2 \rangle \quad ,$$

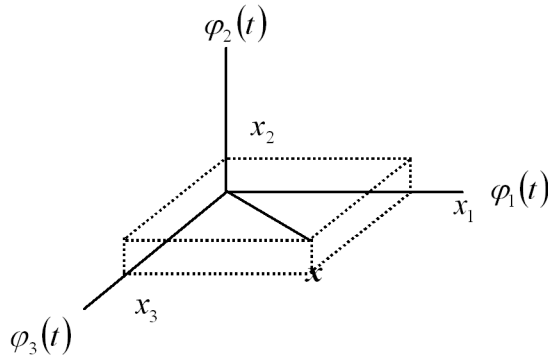


Figure 1.4: Vector space.

where the right hand side is taken as the usual Euclidean inner product in \mathcal{R}^N (discussed later in Definition 1.1.6). This decomposition of continuous-time functions extends to random processes using what is known as a “Karhunen-Loeve expansion.” The basis functions also extend for all time, i.e. on the infinite time interval $(-\infty, \infty)$, in which case the inner product becomes $\langle f(t), g(t) \rangle = \int_{-\infty}^{\infty} f(t)g(t)dt$.

Decomposition of random processes is fundamental to demodulation and detection in the presence of noise. Modulation constructively assembles random signals for the communication system from a set of basis functions $\{\varphi_n(t)\}$ and a set of signal points $\{\mathbf{x}_i\}$. The chosen basis functions and signal points typically satisfy physical constraints of the system and determine performance in the presence of noise.

1.1.2 Synthesis of the Modulated Waveform

The description of modulation begins with the definition of a **data symbol**:

Definition 1.1.1 (Data Symbol) *A data symbol is defined as any N -dimensional real vector*

$$\mathbf{x} \triangleq \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} . \quad (1.1)$$

The data symbol is in lower-case boldface, indicating a vector, to distinguish it from its components, shown in lowercase Roman to indicate scalars. Unless specified otherwise, all quantities shall be real-valued in this chapter. Extensions of the definitions to complex-valued quantities occurs in succeeding chapters as necessary. The synthesis of a modulated waveform uses a set of **orthonormal basis functions**.

Definition 1.1.2 (Orthonormal Basis Functions) *A set of N functions $\{\varphi_n(t)\}$ constitute an N -dimensional **orthonormal basis** if they satisfy the following property:*

$$\int_{-\infty}^{\infty} \varphi_m(t)\varphi_n(t)dt = \delta_{mn} = \begin{cases} 1 & m = n \\ 0 & m \neq n \end{cases} . \quad (1.2)$$

The discrete-time function δ_{mn} will be called the **discrete delta function**².

The construction of a **modulated waveform** $x(t)$ appears in Figure 1.5:

Definition 1.1.3 (Modulated Waveform) *A modulated waveform, corresponding to the data symbol \mathbf{x} , for the orthonormal basis $\varphi_n(t)$ is defined as*

$$x(t) \triangleq \sum_{n=1}^N x_n \varphi_n(t) , \quad (1.3)$$

² δ_{mn} is also called a “Kronecker” delta.

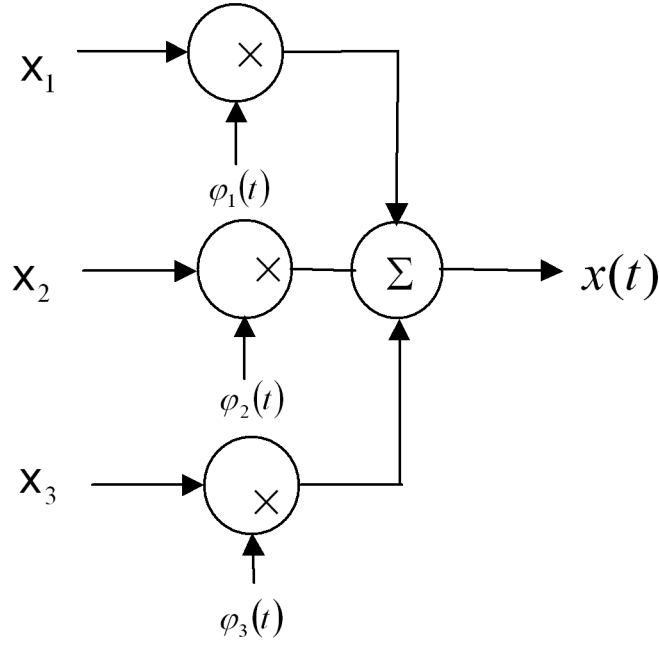


Figure 1.5: Modulator.

Thus, the modulated signal $x(t)$ is formed by multiplying each of the components of the vector \mathbf{x} by the corresponding basis function and summing the continuous-time waveforms, as shown in Figure 1.5. There are many possible choices for the basis functions $\varphi_n(t)$, and correspondingly many possible modulated waveforms $x(t)$ for the same vector \mathbf{x} . The specific choice of basis functions used in a communication system depends on physical limitations of the system.

In practice, a modulator can construct a modulated waveform from any set of data symbols, leading to the concept of a **signal constellation**:

Definition 1.1.4 A **signal constellation** is a set of M vectors, $\{\mathbf{x}_i\}$ $i = 0, \dots, M-1$. The corresponding set of modulated waveforms $\{x_i(t)\}$ $i = 0, \dots, M-1$ is a **signal set**.

Each distinct point in the signal constellation corresponds to a different modulated waveform, but all the waveforms share the same set of basis functions. The component of the i^{th} vector \mathbf{x}_i along the n^{th} basis function $\varphi_n(t)$ is denoted x_{in} . The occurrence of a particular data symbol in the constellation determines the probability of the i^{th} vector (and thus of the i^{th} waveform), $p_{\mathbf{x}}(i)$.

The power available in any physical communication system limits the average amount of energy required to transmit each successive data symbol. Thus, an important concept for a signal constellation (set) is its average energy:

Definition 1.1.5 (Average Energy) The **average energy** of a signal constellation is defined by

$$\mathcal{E}_{\mathbf{x}} \triangleq E[|\mathbf{x}|^2] = \sum_{i=0}^{M-1} \|\mathbf{x}_i\|^2 p_{\mathbf{x}}(i) \quad , \quad (1.4)$$

where $\|\mathbf{x}_i\|^2$ is the squared-length of the vector \mathbf{x}_i , $\|\mathbf{x}_i\|^2 \triangleq \sum_{n=1}^N x_{in}^2$. “E” denotes expected or mean value. (This definition assumes there are only M possible waveforms and $\sum_{i=0}^{M-1} p_{\mathbf{x}}(i) = 1$.)

The average energy is also closely related to the concept of **average power**, which is

$$P_{\mathbf{x}} \triangleq \frac{\mathcal{E}_{\mathbf{x}}}{T} \quad , \quad (1.5)$$

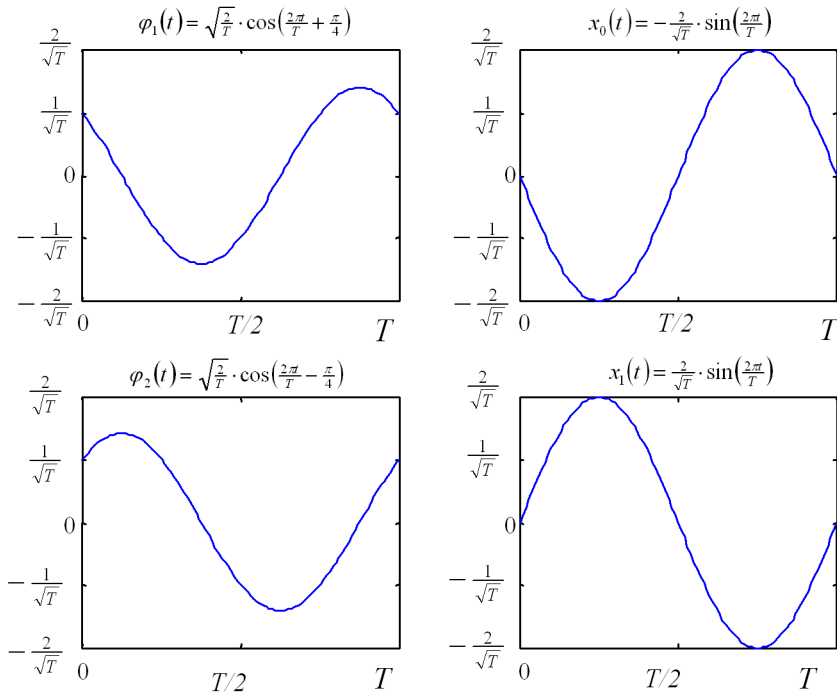


Figure 1.6: BPSK basis functions and waveforms.

corresponding to the amount of energy per symbol period.

The minimization of $\mathcal{E}_{\mathbf{x}}$ places signal-constellation points near the origin; however, the distance between points shall relate to the probability of correctly detecting the symbols in the presence of noise. The geometric problem of optimally arranging points in a vector space with minimum average energy while maintaining at least a minimum distance between each pair of points is the well-studied sphere-packing problem, said geometric viewpoint of communication appeared first in Shannon’s 1948 seminal famous work, *A Mathematical Theory of Communication* (Bell Systems Technical Journal).

The following example at this point illustrates the utility of the basis-function concept:

EXAMPLE 1.1.2 A commonly used and previously discussed transmission method is Binary Phase-Shift Keying (BPSK), used in some satellite and deep-space transmissions as well as a number of simple transmission systems. A more general form of the basis functions, which are parameterized by variable T , is $\varphi_1(t) = \sqrt{\frac{2}{T}} \cos\left[\frac{2\pi t}{T} + \frac{\pi}{4}\right]$ and $\varphi_2(t) = \sqrt{\frac{2}{T}} \cos\left[\frac{2\pi t}{T} - \frac{\pi}{4}\right]$ for $0 \leq t \leq T$ and 0 elsewhere. These two basis functions ($N = 2$), $\varphi_1(t)$ and $\varphi_2(t)$, are shown in Figure 1.6. The two basis functions are orthogonal to each other and both have unit energy, thus satisfying the orthonormality condition. The two possible modulated waveforms transmitted during the interval $[0, T]$ also appear in Figure 1.6, where $x_0(t) = \varphi_1(t) - \varphi_2(t)$ and $x_1(t) = \varphi_2(t) - \varphi_1(t)$. Thus, the data symbols associated with the continuous waveforms are $\mathbf{x}_0 = [1 \ -1]'$ and $\mathbf{x}_1 = [-1 \ 1]'$ (a prime denotes transpose). The signal constellation appears in Figure 1.7. The resulting waveforms are $x_0(t) = -\frac{2}{\sqrt{T}} \sin\left(\frac{2\pi t}{T}\right)$ and $x_1(t) = \frac{2}{\sqrt{T}} \sin\left(\frac{2\pi t}{T}\right)$. This type of modulation is called “binary phase-shift keying,” because the two waveforms are shifted in phase from each other. Since only two possible waveforms are transmitted during each T second time interval, the information rate is $\log_2(2) = 1$ bit per T seconds. Thus to transmit at 1 Mbps, T must equal $1 \mu\text{s}$. (Additional scaling may be used to adjust the BPSK transmit power/energy level to some desired value, but this simply scales all possible constellation points and transmit signals by the same constant value.)

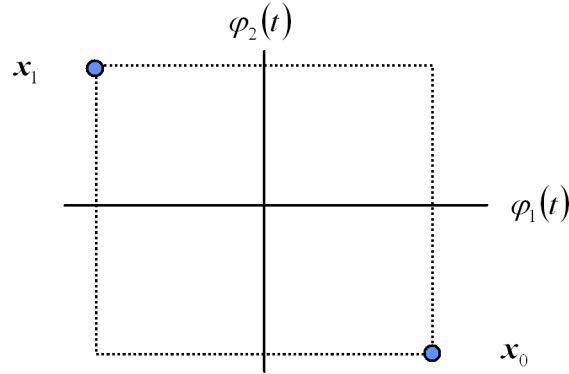


Figure 1.7: BPSK and FM/Manchester signal constellation.

Another set of basis functions is known as “FM code” (FM is “Frequency Modulation”) in the storage industry and also as “Manchester Encoding” in data communications. This method is used in many commercial disk storage products and also in what is known as “10BT or Ethernet” (commonly used in local area networks for the internet). The basis functions are approximated in Figure 1.8 – in practice, the sharp edges are somewhat smoother depending on the specific implementation. The two basis functions again satisfy the orthonormality condition. The data rate equals one bit per T seconds; for a data transfer rate into the disk of 24 MBytes/s or 192 Mbps, $T = 1/(192MHz)$; for a data rate of 10 Mbps in “Ethernet,” $T = 100$ ns. Again for the FM/Manchester example, only two signal points are used, $\mathbf{x}_0 = [1 \ -1]'$ and $\mathbf{x}_1 = [-1 \ 1]'$, with the same constellation shown in Figure 1.7, although the basis functions differ from the previous example. The resulting modulated waveforms appear in Figure 1.8 and correspond to the write currents that are applied to the head in the storage system. (Additional scaling may be used to adjust either the FM or Ethernet transmit power/energy level to some desired value, but this simply scales all possible constellation points and transmit signals by the same constant value.)

The common vector space representation (i.e. signal constellation) of the “Ethernet” and “BPSK” examples allows the performance of a detector to be analyzed for either system in the same way, despite the gross differences in the overall systems.

In either of the systems in Example 1.1.2, a more compact representation of the signals with only one basis function is possible. (As an exercise, the reader should conjecture what this basis function could be and what the associated signal constellation would be.) Appendix A considers the construction of a minimal set of basis functions for a given set of modulated waveforms.

Two more examples briefly illustrate vector components x_n that are not necessarily binary-valued.

EXAMPLE 1.1.3 (ISDN - 2B1Q)³ ISDN digital phone-line service uses $M = 4$ waveforms while the number of basis functions $N = 1$. Thus, the ISDN system transmits 2 bits of information per T seconds of channel use. ISDN uses a basis function that is roughly approximated⁴ by $\varphi_1(t) = \sqrt{\frac{1}{T}}\text{sinc}(\frac{t}{T})$, where $1/T = 80\text{kHz}$, and $\text{sinc}(x) \triangleq \frac{\sin(\pi x)}{\pi x}$. This basis function is not time limited to the interval $[0, T]$. The associated signal constellation appears in Figure 1.9. 2 bits are transmitted using one 4-level (or “quaternary”) symbol every T seconds, hence the name “2B1Q.”

Telephone companies also often transmit the data rate 1.544 Mbps on twisted pairs (such a signal often carries twenty-four 64 kbps digital voice signals plus overhead signaling information of 8 kbps). A method, known as HDSL (High-bit-rate Digital Subscriber Lines), uses

³ISDN stands for Integrated Services Digital Network, an all digital communications standard established by the CCITT for the public telephone network to carry voice and data services simultaneously. It has largely yielded to more sophisticated transmission at higher rates, known as DSL, but provides a good introductory example.

⁴Actually $1/\sqrt{T}\text{sinc}(t/T)$, or some other “Nyquist” pulse shape is used, see Chapter 3 on Intersymbol Interference.

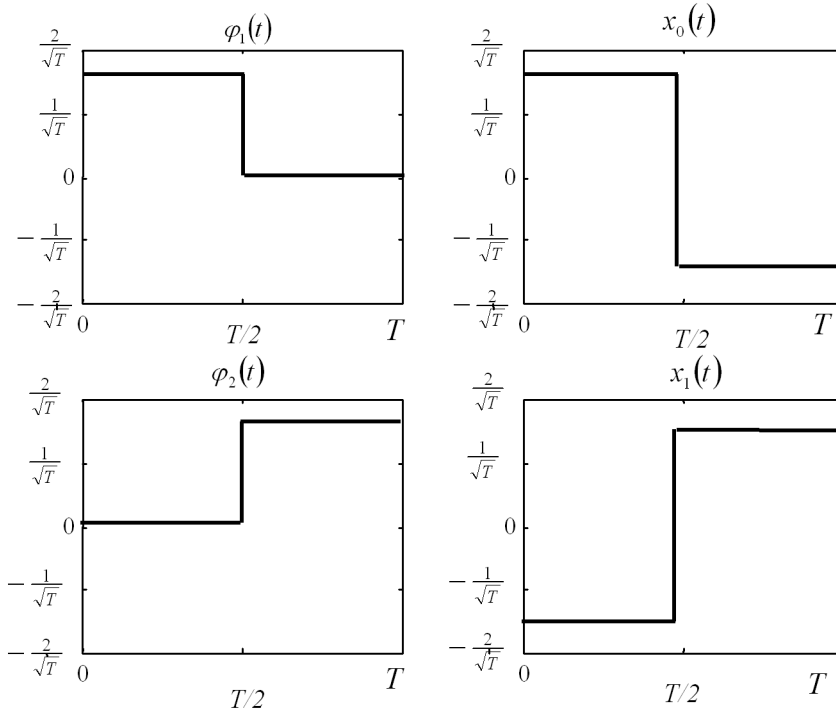


Figure 1.8: Manchester/FM (“Ethernet”) basis functions and waveforms.

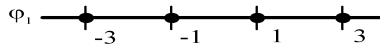


Figure 1.9: 2B1Q signal constellation.

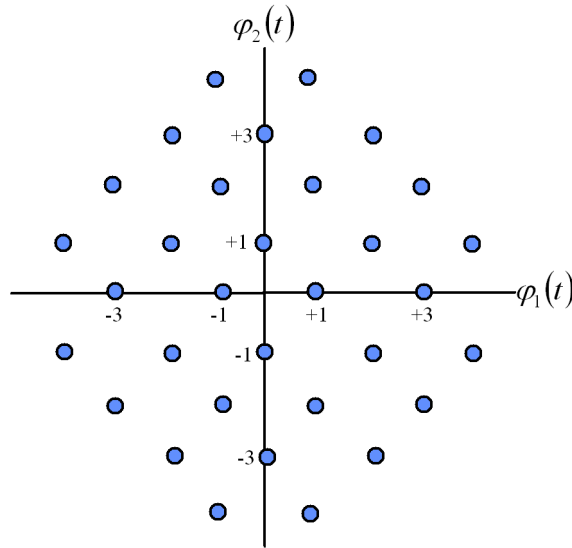


Figure 1.10: 32 Cross signal constellation.

2B1Q with $1/T = 392$ kHz, and thus transmits a data rate of 784 kbps on each of two phone lines for a total of 1.568 Mbps (1.544 Mbps plus 24 kbps of additional HDSL management overhead).

EXAMPLE 1.1.4 (V.32 - 32CR)⁵ Consider a signal set with 32 waveforms ($M = 32$) and with 2 basis functions ($N = 2$) for transmission of 32 signals per channel use. The CCITT V.32-compatible 9600bps voiceband modems use basis functions that are equivalent to $\varphi_1(t) = \sqrt{\frac{2}{T}} \cos \frac{\pi t}{T}$ and $\varphi_2(t) = \sqrt{\frac{2}{T}} \sin \frac{\pi t}{T}$ for $0 \leq t \leq T$ and 0 elsewhere. A raw bit rate of 12.0Kbps⁶ is achieved with a symbol rate of $1/T = 2400$ Hz. The signal constellation is shown in Figure 1.10; the 32 points are arranged in a rotated cross pattern, called 32 CR or 32 cross.

5 bits are transformed into 1 of 32 possible 2-dimensional symbols, hence the extension in the name V.32.

The last two examples also emphasize another tacit advantage of the vector representation, namely that the details of the rates and carrier frequencies in the modulation format are implicit in the normalization of the basis functions, and they do not appear in the description of the signal constellation.

1.1.3 Vector-Space Interpretation of the Modulated Waveforms

A concept that arises frequently in transmission analysis is the inner product of two time functions and/or of two N -dimensional vectors:

Definition 1.1.6 (Inner Product) *The inner product of two (real) functions of time $u(t)$ and $v(t)$ is defined by*

$$\langle u(t), v(t) \rangle \triangleq \int_{-\infty}^{\infty} u(t)v(t)dt . \quad (1.6)$$

⁵The CCITT has published a set of modem standards numbered V.XX.

⁶The actual user information rate is usually 9600 bps with the extra bits used for error-correction purposes as shown in Chapter 8.

The inner product of two (real) vectors \mathbf{u} and \mathbf{v} is defined by

$$\langle \mathbf{u}, \mathbf{v} \rangle \triangleq \mathbf{u}^* \mathbf{v} = \sum_{n=1}^N u_n v_n, \quad (1.7)$$

where $*$ denotes vector transpose (and conjugate vector transpose in Chapter 2 and beyond).

The two inner products in the above definition are equal under the conditions in the following theorem:

Theorem 1.1.1 (Invariance of the Inner Product) *If there exists a set of basis functions $\varphi_n(t)$, $n = 1, \dots, N$ for some N such that $u(t) = \sum_{n=1}^N u_n \varphi_n(t)$ and $v(t) = \sum_{n=1}^N v_n \varphi_n(t)$ then*

$$\langle u(t), v(t) \rangle = \langle \mathbf{u}, \mathbf{v} \rangle. \quad (1.8)$$

where

$$\mathbf{u} \triangleq \begin{bmatrix} u_1 \\ \vdots \\ u_N \end{bmatrix} \quad \text{and} \quad \mathbf{v} \triangleq \begin{bmatrix} v_1 \\ \vdots \\ v_N \end{bmatrix}. \quad (1.9)$$

The proof follows from

$$\langle u(t), v(t) \rangle = \int_{-\infty}^{\infty} u(t)v(t)dt = \int_{-\infty}^{\infty} \sum_{n=1}^N \sum_{m=1}^N u_n v_m \varphi_n(t) \varphi_m(t) dt \quad (1.10)$$

$$= \sum_{n=1}^N \sum_{m=1}^N u_n v_m \int_{-\infty}^{\infty} \varphi_n(t) \varphi_m(t) dt = \sum_{m=1}^N \sum_{n=1}^N u_n v_m \delta_{nm} = \sum_{n=1}^N u_n v_n \quad (1.11)$$

$$= \langle \mathbf{u}, \mathbf{v} \rangle \quad \mathbf{QED}. \quad (1.12)$$

Thus the inner product is “invariant” to the choice of basis functions and only depends on the components of the time functions along each of the basis functions. While the inner product is invariant to the choice of basis functions, the component values of the data symbols depend on basis functions. For example, for the V.32 example, one could recognize that the integral $\frac{2}{T} \int_0^T [2 \cos(\frac{\pi t}{T}) + \sin(\frac{\pi t}{T})] \cdot [\cos(\frac{\pi t}{T}) + 2 \sin(\frac{\pi t}{T})] dt = 2 \cdot 1 + 1 \cdot 2 = 4$.

Parseval’s Identity is a special case (with $\mathbf{x} = \mathbf{u} = \mathbf{v}$) of the invariance of the inner product.

Theorem 1.1.2 (Parseval’s Identity) *The following relation holds true for any modulated waveform*

$$\mathcal{E}_{\mathbf{x}} = E [\|\mathbf{x}\|^2] = E \left[\int_{-\infty}^{\infty} x^2(t) dt \right]. \quad (1.13)$$

The proof follows from the previous Theorem 1.1.1 with $\mathbf{u} = \mathbf{v} = \mathbf{x}$

$$E [\langle u(t), v(t) \rangle] = E [\langle \mathbf{x}, \mathbf{x} \rangle] \quad (1.14)$$

$$= E \left[\sum_{n=1}^N x_n x_n \right] \quad (1.15)$$

$$= E [\|\mathbf{x}\|^2] \quad (1.16)$$

$$= \mathcal{E}_{\mathbf{x}} \quad \mathbf{QED}. \quad (1.17)$$

Parseval’s Identity implies that the average energy of a signal constellation is invariant to the choice of basis functions, as long as they satisfy the orthonormality condition of Equation (1.2). As another V.32 example, one could recognize that the energy of the [2,1] point is $\frac{2}{T} \int_0^T [2 \cos(\frac{2\pi t}{T}) + \sin(\frac{2\pi t}{T})]^2 dt = 2 \cdot 2 + 1 \cdot 1 = 5$.

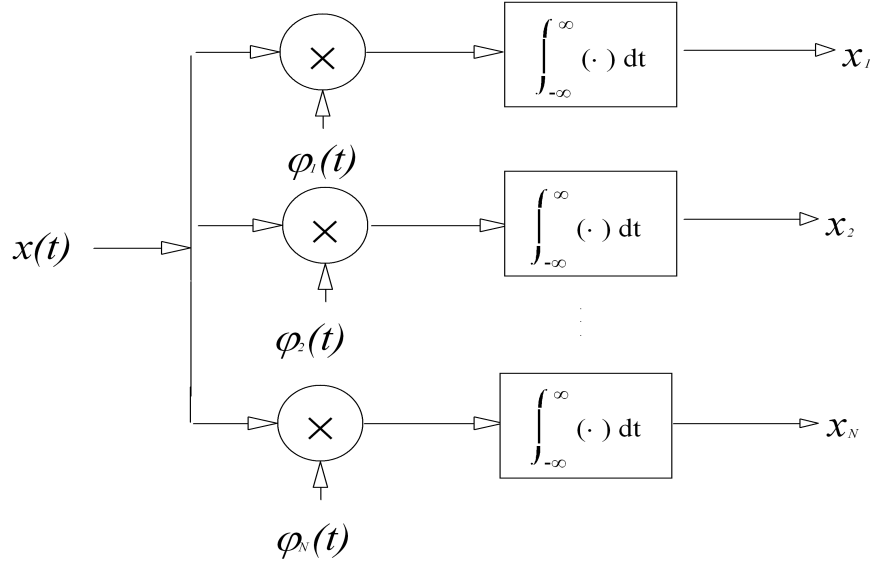


Figure 1.11: The correlative demodulator.

The individual basis functions themselves have a trivial vector representation; namely $\varphi_n(t)$ is represented by $\boldsymbol{\varphi}_n = [0 \ 0 \ , \dots \ , 1 \ , \dots \ , 0]^*$, where the 1 occurs in the n^{th} position. Thus, the data symbol \boldsymbol{x}_i has a representation in terms of the unit basis vectors $\boldsymbol{\varphi}_n$ that is

$$\boldsymbol{x}_i = \sum_{n=1}^N x_{in} \boldsymbol{\varphi}_n \quad . \quad (1.18)$$

The data-symbol component x_{in} can be determined as

$$x_{in} = \langle \boldsymbol{x}_i, \boldsymbol{\varphi}_n \rangle , \quad (1.19)$$

which, using the invariance of the inner product, becomes

$$x_{in} = \langle x_i(t), \varphi_n(t) \rangle = \int_{-\infty}^{\infty} x_i(t) \varphi_n(t) dt \quad n = 1, \dots, N \quad . \quad (1.20)$$

Thus any set of modulated waveforms $\{x_i(t)\}$ can be interpreted as a vector signal constellation, with the components of any particular vector \boldsymbol{x}_i given by Equation (1.20). In effect, x_{in} is the projection of the i^{th} modulated waveform on the n^{th} basis function. The Gram-Schmidt procedure can be used to determine the minimum number of basis functions needed to represent any signal in the signal set, as discussed in Appendix A of this chapter.

1.1.4 Demodulation

As in (1.20), the data symbol vector \boldsymbol{x} can be recovered, component-by-component, by computing the inner product of $x(t)$ with each of the N basis functions. This recovery is called **correlative demodulation** because the modulated signal, $x(t)$, is “correlated” with each of the basis functions to determine \boldsymbol{x} , as is illustrated in Figure 1.11. The modulated signal, $x(t)$, is first multiplied by each of the basis functions in parallel, and the outputs of the multipliers are then passed into a bank of N integrators to produce the components of the data symbol vector \boldsymbol{x} . Practical realization of the multipliers and integrators may be difficult. Any physically implementable set of basis functions can only exist over a

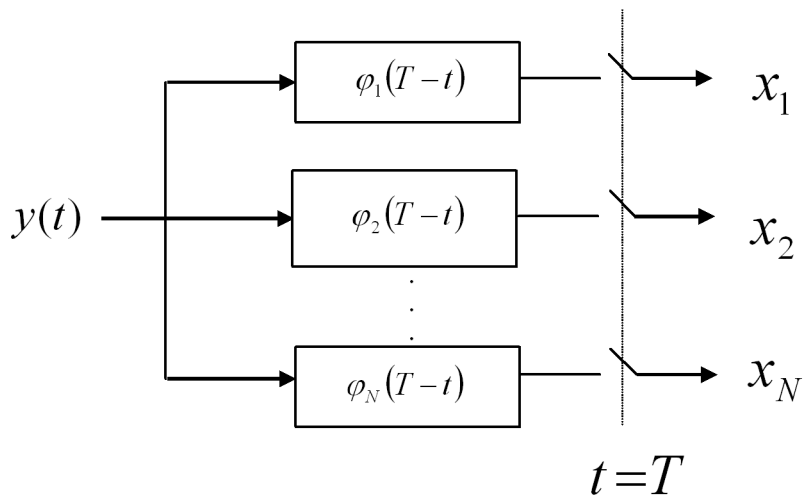


Figure 1.12: The matched-filter demodulator.

finite interval in time, call it T , the **symbol period**.⁷ Then the computation of x_n alternately becomes

$$x_n = \int_0^T x(t)\varphi_n(t)dt \quad . \quad (1.21)$$

The computation in (1.21) is more easily implemented by noting that it is equal to

$$x(t) * \varphi_n(T-t)|_{t=T} \quad , \quad (1.22)$$

where $*$ indicates convolution. The component of the modulated waveform $x(t)$ along the n^{th} basis function is equivalently the convolution (filter) of the waveform $x(t)$ with a filter $\varphi_n(T-t)$ at output sample time T . Such **matched-filter demodulation** is “matched” to the corresponding modulator basis function. Matched-filter demodulation is illustrated in Figure 1.12.

Figure 1.12 illustrates a conversion between the data symbol and the corresponding modulated waveform such that the modulated waveform can be represented by a finite (or countably infinite as $N \rightarrow \infty$) set of components along an orthonormal set of basis functions. The coming sections use this concept to analyze the performance of a particular modulation scheme on the AWGN channel.

1.2 Discrete Data Detection

In practice, the channel output waveform $y(t)$ is not equal to the modulated signal $x(t)$. In many cases, the “essential” information of the channel output $y(t)$ is captured by a finite set of vector components, i.e. a vector \mathbf{y} generated by the demodulation described in Section 1.1. Specific important examples appear later in this chapter, but presently the analysis shall presume the existence of the vector \mathbf{y} and proceed to study the **detector** for the channel. The detector decides which of the discrete channel input vectors \mathbf{x}_i $i = 0, \dots, M-1$ was transmitted based on the observation of the channel output vector \mathbf{y} .

1.2.1 The Vector Channel Model

The vector channel model appears in Figure 1.13. This model suppresses all continuous-time waveforms,

⁷The restriction to a finite time interval is later removed with the introduction of “Nyquist” Pulse shapes in Chapter 3, and the term “symbol period” will be correspondingly reinterpreted.

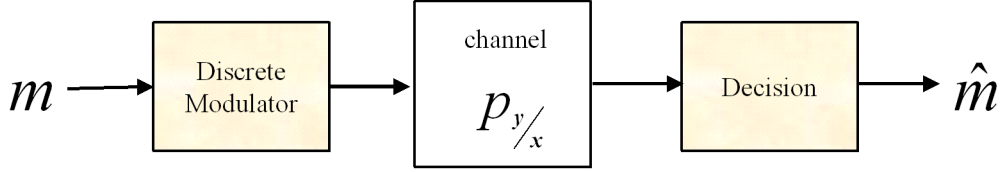


Figure 1.13: Vector channel model.

and the channel produces a discrete vector output given a discrete vector input. The detector chooses a message m_i from among the set of M possible messages $\{m_i\} i = 0, \dots, M-1$ transmitted over the vector channel. The encoder formats the messages for transmission over the vector channel by translating the message m_i into \mathbf{x}_i , an N -dimensional real data symbol chosen from a signal constellation. The encoders of this text are one-to-one mappings between the message set and the signal-constellation vectors. The channel-input vector \mathbf{x} corresponds to a channel-output vector \mathbf{y} , an N -dimensional real vector. (Thus, the transformation of $y(t) \rightarrow \mathbf{y}$ is here assumed to occur within the channel.) The conditional probability of the output vector \mathbf{y} given the input vector \mathbf{x} , $p_{\mathbf{y}|\mathbf{x}}$, completely describes the discrete version of the channel. The decision device then translates the output vector \mathbf{y} into an estimate of the transmitted message $\hat{\mathbf{x}}$. A decoder (which is part of the decision device) reverses the process of the encoder and converts the detector output $\hat{\mathbf{x}}$ into the message decision \hat{m} .

The particular message vector corresponding to m_i is \mathbf{x}_i , and its n^{th} component is x_{in} . The n^{th} component of \mathbf{y} is denoted y_n , $n = 1, \dots, N$. In the vector channel, \mathbf{x} is a random vector, with discrete probability mass function $p_{\mathbf{x}}(i) i = 0, \dots, M-1$.

The output random vector \mathbf{y} may have a continuous probability density or a discrete probability mass function $p_{\mathbf{y}}(\mathbf{v})$, where \mathbf{v} is a dummy variable spanning all the possible N -dimensional outputs for \mathbf{y} . This density is a function of the input and channel transition probability density functions:

$$p_{\mathbf{y}}(\mathbf{v}) = \sum_{i=0}^{M-1} p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \cdot p_{\mathbf{x}}(i) \quad . \quad (1.23)$$

The average energy of the channel input symbols is

$$\mathcal{E}_{\mathbf{x}} = \sum_{i=0}^{M-1} \|\mathbf{x}_i\|^2 \cdot p_{\mathbf{x}}(i) \quad . \quad (1.24)$$

The corresponding average energy for the channel-output vector is

$$\mathcal{E}_{\mathbf{y}} = \sum_{\mathbf{v}} \|\mathbf{v}\|^2 \cdot p_{\mathbf{y}}(\mathbf{v}) \quad . \quad (1.25)$$

An integral replaces⁸ the sum in (1.25) for the case of a continuous density function $p_{\mathbf{y}}(\mathbf{v})$.

As an example, consider the simple additive noise channel $\mathbf{y} = \mathbf{x} + \mathbf{n}$. In this case $p_{\mathbf{y}|\mathbf{x}} = p_{\mathbf{n}}(\mathbf{y} - \mathbf{x})$, where $p_{\mathbf{n}}(\bullet)$ is the noise density, when \mathbf{n} is independent of the input \mathbf{x} .

1.2.2 Optimum Data Detection

For the channel of Figure 1.13, the probability of error is defined as the probability that the decoded message \hat{m} is not equal to the message that was transmitted:

Definition 1.2.1 (Probability of Error) *The Probability of Error is defined as*

$$P_e \triangleq P\{\hat{m} \neq m\} \quad . \quad (1.26)$$

⁸The replacement of a continuous probability density function by a discrete probability mass function is, in strictest mathematical terms, not advisable; however, we do so here, as this particular substitution prevents a preponderance of additional notation, and it has long been conventional in the data transmission literature. The reader is thus forewarned to keep the continuous or discrete nature of the probability density in mind in the analysis of any particular vector channel.

The corresponding probability of being correct is therefore

$$P_c = 1 - P_e = 1 - P\{\hat{m} \neq m\} = P\{\hat{m} = m\} . \quad (1.27)$$

The optimum data detector chooses \hat{m} to minimize P_e , or equivalently, to maximize P_c . The probability of being correct is a function of the particular transmitted message, m_i .

The MAP Detector

The probability of the decision $\hat{m} = m_i$ being correct, given the channel output vector $\mathbf{y} = \mathbf{v}$, is

$$P_c(\hat{m} = m_i, \mathbf{y} = \mathbf{v}) = P_{m|\mathbf{y}}(m_i|\mathbf{v}) \cdot p_{\mathbf{y}}(\mathbf{v}) = P_{\mathbf{x}|\mathbf{y}}(i|\mathbf{v}) \cdot p_{\mathbf{y}}(\mathbf{v}) . \quad (1.28)$$

Thus the optimum decision device observes the particular received output $\mathbf{y} = \mathbf{v}$ and, as a function of that output, chooses $\hat{m} = m_i$ $i = 0, \dots, M-1$ to maximize the probability of a correct decision in (1.28). This quantity is referred to as the *à posteriori* probability for the vector channel. Thus, the optimum detector for the vector channel in Figure 1.13 is called the *Maximum à Posteriori (MAP)* detector:

Definition 1.2.2 (MAP Detector) *The Maximum à Posteriori Detector is defined as the detector that chooses the index i to maximize the à posteriori probability $p_{\mathbf{x}|\mathbf{y}}(i|\mathbf{v})$ given a received vector $\mathbf{y} = \mathbf{v}$.*

The MAP detector thus simply chooses the index i with the highest conditional probability $p_{\mathbf{x}|\mathbf{y}}(i|\mathbf{v})$. For every possible received vector \mathbf{y} the designer of the detector can calculate the corresponding best index i , which depends on the input distribution $p_{\mathbf{x}}(i)$. The *à posteriori* probabilities can be rewritten in terms of the *a priori* probabilities $p_{\mathbf{x}}$ and the channel transition probabilities $p_{\mathbf{y}|\mathbf{x}}$ by recalling the identity⁹,

$$p_{\mathbf{x}|\mathbf{y}}(i|\mathbf{v}) \cdot p_{\mathbf{y}}(\mathbf{v}) = p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \cdot p_{\mathbf{x}}(i) . \quad (1.29)$$

Thus,

$$P_{\mathbf{x}|\mathbf{y}}(i|\mathbf{v}) = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \cdot p_{\mathbf{x}}(i)}{p_{\mathbf{y}}(\mathbf{v})} = \frac{p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \cdot p_{\mathbf{x}}(i)}{\sum_{j=0}^{M-1} p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|j) p_{\mathbf{x}}(j)} , \quad (1.30)$$

for $p_{\mathbf{y}}(\mathbf{v}) \neq 0$. If $p_{\mathbf{y}}(\mathbf{v}) = 0$, then that particular output does not contribute to P_e and therefore is not of further concern. When maximizing (1.30) over i , the denominator $p_{\mathbf{y}}(\mathbf{v})$ is a constant that is ignored.

Thus, Rule 1.2.1 below summarizes the following **MAP** detector rule in terms of the known probability densities of the channel ($p_{\mathbf{y}|\mathbf{x}}$) and of the input vector ($p_{\mathbf{x}}$):

Rule 1.2.1 (MAP Detection Rule)

$$\hat{m} \Rightarrow m_i \text{ if } p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \cdot p_{\mathbf{x}}(i) \geq p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|j) \cdot p_{\mathbf{x}}(j) \quad \forall j \neq i \quad (1.31)$$

If equality holds in (1.31), then the decision can be assigned to either message m_i or m_j without changing the minimized probability of error.

The Maximum Likelihood (ML) Detector

If all transmitted messages are of equal probability, that is if

$$p_{\mathbf{x}}(i) = \frac{1}{M} \quad \forall i = 0, \dots, M-1 , \quad (1.32)$$

then the MAP Detection Rule becomes the Maximum Likelihood Detection Rule:

⁹The more general form of this identity is called “Bayes Theorem”, [2].

Rule 1.2.2 (ML Detection Rule)

$$\hat{m} \Rightarrow m_i \text{ if } p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \geq p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|j) \quad \forall j \neq i \quad . \quad (1.33)$$

If equality holds in (1.33), then the decision can be assigned to either message m_i or m_j without changing the probability of error.

As with the MAP detector, the ML detector also chooses an index i for each possible received vector $\mathbf{y} = \mathbf{v}$, but this index now only depends on the channel transition probabilities and is independent of the input distribution (by assumption). The ML detector essentially cancels the $1/M$ factor on both sides of (1.31) to get (1.33). This type of detector only minimizes P_e when the input data symbols have equal probability of occurrence. As this requirement is often met in practice, ML detection is often used. Even when the input distribution is not uniform, ML detection is still often employed as a detection rule, because the input distribution may be unknown and thus assumed to be uniform. The **Minimax Theorem** sometimes justifies this uniform assumption:

Theorem 1.2.1 (Minimax Theorem) *The ML detector minimizes the maximum possible average probability of error when the input distribution is unknown if the conditional probability of error $P_{e,ML/m=m_i}$ is independent of i .*

Proof:

First, if $P_{e,ML/i}$ is independent of i , then

$$\begin{aligned} P_{e,ML} &= \sum_{i=0}^{M-1} p_{\mathbf{x}}(i) \cdot P_{e,ML/i} \\ &= P_{e,ML/i} \end{aligned}$$

And so,

$$\begin{aligned} \max_{\{p_{\mathbf{x}}\}} P_{e,ML} &= \max_{\{p_{\mathbf{x}}\}} \sum_{i=0}^{M-1} p_{\mathbf{x}}(i) \cdot P_{e,ML/i} \\ &= P_{e,ML} \sum_{i=0}^{M-1} p_{\mathbf{x}}(i) \\ &= P_{e,ML} \end{aligned}$$

Now, let R be any receiver other than the ML receiver. Then,

$$\begin{aligned} \max_{\{p_{\mathbf{x}}\}} P_{e,R} &= \max_{\{p_{\mathbf{x}}\}} \sum_{i=0}^{M-1} p_{\mathbf{x}}(i) \cdot P_{e,R/i} \\ &\geq \sum_{i=0}^{M-1} \frac{1}{M} P_{e,R/i} \quad (\text{Since } \max_{\{p_{\mathbf{x}}\}} P_{e,R} \geq P_{e,R} \text{ for given } \{p_{\mathbf{x}}\}.) \\ &\geq \sum_{i=0}^{M-1} \frac{1}{M} P_{e,ML/i} \quad (\text{Since the ML minimizes } P_e \text{ when } p_{\mathbf{x}}(i) = \frac{1}{M} \text{ for } i = 0, \dots, M-1.) \\ &= P_{e,ML} \end{aligned}$$

So,

$$\begin{aligned} \max_{\{p_{\mathbf{x}}\}} P_{e,R} &\geq P_{e,ML} \\ &= \max_{\{p_{\mathbf{x}}\}} P_{e,ML} \end{aligned}$$

The ML receiver minimizes the maximum P_e over all possible receivers. **QED.**

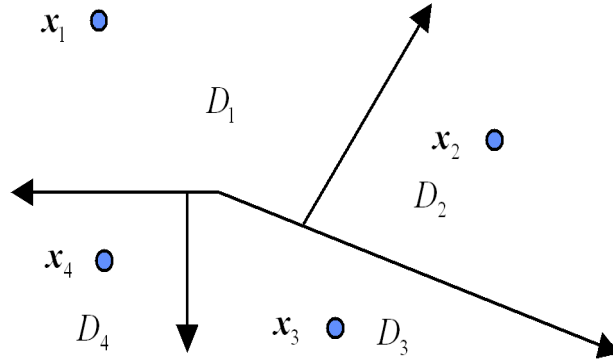


Figure 1.14: Decision regions.

The condition of symmetry imposed by the above Theorem is not always satisfied in practical situations; but the likelihood of an application where both the inputs are nonuniform in distribution and the ML conditional error probabilities are not symmetric is rare. Thus, ML receivers have come to be of nearly ubiquitous use in place of MAP receivers.

1.2.3 Decision Regions

In the case of either the MAP Rule in (1.31) or the ML Rule in (1.33), each and every possible value for the channel output \mathbf{y} maps into one of the M possible transmitted messages. Thus, the vector space for \mathbf{y} is partitioned into M regions corresponding to the M possible decisions. Simple communication systems have well-defined boundaries (to be shown later), so the decision regions often coincide with intuition. Nevertheless, in some well-designed communications systems, the decoding function and the regions can be more difficult to visualize.

Definition 1.2.3 (Decision Region) *The decision region using a MAP detector for each message m_i , $i = 0, \dots, M - 1$ is defined as*

$$\mathcal{D}_i \triangleq \{ \mathbf{v} \mid p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \cdot p_{\mathbf{x}}(i) \geq p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|j) \cdot p_{\mathbf{x}}(j) \quad \forall j \neq i \} \quad . \quad (1.34)$$

With uniformly distributed input messages, the decision regions reduce to

$$\mathcal{D}_i \triangleq \{ \mathbf{v} \mid p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) \geq p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|j) \quad \forall j \neq i \} \quad . \quad (1.35)$$

In Figure (1.14), each of the four different two-dimensional transmitted vectors \mathbf{x}_i (corresponding to the messages m_i) has a surrounding decision region in which any received value for $\mathbf{y} = \mathbf{v}$ is mapped to the message m_i . In general, the regions need not be connected, and although such situations are rare in practice, they can occur (see Problem 1.12). Section 1.3 illustrates several examples of decision regions for the AWGN channel.

1.2.4 Irrelevant Components of the Channel Output

The discrete channel-output vector \mathbf{y} may contain information that does not help determine which of the M messages has been transmitted. These irrelevant components may be discarded without loss of performance, i.e. the input detected and the associated probability of error remain unchanged. Let us

presume the L -dimensional channel output \mathbf{y} can be separated into two sets of dimensions, those which do carry useful information \mathbf{y}_1 and those which do not carry useful information \mathbf{y}_2 . That is,

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} . \quad (1.36)$$

Theorem 1.2.2 summarizes the condition on \mathbf{y}_2 that guarantees irrelevance [1]:

Theorem 1.2.2 (Theorem on Irrelevance) *If*

$$p_{\mathbf{x}|(\mathbf{y}_1, \mathbf{y}_2)} = p_{\mathbf{x}|\mathbf{y}_1} \quad (1.37)$$

or equivalently for the ML receiver,

$$p_{\mathbf{y}_2|(\mathbf{y}_1, \mathbf{x})} = p_{\mathbf{y}_2|\mathbf{y}_1} \quad (1.38)$$

then \mathbf{y}_2 is not needed in the optimum receiver, that is, \mathbf{y}_2 is irrelevant.

Proof: For a MAP receiver, then clearly the value of \mathbf{y}_2 does not affect the maximization of $p_{\mathbf{x}|(\mathbf{y}_1, \mathbf{y}_2)}$ if $p_{\mathbf{x}|(\mathbf{y}_1, \mathbf{y}_2)} = p_{\mathbf{x}|\mathbf{y}_1}$ and thus \mathbf{y}_2 is irrelevant to the optimum receiver's decision.

Equation (1.37) can be written as

$$\frac{P(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2)}{P(\mathbf{y}_1, \mathbf{y}_2)} = \frac{P(\mathbf{x}, \mathbf{y}_1)}{P_{\mathbf{y}_1}} \quad (1.39)$$

or equivalently via "cross multiplication"

$$\frac{P(\mathbf{x}, \mathbf{y}_1, \mathbf{y}_2)}{P(\mathbf{x}, \mathbf{y}_1)} = \frac{P(\mathbf{y}_1, \mathbf{y}_2)}{P_{\mathbf{y}_1}} , \quad (1.40)$$

which is the same as (1.38). **QED.**

The reverse of the theorem of irrelevance is not necessarily true, as can be shown by counterexamples. Two examples (due to Wozencraft and Jacobs, [1]) reinforce the concept of irrelevance. In these examples, the two noise signals \mathbf{n}_1 and \mathbf{n}_2 are independent and a uniformly distributed input is assumed:

EXAMPLE 1.2.1 (Extra Irrelevant Noise) Suppose \mathbf{y}_1 is the noisy channel output shown in Figure 1.15. In the first example, $p_{\mathbf{y}_2|\mathbf{y}_1, \mathbf{x}} = p_{\mathbf{n}_2} = p_{\mathbf{y}_2|\mathbf{y}_1}$, thus satisfying the condition for \mathbf{y}_2 to be ignored, as might be obvious upon casual inspection. The extra independent noise signal \mathbf{n}_2 tells the receiver nothing given \mathbf{y}_1 about the transmitted message \mathbf{x} . In the second example, the irrelevance of \mathbf{y}_2 given \mathbf{y}_1 is not quite as obvious as the signal is present in both the received channel output components. Nevertheless, $p_{\mathbf{y}_2|\mathbf{y}_1, \mathbf{x}} = p_{\mathbf{n}_2}(\mathbf{v}_2 - \mathbf{v}_1) = p_{\mathbf{y}_2|\mathbf{y}_1}$.

Of course, in some cases the output component \mathbf{y}_2 should not be discarded. A classic example is the following case of "noise cancelation."

EXAMPLE 1.2.2 (Noise Cancelation) Suppose \mathbf{y}_1 is the noisy channel output shown in Figure 1.16 while \mathbf{y}_2 may appear to contain only useless noise, it is in fact possible to reduce the effect of \mathbf{n}_1 in \mathbf{y}_1 by constructing an estimate of \mathbf{n}_1 using \mathbf{y}_2 . Correspondingly, $p_{\mathbf{y}_2|\mathbf{y}_1, \mathbf{x}} = p_{\mathbf{n}_2}(\mathbf{v}_2 - (\mathbf{v}_1 - \mathbf{x}_i)) \neq p_{\mathbf{y}_2|\mathbf{y}_1}$.

Reversibility

An important result in digital communication is the **Reversibility Theorem**, which will be used several times over the course of this book. This theorem is, in effect, a special case of the Theorem on Irrelevance:

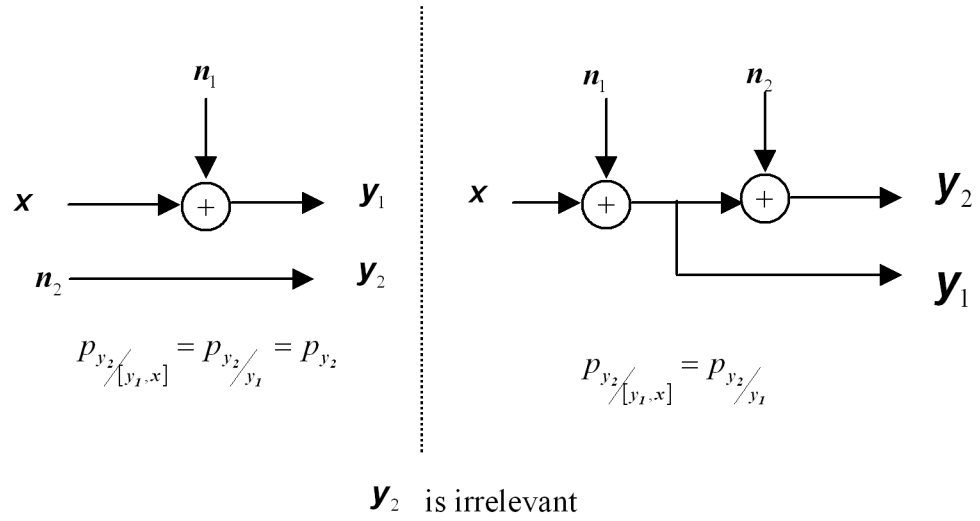


Figure 1.15: Extra irrelevant noise.

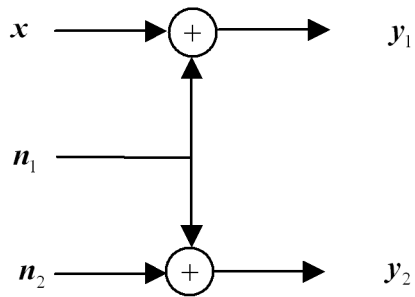


Figure 1.16: Noise can be partially canceled.

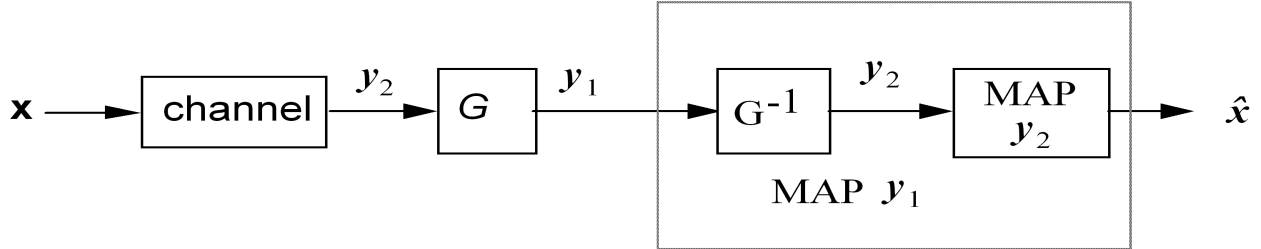


Figure 1.17: Reversibility theorem illustration.

Theorem 1.2.3 (Reversibility Theorem) *The application of an invertible transformation on the channel output vector \mathbf{y} does not affect the performance of the MAP detector.*

Proof: Using the Theorem on Irrelevance, if the channel output is \mathbf{y}_2 and the result of the invertible transformation is $\mathbf{y}_1 = G(\mathbf{y}_2)$, with inverse $\mathbf{y}_2 = G^{-1}(\mathbf{y}_1)$ then $[\mathbf{y}_1 \ \mathbf{y}_2] = [\mathbf{y}_1 \ G^{-1}(\mathbf{y}_1)]$. Then, $p_{\mathbf{x}/(\mathbf{y}_1, \mathbf{y}_2)} = p_{\mathbf{x}/\mathbf{y}_1}$, which is definition of irrelevance. Thus, either of \mathbf{y}_1 or \mathbf{y}_2 is sufficient to detect \mathbf{x} optimally. **QED.**

Equivalently, Figure 1.17 illustrates the reversibility theorem by constructing a MAP receiver for the output of the invertible transformation \mathbf{y}_1 as the cascade of the inverse filter G^{-1} and the MAP receiver for the input of the invertible transformation \mathbf{y}_2 .

1.3 The Additive White Gaussian Noise (AWGN) Channel

Perhaps the most important, and certainly the most analyzed, digital communication channel is the AWGN channel shown in Figure 1.18. This channel passes the sum of the modulated signal $x(t)$ and an uncorrelated Gaussian noise $n(t)$ to the output. The Gaussian noise is assumed to be uncorrelated with itself (or “white”) for any non-zero time offset τ , that is

$$E[n(t)n(t - \tau)] = \frac{\mathcal{N}_0}{2}\delta(\tau) \quad , \quad (1.41)$$

and zero mean, $E[n(t)] = 0$. With these definitions, the Gaussian noise is also strict sense stationary (See Annex C of Chapter 2 for a discussion of stationarity types). The analysis of the AWGN channel is a foundation for the analysis of more complicated channel models in later chapters.

The assumption of white Gaussian noise is valid in the very common situation where the noise is predominantly determined by front-end analog receiver thermal noise. Such noise has a power spectral

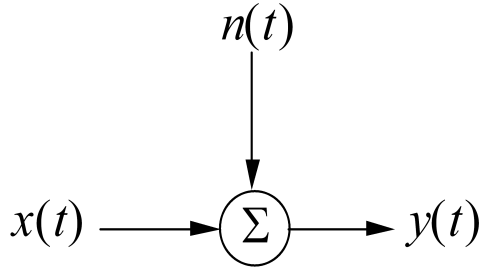


Figure 1.18: AWGN channel.

density given by the Boltzman equation:

$$N(f) = \frac{hf}{e^{\frac{hf}{kT}} - 1} \approx kT \text{ for "small" } f < 10^{12} \text{ ,} \quad (1.42)$$

where Boltzman's constant is $k = 1.38 \times 10^{-23}$ Joules/degree Kelvin, Planck's constant is $h = 6.63 \times 10^{-34}$ Watt-s², and T is the temperature on the Kelvin (absolute) scale. This power spectral density is approximately -174 dBm/Hz ($10^{-17.4}$ mW/Hz) at room temperature (larger in practice). The Gaussian assumption is a consequence of the fact that many small noise sources contribute to this noise, thus invoking the Central Limit Theorem.

1.3.1 Conversion from the Continuous AWGN to a Vector Channel

In the absence of additive noise in Figure 1.18, $y(t) = x(t)$, and the demodulation process in Subsection 1.1.3 would exactly recover the transmitted signal. This section shows that for the AWGN channel, this demodulation process provides sufficient information to determine optimally the transmitted signal. The resulting components $y_l \triangleq \langle y(t), \varphi_l(t) \rangle$, $l = 1, \dots, N$ comprise a vector channel output, $\mathbf{y} = [y_1, \dots, y_N]'$ that is equivalent for detection purposes to $y(t)$. The analysis can thus convert the continuous channel $y(t) = x(t) + n(t)$ to a discrete vector channel model,

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \text{ ,} \quad (1.43)$$

where $\mathbf{n} \triangleq [n_1 \ n_2 \ \dots \ n_N]$ and $n_l \triangleq \langle n(t), \varphi_l(t) \rangle$. The vector channel output is the sum of the vector equivalent of the modulated signal and the vector equivalent of the demodulated noise. Nevertheless, the exact noise sample function may not be reconstructed from \mathbf{n} ,

$$n(t) \neq \sum_{l=1}^N n_l \varphi_l(t) \triangleq \hat{n}(t) \text{ ,} \quad (1.44)$$

or equivalently,

$$y(t) \neq \sum_{l=1}^N y_l \varphi_l(t) \triangleq \hat{y}(t) \text{ .} \quad (1.45)$$

There may exist a component of $n(t)$ that is orthogonal to the space spanned by the basis functions $\{\varphi_1(t) \dots \varphi_N(t)\}$. This unrepresented noise component is

$$\tilde{n}(t) \triangleq n(t) - \hat{n}(t) = y(t) - \hat{y}(t) \text{ .} \quad (1.46)$$

A lemma quickly follows:

Lemma 1.3.1 (Uncorrelated noise samples) *The noise samples in the demodulated noise vector are independent for AWGN and of equal variance $\frac{\mathcal{N}_0}{2}$.*

Proof: Write

$$E [n_k n_l] = E \left[\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} n(t) n(s) \varphi_k(t) \varphi_l(s) dt ds \right] \quad (1.47)$$

$$= \frac{\mathcal{N}_0}{2} \int_{-\infty}^{\infty} \varphi_k(t) \varphi_l(t) dt \quad (1.48)$$

$$= \frac{\mathcal{N}_0}{2} \delta_{kl} \quad \mathbf{QED}. \quad (1.49)$$

The development of the MAP detector could have replaced \mathbf{y} by $y(t)$ everywhere and the development would have proceeded identically with the tacit inclusion of the time variable t in the probability densities (and also assuming stationarity of $y(t)$ as a random process). The Theorem of Irrelevance would hold with $[\mathbf{y}_1 \ \mathbf{y}_2]$ replaced by $[\hat{y}(t) \ \tilde{n}(s)]$, as long as the relation (1.38) holds for any pair of time instants t and s . In a non-mathematical sense, the unrepresented noise is useless to the receiver, so there is nothing of value lost in the vector demodulator, even though some of the channel output noise is not represented. The following algebra demonstrates that $\tilde{n}(s)$ is irrelevant:

First,

$$E [\tilde{n}(s) \cdot \hat{n}(t)] = E \left[\tilde{n}(s) \cdot \sum_{l=1}^N n_l \varphi_l(t) \right] = \sum_{l=1}^N \varphi_l(t) E [\tilde{n}(s) \cdot n_l] \quad (1.50)$$

and,

$$E [\tilde{n}(s) \cdot n_l] = E [(n(s) - \hat{n}(s)) \cdot n_l] \quad (1.51)$$

$$= E \left[\int_{-\infty}^{\infty} n(s) \varphi_l(\tau) n(\tau) d\tau \right] - E \left[\sum_{k=1}^N n_k n_l \varphi_k(s) \right] \quad (1.52)$$

$$= \frac{\mathcal{N}_0}{2} \int_{-\infty}^{\infty} \delta(s - \tau) \varphi_l(\tau) d\tau - \frac{\mathcal{N}_0}{2} \varphi_l(s) \quad (1.53)$$

$$= \frac{\mathcal{N}_0}{2} [\varphi_l(s) - \varphi_l(s)] = 0 \quad (1.54)$$

Second,

$$\mathbf{P}_{\mathbf{x}|\hat{y}(t), \tilde{n}(s)} = \frac{\mathbf{P}_{\mathbf{x}, \hat{y}(t), \tilde{n}(s)}}{\mathbf{P}_{\hat{y}(t), \tilde{n}(s)}} \quad (1.55)$$

$$= \frac{\mathbf{P}_{\mathbf{x}, \hat{y}(t)} \cdot \mathbf{P}_{\tilde{n}(s)}}{\mathbf{P}_{\hat{y}(t)} \cdot \mathbf{P}_{\tilde{n}(s)}} \quad (1.56)$$

$$= \frac{\mathbf{P}_{\mathbf{x}, \hat{y}(t)}}{\mathbf{P}_{\hat{y}(t)}} \quad (1.57)$$

$$= \mathbf{P}_{\mathbf{x}|\hat{y}(t)} \quad (1.58)$$

Equation (1.58) satisfies the theorem of irrelevance, and thus the receiver need only base its decision on $\hat{y}(t)$, or equivalently, only on the received vector \mathbf{y} . The vector AWGN channel is equivalent to the continuous-time AWGN channel.

Rule 1.3.1 (The Vector AWGN Channel) *The vector AWGN channel is given by*

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad (1.59)$$

and is equivalent to the channel illustrated in Figure 1.18. The noise vector \mathbf{n} is an N -dimensional Gaussian random vector with zero mean, equal-variance, uncorrelated components in each dimension. The noise distribution is

$$\mathbf{p}_{\mathbf{n}}(\mathbf{u}) = (\pi \mathcal{N}_0)^{-\frac{N}{2}} \cdot e^{-\frac{1}{\mathcal{N}_0} \|\mathbf{u}\|^2} = (2\pi\sigma^2)^{-\frac{N}{2}} \cdot e^{-\frac{1}{2\sigma^2} \|\mathbf{u}\|^2} \quad (1.60)$$

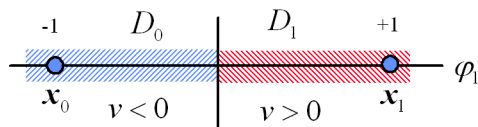


Figure 1.19: Binary ML detector.

Application of $y(t)$ to either the correlative demodulator of Figure 1.11 or to the matched-filter demodulator of Figure 1.12, generates the desired vector channel output \mathbf{y} at the demodulator output. The following section specifies the decision process that produces an estimate of the input message, given the output \mathbf{y} , for the AWGN channel.

1.3.2 Optimum Detection with the AWGN Channel

For the vector AWGN channel in (1.59),

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{v}|i) = p_{\mathbf{n}}(\mathbf{v} - \mathbf{x}_i) \quad , \quad (1.61)$$

where $p_{\mathbf{n}}$ is the vector noise distribution in (1.60). Thus for AWGN the MAP Decision Rule becomes

$$\hat{m} \Rightarrow m_i \text{ if } e^{-\frac{1}{\mathcal{N}_0}\|\mathbf{v}-\mathbf{x}_i\|^2} \cdot p_{\mathbf{x}}(i) \geq e^{-\frac{1}{\mathcal{N}_0}\|\mathbf{v}-\mathbf{x}_j\|^2} \cdot p_{\mathbf{x}}(j) \quad \forall j \neq i \quad , \quad (1.62)$$

where the common factor of $(\pi\mathcal{N}_0)^{-\frac{N}{2}}$ has been canceled from each side of (1.62). As noted earlier, if equality holds in (1.62), then the decision can be assigned to any of the corresponding messages without change in minimized probability of error. The log of (1.62) is the preferred form of the MAP Decision Rule for the AWGN channel:

Rule 1.3.2 (AWGN MAP Detection Rule)

$$\hat{m} \Rightarrow m_i \text{ if } \|\mathbf{v} - \mathbf{x}_i\|^2 - \mathcal{N}_0 \ln\{p_{\mathbf{x}}(i)\} \leq \|\mathbf{v} - \mathbf{x}_j\|^2 - \mathcal{N}_0 \ln\{p_{\mathbf{x}}(j)\} \quad \forall j \neq i \quad (1.63)$$

If the channel input messages are equally likely, the \ln terms on both sides of (1.63) cancel, yielding the AWGN ML Detection Rule:

Rule 1.3.3 (AWGN ML Detection Rule)

$$\hat{m} \Rightarrow m_i \text{ if } \|\mathbf{v} - \mathbf{x}_i\|^2 \leq \|\mathbf{v} - \mathbf{x}_j\|^2 \quad \forall j \neq i \quad . \quad (1.64)$$

The ML detector for the AWGN channel in (1.64) has the intuitively appealing physical interpretation that the decision $\hat{m} = m_i$ corresponds to choosing the data symbol \mathbf{x}_i that is closest, in terms of the Euclidean distance, to the received vector channel output $\mathbf{y} = \mathbf{v}$. Without noise, the received vector is $\mathbf{y} = \mathbf{x}_i$ the transmitted symbol, but the additive Gaussian noise results in a received symbol most likely in the neighborhood of \mathbf{x}_i . The Gaussian shape of the noise implies the probability of a received point decreases as the distance from the transmitted point increases. As an example consider the decision regions for binary data transmission over the AWGN channel illustrated in Figure 1.19. The ML receiver decides \mathbf{x}_1 if $\mathbf{y} = \mathbf{v} \geq 0$ and \mathbf{x}_0 if $\mathbf{y} = \mathbf{v} < 0$. (One might have guessed this answer without need for theory.) With d defined as the distance $\|\mathbf{x}_1 - \mathbf{x}_0\|$, the decision regions are offset in the MAP detector by $\frac{\sigma^2}{d} \ln\{\frac{p_{\mathbf{x}}(j)}{p_{\mathbf{x}}(i)}\}$ with the decision boundary shifting towards the data symbol of lesser probability, as illustrated in Figure 1.20. Unlike the ML detector, the MAP detector accounts for the a priori message probabilities. The decision region for the more likely symbol is extended by shifting the boundary towards the less likely symbol. Figure 1.21 illustrates the decision region for a two-dimensional example of the QPSK signal set, which uses the same basis functions as the V.32 example (Example 1.1.4). The points in the signal constellation are all assumed to be equally likely.

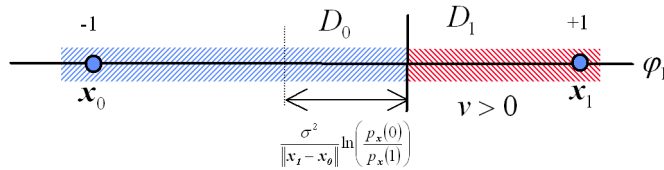


Figure 1.20: Binary MAP detector.

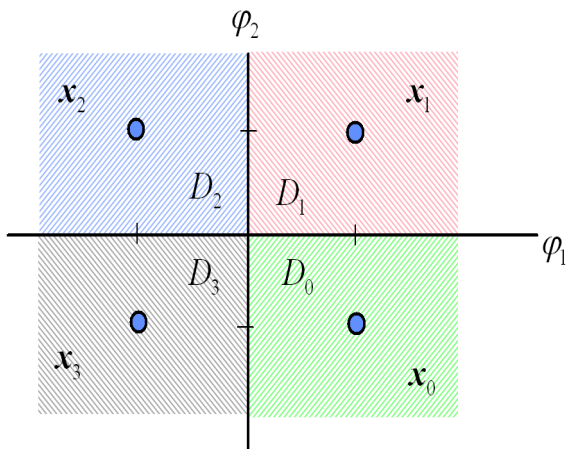


Figure 1.21: QPSK decision regions.

General Receiver Implementation

While the decision regions in the above examples appear simple to implement, in a digital system, the implementation may be more complex. This section investigates general receiver structures and the detector implementation.

The MAP detector minimizes the quantity (the quantity \mathbf{y} now replaces \mathbf{v} averting strict mathematical notation, because probability density functions are used less often in the subsequent analysis):

$$\|\mathbf{y} - \mathbf{x}_i\|^2 - \mathcal{N}_0 \ln\{p_{\mathbf{x}}(i)\} \quad (1.65)$$

over the M possible messages, indexed by i . The quantity in (1.65) expands to

$$\|\mathbf{y}\|^2 - 2\langle \mathbf{y}, \mathbf{x}_i \rangle + \|\mathbf{x}_i\|^2 - \mathcal{N}_0 \ln\{p_{\mathbf{x}}(i)\} \quad (1.66)$$

Minimization of (1.66) can ignore the $\|\mathbf{y}\|^2$ term. The MAP decision rule then becomes

$$\hat{m} \Rightarrow m_i \text{ if } \langle \mathbf{y}, \mathbf{x}_i \rangle + c_i \geq \langle \mathbf{y}, \mathbf{x}_j \rangle + c_j \quad \forall j \neq i \quad (1.67)$$

where c_i is the constant (independent of \mathbf{y})

$$c_i \triangleq \frac{\mathcal{N}_0}{2} \ln\{p_{\mathbf{x}}(i)\} - \frac{\|\mathbf{x}_i\|^2}{2} \quad (1.68)$$

A system design can precompute the constants $\{c_i\}$ from the transmitted symbols $\{\mathbf{x}_i\}$ and their probabilities $p_{\mathbf{x}}(i)$. The detector thus only needs to implement the M inner products, $\langle \mathbf{y}, \mathbf{x}_i \rangle$ $i = 0, \dots, M-1$. When all the data symbols have the same energy ($\mathcal{E}_{\mathbf{x}} = \|\mathbf{x}_i\|^2 \forall i$) and are equally probable (i.e. MAP = ML), then the constant c_i is independent of i and can be eliminated from (1.67). The ML detector thus chooses the \mathbf{x}_i that maximizes the inner product (or correlation) of the received value for $\mathbf{y} = \mathbf{v}$ with \mathbf{x}_i over i .

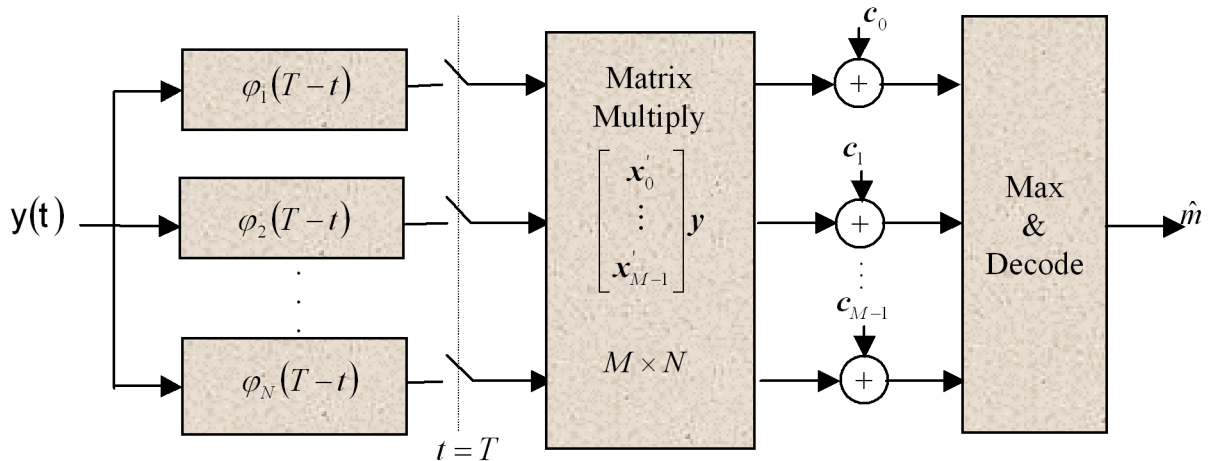


Figure 1.22: Basis detector.

There exist two common implementations of the MAP receiver in (1.67). The first, shown in Figure 1.22, called a “basis detector,” computes \mathbf{y} using a matched filter demodulator. This MAP receiver computes the M inner products of (1.67) digitally (an $M \times N$ matrix multiply with \mathbf{y}), adds the constant c_i of (1.68), and picks the index i with maximum result. Finally, a decoder translates the index i into the desired message m_i . Often in practice, the signal constellation is such (see Section 1.6 for examples) that the max and decode function reduces to simple truncation of each component in the received vector \mathbf{y} .

The second form of the demodulator eliminates the matrix multiply in Figure 1.22 by recalling the inner product equivalences between the discrete vectors \mathbf{x}_i, \mathbf{y} and the continuous time functions $x_i(t)$ and $y(t)$. That is

$$\langle \mathbf{y}, \mathbf{x}_i \rangle = \int_0^T y(t)x_i(t)dt = \langle y(t), x_i(t) \rangle . \quad (1.69)$$

Equivalently,

$$\langle \mathbf{y}, \mathbf{x}_i \rangle = y(t) * x_i(T-t)|_{t=T} \quad (1.70)$$

where $*$ indicates convolution. This type of detector is called a “signal detector” and appears in Figure 1.23.

EXAMPLE 1.3.1 (pattern recognition as a signal detector) Pattern recognition is a digital signal processing procedure that is used to detect whether a certain signal is present. An example occurs when an aircraft takes electronic pictures of the ground and the corresponding electrical signal is analyzed to determine the presence of certain objects. This is a communication channel in disguise where the two inputs are the usual terrain of the ground and the terrain of the ground including the object to be detected. A signal detector consisting of two filters that are essentially the time reverse of each of the possible input signals, with a comparison of the outputs (after adding any necessary constants), allows detection of the presence of the object or pattern. There are many other examples of pattern recognition in voice/command recognition or authentication, written character scanning, and so on.

The above example/discussion illustrates that many of the principles of digital communication theory are common to other fields of digital signal processing and science.

1.3.3 Signal-to-Noise Ratio (SNR) Maximization with a Matched Filter

SNR is a good measure for a system’s performance, describing the ratio of signal power (message) to unwanted noise power. The SNR at the output of a filter is defined as the ratio of the modulated

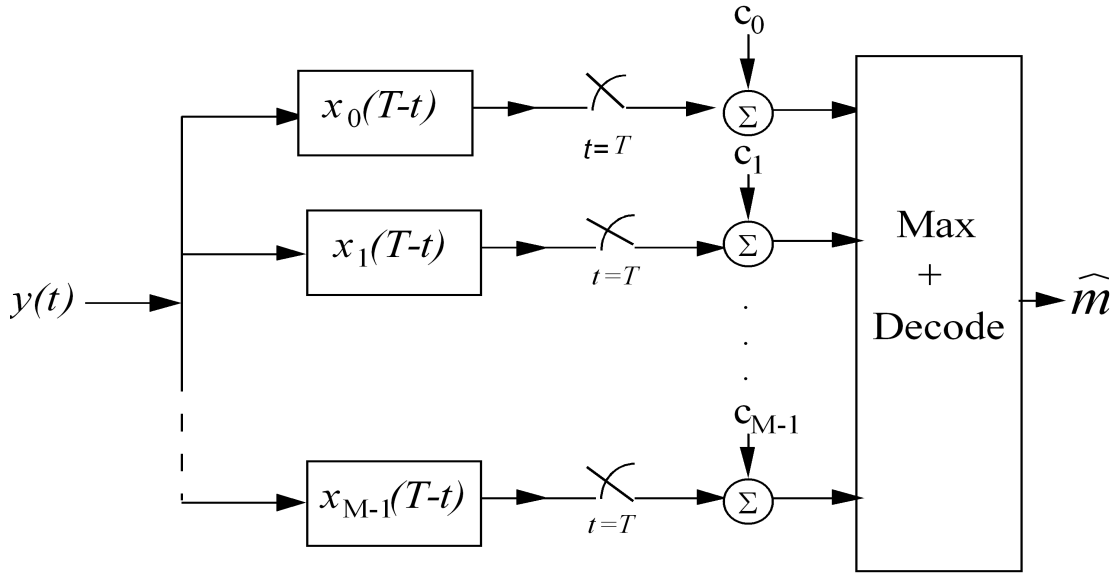


Figure 1.23: Signal detector.

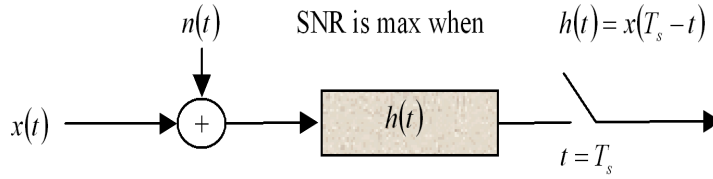


Figure 1.24: SNR maximization by matched filter.

signal's energy to the mean-square value of the noise. The SNR can be defined for both continuous- and discrete-time processes; the discrete SNR is SNR of the samples of the received and filtered waveform. The matched filters shown in Figure 1.23 satisfy the **SNR maximization property**, which the following theorem summarizes:

Theorem 1.3.1 (SNR Maximization) *For the system shown in Figure 1.24, the filter $h(t)$ that maximizes the signal-to-noise ratio at sample time T_s is given by the matched filter $h(t) = x(T_s - t)$.*

Proof: Compute the SNR at sample time $t = T_s$ as follows.

$$\text{Signal Energy} = [x(t) * h(t)|_{t=T_s}]^2 \quad (1.71)$$

$$= \left[\int_{-\infty}^{\infty} x(t) \cdot h(T_s - t) dt \right]^2 = [\langle x(t), h(T_s - t) \rangle]^2 \quad (1.72)$$

The sampled noise at the matched filter output has energy or mean-square

$$\text{Noise Energy} = E \left[\int_{-\infty}^{\infty} n(t)h(T_s - t)dt \int_{-\infty}^{\infty} n(s)h(T_s - s)ds \right] \quad (1.73)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{N_0}{2} \delta(t-s) h(T_s-t) h(T_s-s) dt ds \quad (1.74)$$

$$= \frac{N_0}{2} \int_{-\infty}^{\infty} h^2(T_s-t) dt \quad (1.75)$$

$$(1.76)$$

$$= \frac{N_0}{2} \|h\|^2 \quad (1.77)$$

The signal-to-noise ratio, defined as the ratio of the signal power in (1.72) to the noise power in (1.77), equals

$$\text{SNR} = \frac{2}{N_0} \cdot \frac{[\langle x(t), h(T_s-t) \rangle]^2}{\|h\|^2} \quad (1.78)$$

The ‘‘Cauchy-Schwarz Inequality’’ states that

$$[\langle x(t), h(T_s-t) \rangle]^2 \leq \|x\|^2 \|h\|^2 \quad (1.79)$$

with equality if and only if $x(t) = kh(T_s-t)$, where k is some arbitrary constant. Thus, by inspection, (1.78) is maximized over all choices for $h(t)$ when $h(t) = x(T_s-t)$. The filter $h(t)$ is ‘‘matched’’ to $x(t)$, and the corresponding maximum SNR (for any k) is

$$\text{SNR}_{max} = \frac{2}{N_0} \|x\|^2 \quad (1.80)$$

An example of the use of the SNR maximization property of the matched filter occurs in time-delay estimation, which is used for instance in radar:

EXAMPLE 1.3.2 (Time-delay estimation) Radar systems emit electromagnetic pulses and measure reflection of those pulses off objects within range of the radar. The distance of the object is determined by the delay of the reflected energy, with longer delay corresponding to longer distance. By processing the received signal at the radar with a filter matched to the radar pulse shape, the signal level measured in the presence of a presumably fixed background white noise will appear largest relative to the noise. Thus, the ability to determine the exact time instant at which the maximum pulse returned is improved by the use of the matched filter, allowing more accurate estimation of the position of the object.

1.4 Error Probability for the AWGN Channel

This section discusses the computation of the average probability of error of decoding the transmitted message incorrectly on an AWGN channel. From the previous section, the AWGN channel is equivalent to a vector channel with output given by

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad (1.81)$$

The computation of P_e often assumes that the inputs \mathbf{x}_i are equally likely, or $p_{\mathbf{x}}(i) = \frac{1}{M}$. Under this assumption, the optimum detector is the ML detector, which has decision rule

$$\hat{m} \Rightarrow m_i \text{ if } \|\mathbf{v} - \mathbf{x}_i\|^2 \leq \|\mathbf{v} - \mathbf{x}_j\|^2 \quad \forall j \neq i \quad (1.82)$$

The P_e associated with this rule depends on the signal constellation $\{\mathbf{x}_i\}$ and the noise variance $\frac{N_0}{2}$. Two general invariance theorems in Subsection 1.4.1 facilitate the computation of P_e . The exact P_e ,

$$P_e = \frac{1}{M} \cdot \sum_{i=0}^{M-1} P_{e/i} \quad (1.83)$$

$$= 1 - \frac{1}{M} \cdot \sum_{i=0}^{M-1} P_{c/i} \quad (1.84)$$

may be difficult to compute, so convenient and accurate bounding procedures in Subsections 1.4.2 through 1.4.4 can alternately approximate P_e .

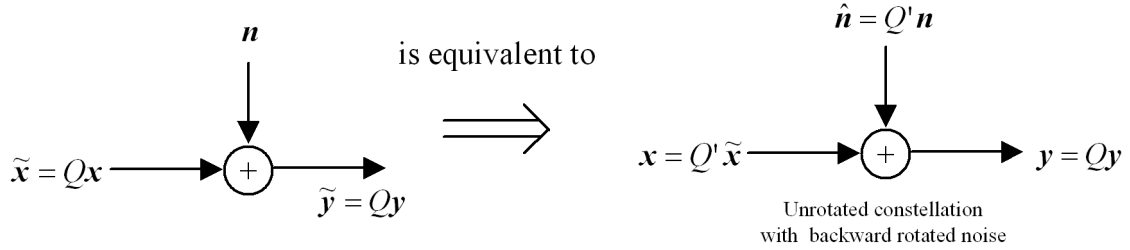


Figure 1.25: Rotational invariance with AWGN.

1.4.1 Invariance to Rotation and Translation

The orientation of the signal constellation with respect to the coordinate axes and to the origin does not affect the P_e . This result follows because (1) the error depends only on relative distances between points in the signal constellation, and (2) AWGN is spherically symmetric in all directions. First, the probability of error for the ML receiver is invariant to any rotation of the signal constellation, as summarized in the following theorem:

Theorem 1.4.1 (Rotational Invariance) *If all the data symbols in a signal constellation are rotated by an orthogonal transformation, that is $\tilde{\mathbf{x}}_i \leftarrow Q\mathbf{x}_i$ for all $i = 0, \dots, M-1$ (where Q is an $N \times N$ matrix such that $QQ' = Q'Q = I$), then the probability of error of the ML receiver remains unchanged on an AWGN channel.*

Proof: The AWGN remains statistically equivalent after rotation by Q' . In particular consider $\tilde{\mathbf{n}} = Q'\mathbf{n}$, a rotated Gaussian random vector. ($\tilde{\mathbf{n}}$ is Gaussian since a linear combination of Gaussian random variables remains a Gaussian random variable). A Gaussian random vector is completely specified by its mean and covariance matrix: The mean is $E[\tilde{\mathbf{n}}] = 0$ since $E[n_i] = 0, \forall i = 0, \dots, N-1$. The covariance matrix is $E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}'] = Q'E[\mathbf{n}\mathbf{n}']Q = \frac{N_0}{2}I$. Thus, $\tilde{\mathbf{n}}$ is statistically equivalent to \mathbf{n} . The channel output for the rotated signal constellation is now $\tilde{\mathbf{y}} = \tilde{\mathbf{x}} + \mathbf{n}$ as illustrated in Figure 1.25. The corresponding decision rule is based on the distance from the received signal sample $\tilde{\mathbf{y}} = \tilde{\mathbf{v}}$ to the rotated constellation points $\tilde{\mathbf{x}}_i$.

$$\|\tilde{\mathbf{v}} - \tilde{\mathbf{x}}_i\|^2 = (\tilde{\mathbf{v}} - \tilde{\mathbf{x}}_i)' (\tilde{\mathbf{v}} - \tilde{\mathbf{x}}_i) \quad (1.85)$$

$$= (\mathbf{v} - \mathbf{x}_i)' Q'Q (\mathbf{v} - \mathbf{x}_i) \quad (1.86)$$

$$= \|\mathbf{v} - \mathbf{x}_i\|^2, \quad (1.87)$$

where $\mathbf{y} = \mathbf{x} + Q\mathbf{n}$. Since $\tilde{\mathbf{n}} = Q'\mathbf{n}$ has the same distribution as \mathbf{n} , and the distances measured in (1.87) are the same as in the original unrotated signal constellation, the ML detector for the rotated constellation is the same as the ML detector for the original (unrotated) constellation in terms of all distances and noise variances. Thus, the probability of error must be identical. **QED.**

An example of the QPSK constellation appears in Figure 1.21, where $N = 2$. With Q be a 45° rotation matrix,

$$Q = \begin{bmatrix} \cos \frac{\pi}{4} & \sin \frac{\pi}{4} \\ -\sin \frac{\pi}{4} & \cos \frac{\pi}{4} \end{bmatrix}, \quad (1.88)$$

then the rotated constellation and decision regions are shown in Figure 1.26. From Figure 1.26, clearly the rotation has not changed the detection problem and has only changed the labeling of the axes, effectively giving another equivalent set of orthonormal basis functions. Since rotation does not change the squared length of any of the data symbols, the average energy remains unchanged. The invariance does depend on the noise components being uncorrelated with one another, and of equal variance, as in (1.49); for other noise correlations (i.e., $n(t)$ not white, see Section 1.7) rotational invariance does not

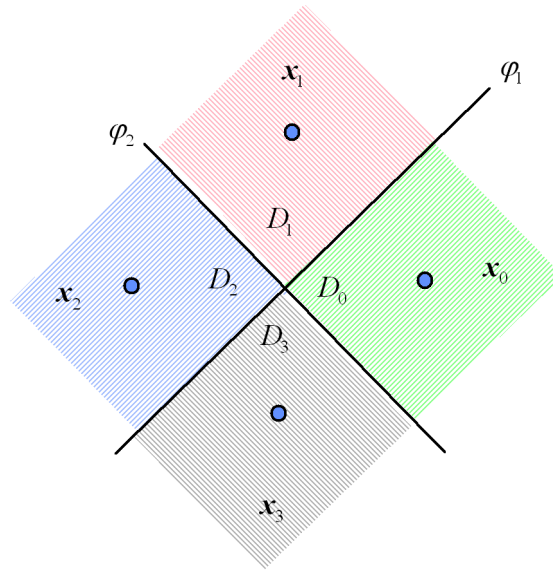


Figure 1.26: QPSK rotated by 45° .

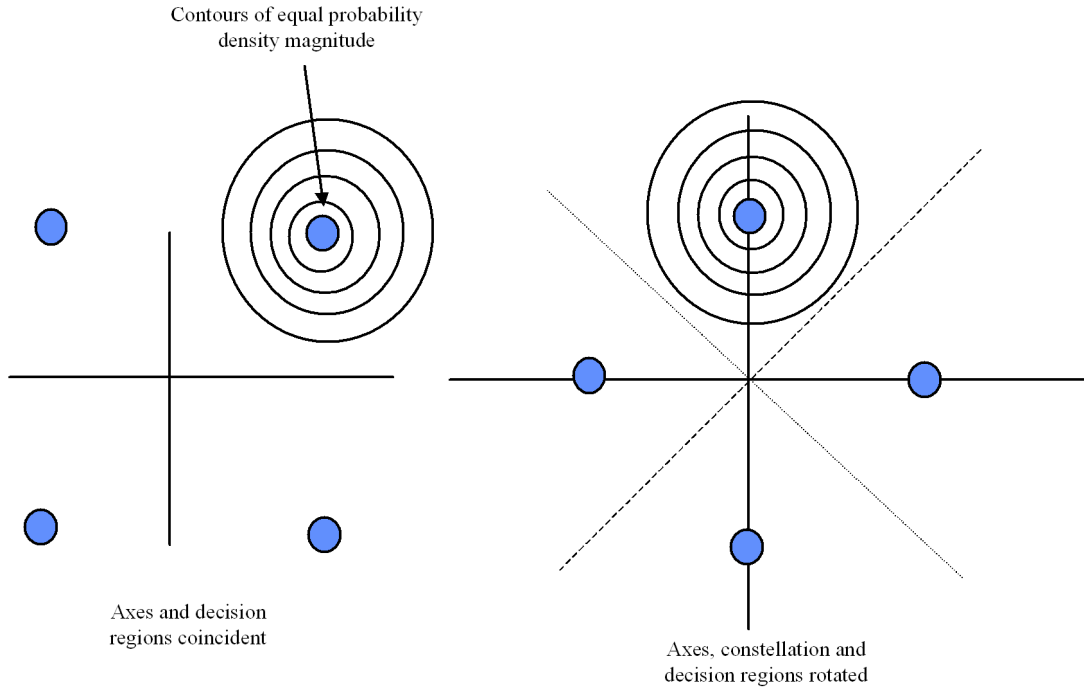


Figure 1.27: Rotational invariance summary.

hold. Rotational invariance is summarized in Figure 1.27. Each of the three diagrams shown in figures 1.26 and 1.27 have identical P_e when used with identical AWGN.

The probability of error is also invariant to translation by a constant vector amount for the AWGN, because again P_e depends only on relative distances and the noise remains unchanged.

Theorem 1.4.2 (Translational Invariance) *If all the data symbols in a signal constellation are translated by a constant vector amount, that is $\bar{\mathbf{x}}_i \leftarrow \mathbf{x}_i - \mathbf{a}$ for all $i = 0, \dots, M - 1$, then the probability of error of the ML detector remains unchanged on an AWGN channel.*

Proof: Note that the constant vector \mathbf{a} is common to both \mathbf{y} and to \mathbf{x} , and thus subtracts from $\|(\mathbf{v} - \mathbf{a}) - (\mathbf{x}_i - \mathbf{a})\|^2 = \|\mathbf{v} - \mathbf{x}_i\|^2$, so (1.82) remains unchanged. **QED.**

An important use of the Theorem of Translational Invariance is the **minimum energy translate** of a signal constellation:

Definition 1.4.1 (Minimum Energy Translate) *The minimum energy translate of a signal constellation is defined as that constellation obtained by subtracting the constant vector $E\{\mathbf{x}\}$ from each data symbol in the constellation.*

To show that the minimum energy translate has the minimum energy among all possible translations of the signal constellation, write the average energy of the translated signal constellation as

$$\mathcal{E}_{\mathbf{x}-\mathbf{a}} = \sum_{i=0}^{M-1} \|\mathbf{x}_i - \mathbf{a}\|^2 p_{\mathbf{x}}(i) \quad (1.89)$$

$$\begin{aligned} &= \sum_{i=0}^{M-1} [\|\mathbf{x}_i\|^2 - 2\langle \mathbf{x}_i, \mathbf{a} \rangle + \|\mathbf{a}\|^2] p_{\mathbf{x}}(i) \\ &= \mathcal{E}_{\mathbf{x}} + \|\mathbf{a}\|^2 - 2\langle E\{\mathbf{x}\}, \mathbf{a} \rangle \end{aligned} \quad (1.90)$$

From (1.90), the energy $\mathcal{E}_{\mathbf{x}-\mathbf{a}}$ is minimized over all possible translates \mathbf{a} if and only if $\mathbf{a} = E\{\mathbf{x}\}$, so

$$\min \mathcal{E}_{\mathbf{x}-\mathbf{a}} = \sum_{i=0}^{M-1} [\|\mathbf{x}_i - E\{\mathbf{x}\}\|^2 p_{\mathbf{x}}(i)] = \mathcal{E}_{\mathbf{x}} - [E(\mathbf{x})]^2 \quad (1.91)$$

Thus, as transmitter energy (or power) is often a quantity to be preserved, the engineer can always translate the signal constellation by $E\{\mathbf{x}\}$, to minimize the required energy without affecting performance. (However, there may be practical reasons, such as complexity and synchronization, where this translation is avoided in some designs.)

1.4.2 Union Bounding

Specific examples of calculating P_e appear in the next two subsections. This subsection illustrates this calculation for binary signaling in N dimensions for use in probability-of-error bounds.

Suppose a system has two signals in N dimensions, as illustrated for $N = 1$ dimension in Figure 1.19 with an AWGN channel. Then the probability of error for the ML detector is the probability that the component of the noise vector \mathbf{n} along the line connecting the two data symbols is greater than half the distance along this line. In this case, the noisy received vector \mathbf{y} lies in the incorrect decision region, resulting in an error. Since the noise is white Gaussian, its projection in any dimension, in particular, the segment of the line connecting the two data symbols, is of variance $\sigma^2 = \frac{N_0}{2}$, as was discussed in the proof of Theorem 1.4.1. Thus,

$$P_e = P\{\langle \mathbf{n}, \boldsymbol{\varphi} \rangle \geq \frac{d}{2}\} \quad (1.92)$$

where $\boldsymbol{\varphi}$ is a unit norm vector along the line between \mathbf{x}_0 and \mathbf{x}_1 and $d \triangleq \|\mathbf{x}_0 - \mathbf{x}_1\|$. This error probability is

$$P_e = \int_{\frac{d}{2}}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}u^2} du$$

$$\begin{aligned}
&= \int_{\frac{d}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \\
&= Q\left[\frac{d}{2\sigma}\right] .
\end{aligned} \tag{1.93}$$

The Q-function is defined in Appendix B of this chapter. As $\sigma^2 = \frac{N_0}{2}$, (1.93) can also be written

$$P_e = Q\left[\frac{d}{\sqrt{2N_0}}\right] . \tag{1.94}$$

Minimum Distance

Every signal constellation has an important characteristic known as the minimum distance:

Definition 1.4.2 (Minimum Distance, d_{\min}) *The minimum distance, $d_{\min}(\mathbf{x})$ is defined as the minimum distance between any two data symbols in a signal constellation $\mathbf{x} \triangleq \{\mathbf{x}_i\}_{i=0,\dots,M-1}$. The argument (\mathbf{x}) is often dropped when the specific signal constellation is obvious from the context, thus leaving*

$$d_{\min} \triangleq \min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\| \quad \forall i, j . \tag{1.95}$$

Equation (1.93) is useful in the proof of the following theorem for the probability of error of a ML detector for any signal constellation with M data symbols:

Theorem 1.4.3 (Union Bound) *The probability of error for the ML detector on the AWGN channel, with an M -point signal constellation with minimum distance d_{\min} , is bounded by*

$$P_e \leq (M-1)Q\left[\frac{d_{\min}}{2\sigma}\right] . \tag{1.96}$$

The proof of the Union Bound defines an “error event” ε_{ij} as the event where the ML detector chooses $\hat{\mathbf{x}} = \mathbf{x}_j$ while \mathbf{x}_i is the correct transmitted data symbol. The conditional probability of error given that \mathbf{x}_i was transmitted is then

$$P_{e/i} = P\{\varepsilon_{i0} \cup \varepsilon_{i1} \dots \cup \varepsilon_{i,i-1} \cup \varepsilon_{i,i+1} \cup \dots \cup \varepsilon_{i,M-1}\} = P\left\{\bigcup_{\substack{j=0 \\ (j \neq i)}}^{M-1} \varepsilon_{ij}\right\} . \tag{1.97}$$

Because the error events in (1.97) are mutually exclusive (meaning if one occurs, the others cannot), the probability of the union is the sum of the probabilities,

$$P_{e/i} = \sum_{\substack{j=0 \\ (j \neq i)}}^{M-1} P\{\varepsilon_{ij}\} \leq \sum_{\substack{j=0 \\ (j \neq i)}}^{M-1} P_2(\mathbf{x}_i, \mathbf{x}_j) , \tag{1.98}$$

where

$$P_2(\mathbf{x}_i, \mathbf{x}_j) \triangleq P\{\mathbf{y} \text{ is closer to } \mathbf{x}_j \text{ than to } \mathbf{x}_i\} , \tag{1.99}$$

because

$$P\{\varepsilon_{ij}\} \leq P_2(\mathbf{x}_i, \mathbf{x}_j) . \tag{1.100}$$

As illustrated in Figure 1.28, $P\{\varepsilon_{ij}\}$ is the probability the received vector \mathbf{y} lies in the shaded decision region for \mathbf{x}_j given the symbol \mathbf{x}_i was transmitted. The incorrect decision region for the probability $P_2(\mathbf{x}_i, \mathbf{x}_j)$ includes part (shaded red in Figure 1.28) of the region for $P\{\varepsilon_{ik}\}$, which explains the inequality in Equation (1.100). Thus, the union bound overestimates $P_{e/i}$ by integrating pairwise on overlapping half-planes.

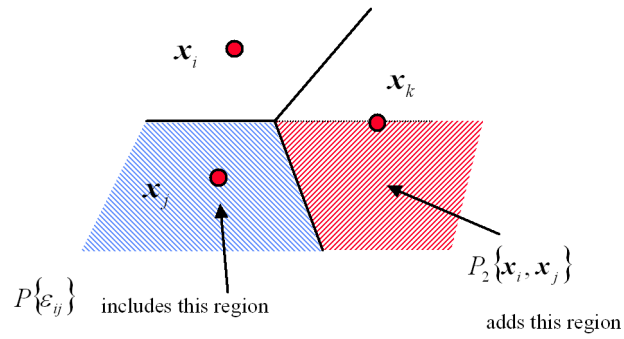


Figure 1.28: Probability of error regions.

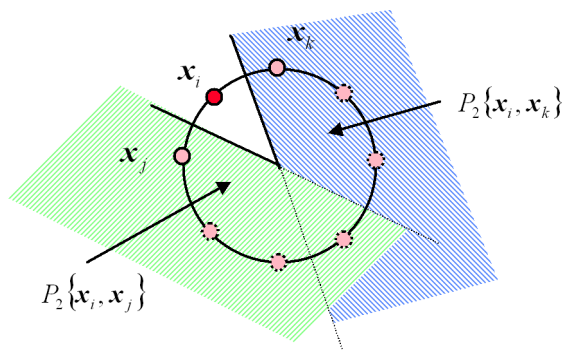


Figure 1.29: NNUB PSK constellation.

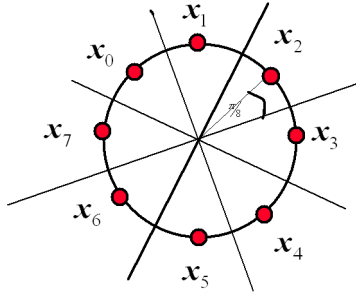


Figure 1.30: 8 Phase Shift Keying.

Using the result in (1.93),

$$P_2(\mathbf{x}_i, \mathbf{x}_j) = Q \left[\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma} \right] . \quad (1.101)$$

Substitution of (1.101) into (1.98) results in

$$P_{e/i} \leq \sum_{\substack{j=0 \\ (j \neq i)}}^{M-1} Q \left[\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma} \right] , \quad (1.102)$$

and thus averaging over all transmitted symbols

$$P_e \leq \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ (j \neq i)}}^{M-1} Q \left[\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma} \right] p_{\mathbf{x}}(i) . \quad (1.103)$$

$Q(x)$ is monotonically decreasing in x , and thus since $d_{\min} \leq \|\mathbf{x}_i - \mathbf{x}_j\|$,

$$Q \left[\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma} \right] \leq Q \left[\frac{d_{\min}}{2\sigma} \right] . \quad (1.104)$$

Substitution of (1.104) into (1.103), and recognizing that d_{\min} is not a function of the indices i or j , one finds the desired result

$$P_e \leq \sum_{i=0}^{M-1} (M-1) Q \left[\frac{d_{\min}}{2\sigma} \right] p_{\mathbf{x}}(i) = (M-1) Q \left[\frac{d_{\min}}{2\sigma} \right] . \quad (1.105)$$

QED.

Since the constellation contains M points, the factor $M-1$ equals the maximum number of neighboring constellation points that can be at distance d_{\min} from any particular constellation point.

Examples

The union bound can be tight (or exact) in some cases, but it is not always a good approximation to the actual P_e , especially when M is large. Two examples for $M=8$ show situations where the union bound is a poor approximation to the actual probability of error. These two examples also naturally lead to the “nearest neighbor” bound of the next subsection.

EXAMPLE 1.4.1 (8PSK) The constellation in Figure 1.30 is often called “eight phase” or “8PSK”. For the maximum likelihood detector, the 8 decision regions correspond to sectors

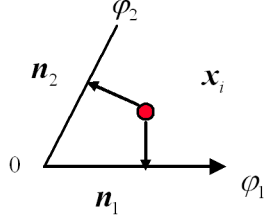


Figure 1.31: 8PSK P_e bounding.

bounded by straight lines emanating from the origin as shown in Figure 1.29. The union bound for 8PSK equals

$$P_e \leq 7Q \left[\frac{\sqrt{\mathcal{E}_x} \sin(\frac{\pi}{8})}{\sigma} \right] , \quad (1.106)$$

and $d_{\min} = 2\sqrt{\mathcal{E}_x} \sin(\frac{\pi}{8})$.

Figure 1.31 magnifies the detection region for one of the 8 data symbols. By symmetry the analysis would proceed identically, no matter which point is chosen, so $P_{e/i} = P_e$. An error can occur if the component of the additive white Gaussian noise, along either of the two directions shown, is greater than $d_{\min}/2$. These two events are not mutually exclusive, although the variance of the noise along either vector (with unit vectors along each defined as φ_1 and φ_2) is σ^2 . Thus,

$$P_e = P\{(\|\langle \mathbf{n}, \varphi_1 \rangle\| > \frac{d_{\min}}{2}) \cup (\|\langle \mathbf{n}, \varphi_2 \rangle\| > \frac{d_{\min}}{2})\} \quad (1.107)$$

$$\leq P\{n_1 > \frac{d_{\min}}{2}\} + P\{n_2 > \frac{d_{\min}}{2}\} \quad (1.108)$$

$$= 2Q \left[\frac{d_{\min}}{2\sigma} \right] , \quad (1.109)$$

which is a tighter “union bound” on the probability of error. Also

$$P\{\|\mathbf{n}_1\| > \frac{d_{\min}}{2}\} \leq P_e , \quad (1.110)$$

yielding a lower bound on P_e , thus the upper bound in (1.109) is tight. This bound is graphically illustrated in Figure 1.29. The bound in (1.109) overestimates the P_e by integrating the two half planes, which overlap as clearly depicted in the doubly shaded region of figure 1.28. The lower bound of (1.110) only integrates over one half plane that does not completely cover the shaded region. The multiplier in front of the Q function in (1.109) equals the number of “nearest neighbors” for any one data symbol in the 8PSK constellation.

The following second example illustrates problems in applying the union bound to a 2-dimensional signal constellation with 8 or more signal points on a rectangular grid (or lattice):

EXAMPLE 1.4.2 (8AMPM) Figure 1.32 illustrates an 8-point signal constellation called “8AMPM” (amplitude-modulated phase modulation), or “8 Square”. The union bound for P_e yields

$$P_e \leq 7Q \left[\frac{\sqrt{2}}{\sigma} \right] . \quad (1.111)$$

By rotational invariance the rotated 8AMPM constellation shown in Figure 1.33 has the same P_e as the unrotated constellation. The decision boundaries shown are pessimistic at

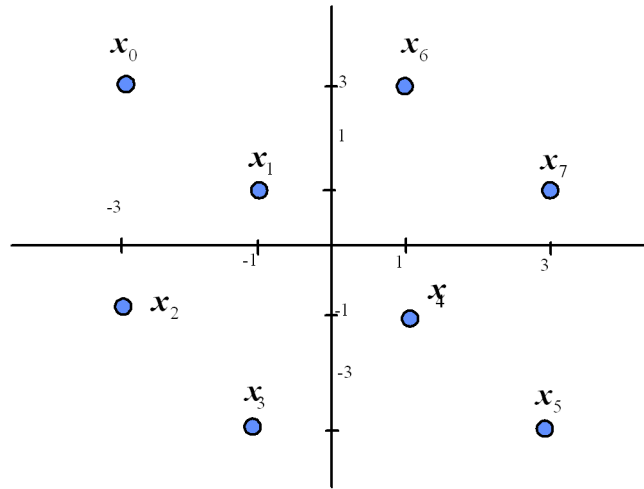


Figure 1.32: 8AMPM signal constellation.

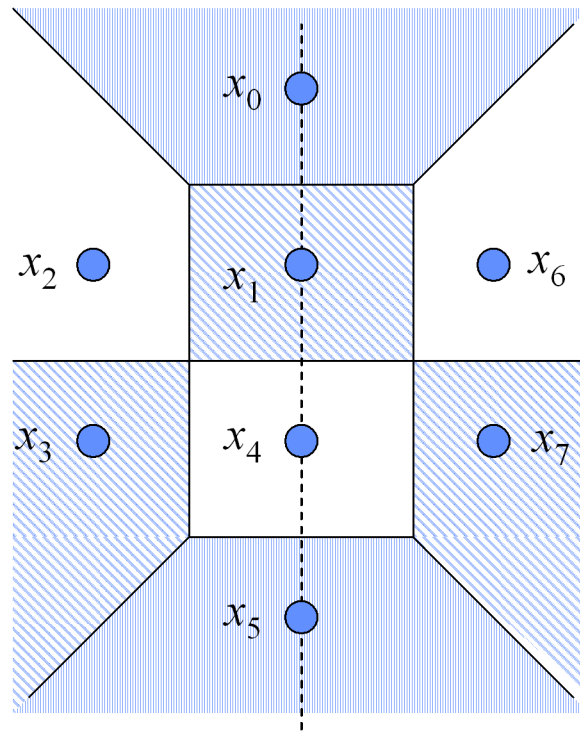


Figure 1.33: 8AMPM rotated by 45° with decision regions.

the corners of the constellation, so the P_e derived from them will be an upper bound. For notational brevity, let $Q \triangleq Q[d_{\min}/2\sigma]$. The probability of a correct decision for 8AMPM is

$$P_c = \sum_{i=0}^7 P_{c/i} \cdot p_{\mathbf{x}}(i) = \sum_{i \neq 1,4} P_{c/i} \cdot \frac{1}{8} + \sum_{i=1,4} P_{c/i} \cdot \frac{1}{8} \quad (1.112)$$

$$> \frac{6}{8}(1-Q)(1-2Q) + \frac{2}{8}(1-2Q)^2 \quad (1.113)$$

$$= \frac{3}{4}(1-3Q+2Q^2) + \frac{1}{4}(1-4Q+4Q^2) \quad (1.114)$$

$$= 1 - 3.25Q + 2.5Q^2 \quad (1.115)$$

Thus P_e is upper bounded by

$$P_e = 1 - P_c < 3.25Q \left[\frac{d_{\min}}{2\sigma} \right] \quad (1.116)$$

which is tighter than the union bound in (1.111). As M increases for constellations like 8AMPM, the accuracy of the union bound degrades, since the union bound calculates P_e by pairwise error events and thus redundantly includes the probabilities of overlapping half-planes. It is desirable to produce a tighter bound. The multiplier on the Q-function in (1.116) is the average number of nearest neighbors (or decision boundaries) $= \frac{1}{4}(4+3+3+3) = 3.25$ for the constellation. This rule of thumb, the Nearest-Neighbor Union bound (NNUB), often used by practicing data transmission engineers, is formalized in the next subsection.

1.4.3 The Nearest Neighbor Union Bound

The **Nearest Neighbor Union Bound** (NNUB) provides a tighter bound on the probability of error for a signal constellation by lowering the multiplier of the Q-function. The factor $(M-1)$ in the original union bound is often too large for accurate performance prediction as in the preceding section's two examples. The NNUB requires more computation. However, it is easily approximated.

The development of this bound uses the average number of nearest neighbors:

Definition 1.4.3 (Average Number of Nearest Neighbors) *The average number of neighbors, N_e , for a signal constellation is defined as*

$$N_e = \sum_{i=0}^{M-1} N_i \cdot p_{\mathbf{x}}(i) \quad (1.117)$$

where N_i is the number of neighboring constellation points of the point \mathbf{x}_i , that is the number of other signal constellation points sharing a common decision region boundary with \mathbf{x}_i . Often, N_e is approximated by

$$N_e \approx \sum_{i=0}^{M-1} \tilde{N}_i \cdot p_{\mathbf{x}}(i) \quad (1.118)$$

where \tilde{N}_i is the set of points at minimum distance from \mathbf{x}_i , whence the often used name “nearest” neighbors. This approximation is often very tight and facilitates computation of N_e when signal constellations are complicated (i.e., coding is used - see Chapters 6, 7, and 8).

Thus, N_e also measures the average number of sides of the decision regions surrounding any point in the constellation. These decision boundaries can be at different distances from any given point and thus might best not be called “nearest.” N_e is used in the following theorem:

Theorem 1.4.4 (Nearest Neighbor Union Bound) *The probability of error for the ML detector on the AWGN channel, with an M -point signal constellation with minimum distance d_{\min} , is bounded by*

$$P_e \leq N_e \cdot Q \left[\frac{d_{\min}}{2\sigma} \right] . \quad (1.119)$$

In the case that N_e is approximated by counting only “nearest” neighbors, then the NNUB becomes an approximation to probability of symbol error, and not necessary an upper bound.

Proof: Note that for each signal point, the distance to each decision-region boundary must be at least $d_{\min}/2$. The probability of error for point \mathbf{x}_i , $P_{e/i}$ is upper bounded by the union bound as

$$P_{e/i} \leq N_i \cdot Q \left[\frac{d_{\min}}{2\sigma} \right] . \quad (1.120)$$

Thus,

$$P_e = \sum_{i=0}^{M-1} P_{e/i} \cdot p_{\mathbf{x}}(i) \leq Q \left[\frac{d_{\min}}{2\sigma} \right] \sum_{i=0}^{M-1} N_i \cdot p_{\mathbf{x}}(i) = N_e \cdot Q \left[\frac{d_{\min}}{2\sigma} \right] . \quad (1.121)$$

QED.

The previous Examples 1.4.1 and 1.4.2 show that the Q -function multiplier in each case is exactly N_e for that constellation.

As signal set design becomes more complicated in Chapters 7 and 8, the number of nearest neighbors is commonly taken as only those neighbors who also are at minimum distance, and N_e is then approximated by (1.118). With this approximation, the P_e expression in the NNUB consequently becomes only an approximation rather than a strict upper bound.

1.4.4 Alternative Performance Measures

The optimum receiver design minimizes the symbol error probability P_e . Other closely related measures of performance can also be used.

An important measure used in practical system design is the **Bit Error Rate**. Most digital communication systems encode the message set $\{m_i\}$ into bits. Thus engineers are interested in the average number of bit errors expected. The bit error probability will depend on the specific binary labeling applied to the signal points in the constellation. The quantity $n_b(i, j)$ denotes the number of bit errors corresponding to a symbol error when the detector incorrectly chooses m_j instead of m_i , while $P\{\varepsilon_{ij}\}$ denotes the probability of this symbol error.

The bit error rate P_b obeys the following bound:

Definition 1.4.4 (Bit Error Rate) *The bit error rate is*

$$P_b \triangleq \sum_{i=0}^{M-1} \sum_{\substack{j \\ j \neq i}} p_{\mathbf{x}}(i) P\{\varepsilon_{ij}\} n_b(i, j) \quad (1.122)$$

where $n_b(i, j)$ is the number of bit errors for the particular choice of encoder when symbol i is erroneously detected as symbol j . This quantity, despite the label using P , is not strictly a probability.

The bit error rate will always be approximated for the AWGN in this text by:

$$P_b \approx \sum_{i=0}^{M-1} \sum_{j=1}^{N_i} p_{\mathbf{x}}(i) P\{\varepsilon_{ij}\} n_b(i, j) \quad (1.123)$$

$$\begin{aligned}
&\leq Q \left[\frac{d_{\min}}{2\sigma} \right] \sum_{i=0}^{M-1} p_{\mathbf{x}}(i) \sum_{j=1}^{N_i} n_b(i, j) \\
P_b &\lesssim Q \left[\frac{d_{\min}}{2\sigma} \right] \sum_{i=0}^{M-1} p_{\mathbf{x}}(i) n_b(i) \\
&\lesssim N_b \cdot Q \left[\frac{d_{\min}}{2\sigma} \right]
\end{aligned} \tag{1.124}$$

where

$$n_b(i) \triangleq \sum_{j=1}^{N_i} n_b(i, j) \quad , \tag{1.125}$$

and the **Average Total Bit Errors per Error Event**, N_b , is defined as:

$$N_b = \sum_{i=0}^{M-1} p_{\mathbf{x}}(i) \cdot n_b(i) \quad . \tag{1.126}$$

An expression similar to the NNUB for P_b is

$$P_b \approx N_b Q \left[\frac{d_{\min}}{2\sigma} \right] \quad , \tag{1.127}$$

where the approximation comes from Equation (1.123), which is an approximation because of the reduction in the number of included terms in the sum over other points. The accuracy of this approximation is good as long as those terms corresponding to distant neighbors have small value in comparison to nearest neighbors, which is a reasonable assumption for good constellation designs. The bit error rate is sometimes a more uniform measure of performance because it is independent of M and N . On the other hand, P_e is a block error probability (with block length N) and can correspond to more than one bit in error (if $M > 2$) over N dimensions. Both P_e and P_b depend on the same distance-to-noise ratio (the argument of the Q function). While the notation for P_b is commonly expressed with a P , the bit error rate is not a probability and could exceed unity in value in aberrant cases. A better measure that is a probability is to normalize the bit-error rate by the number of bits per symbol: Normalization of P_b produces a probability measure because it is the average number of bit errors divided by the number of bits over which those errors occur - this probability is the desired probability of bit error:

Lemma 1.4.1 (Probability of bit error \bar{P}_b .) *The probability of bit error is defined by*

$$\bar{P}_b = \frac{P_b}{b} \quad . \tag{1.128}$$

The corresponding average total number of bit errors per bit is

$$\bar{N}_b \triangleq \frac{N_b}{b} \quad . \tag{1.129}$$

The bit error rate can exceed one, but the probability of bit error never exceeds one.

Furthermore, comparison of values of P_e between systems of different dimensionality is not fair (for instance to compare a 2B1Q system operating at $P_e = 10^{-7}$ against a multi-dimensional design consisting of 10 successive 2B1Q dimensions decoded jointly as a single symbol also with $P_e = 10^{-7}$, the latter system really has 10^{-8} errors per dimension and so is better.) A more fair measure of symbol error probability normalizes the measure by the dimensionality (or number of bits per symbol) of the system to compare systems with different block lengths.

Definition 1.4.5 (Normalized Error Probability \bar{P}_e .) *The normalized error probability is defined by*

$$\bar{P}_e \triangleq \frac{P_e}{N} \quad . \tag{1.130}$$

The normalized average number of nearest neighbors is:

Definition 1.4.6 (Normalized Number of Nearest Neighbors) *The normalized number of nearest neighbors, \bar{N}_e , for a signal constellation is defined as*

$$\bar{N}_e = \sum_{i=0}^{M-1} \frac{N_i}{N} \cdot p_{\mathbf{x}}(i) = \frac{N_e}{N} \quad . \quad (1.131)$$

Thus, the NNUB is

$$\bar{P}_e \leq \bar{N}_e \cdot Q \left[\frac{d_{\min}}{2\sigma} \right] \quad . \quad (1.132)$$

EXAMPLE 1.4.3 (8AMPM) The average number of bit errors per error event for 8AMPM using the octal labeling indicated by the subscripts in Figure 1.32 is computed by

$$\begin{aligned} N_b &= \sum_{i=0}^7 \frac{1}{8} n_b(i) \\ &= \frac{1}{8} [(1 + 1 + 2) + (3 + 1 + 2 + 2) + \end{aligned} \quad (1.133)$$

$$(2 + 1 + 1) + (1 + 2 + 3) + (3 + 2 + 2 + 1) + \quad (1.134)$$

$$+(1 + 1 + 2) + (3 + 1 + 2) + (1 + 2 + 1)] \quad (1.135)$$

$$= \frac{44}{8} = 5.5 \quad . \quad (1.136)$$

Then

$$P_b \approx 5.5 \cdot Q \left[\frac{d_{\min}}{2\sigma} \right] \quad . \quad (1.137)$$

Also,

$$\bar{N}_e = \frac{3.25}{2} = 1.625 \quad (1.138)$$

so that

$$\bar{P}_e \leq 1.625 \cdot Q \left[\frac{d_{\min}}{2\sigma} \right] \quad , \quad (1.139)$$

and

$$\bar{P}_b \approx \frac{5.5}{3} \cdot Q \left[\frac{d_{\min}}{2\sigma} \right] \quad . \quad (1.140)$$

Thus the bit error rate is somewhat higher than the normalized symbol error rate. Careful assignment of bits to symbols can reduce the bit error rate slightly.

1.4.5 Block Error Measures

Higher-level engineering of communication systems may desire knowledge of message errors within packets of several messages cascaded into a larger message. An entire packet may be somewhat useless if any part of it is in error. Thus, the concept of a symbol from this perspective of analysis may be the entire packet of messages. The **probability of “block” or “packet” error** is truly identical to the probability of symbol error already analyzed as long as the entire packet is considered as a single symbol.

If a packet contains B bits, each of which independently has probability of bit error \bar{P}_b , then the probability of packet (or block) error is often approximated by $B \cdot \bar{P}_b$. Clearly then if one says the probability of packet error is 10^{-7} and there are 125 bytes per packet, or 1000 bits per packet, then the probability of bit error would then be 10^{-10} . Low packet error rate is thus a more stringent criterion on the detector performance than is low probability of bit error. Nonetheless, analysis can proceed exactly as in this section. As B increases, the approximation above of $P_e = B \cdot \bar{P}_b$ can become inaccurate as we see below.

An **errored second** is often used in telecommunications as a measure of performance. An errored second is any second in which any bit error occurs. Obviously, fewer errored seconds is better. A given fixed number of error seconds translates into increasingly lower probability of bit error as the data rate of the channel increases. An **error-free second** is a second in which no error occurs. If a second contains B independent bits, then the exact probability of an error-free second is

$$P_{efs} = (1 - \bar{P}_b)^B \quad (1.141)$$

while the exact probability of an errored second is

$$P_e = 1 - P_{efs} = \sum_{i=1}^B \binom{B}{i} (1 - \bar{P}_b)^{B-i} \bar{P}_b^i \quad (1.142)$$

Dependency between bits and bit errors will change the exact nature of the above formulae, but is usually ignored in calculations. More common in telecommunications is the derived concept of **percentage error free seconds** which is the percentage of seconds that are error free. Thus, if a detector has $\bar{P}_b = 10^{-7}$ and the data rate is 10 Mbps, then one might naively guess that almost every second contains errors according to $P_e = B \cdot \bar{P}_b$, and the percentage of error free seconds is thus very low. To be exact, $P_{efs} = (1 - 10^{-7})^{10^7} = .368$, so that the link has 36.8% error free seconds, so actually about 63% of the seconds have errors. Typically large telecommunications networks strive for **five nines** reliability, which translates into 99.999% error free seconds. At 10 Mbps, this means that the detector has $\bar{P}_b = 1 - e^{10^{-7} \ln(.99999)} = 2.3 \cdot 10^{-12}$. At lower data rates, five nines is less stringent on the channel error probability. Data networks today, often designed for bit error rates above 10^{-12} operate at 10 Mbps with external “error detection and retransmission” protocols. Retransmission may not be acceptable for continuous signals like voice or video, so that five nines reliability is often not possible on the data network – this has become a key issue in the convergence of telecommunications networks (designed with 5 nines reliability normally for voice transmission over the last 5 decades) and data networks, designed with higher error rates for data transmission over the last 3 decades. (Often though, the data network probability error is much better than the specification, so systems may work fine without any true understanding of the designers as to exactly why. However, Voice (“VoIP”) and video (“IPTV”) signals on data networks often exhibit quality issues, which is a function of the packet error rate being too high and also of the failure of retransmission approaches to restore the quality.)

In any case, the probability of symbol and bit error are fundamental to all other measures of network performance and can be used by the serious communication engineer to evaluate carefully a system’s performance.

1.5 General Classes of Constellations and Modulation

This section describes three classes of modulation that abound in digital data transmission. Each of these three classes represent different geometric approaches to constellation construction. Three successive subsections examine the choice of basis functions for each modulation class and develop corresponding general expressions for the average probability of error P_e for these modulation classes when used on the AWGN channel. Subsection 1.5.1 discusses cubic constellations (Section 1.6 also investigates some important extensions to the cubic constellations). Subsection 1.5.2 examines orthogonal constellations, while Subsection 1.5.3 studies circular constellations.

To compare constellations, some constellation measures are developed first. The cost of modulation depends upon transmitted power (energy per unit time). A unit of time translates to a number of dimensions, given a certain system bandwidth, so the energy per dimension is essentially a measure of power. Given a wider bandwidth, the same unit in time and power will correspond to proportionately more dimensions, but a lower power spectral density. While somewhat loosely defined, a system with symbol period T and bandwidth¹⁰ W , has a number of dimensions available for signal constellation construction

$$N = 2 \cdot W \cdot T \quad \text{dimensions.} \quad (1.143)$$

¹⁰It is theoretically not possible to have finite bandwidth and finite time extent, but in practice this can be approximated closely.

The reasons for this approximation will become increasingly apparent, but all the methods of this section will follow this simple rule when the reasonable and obvious definition of bandwidth is applied. Systems in practice all follow this rule (or have fewer dimensions than this practical maximum) even though it may be possible to construct signal sets with slightly more dimensions theoretically. The number of dimensions in any case is a combined measure of the system resources of bandwidth and time - thus, performance measures and energy are thus often normalized by N for fair comparison. The data rate concept thus generalizes to the **number of bits per dimension**:

Definition 1.5.1 (Average Number of Bits Per Dimension) *The average number of bits per dimension, \bar{b} , for a signal constellation \mathbf{x} , is*

$$\bar{b} \triangleq \frac{b}{N} . \quad (1.144)$$

*The related quantity, **data rate**, is*

$$R = \frac{b}{T} . \quad (1.145)$$

Using (1.143), one can compute that

$$2\bar{b} = \frac{R}{W} , \quad (1.146)$$

the **spectral efficiency** of a modulation method which is often used by transmission engineers to describe an efficiency of transmission (how much data rate per unit of bandwidth). Spectral efficiency is often described in terms of the unit bits/second/Hz, which is really a measure of double the number of bits/dimension. Engineers often abbreviated the term bits/second/Hz to say bits/Hz, which is an (unfortunately) often used and confusing term because the units are incorrect. Nonetheless, experienced engineers automatically translate the verbal abbreviation bits/Hz to the correct units and interpretation, bits-per-second/Hz, or simply double the number of bits/dimension.

The concept of power also generalizes to **energy per dimension**:

Definition 1.5.2 (Average Energy Per Dimension) *The average energy per dimension, $\bar{\mathcal{E}}_{\mathbf{x}}$, for a signal constellation \mathbf{x} , is*

$$\bar{\mathcal{E}}_{\mathbf{x}} \triangleq \frac{\mathcal{E}_{\mathbf{x}}}{N} . \quad (1.147)$$

*A related quantity is the **average power**,*

$$P_{\mathbf{x}} = \frac{\mathcal{E}_{\mathbf{x}}}{T} . \quad (1.148)$$

Clearly N cannot exceed the actual number of dimensions in the constellation, but the constellation may require fewer dimensions for a complete representation. For example the two-dimensional constellation in Figure 1.7 can be described using only one basis vector simply by rotating the constellation by 45 degrees. The average power, which was also defined earlier, is a scaled quantity, but consistently defined for all constellations. In particular, the normalization of basis functions often absorbs gain into the signal constellation definition that may tacitly conceal complicated calculations based on transmission-channel impedance, load matching, and various non-trivially calculated analog effects. These effects can also be absorbed into bandlimited channel models as is the case in Chapters 2, 3, 4, 10 and 11.

The energy per dimension allows the comparison of constellations with different dimensionality. The smaller the $\bar{\mathcal{E}}_{\mathbf{x}}$ for a given \bar{P}_e and \bar{b} , the better the design. The concatenation of two successively transmitted N -dimensional signals taken from the same N -dimensional signal constellation as a single $2N$ -dimensional signal causes the resulting $2N$ -dimensional constellation, formed as a Cartesian product of the constituent N -dimensional constellations, to have the same average energy per dimension as the N -dimensional constellation. Thus, simple concatenation of a signal set with itself does not improve the design. However, careful packing of signals in increasingly larger dimensional signals sets can lead to a reduction in the energy per dimension required to transmit a given set of messages, which will be of interest in this section and throughout this text.

The average power is the usual measure of energy per unit time and is useful when sizing the power requirements of a modulator or in determining scale constants for analog filter/driver circuits in the actual implementation. The power can be set equal to the square of the voltage over the load resistance.

The **noise energy per dimension** for an N -dimensional AWGN channel is

$$\bar{\sigma}^2 = \frac{\sum_{l=1}^N \sigma^2}{N} = \sigma^2 = \frac{\mathcal{N}_0}{2} . \quad (1.149)$$

While AWGN is inherently infinite dimensional, by the theorem of irrelevance, a computation of probability of error need only consider the noise components in the N dimensions of the signal constellation.

For AWGN channels, the **signal-to-noise ratio (SNR)** is used often by this text to characterize the channel:

Definition 1.5.3 (SNR) *The SNR is*

$$SNR = \frac{\bar{\mathcal{E}}_{\mathbf{x}}}{\sigma^2} \quad (1.150)$$

As shown in Section 1.4, the performance of a constellation in the presence of AWGN depends on the minimum distance between any two vectors in the constellation. Increasing the distance between points in a particular constellation increases the average energy per dimension of the constellation. The ‘‘Constellation Figure of Merit.’’¹¹ combines the energy per dimension and the minimum distance measures:

Definition 1.5.4 (Constellation Figure of Merit - CFM) *The constellation figure of merit, $\zeta_{\mathbf{x}}$ for a signal constellation \mathbf{x} , is*

$$\zeta_{\mathbf{x}} \triangleq \frac{\left(\frac{d_{\min}}{2}\right)^2}{\bar{\mathcal{E}}_{\mathbf{x}}} , \quad (1.151)$$

a unit-less quantity, defined only when $\bar{b} \geq 1$.

The CFM $\zeta_{\mathbf{x}}$ will measure the quality of any constellation used with an AWGN channel. A higher CFM $\zeta_{\mathbf{x}}$ generally results in better performance. The CFM should only be used to compare systems with equal numbers of bits per dimension $\bar{b} = b/N$, but can be used to compare systems of different dimensionality.

A different measure, known as the ‘‘energy per bit,’’ measures performance in systems with low average bit rate of $\bar{b} \leq 1$ (see Chapter 7).

Definition 1.5.5 (Energy Per Bit) *The energy per bit, \mathcal{E}_b , in a signal constellation $\{\mathbf{x}\}$ is:*

$$\mathcal{E}_b = \frac{\mathcal{E}_{\mathbf{x}}}{b} = \frac{\bar{\mathcal{E}}_{\mathbf{x}}}{\bar{b}} . \quad (1.152)$$

This measure is only defined when $\bar{b} \leq 1$ and has no meaning in other contexts.

Fair comparisons of modulation types consider the following parameters:

1. data rate R
2. power P_x
3. total bandwidth needed for all basis functions W
4. symbol period T
5. probability of error, \bar{P}_b or possibly P_e .

¹¹G. D. Forney, Jr., 8/89 IEEE Journal on Selected Areas in Communication.

Any 4 of these parameters may be held constant for two compared modulation methods, while the 5th varies and determines the better method. This set of 5 parameters can be reduced to 3 by normalizing to the number of dimensions:

1. bits per dimension \bar{b}
2. energy per dimension $\bar{\mathcal{E}}_{\mathbf{x}}$
3. normalized probability of symbol error, \bar{P}_e .

Any two can be fixed while the third is varied for the comparison. This is one of the advantages of the concept of normalizing to a number of dimensions. The constellation figure of merit presumes \bar{b} fixed and then looks at the ratio of d_{\min}^2 to $\bar{\mathcal{E}}_{\mathbf{x}}$, essentially holding $\bar{\mathcal{E}}_{\mathbf{x}}$ fixed and looking at \bar{P}_e (equivalent to d_{\min} if nearest neighbors are ignored on the AWGN). The normalization essentially prevents an excess of symbol period or bandwidth from letting one modulation method look better than another, tacitly including the third and fourth parameters (bandwidth and symbol period) from the list of parameters in a comparison.

Definition 1.5.6 (margin) *The margin of a transmission system is the amount by which the argument of the Q-function can be reduced while retaining the probability of error below a specified maximum that is associated with the margin.*

Margin's are often quoted in transmission design as they give a level of confidence to designers that unforeseen noise increases or signal attenuation will not cause the system performance to become unacceptable.

EXAMPLE 1.5.1 (Margin in DSL) Digital Subscriber Line systems deliver 100's of kilobits to 10's of megabits of data over telephone lines and use sophisticated adaptive modulation systems described later in Chapters 4 and 5. The two modems are located at the ends of the telephone line at the telephone-company central office and at the customer's premise. However, they ultimately also have probability of error specified by a relation of the form $N_e \cdot Q(d_{\min}/2\sigma)$. Because noise sources can be unpredictable on telephone lines, which tend to sense everything from other phone lines' signals to radio signals to refrigerator doors and fluorescent and other lights, and because customer-location additional wiring to the modem can be poor grade or long, a margin of at least 6 dB is mandated at the data rate of service offered if the customer is to be allowed service. This 6 dB essentially allows performance to be degraded by a combined factor of 4 in increased noise before costly manual maintenance or repair service would be necessary.

1.5.1 Cubic Constellations

Cubic constellations are commonly used on simple data communication channels. Some examples of cubic signal constellations are shown in Figure 1.34 for $N = 1, 2,$ and 3 . The construction of a cubic constellation directly maps a sequence of $N = b$ bits into the components of the basis vectors in a corresponding N -dimensional signal constellation. For example, the bit stream $\dots 010010\dots$ may be grouped as the sequence of two-dimensional vectors $\dots (01)(00)(10)\dots$. The resulting constellation is uniformly scaled in all dimensions, and may be translated or rotated in the N -dimensional space it occupies.

The simplest cubic constellation appears in Figure 1.34, where $N = b = \bar{b} = 1$. This constellation is known as "binary signaling", since only two possible signals are transmitted using one basis function $\varphi_1(t)$. Several examples of binary signaling are described next.

Binary Antipodal Signaling

In binary antipodal signaling, the two possible values for $\mathbf{x} = x_1$ are equal in magnitude but opposite in sign, e.g. $x_1 = \pm \frac{d}{2}$. As for all binary signaling methods, the average probability of error is

$$P_e = P_b = Q\left[\frac{d_{\min}}{2\sigma}\right] . \quad (1.153)$$

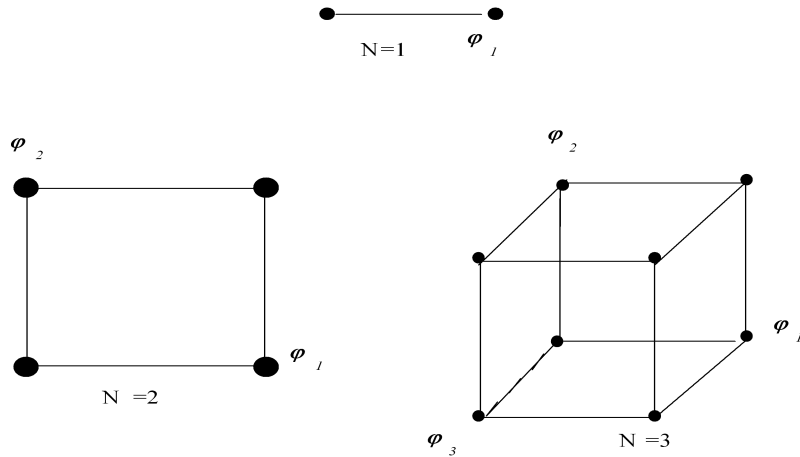


Figure 1.34: Cubic constellations for $N = 1, 2,$ and $3.$

The CFM for binary antipodal signaling equals $\zeta_{\mathbf{x}} = (d/2)^2 / [(d/2)^2] = 1.$

Particular types of binary antipodal signaling differ only in the choice of the basis function $\varphi_1(t).$ In practice, these basis functions may include “Nyquist” pulse shaping waveforms to avoid intersymbol interference. Chapter 3 further discusses waveform shaping. Besides the time-domain shaping, the basis function $\varphi_1(t)$ shapes the power spectral density of the resultant modulated waveform. Thus, different basis functions may require different bandwidths.

Binary Phase Shift Keying

Binary Phase Shift Keying (BPSK) uses a sinusoid to modulate the sequence of data symbols $\{\pm\sqrt{\mathcal{E}_x}\}.$

$$\varphi_1(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin \frac{2\pi t}{T} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.154)$$

This representation uses the minimum number of basis functions $N = 1$ to represent BPSK, rather than $N = 2$ as in Example 1.1.2.

Bipolar (NRZ) transmission

Bipolar signaling, also known as “baseband binary” or “Non-Return-to-Zero (NRZ)” signaling, uses a square pulse to modulate the sequence of data symbols $\{\pm\sqrt{\mathcal{E}_x}\}.$

$$\varphi_1(t) = \begin{cases} \frac{1}{\sqrt{T}} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.155)$$

Manchester Coding (Bi-Phase Level)

Manchester Coding, also known as “biphase level” (BPL) or, in magnetic and optical recording, as “frequency modulation,” uses a sequence of two opposite phase square pulses to modulate each data symbol. In NRZ signaling long runs of the same bit result in a constant output signal with no transitions until the bit changes. Since timing recovery circuits usually require some transitions, Manchester or BPL guarantees a transition occurs in the middle of each bit (or symbol) period $T.$ The basis function is:

$$\varphi_1(t) = \begin{cases} \frac{1}{\sqrt{T}} & 0 \leq t < T/2 \\ -\frac{1}{\sqrt{T}} & T/2 \leq t < T \\ 0 & \text{elsewhere} \end{cases} \quad (1.156)$$

The power spectral density of the modulated signal is related to the Fourier transform $\Phi_1(f)$ of the pulse $\varphi_1(t)$. The Fourier transform of the NRZ square pulse is a sinc function with zero crossings spaced at $\frac{1}{T}$ Hz. The basis function for BPL in Equation (1.156) requires approximately twice the bandwidth of the basis function for NRZ in Equation (1.155), because the Fourier transform of the biphasic pulse is a sinc function with zero crossings spaced at $\frac{2}{T}$ Hz. Similarly BPSK requires double the bandwidth of NRZ. Both BPSK and BPL are referred to as “rate 1/2” transmission schemes, because for the same bandwidth they permit only half the transmitted transmission rate compared with NRZ.

On-Off Keying (OOK)

On-Off Keying, used in direct-detection optical data transmission, as well as in “gate-to-gate” transmission in most digital circuits, uses the same basis function as bipolar transmission.

$$\varphi_1(t) = \begin{cases} \frac{1}{\sqrt{T}} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.157)$$

Unlike bipolar transmission, however, one of the levels for x_1 is zero, while the other is nonzero ($\sqrt{2\mathcal{E}_x}$). Because of the asymmetry, this method includes a DC offset, i.e. a nonzero mean value. The CFM is $\zeta_x = .5$, and thus OOK is 3dB inferior to any type of binary antipodal transmission. The comparison between signal constellations is $10 \log_{10} [\zeta_{x,OOK}/\zeta_{x,NRZ}] = 10 \log_{10}(0.5) = -3$ dB.

As for any binary signaling method, OOK has

$$P_e = P_b = Q \left[\frac{d_{\min}}{2\sigma} \right] \quad (1.158)$$

Vertices of a Hypercube (Block Binary)

Binary signaling in one dimension generalizes to the corners of a hypercube in N -dimensions, hence the name “cubic constellations.” The hypercubic constellations all transmit an average of $\bar{b} = 1$ bit per dimension. For two dimensions, the most common signal set is QPSK.

Quadrature Phase Shift Keying (QPSK)

The basis functions for the two dimensional QPSK constellation are

$$\varphi_1(t) = \begin{cases} \sqrt{\frac{2}{T}} \cos \frac{2\pi t}{T} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.159)$$

$$\varphi_2(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin \frac{2\pi t}{T} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.160)$$

The transmitted signal is a linear combination of both an inphase (cos) component and a quadrature (sin) component. The four possible data symbols are

$$[x_1 \ x_2] = \left\{ \begin{array}{l} \sqrt{\frac{\mathcal{E}_x}{2}} [-1 \ -1] \\ \sqrt{\frac{\mathcal{E}_x}{2}} [-1 \ +1] \\ \sqrt{\frac{\mathcal{E}_x}{2}} [+1 \ -1] \\ \sqrt{\frac{\mathcal{E}_x}{2}} [+1 \ +1] \end{array} \right\} \quad (1.161)$$

The additional basis function does not require any extra bandwidth with respect to BPSK, and the average energy \mathcal{E}_x remains unchanged. While the minimum distance d_{\min}^2 has decreased by a factor of two, the number of dimensions has doubled, thus the CFM for QPSK is $\zeta_x = 1$ again, as with BPSK.

For performance evaluation, it is easier to compute the average probability of a correct decision P_c rather than P_e for maximum likelihood detection on the AWGN channel with

equally probable signals. By symmetry of the signal constellation, $P_{c|i}$ is identical $\forall i = 0, \dots, 3$.

$$\begin{aligned} P_c &= \sum_{i=0}^3 P_{c|i} p_{\mathbf{x}}(i) = P_{c/i} \\ &= \left(1 - Q\left[\frac{d_{\min}}{2\sigma}\right]\right) \left(1 - Q\left[\frac{d_{\min}}{2\sigma}\right]\right) \end{aligned} \quad (1.162)$$

$$= 1 - 2Q\left[\frac{d_{\min}}{2\sigma}\right] + \left(Q\left[\frac{d_{\min}}{2\sigma}\right]\right)^2. \quad (1.163)$$

To prove the step from the first to second line, note that the noise in the two dimensions is independent. The probability of a correct decision requires both noise components to fall within the decision region, which gives the product in (1.162). Thus

$$P_e = 1 - P_c \quad (1.164)$$

$$= 2Q\left[\frac{d_{\min}}{2\sigma}\right] - \left(Q\left[\frac{d_{\min}}{2\sigma}\right]\right)^2 < 2Q\left[\frac{d_{\min}}{2\sigma}\right], \quad (1.165)$$

where $d_{\min} = \sqrt{2\mathcal{E}_{\mathbf{x}}} = 2\bar{\mathcal{E}}_{\mathbf{x}}^{1/2}$. For reasonable error rates ($P_e < 10^{-2}$), the $\left(Q\left[\frac{d_{\min}}{2\sigma}\right]\right)^2$ term in (1.165) is negligible, and the bound on the right, which is also the NNUB, is tight. With a “reasonable” mapping of bits to data symbols (e.g. the Gray code $0 \rightarrow -1$ and $1 \rightarrow +1$), the probability of a bit error $\bar{P}_b = \bar{P}_e$ for QPSK. P_e for QPSK is twice P_e for BPSK, but \bar{P}_e is the same for both systems. Comparing \bar{P}_e is usually more informative.

Block Binary

For hypercubic signal constellations in three or more dimensions, $N \geq 3$, the signal points are the vertices of a hypercube centered on the origin. In this case, the probability of error generalizes to

$$P_e = 1 - \left(1 - Q\left[\frac{d_{\min}}{2\sigma}\right]\right)^N < NQ\left[\frac{d_{\min}}{2\sigma}\right]. \quad (1.166)$$

where $d_{\min} = 2\bar{\mathcal{E}}_{\mathbf{x}}^{1/2}$. The basis functions are usually given by $\varphi_n(t) = \varphi(t - nT)$, where $\varphi(t)$ is the square pulse given in (1.155). The transmission of one symbol with the hypercubic constellation requires a time interval of length NT . Alternatively, scaling of the basis functions in time can retain a symbol period of length T , but the narrower pulse will require N times the bandwidth as the T width pulses. For this case again $\zeta_{\mathbf{x}} = 1$. As $N \rightarrow \infty$, $P_e \rightarrow 1$. While the probability of any single dimension being correct remains constant and less than one, as N increases, the probability of all dimensions being correct decreases.

Ignoring the higher order terms $Q^i, i \geq 2$, the average probability of error is approximately $\bar{P}_e \approx Q(d_{\min}/(2\sigma))$, which equals \bar{P}_e for binary antipodal signaling. This example illustrates that increasing dimensionality does not always reduce the probability of error unless the signal constellation has been carefully designed. As block binary constellations are just a concatenation of several binary transmissions, the receiver can equivalently decode each of the independent dimensions separately. However, with a careful selection of the transmitted signal constellation, it is possible to drive the probability of both a message error P_e and a bit error P_b to zero with increasing dimensionality N , as long as the average number of transmitted bits per unit time does not exceed a fundamental rate known as the “capacity” of the communication channel. (Chapter 8)

1.5.2 Orthogonal Constellations

In orthogonal signal sets, the dimensionality increases linearly with the number of points $M = \alpha N$ in the signal constellation, which results in a decrease in the number of bits per dimension $\bar{b} = \frac{\log_2(M)}{N} = \frac{\log_2(\alpha N)}{N}$.

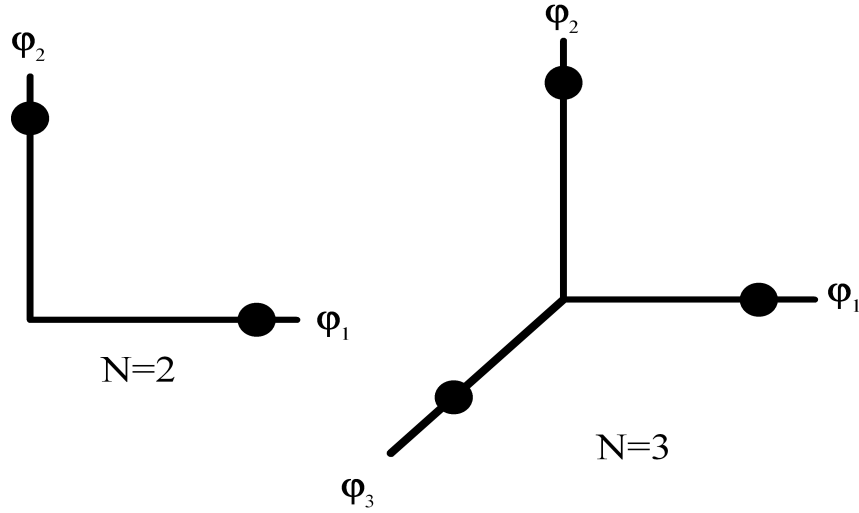


Figure 1.35: Block orthogonal constellations for $N = 2$, and 3 .

Block Orthogonal

Block orthogonal signal constellations have a dimension, or basis function, for each signal point. The block orthogonal signal set thus consists of $M = N$ orthogonal signals $x_i(t)$, that is

$$\langle x_i(t), x_j(t) \rangle = \mathcal{E}_x \delta_{ij} \quad . \quad (1.167)$$

Block orthogonal signal constellations appear in Figure 1.35 for $N = 2$ and 3 . The signal constellation vectors are, in general,

$$\mathbf{x}_i = \left[0 \dots 0 \sqrt{\mathcal{E}_x} 0 \dots 0 \right] = \sqrt{\mathcal{E}_x} \boldsymbol{\varphi}_{i+1} \quad . \quad (1.168)$$

The CFM should not be used on block orthogonal signal sets because $\bar{b} < 1$.

As examples of block orthogonal signaling, consider the following two dimensional signal sets.

Return to Zero (RZ) Signaling

RZ uses the following two basis functions for the two-dimensional signal constellation shown in Figure 1.35:

$$\varphi_1(t) = \begin{cases} \frac{1}{\sqrt{T}} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.169)$$

$$\varphi_2(t) = \begin{cases} \frac{1}{\sqrt{T}} & 0 \leq t < T/2 \\ -\frac{1}{\sqrt{T}} & T/2 \leq t < T \\ 0 & \text{elsewhere} \end{cases} \quad (1.170)$$

“Return to zero” indicates that the transmitted voltage (i.e. the real value of the signal waveform) always returns to the same value at the beginning of any symbol interval.

As for any binary signal constellation,

$$P_b = P_e = Q \left[\frac{d_{\min}}{2\sigma} \right] = Q \left[\sqrt{\frac{\mathcal{E}_x}{2\sigma^2}} \right] \quad . \quad (1.171)$$

RZ is 3 dB inferior to binary antipodal signaling, and uses twice the bandwidth of NRZ.

Frequency Shift Keying (FSK)

Frequency shift keying uses the following two basis functions for the two dimensional signal constellation shown in Figure 1.35.

$$\varphi_1(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin \frac{\pi t}{T} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.172)$$

$$\varphi_2(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin \frac{2\pi t}{T} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.173)$$

The term “frequency-shift” indicates that the sequence of “1’s” and “0’s” in the transmitted data shifts between two different frequencies, $1/(2T)$ and $1/T$.

As for any binary signal constellation,

$$P_b = P_e = Q \left[\frac{d_{\min}}{2\sigma} \right] = Q \left[\sqrt{\frac{\mathcal{E}_x}{2\sigma^2}} \right] . \quad (1.174)$$

FSK is also 3 dB inferior to binary antipodal signaling.

FSK can be extended to higher dimensional signal sets $N > 2$ by adding the following basis functions ($i \geq 3$):

$$\varphi_i(t) = \begin{cases} \sqrt{\frac{2}{T}} \sin \frac{i\pi t}{T} & 0 \leq t \leq T \\ 0 & \text{elsewhere} \end{cases} \quad (1.175)$$

The required bandwidth necessary to realize the additional basis functions grows linearly with N for this FSK extension.

P_e Computation for Block Orthogonal

The computation of P_e for block orthogonal signaling returns to the discussion of the signal detector in Figure 1.23. Because all the signals are equally likely and of equal energy, the constants c_i can be omitted (because they are all the same constant $c_i = c$). In this case, the MAP receiver becomes

$$\hat{m} \Rightarrow m_i \text{ if } \langle \mathbf{y}, \mathbf{x}_i \rangle \geq \langle \mathbf{y}, \mathbf{x}_j \rangle \forall j \neq i, \quad (1.176)$$

By the symmetry of the block orthogonal signal constellation, $P_{e/i} = P_e$ or $P_{c/i} = P_c$ for all i . For convenience, the analysis calculates $P_c = P_{c|i=0}$, in which case the i^{th} elements of \mathbf{y} are

$$y_0 = \sqrt{\mathcal{E}_x} + n_0 \quad (1.177)$$

$$y_i = n_i \quad \forall i \neq 0 . \quad (1.178)$$

If a decision is made that message 0 was sent, then $\langle \mathbf{y}, \mathbf{x}_0 \rangle \geq \langle \mathbf{y}, \mathbf{x}_i \rangle$ or equivalently $y_0 \geq y_i \forall i \neq 0$. The probability of this decision being correct is

$$P_{c/0} = P\{y_0 \geq y_i \forall i \neq 0 \mid \text{given 0 was sent}\} . \quad (1.179)$$

If y_0 takes on a particular value v , then since $y_i = n_i \forall i \neq 0$ and since all the noise components are independent,

$$P_{c/0, y_0=v} = P\{n_i \leq v, \forall i \neq 0\} \quad (1.180)$$

$$= \prod_{i=1}^{N-1} P\{n_i \leq v\} \quad (1.181)$$

$$= [1 - Q(v/\sigma)]^{N-1} . \quad (1.182)$$

The last equation uses the fact that the n_i are independent, identically distributed Gaussian random variables $N(0, \sigma^2)$. Finally, recalling that y_0 is also a Gaussian random variable $N(\sqrt{\mathcal{E}_x}, \sigma^2)$.

$$P_c = P_{c/0} = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(v-\sqrt{\mathcal{E}_x})^2} [1 - Q(v/\sigma)]^{N-1} dv , \quad (1.183)$$

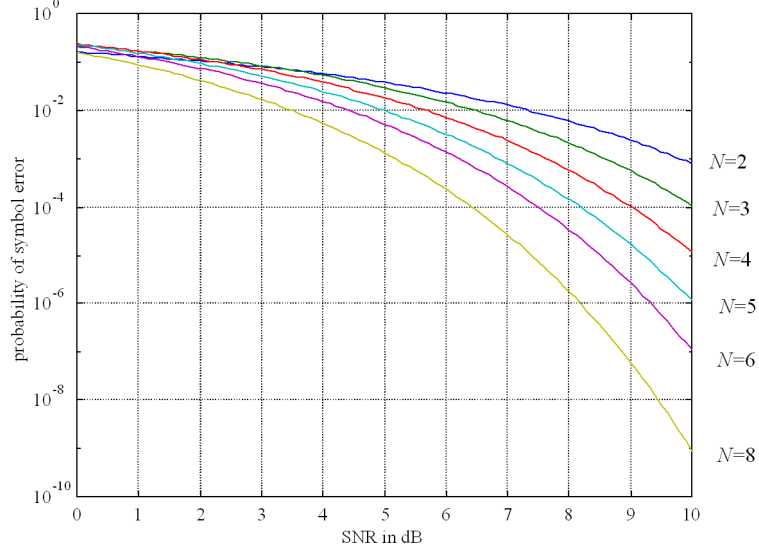


Figure 1.36: Plot of probability of symbol error for orthogonal constellations.

yielding

$$P_e = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(v-\sqrt{\mathcal{E}_x})^2} [1 - Q(v/\sigma)]^{N-1} dv \quad (1.184)$$

This function must be evaluated numerically using a computer.

A simpler calculation yields the NNUB, which also coincides with the union bound because the number of nearest neighbors $M-1$ equals the total number of neighbors to any point for block orthogonal signaling. The NNUB is given by

$$P_e \leq (M-1)Q\left[\frac{d_{\min}}{2\sigma}\right] = (M-1)Q\left[\sqrt{\frac{\mathcal{E}_x}{2\sigma^2}}\right] \quad (1.185)$$

A plot of performance for several values of N appears in Figure 1.36. As N gets large, performance improves without increase of SNR, but at the expense of a lower \bar{b} .

Simplex Constellation

For block orthogonal signaling, the mean value of the signal constellation is nonzero, that is $E[\mathbf{x}] = (\sqrt{\mathcal{E}_x}/M)[1 \ 1 \ \dots \ 1]$. Translation of the constellation by $-E[\mathbf{x}]$ minimizes the energy in the signal constellation without changing the average error probability. The translated signal constellation, known as the simplex constellation, is

$$\mathbf{x}_i^s = \left[-\frac{\sqrt{\mathcal{E}_x}}{M}, \dots, -\frac{\sqrt{\mathcal{E}_x}}{M}, \sqrt{\mathcal{E}_x}\left(1 - \frac{1}{M}\right), -\frac{\sqrt{\mathcal{E}_x}}{M}, \dots, -\frac{\sqrt{\mathcal{E}_x}}{M} \right]' \quad (1.186)$$

where the $\sqrt{\mathcal{E}_x}(1 - \frac{1}{M})$ occurs in the i^{th} position. The superscript s distinguishes the simplex constellation $\{\mathbf{x}_i^s\}$ from the block orthogonal constellation $\{\mathbf{x}_i\}$ from which the simplex constellation is constructed. The energy of the simplex constellation equals

$$\mathcal{E}_{\mathbf{x}}^s = \frac{M-1}{M}\mathcal{E}_{\mathbf{x}} \quad (1.187)$$

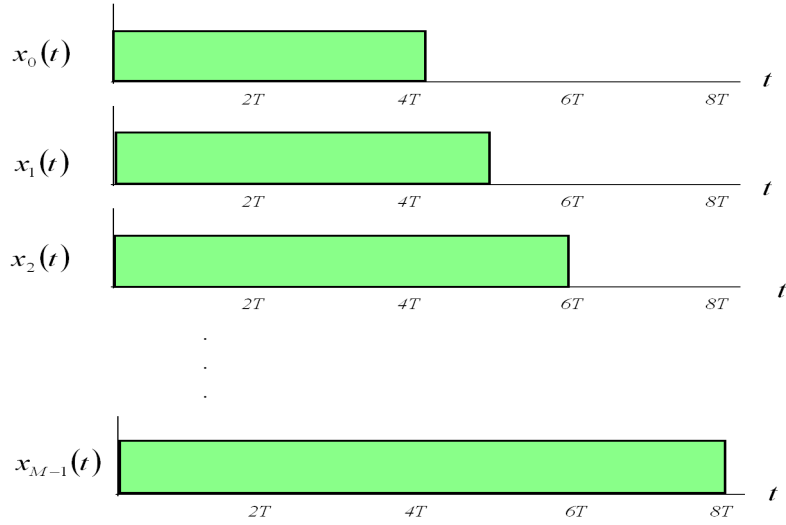


Figure 1.37: Pulse Duration Modulation (PDM).

which provides significant energy savings for small M . The set of data symbols, however, are no longer orthogonal.

$$\langle \mathbf{x}_i^s, \mathbf{x}_j^s \rangle = (\mathbf{x}_i - E[\mathbf{x}])' (\mathbf{x}_j - E[\mathbf{x}]) \quad (1.188)$$

$$= \mathcal{E}_{\mathbf{x}} \delta_{ij} - \langle E[\mathbf{x}], (\mathbf{x}_i + \mathbf{x}_j) \rangle + \frac{\mathcal{E}_{\mathbf{x}}}{M} \quad (1.189)$$

$$= \mathcal{E}_{\mathbf{x}} \delta_{ij} - 2 \frac{\mathcal{E}_{\mathbf{x}}}{M} + \frac{\mathcal{E}_{\mathbf{x}}}{M} \quad (1.190)$$

$$= \mathcal{E}_{\mathbf{x}} \delta_{ij} - \frac{\mathcal{E}_{\mathbf{x}}}{M} . \quad (1.191)$$

By the theorem of translational invariance, P_e of the simplex signal set equals P_e of the block orthogonal signal set given in (1.184) and bounded in (1.185).

Pulse Duration Modulation

Another signal set in which the signals are not orthogonal, as usually described, is pulse duration modulation (PDM). The number of signals points in the PDM constellation increases linearly with the number of dimensions, as for orthogonal constellations. PDM is commonly used, with some modifications, in read-only optical data storage (i.e. compact disks and CD-ROM). In optical data storage, data is recorded by the length of a hole or “pit” burned into the storage medium. The signal set can be constructed as illustrated in Figure 1.37. The minimum width of the pit ($4T$ in the figure) is much larger than the separation (T) between the different PDM signal waveforms. The signal set is evidently not orthogonal.

A second performance-equivalent set of waveforms, known as a **Pulse Position Modulation (PPM)**, appears in Figure 1.38. The PPM constellation is a block orthogonal constellation, which has the previously derived P_e . The average energy of the PDM constellation clearly exceeds that of the PPM constellation, which in turn exceeds that of a corresponding Simplex constellation. Nevertheless, constellation energy minimization is usually not important when PDM is used; for example in optical storage, the optical channel physics mandate the minimum “pit” duration and the resultant “energy” increase is not of concern.

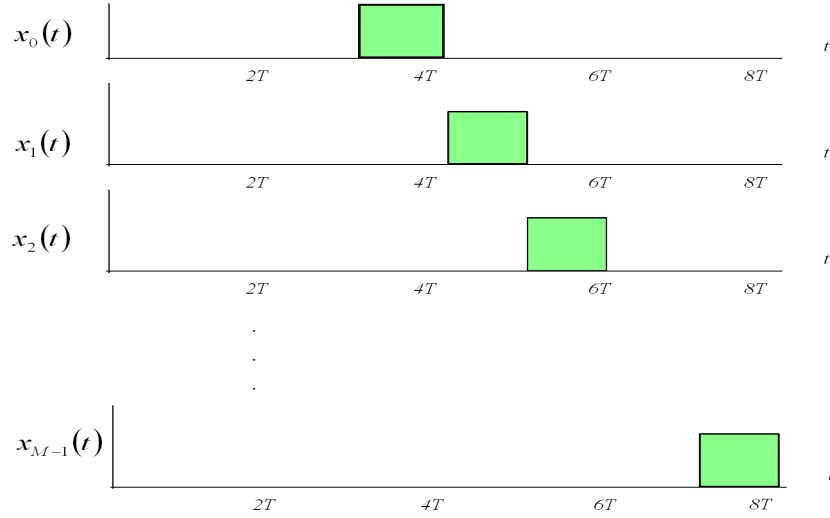


Figure 1.38: Pulse-Position Modulation (PPM).

Biorthogonal Signal Constellations

A variation on block orthogonal signaling is the biorthogonal signal set, which doubles the size of the signal set from $M = N$ to $M = 2N$ by including the negative of each of the data symbol vectors in the signal set. From this perspective, QPSK is both a biorthogonal signal set and a cubic signal set.

The probability of error analysis for biorthogonal constellations parallels that for block orthogonal signal sets. As with orthogonal signaling, because all the signals are equally likely and of equal energy, the constants c_i in the signal detector in Figure 1.23 can be omitted, and the MAP receiver becomes

$$\hat{m} \Rightarrow m_i \text{ if } \langle \mathbf{y}, \mathbf{x}_i \rangle \geq \langle \mathbf{y}, \mathbf{x}_j \rangle \quad \forall j \neq i \quad . \quad (1.192)$$

By symmetry $P_{e/i} = P_e$ or $P_{c/i} = P_c$ for all i . Let $i = 0$. Then

$$y_0 = \sqrt{\mathcal{E}_x} + n_0 \quad (1.193)$$

$$y_i = n_i \quad \forall i \neq 0 \quad . \quad (1.194)$$

If x_0 was sent, then a correct decision is made if $\langle \mathbf{y}, \mathbf{x}_0 \rangle \geq \langle \mathbf{y}, \mathbf{x}_i \rangle$ or equivalently if $y_0 \geq |y_i| \quad \forall i \neq 0$. Thus

$$P_{c/0} = P\{y_0 \geq |y_i|, \quad \forall i \neq 0 | 0 \text{ was sent}\} \quad . \quad (1.195)$$

Suppose y_0 takes on a particular value $v \in [0, \infty)$, then since the noise components n_i are iid

$$P_{c/0, y_0=v} = \prod_{i=1}^{N-1} P\{|n_i| \leq v\} \quad (1.196)$$

$$= [1 - 2Q(v/\sigma)]^{N-1} \quad . \quad (1.197)$$

If $y_0 < 0$, then an incorrect decision is guaranteed if symbol zero was sent. (The reader should visualize the decision regions for this constellation). Thus

$$P_c = P_{c/o} = \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(v - \sqrt{\mathcal{E}_x})^2} [1 - 2Q(v/\sigma)]^{N-1} dv \quad , \quad (1.198)$$

yielding

$$P_e = 1 - \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(v - \sqrt{\mathcal{E}_x})^2} [1 - 2Q(v/\sigma)]^{N-1} dv \quad . \quad (1.199)$$

This function can be evaluated numerically using a computer.

Using the NNUB, which is slightly tighter than the union bound because the number of nearest neighbors is $M - 2$ for biorthogonal signaling,

$$P_e \leq (M - 2)Q \left[\frac{d_{\min}}{2\sigma} \right] = 2(N - 1)Q \left[\sqrt{\frac{\mathcal{E}_x}{2\sigma^2}} \right] . \quad (1.200)$$

1.5.3 Circular Constellations - M -ary Phase Shift Keying

Examples of Phase Shift Keying appeared in Figures 1.21 and 1.30. In general, M -ary PSK places the data symbol vectors at equally spaced angles (or phases) around a circle of radius $\sqrt{\mathcal{E}_x}$ in two dimensions. Only the phase of the signal changes with the transmitted message, while the amplitude of the signal envelope remains constant, thus the origin of the name.

PSK is often used on channels with nonlinear amplitude distortion where signals that include information content in the time varying amplitude would otherwise suffer performance degradation from nonlinear amplitude distortion. The minimum distance for M -ary PSK is given by

$$d_{\min} = 2\sqrt{\mathcal{E}_x} \sin \frac{\pi}{M} = 2\sqrt{2\mathcal{E}_x} \sin \frac{\pi}{M} . \quad (1.201)$$

The CFM is

$$\zeta_x = 2 \sin^2 \left(\frac{\pi}{M} \right) , \quad (1.202)$$

which is inferior to block binary signaling for any constellation with $M > 4$. The NNUB on error probability is tight and equal to

$$P_e < 2Q \left[\frac{\sqrt{\mathcal{E}_x} \sin \frac{\pi}{M}}{\sigma} \right] , \quad (1.203)$$

for all M .

1.6 Rectangular (and Hexagonal) Signal Constellations

This section studies several popular signal constellations for data transmission. These constellations use equally spaced points on a one- or two-dimensional lattice. This study introduces and uses some basic concepts, namely SNR, the shaping gain, the continuous approximation, and the peak-to-average power ratio. These concepts and/or their results measure the quality of signal constellations and are fundamental to understanding performance.

Subsection 1.6.1 studies pulse amplitude modulation (PAM), while Subsection 1.6.2 studies quadrature amplitude modulation (QAM). Subsection 1.6.3 discusses several measures of constellation performance.

The Continuous Approximation

For constellations that possess a degree of geometric uniformity, the **continuous approximation** computes the average energy of a signal constellation by replacing the discrete sum of energies with a continuous integral. The discrete probability distribution of the signal constellation is approximated by a continuous distribution that is uniform over a region defined by the signal points. In defining the region of the constellation, each point in the constellation is associated with an identical fundamental volume, \mathcal{V}_x . Each point in the constellation is centered within an N -dimensional decision region (or **Voronoi Region**) with volume equal to the fundamental volume. This region is the decision region for internal points in the constellation that have a maximum number of nearest neighbors. The union of these Voronoi regions for the points is called the Voronoi Region \mathcal{D}_x of the constellation and has volume $M\mathcal{V}_x$.

The discrete distribution of the constellation is approximated by a uniform distribution over the Voronoi Region with probability density $p_x(\mathbf{u}) = \frac{1}{M\mathcal{V}_x} \forall \mathbf{u}$ in \mathcal{D}_x .

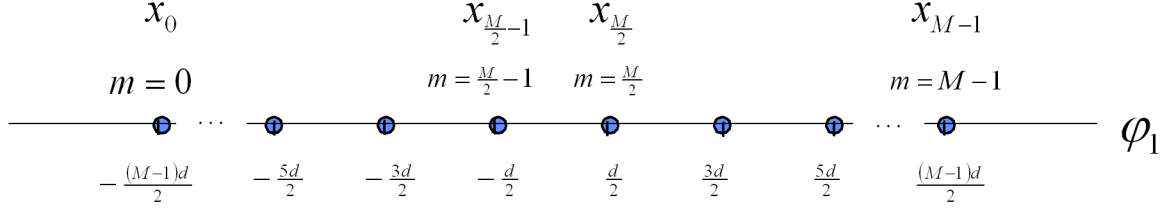


Figure 1.39: PAM constellation.

Definition 1.6.1 (Continuous Approximation) *The continuous approximation to the average energy equals*

$$\mathcal{E}_{\mathbf{x}} \approx \tilde{\mathcal{E}}_{\mathbf{x}} = \int_{\mathcal{D}_{\mathbf{x}}} \|\mathbf{u}\|^2 \frac{1}{M\mathcal{V}_{\mathbf{x}}} d\mathbf{u} \quad , \quad (1.204)$$

where the N -dimensional integral covers the Voronoi region $\mathcal{D}_{\mathbf{x}}$.

For large size signal sets with regular spacing between points, the error in using the continuous approximation is small, as several examples will demonstrate in this section.

1.6.1 Pulse Amplitude Modulation (PAM)

Pulse amplitude modulation, or amplitude shift keying (ASK), is a one-dimensional modulated signal set with $M = 2^b$ signals in the constellation, for $b \in \mathcal{Z}^+$, the set of positive integers. Figure 1.39 illustrates the PAM constellation. The basis function can be any unit-energy function, but often $\varphi_1(t)$ is

$$\varphi_1(t) = \frac{1}{\sqrt{T}} \text{sinc} \left(\frac{t}{T} \right) \quad (1.205)$$

or another ‘‘Nyquist’’ pulse shape (see Chapter 3). The data-symbol amplitudes are $\pm \frac{d}{2}, \pm \frac{3d}{2}, \pm \frac{5d}{2}, \dots, \pm \frac{(M-1)d}{2}$, and all input levels are equally likely. The minimum distance between points in a PAM constellation abbreviates as

$$d_{\min} = d \quad . \quad (1.206)$$

Both binary antipodal and ‘‘2B1Q’’ are examples of PAM signals.

The average energy of a PAM constellation is

$$\mathcal{E}_{\mathbf{x}} = \bar{\mathcal{E}}_{\mathbf{x}} = \frac{1}{M} (2) \sum_{k=1}^{M/2} \left(\frac{2k-1}{2} \right)^2 d^2 \quad (1.207)$$

$$= \frac{d^2}{2M} \sum_{k=1}^{M/2} (4k^2 - 4k + 1) \quad (1.208)$$

$$= \frac{d^2}{2M} \left[4 \left(\frac{(M/2)^3}{3} + \frac{(M/2)^2}{2} + \frac{(M/2)}{6} \right) - 4 \left(\frac{(M/2)^2}{2} + \frac{(M/2)}{2} \right) + \frac{M}{2} \right] \quad (1.209)$$

$$= \frac{d^2}{2M} \left[\frac{M^3}{6} - \frac{M}{6} \right] \quad . \quad (1.210)$$

This PAM average energy is expressed in terms of the minimum distance and constellation size as

$$\mathcal{E}_{\mathbf{x}} = \bar{\mathcal{E}}_{\mathbf{x}} = \frac{d^2}{12} [M^2 - 1] \quad . \quad (1.211)$$

The PAM minimum distance is a function of \mathcal{E}_x and M :

$$d = \sqrt{\frac{12\mathcal{E}_x}{M^2 - 1}}. \quad (1.212)$$

Finally, given distance and average energy,

$$\bar{b} = \log_2 M = \frac{1}{2} \log \left(12 \frac{\bar{\mathcal{E}}_x}{d^2} + 1 \right). \quad (1.213)$$

Figure 1.39 shows that the decision region for an interior point of PAM extends over a length d interval centered on that point. The Voronoi region of the constellation thus extends for an interval of Md over $[-L, L]$ where $L = \frac{Md}{2}$. The continuous approximation for PAM assumes a uniform distribution on this interval $(-L, L)$, and thus approximates the average energy of the constellation as

$$\mathcal{E}_x = \bar{\mathcal{E}}_x \approx \int_{-L}^L \frac{x^2}{2L} dx = \frac{L^2}{3} = \frac{M^2 d^2}{12}. \quad (1.214)$$

The approximation for the average energy does not include the constant term $-\frac{d^2}{12}$, which becomes insignificant as M becomes large.

Since $M = 2^{\bar{b}}$, then $M^2 = 4^{\bar{b}}$, leaving alternative relations ($\bar{b} = b$ for $N = 1$) for (1.211) and (1.212)

$$\mathcal{E}_x = \bar{\mathcal{E}}_x = \frac{d^2}{12} [4^b - 1] = \frac{d^2}{12} [4^{\bar{b}} - 1], \quad (1.215)$$

and

$$d = \sqrt{\frac{12\mathcal{E}_x}{4^{\bar{b}} - 1}}. \quad (1.216)$$

The following recursion derives from increasing the number of bits, $b = \bar{b}$, in a PAM constellation while maintaining constant minimum distance between signal points:

$$\bar{\mathcal{E}}_x(b+1) = 4\bar{\mathcal{E}}_x(b) + \frac{d^2}{4}. \quad (1.217)$$

Thus for moderately large b , the required signal energy increases by a factor of 4 for each additional bit of information in the signal constellation. This corresponds to an increase of 6dB per bit, a measure commonly quoted by communication engineers as the required SNR increase for a transmission scheme to support an additional bit-per-dimension of information.

The PAM probability of correct symbol detection is

$$P_c = \sum_{i=0}^{M-1} P_{c|i} p_{\mathbf{x}}(i) \quad (1.218)$$

$$= \frac{M-2}{M} \left(1 - 2Q \left[\frac{d_{\min}}{2\sigma} \right] \right) + \frac{2}{M} \left(1 - Q \left[\frac{d_{\min}}{2\sigma} \right] \right) \quad (1.219)$$

$$= 1 - \left(\frac{2M-4+2}{M} \right) Q \left[\frac{d_{\min}}{2\sigma} \right] \quad (1.220)$$

$$= 1 - 2 \left(1 - \frac{1}{M} \right) Q \left[\frac{d_{\min}}{2\sigma} \right] \quad (1.221)$$

Thus, the PAM probability of symbol error is

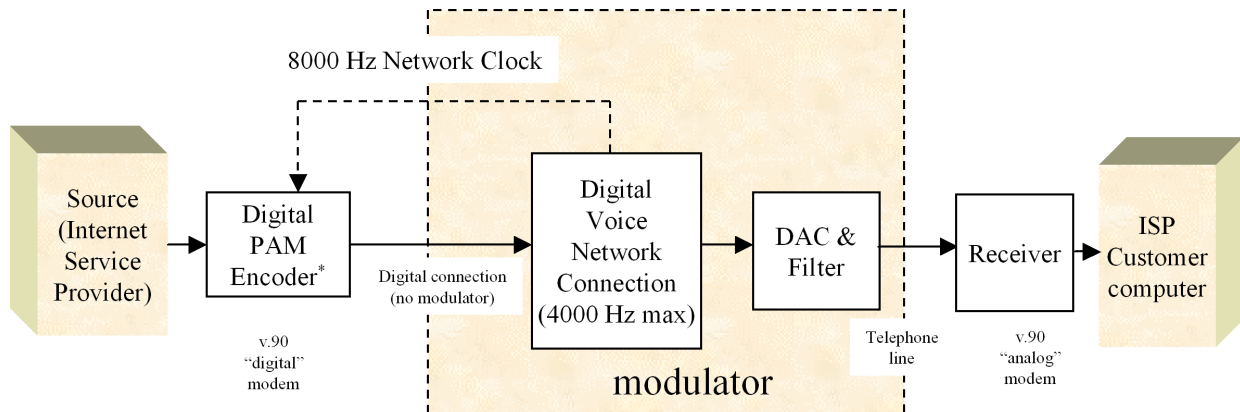
$$P_e = \bar{P}_e = 2 \left(1 - \frac{1}{2^{\bar{b}}} \right) Q \left[\frac{d_{\min}}{2\sigma} \right] < 2Q \left[\frac{d_{\min}}{2\sigma} \right]. \quad (1.222)$$

The average number of nearest neighbors for the constellation is $2(1 - 1/M)$; thus, the NNUB is exact for PAM. Thus

$$P_e = 2 \left(1 - \frac{1}{M} \right) Q \left(\sqrt{\frac{3}{M^2 - 1} SNR} \right) \quad (1.223)$$

$b = \bar{b}$	M	$\frac{d}{2\sigma}$ for $\bar{P}_e = 10^{-6} \approx 2Q\left(\frac{d_{\min}}{2\sigma}\right)$	SNR = $\frac{(M^2-1) \cdot 10^{1.37}}{3}$	SNR increase = $\frac{M^2-1}{(M-1)^2-1}$
1	2	13.7dB	13.7dB	—
2	4	13.7dB	20.7dB	7dB
3	8	13.7dB	27.0dB	6.3dB
4	16	13.7dB	33.0dB	6.0dB
5	32	13.7dB	39.0dB	6.0dB

Table 1.1: PAM constellation energies.



*the PAM levels are not equally spaced in v.90 – instead they are chosen to match the levels of the so-called “voice codecs” (μ -law in US, A-law internationally) that are matched to best capture of analog voice signals and unfortunately all that is available for the v.90 modem path even if worse than equally spaced PAM for data traffic. 7 effective bits/symbol are transmitted, but an 8th redundant bit is also sent for use with coding as in Chapters 8-10.

Figure 1.40: 56K Modem PAM example.

For $P_e = 10^{-6}$, one determines that $\frac{d}{2\sigma} \approx 4.75$ (13.5dB). Table 1.1 relates $b = \bar{b}$, M , $\frac{d}{2\sigma}$, the SNR, and the required increase in SNR (or equivalently in $\bar{\mathcal{E}}_{\mathbf{x}}$) to transmit an additional bit of information at a probability of error $P_e = 10^{-6}$. Table 1.1 shows that for $b = \bar{b} > 2$, the approximation of 6dB per bit is very accurate.

Pulse amplitude constellations with $b > 2$ are typically known as 3B1O - three bits per octal signal (for 8 PAM) and 4B1H (4 bits per hexadecimal signal), but are rare in use with respect to the more popular quadrature amplitude modulation of Section 1.6.2.

EXAMPLE 1.6.1 (56K voiceband modem) The 56 kbps voiceband modem of Figure 1.40 can be viewed as a PAM modem with 128 levels, or $b = 7$, and a symbol rate of 8000 Hz. Thus, the data rate is $R = b/T = 56,000$ bits per second. The 8000Hz is thus consistent with the maximum allowed bandwidth of a voice-only telephone-network connection (4000 Hz) that is imposed by network switches that sample also at 8000Hz and the clock is supplied from the network to the internet service provider’s modem as shown. The modulator is curiously and tacitly implemented by the combination of the network connection and the eventual DAC (digital-to-analog converter) at the beginning of the customer’s analog phone line. A special receiver structure known as a DFE (see Chapter 3) converts the telephone line into an AWGN with an SNR sufficiently high to carry 128PAM. This particular choice of modulator implementation avoids the distortion that would be introduced by an extra ADC

and DAC on the Internet Service Provider's phone line (and which are superfluous as that Internet Service Provider most often has a digital high-speed connection by fiber or other means to the network).

The levels are not equally spaced in most implementations, an artifact only of the fact that the network DAC is not a uniform-DAC, and instead chooses its levels for best voice transmission, which is not the best for data transmission. Yet higher-speed DSL versions of the same system, allow a higher-speed connection through the network, and use a new "DAC" in what is called a DSLAM that replaces the voice connection to the subscriber.

1.6.2 Quadrature Amplitude Modulation (QAM)

QAM is a two dimensional generalization of PAM. The two basis functions are usually

$$\varphi_1(t) = \sqrt{\frac{2}{T}} \operatorname{sinc}\left(\frac{t}{T}\right) \cos \omega_c t \quad , \quad (1.224)$$

$$\varphi_2(t) = -\sqrt{\frac{2}{T}} \operatorname{sinc}\left(\frac{t}{T}\right) \sin \omega_c t \quad . \quad (1.225)$$

The $\operatorname{sinc}(t/T)$ term may be replaced by any Nyquist pulse shape as discussed in Chapter 3. The ω_c is a radian carrier frequency that is discussed for in Chapters 2 and 3; for now, $\omega_c \geq \pi/T$.

The QAM Square Constellation

Figure 1.41 illustrates QAM Square Constellations. These constellations are the Cartesian products¹² of 2-PAM with itself and 4-PAM with itself, respectively. Generally, square M -QAM constellations derive from the Cartesian product of two \sqrt{M} -PAM constellations. For \bar{b} bits per dimension, the $M = 4^{\bar{b}}$ signal points are placed at the coordinates $\pm \frac{d}{2}, \pm \frac{3d}{2}, \pm \frac{5d}{2}, \dots, \pm \frac{(\sqrt{M}-1)d}{2}$ in each dimension.

The average energy of square QAM constellations is easily computed as

$$\mathcal{E}_{M\text{-QAM}} = \mathcal{E}_{\mathbf{x}} = 2\bar{\mathcal{E}}_{\mathbf{x}} = \frac{1}{M} \sum_{i,j=1}^{\sqrt{M}} (x_i^2 + x_j^2) \quad (1.226)$$

$$= \frac{1}{M} \left[\sqrt{M} \sum_{i=1}^{\sqrt{M}} x_i^2 + \sqrt{M} \sum_{j=1}^{\sqrt{M}} x_j^2 \right] \quad (1.227)$$

$$= 2 \frac{1}{\sqrt{M}} \sum_{i=1}^{\sqrt{M}} x_i^2 \quad (1.228)$$

$$= 2\mathcal{E}_{\sqrt{M}\text{-PAM}} \quad (1.229)$$

$$= d^2 \left(\frac{M-1}{6} \right) \quad (1.230)$$

Thus, the average energy per dimension of the M -QAM constellation

$$\bar{\mathcal{E}}_{\mathbf{x}} = d^2 \left(\frac{M-1}{12} \right) \quad , \quad (1.231)$$

equals the average energy of the constituent \sqrt{M} -PAM constellation. The minimum distance $d_{\min} = d$ can be computed from $\mathcal{E}_{\mathbf{x}}$ (or $\bar{\mathcal{E}}_{\mathbf{x}}$) and M by

$$d = \sqrt{\frac{6\mathcal{E}_{\mathbf{x}}}{M-1}} = \sqrt{\frac{12\bar{\mathcal{E}}_{\mathbf{x}}}{M-1}} \quad . \quad (1.232)$$

¹²A Cartesian Product, a product of two sets, is the set of all ordered pairs of coordinates, the first coordinate taken from the first set in the Cartesian product, and the second coordinate taken from the second set in the Cartesian product.

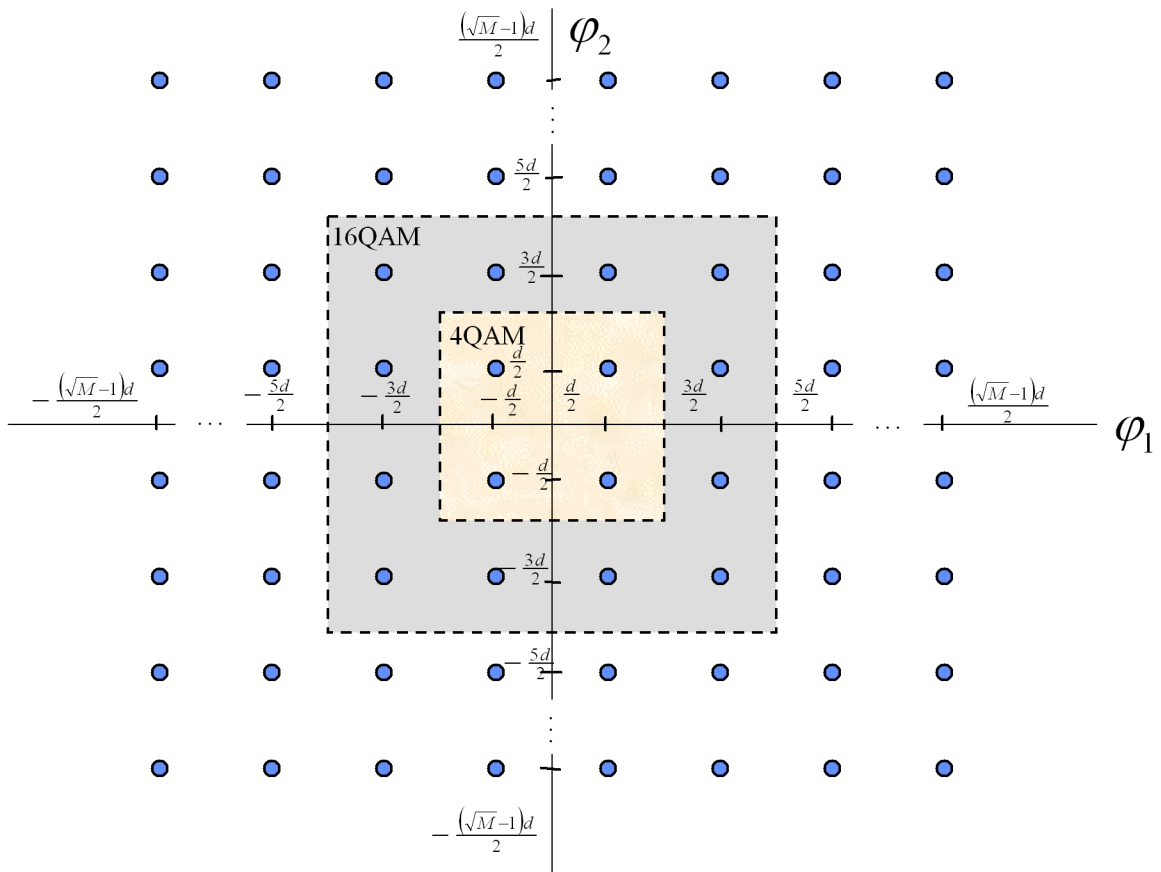


Figure 1.41: QAM constellations.

Since $M = 4^{\bar{b}}$, alternative relations for (1.231) and (1.232) in terms of the average bit rate \bar{b} are

$$\bar{\mathcal{E}}_{\mathbf{x}} = \frac{\mathcal{E}_{\mathbf{x}}}{2} = \frac{d^2}{12} [4^{\bar{b}} - 1] \quad , \quad (1.233)$$

and

$$d = \sqrt{\frac{12\bar{\mathcal{E}}_{\mathbf{x}}}{4^{\bar{b}} - 1}} \quad . \quad (1.234)$$

Finally,

$$\bar{b} = \frac{1}{2} \log_2 \left(\frac{6\mathcal{E}_{\mathbf{x}}}{d^2} + 1 \right) = \frac{1}{2} \log_2 \left(\frac{12\bar{\mathcal{E}}_{\mathbf{x}}}{d^2} + 1 \right) \quad , \quad (1.235)$$

the same as for a PAM constellation.

For large M , $\bar{\mathcal{E}}_{\mathbf{x}} \approx \frac{d^2}{12} M = \frac{d^2}{12} 4^{\bar{b}}$, which is the same as that obtained by using the continuous approximation. The continuous approximation for two dimensional QAM uses a uniform constellation over the square defined by $[\pm L, \pm L]$,

$$\mathcal{E}_{\mathbf{x}} \approx \int_{-L}^L \int_{-L}^L \frac{x^2 + y^2}{4L^2} dx dy = 2 \frac{L^2}{3} \quad , \quad (1.236)$$

or $L = \sqrt{1.5\mathcal{E}_{\mathbf{x}}}$. Since the Voronoi region for each signal point in a QAM constellation has area d^2

$$M \approx \frac{4 \cdot L^2}{d^2} = \frac{6\mathcal{E}_{\mathbf{x}}}{d^2} = \frac{12\bar{\mathcal{E}}_{\mathbf{x}}}{d^2} \quad . \quad (1.237)$$

This result agrees with Equation 1.231 for large M .

As the number of points increases, the energy-computation error caused by using the continuous approximation becomes negligible.

Increasing the number of bits, b , in a QAM constellation while maintaining constant minimum distance, we find the following recursion for the increase in average energy:

$$\mathcal{E}_{\mathbf{x}}(b+1) = 2\mathcal{E}_{\mathbf{x}}(b) + \frac{d^2}{6} \quad . \quad (1.238)$$

Asymptotically the average energy increases by 3dB for each added bit per two dimensional symbol.

The probability of error can be exactly computed for QAM by noting that the conditional probability of a correct decision falls into one of 3 categories:

1. corner points (4 points with only 2 nearest neighbors)

$$P_{c|\text{corner}} = \left(1 - Q \left[\frac{d}{2\sigma} \right] \right)^2 \quad (1.239)$$

2. inner points $(\sqrt{M} - 2)^2$ points with 4 nearest neighbors)

$$P_{c|\text{inner}} = \left(1 - 2Q \left[\frac{d}{2\sigma} \right] \right)^2 \quad (1.240)$$

3. edge points $4(\sqrt{M} - 2)$ points with 3 nearest neighbors)

$$P_{c|\text{edge}} = \left(1 - Q \left[\frac{d}{2\sigma} \right] \right) \left(1 - 2Q \left[\frac{d}{2\sigma} \right] \right) \quad . \quad (1.241)$$

$b = 2\bar{b}$	M	$\frac{d}{2\sigma}$ for $\bar{P}_e = 10^{-6} \approx 2Q\left(\frac{d}{2\sigma}\right)$	SNR = $\frac{(M-1) \cdot 10^{1.37}}{3}$	SNR increase = $\frac{M-1}{(M-1)-1}$	dB/bit
2	4	13.7dB	13.7dB	—	—
4	16	13.7dB	20.7dB	7.0dB	3.5dB
6	64	13.7dB	27.0dB	6.3dB	3.15dB
8	256	13.7dB	33.0dB	6.0dB	3.0dB
10	1024	13.7dB	39.0dB	6.0dB	3.0dB
12	4096	13.7dB	45.0dB	6.0dB	3.0dB
14	16,384	13.7dB	51.0dB	6.0dB	3.0dB

Table 1.2: QAM constellation energies.

The probability of being correct is then (abbreviating $Q \leftarrow Q\left[\frac{d}{2\sigma}\right]$)

$$P_c = \sum_{i=0}^{M-1} P_{c/i} p_{\mathbf{x}}(i) \quad (1.242)$$

$$= \frac{4}{M} (1-Q)^2 + \frac{(\sqrt{M}-2)^2}{M} (1-2Q)^2 + \frac{4(\sqrt{M}-2)}{M} (1-2Q)(1-Q) \quad (1.243)$$

$$= \frac{1}{M} \left[(4-8Q+4Q^2) + (4\sqrt{M}-8)(1-3Q+2Q^2) \right. \quad (1.244)$$

$$\left. + (M-4\sqrt{M}+4)(1-4Q+4Q^2) \right] \quad (1.245)$$

$$= \frac{1}{M} \left[M + (4\sqrt{M}-4M)Q + (4-8\sqrt{M}+4M)Q^2 \right] \quad (1.246)$$

$$= 1 + 4\left(\frac{1}{\sqrt{M}} - 1\right)Q + 4\left(\frac{1}{\sqrt{M}} - 1\right)^2 Q^2 \quad (1.247)$$

Thus, the probability of symbol error is

$$P_e = 4\left(1 - \frac{1}{\sqrt{M}}\right) Q\left[\frac{d}{2\sigma}\right] - 4\left(1 - \frac{1}{\sqrt{M}}\right)^2 \left(Q\left[\frac{d}{2\sigma}\right]\right)^2 < 4\left(1 - \frac{1}{\sqrt{M}}\right) Q\left[\frac{d}{2\sigma}\right] . \quad (1.248)$$

The average number of nearest neighbors for the constellation equals $4(1 - 1/\sqrt{M})$, thus for QAM the NNUB is not exact, but usually tight. The corresponding normalized NNUB is

$$\bar{P}_e \leq 2\left(1 - \frac{1}{2^{\bar{b}}}\right) Q\left[\frac{d}{2\sigma}\right] = 2\left(1 - \frac{1}{2^{\bar{b}}}\right) Q\left[\sqrt{\frac{3}{M-1}} \text{SNR}\right] , \quad (1.249)$$

which equals the PAM result. For $P_e = 10^{-6}$, one determines that $\frac{d}{2\sigma} \approx 4.75$ (13.5dB). Table 1.2 relates \bar{b} , M , $\frac{d}{2\sigma}$, the SNR, and the required increase in SNR (or equivalently in $\bar{\mathcal{E}}_{\mathbf{x}}$) to transmit an additional bit of information.

As with PAM for average bit rates of $\bar{b} > 2$, the approximation of 3dB per bit per two-dimensional additional for the average energy increase is accurate.

The constellation figure of merit for square QAM is

$$\zeta_{\mathbf{x}} = \frac{3}{M-1} = \frac{3}{4^{\bar{b}}-1} = \frac{3}{2^b-1} . \quad (1.250)$$

When b is odd, it is possible to define a SQ QAM constellation by taking every other point from a $b+1$ SQ QAM constellation. (See Problem 1.14.)

A couple of examples illustrate the wide use of QAM transmission.

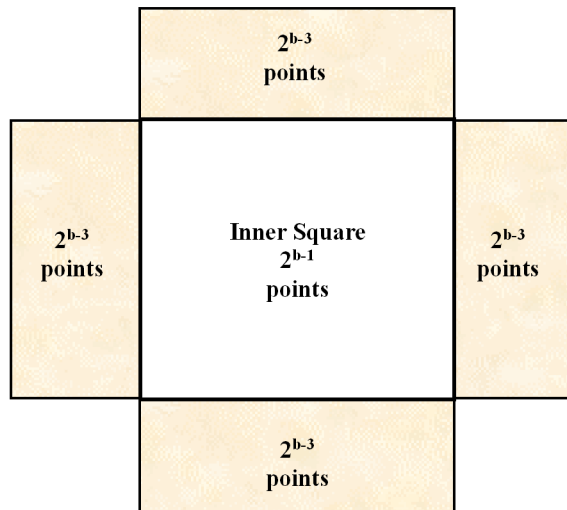


Figure 1.42: Cross constellations.

EXAMPLE 1.6.2 (Cable Modem) Cable modems use existing cable-TV coaxial cables for two-way transmission (presuming the cable TV provider has sent personnel to the various unidirectional blocking points in the network and replaced them with so-called "diplex" filters). Early cable modem conventions (i.e., DOCSIS) use 4QAM in both directions of transmission. The downstream direction from cable TV end to customer is typically at a carrier frequency well above the used TV band, somewhere between 300 MHz and 500 MHz. The upstream direction is below 50 MHz, typically between 5 and 40 MHz. The symbol rate is typically $1/T=2\text{MHz}$ so the data rate is 4 Mbps on any given carrier. Typically about 10 carriers can be used (so 40 Mbps maximum) and each is shared by various subgroups of customers, each customer within a subgroup typically getting 384 kbps in fair systems (although some customers can "hog" all the 4 Mbps and systems for resolving such customer-use are evolving).

EXAMPLE 1.6.3 (Satellite TV Broadcast) Satellite television uses 4QAM in for broadcast transmission at one of 20 carrier frequencies between 12.2 GHz to 12.7 GHz from satellite to customer receiver for some suppliers and satellites. Corresponding carriers between 17.3 and 17.8 GHz are used to send the signals from the broadcaster to the satellite, again with 4 QAM. The symbol rate is $1/T = 19.151\text{ MHz}$, so the aggregate data rate is 38.302 Mbps on any of the 20 carriers. A typical digital TV signal is compressed into about 2-3 Mbps allowing for 4-16 channels per carrier/QAM signal. (Some stations watched by many, for instance sports, may get a larger allocation of bandwidth and carry a higher-quality image than others that are not heavily watched. A high-definition TV channel (there are only 4 presently) requires 20 Mbps if sent with full fidelity. Each carrier is transmitted in a 24 MHz transponder channel on the satellite – these 24 MHz channels were originally used to broadcast a single analog TV channel, modulated via FM unlike terrestrial analog broadcast television (which uses only 6 MHz for analog TV).

QAM Cross Constellations

The QAM cross constellation also allows for odd numbers of bits per symbol in QAM data transmission. To construct a QAM cross constellation with b bits per symbol one augments a square QAM constellation for $b - 1$ bits per symbol by adding 2^{b-1} data symbols that extend the sides of the QAM square. The corners are excluded as shown in Figure 1.42.

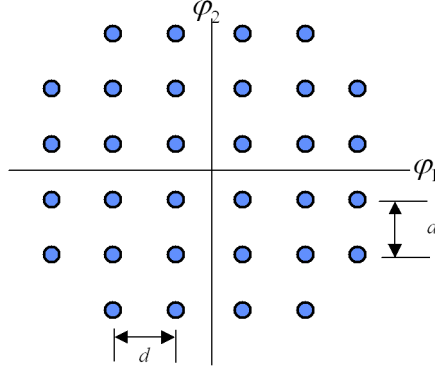


Figure 1.43: 32CR constellation.

One computation of average energy of QAM cross constellations doubles the energy of the two large rectangles ($[2^{\frac{b-3}{2}} + 2^{\frac{b-1}{2}}] \times 2^{\frac{b-1}{2}}$) and then subtracts the energy of the inner square ($2^{\frac{b-1}{2}} \times 2^{\frac{b-1}{2}}$). The energy of the inner square is

$$\mathcal{E}_{\mathbf{x}}(inner) = \frac{d^2}{6}(2^{b-1} - 1) \quad . \quad (1.251)$$

The total sum of energies for all the data symbols in the inner-square-plus-two-side-rectangles is (looking only at one quadrant, and multiplying by 4 because of symmetry)

$$\mathcal{E} = \frac{d^2}{4} (4) \sum_{k=1}^{2^{\frac{b-3}{2}}} \sum_{l=1}^{3 \cdot 2^{\frac{b-5}{2}}} [(2k-1)^2 + (2l-1)^2] \quad (1.252)$$

$$= \frac{d^2}{4} (4) \left[3 \cdot 2^{\frac{b-5}{2}} \left(\frac{2^{\frac{3b-3}{2}} - 2^{\frac{b-1}{2}}}{6} \right) + 2^{\frac{b-3}{2}} \left(\frac{27 \cdot 2^{\frac{3b-9}{2}} - 3 \cdot 2^{\frac{b-3}{2}}}{6} \right) \right] \quad (1.253)$$

$$= \frac{d^2}{4} (4) \left[2^{\frac{b-7}{2}} \left(2^{\frac{3b-3}{2}} - 2^{\frac{b-1}{2}} \right) + 2^{\frac{b-5}{2}} \left(9 \cdot 2^{\frac{3b-9}{2}} - 2^{\frac{b-3}{2}} \right) \right] \quad (1.254)$$

$$= \frac{d^2}{4} [2^{2b-3} - 2^{b-2} + 9 \cdot 2^{2b-5} - 2^{b-2}] \quad (1.255)$$

$$= \frac{d^2}{4} \left[\frac{13}{32} 2^{2b} - 2^{b-1} \right] \quad . \quad (1.256)$$

Then

$$\mathcal{E}_{\mathbf{x}} = \frac{2\mathcal{E} - 2^{b-1}\mathcal{E}_{\mathbf{x}}(inner)}{2^b} = \frac{d^2}{4} \left[\frac{26}{32} 2^b - 1 - \frac{2}{3} 2^{b-2} + \frac{2}{3} \right] \quad (1.257)$$

$$= \frac{d^2}{4} \left[\left(\frac{13}{16} - \frac{1}{6} \right) 2^b - \frac{2}{3} \right] \quad (1.258)$$

$$= \frac{d^2}{4} \left[\frac{31}{48} 2^b - \frac{2}{3} \right] = \frac{d^2}{6} \left[\frac{31}{32} M - 1 \right] \quad (1.259)$$

The minimum distance $d_{\min} = d$ can be computed from $\mathcal{E}_{\mathbf{x}}$ (or $\bar{\mathcal{E}}_{\mathbf{x}}$) and M by

$$d = \sqrt{\frac{6\mathcal{E}_{\mathbf{x}}}{\frac{31}{32}M - 1}} = \sqrt{\frac{12\bar{\mathcal{E}}_{\mathbf{x}}}{\frac{31}{32}M - 1}} = \sqrt{\frac{12\bar{\mathcal{E}}_{\mathbf{x}}}{\frac{31}{32}4^{\frac{b}{2}} - 1}} \quad . \quad (1.260)$$

In (1.259), for large M , $\mathcal{E}_{\mathbf{x}} \approx \frac{31d^2}{192} M = \frac{31d^2}{192} 4^{\frac{b}{2}}$, the same as the continuous approximation.

The following recursion derives from increasing the number of bits, b , in a QAM cross constellation while maintaining constant minimum distance:

$$\mathcal{E}_{\mathbf{x}}(b+1) = 2\mathcal{E}_{\mathbf{x}}(b) + \frac{d^2}{6} . \quad (1.261)$$

As with the square QAM constellation asymptotically the average energy increases by 3 dB for each added bit per two dimensional symbol.

The probability of error can be bounded for QAM Cross by noting that a lower bound on the conditional probability of a correct decision falls into one of two categories:

1. inner points $\left\{2^b - 4 \left(3 \cdot 2^{\frac{b-3}{2}} - 2 \cdot 2^{\frac{b-5}{2}}\right)\right\} = \left\{2^b - 4 \left(2^{\frac{b-1}{2}}\right)\right\}$ with four nearest neighbors

$$P_{c/inner} = \left(1 - 2Q \left[\frac{d_{\min}}{2\sigma}\right]\right)^2 \quad (1.262)$$

2. side points $4 \left(3 \cdot 2^{\frac{b-3}{2}} - 2 \cdot 2^{\frac{b-5}{2}}\right) = 4 \left(2^{\frac{b-1}{2}}\right)$ with three nearest neighbors. (This calculation is only a bound because some of the side points have fewer than three neighbors at distance d_{\min})

$$P_{c/outer} = \left(1 - Q \left[\frac{d_{\min}}{2\sigma}\right]\right) \left(1 - 2Q \left[\frac{d_{\min}}{2\sigma}\right]\right) . \quad (1.263)$$

The probability of a correct decision is then, abbreviating $Q = Q \left[\frac{d_{\min}}{2\sigma}\right]$,

$$P_c \geq \frac{1}{M} \left[4 \left(2^{\frac{b-1}{2}}\right) (1 - Q)(1 - 2Q)\right] \quad (1.264)$$

$$+ \frac{1}{M} \left[\left\{2^b - 4 \left(2^{\frac{b-1}{2}}\right)\right\} (1 - 2Q)^2\right] \quad (1.265)$$

$$= \frac{1}{M} \left[4 \cdot 2^{\frac{b-1}{2}} (1 - 3Q + 2Q^2) + \left[2^b - 4 \cdot 2^{\frac{b-1}{2}}\right] (1 - 4Q + 4Q^2)\right] \quad (1.266)$$

$$= 1 - \left[-2^{\frac{3-b}{2}} + 4\right] Q + \left[2^{\frac{5-b}{2}} - 2 \cdot 2^{\frac{5-b}{2}} + 4\right] Q^2 \quad (1.267)$$

Thus, the probability of symbol error is bounded by

$$P_e \leq 4 \left(1 - \frac{1}{\sqrt{2M}}\right) Q \left[\frac{d_{\min}}{2\sigma}\right] - 4 \left(1 - \sqrt{\frac{2}{M}}\right) \left(Q \left[\frac{d_{\min}}{2\sigma}\right]\right)^2 \quad (1.268)$$

$$< 4 \left(1 - \frac{1}{\sqrt{2M}}\right) Q \left[\frac{d_{\min}}{2\sigma}\right] < 4Q \left[\frac{d_{\min}}{2\sigma}\right] . \quad (1.269)$$

The average number of nearest neighbors for the constellation is $4(1 - 1/\sqrt{2M})$; thus the NNUB is again accurate. The normalized probability of error is

$$\bar{P}_e \leq 2 \left(1 - \frac{1}{2^{b+0.5}}\right) Q \left[\frac{d_{\min}}{2\sigma}\right] , \quad (1.270)$$

which agrees with the PAM result when one includes an additional bit in the constellation, or equivalently an extra .5 bit per dimension. To evaluate (1.270), Equation 1.260 relates that

$$\left(\frac{d_{\min}}{2\sigma}\right)^2 = \frac{3 \text{ SNR}}{\frac{31}{32}M - 1} \quad (1.271)$$

Table 1.3 lists the incremental energies and required SNR for QAM cross constellations in a manner similar to Table 1.2. There are also square constellations for odd numbers of bits that Problem 1.14 addresses.

$b = 2\bar{b}$	M	$\frac{d}{2\sigma}$ for $\bar{P}_e = 10^{-6} \approx 2Q\left(\frac{d_{\min}}{2\sigma}\right)$	$\text{SNR} = \frac{([31/32] \cdot M - 1) \cdot 10^{1.37}}{3}$	$\text{SNR increase} = \frac{[31/32] \cdot M - 1}{[32/32] \cdot (M - 1) - 1}$	dB/bit
5	32	13.7dB	23.7 dB	—	—
7	128	13.7dB	29.8dB	6.1dB	3.05dB
9	512	13.7dB	35.8dB	6.0dB	3.0dB
11	2048	13.7dB	41.8dB	6.0dB	3.0dB
13	8192	13.7dB	47.8dB	6.0dB	3.0dB
15	32,768	13.7dB	53.8dB	6.0dB	3.0dB

Table 1.3: QAM Cross constellation energies.

Vestigial Sideband Modulation (VSB), CAP, and OQAM

From the perspective of performance and constellation design, there are many alternate basis function choices for QAM that are equivalent. These choices sometimes have value from the perspective of implementation considerations. They are all equivalent in terms of the fundamentals of this chapter when implemented for successive transmission of messages over an AWGN. In successive transmission, the basis functions must be orthogonal to one another for all integer translations of the symbol period. Then successive samples at the demodulator output at integer multiples of T will be independent; then, the one-shot optimum receiver can be used repeatedly in succession to detect successive messages without loss of optimality on the AWGN (see Chapter 3 to see when successive transmission can degrade in the presence of “intersymbol interference” on non-AWGN channels with bandwidth limitations in practice.) The PAM basis function always exhibits this desirable translation property on the AWGN, and so do the QAM basis functions as long as $\omega_c \geq \pi/T$. The QAM basis functions are not unique with respect to satisfaction of the translation property, with VSB/SSB, CAP, and OQAM all being variants:

VSB Vestigial sideband modulation (VSB) is an alternative modulation method that is equivalent to QAM. In QAM, typically the same unit-energy basis function ($\sqrt{1/T}\text{sinc}(t/T)$) is “double-side-band modulated” independently by the sine and cosine of a carrier to generate the two QAM basis functions. In VSB, a double bandwidth sinc function and its Hilbert transform (see Chapter 2 for a discussion of Hilbert transforms) are “single-side-band modulated.” The two VSB basis functions are¹³

$$\varphi_1(t) = \sqrt{\frac{1}{T}}\text{sinc}\left(\frac{2t}{T}\right)\cos\omega_c t \quad , \quad (1.272)$$

$$\varphi_2(t) = \sqrt{\frac{1}{T}}\text{sinc}\left(\frac{t}{T}\right)\sin\left(\frac{\pi t}{T}\right)\sin\omega_c t \quad . \quad (1.273)$$

A natural symbol-rate choice for successive transmission with these two basis functions might appear to be $2/T$, twice the rate associated with QAM. However, successive translations of these basic functions by integer multiples of $T/2$ are not orthogonal – that is $\langle \varphi_i(t), \varphi_j(t - T/2) \rangle \neq \delta_{ij}$; however, $\langle \varphi_i(t), \varphi_j(t - kT) \rangle = \delta_{ij}$ for any integer k . Thus, the symbol rate for successive orthogonal transmissions needs to be $1/T$.

VSB designers often prefer to exploit the observation that $\langle \varphi_1(t), \varphi_2(t - kT/2) \rangle = 0$ for all odd integers k to implement the VSB transmission system as one-dimensional time-varying at rate $2/T$ dimensions per second. The first and second dimensions are alternately implemented at an aggregate rate of $2/T$ dimensions per second. The optimum receiver consists of two matched filters to the two basis functions, which have their outputs each sampled at rate $1/T$ (staggered by $T/2$ with respect to one another), and these samples interleaved to form a single one-dimensional sample stream for the detector. Nonetheless, those same designers call the VSB constellations by two-dimensional names.

¹³This simple description is actually single-side-band (SSB), a special case of VSB. VSB uses practical realizable functions instead of the unrealizable sinc functions that simplify fundamental developments in Chapter 1.

Thus, one may hear of 16 VSB or 64 VSB, which are equivalent to 16 QAM (or 4 PAM) and 64 QAM (8 PAM) respectively. VSB transmission may be more convenient for upgrading existing analog systems that are already VSB (i.e., commercial television) to digital systems that use the same bandwidths and carrier frequencies - that is where the carrier frequencies are not centered within the existing band. VSB otherwise has no fundamental advantages or differences from QAM.

CAP Carrierless Amplitude/Phase (CAP) transmission systems are also very similar to QAM. The basis functions of QAM are time-varying when ω_c is arbitrary - that is, the basis functions on subsequent transmissions may differ. CAP is a method that can eliminate this time variation for any choice of carrier-frequency, making the combined transmitter implementation appear “carrierless” and thus time-invariant. CAP has the same one-shot basis functions as QAM, but also has a time-varying encoder constellation when used for successive transmission of two-dimensional symbols. The time-varying CAP encoder implements a sequence of additional two-dimensional constellation rotations that are known and easily removed at the receiver after the demodulator and just before the detector. The time-varying encoder usually selects the sequence of rotations so that the phase (argument of sines and cosines) of the carrier is the same at the beginning of each symbol period, regardless of the actual carrier frequency. Effectively, all carrier frequencies thus appear the same, hence the term “carrierless.” The sequence of rotations has an angle that increases linearly with time and can often be very easily implemented (and virtually omitted when differential encoding - see Chapter 4 - is implemented). See Section 2.4.

OQAM Offset QAM (OQAM) or “staggered” QAM uses the alternative basis functions

$$\varphi_1(t) = \sqrt{\frac{2}{T}} \operatorname{sinc}\left(\frac{t}{T}\right) \cos\left(\frac{\pi t}{T}\right) \quad (1.274)$$

$$\varphi_2(t) = -\sqrt{\frac{2}{T}} \operatorname{sinc}\left(\frac{t-T/2}{T}\right) \sin\left(\frac{\pi t}{T}\right) \quad (1.275)$$

effectively “offsetting” the two dimensions by $T/2$. For one-shot transmission, such offset has no effect (the receiver matched filters effectively re-align the two dimensions) and OQAM and QAM are the same. For successive transmission, the derivative (rate of change) of $x(t)$ is less for OQAM than for QAM, effectively reducing the bandwidth of transmitted signals when the sinc functions cannot be perfectly implemented. OQAM signals will never take the value $x(t) = 0$, while this value is instantaneously possible with QAM - thus nonlinear transmitter/receiver amplifiers are not as stressed by OQAM. There is otherwise no difference between OQAM and QAM.

The Gap

The **gap**, Γ , is an approximation introduced by Forney for constellations with $\bar{b} \geq 1/2$ that is empirically evident in the PAM and QAM tables. Specifically, if one knows the SNR for an AWGN channel, the number of bits that can be transmitted with PAM or QAM according to

$$\bar{b} = \frac{1}{2} \log_2 \left(1 + \frac{\text{SNR}}{\Gamma} \right) . \quad (1.276)$$

At error rate $\bar{P}_e = 10^{-6}$, the gap is 8.8 dB. For $\bar{P}_e = 10^{-7}$, the gap is 9.5 dB. If the designer knows the SNR and his desired performance level (\bar{P}_e) or equivalently the gap, then the number of bits per dimension (and thus the achievable data rate $R = b/T$) are immediately computed. Chapters 8-10 will introduce more sophisticated encoder designs where the gap can be reduced, ultimately to 0 dB, enabling a highest possible data rate of $.5 \log_2(1 + \text{SNR})$, sometimes known as the “channel capacity” of the AWGN channel. QAM and PAM are thus about 9 dB away in terms of efficient use of SNR from ultimate limits.

1.6.3 Constellation Performance Measures

Having introduced many commonly used signal constellations for data transmission, several performance measures compare coded systems based on these constellations.

Coding Gain

Of fundamental importance to the comparison of two systems that transmit the same number of bits per dimension is the **coding gain**, which specifies the improvement of one constellation over another when used to transmit the same information.

Definition 1.6.2 (Coding Gain) *The coding gain (or loss), γ , of a particular constellation with data symbols $\{\mathbf{x}_i\}_{i=0,\dots,M-1}$ with respect to another constellation with data symbols $\{\tilde{\mathbf{x}}_i\}_{i=0,\dots,M-1}$ is defined as*

$$\gamma \triangleq \frac{\left(d_{\min}^2(\mathbf{x}) / \bar{\mathcal{E}}_{\mathbf{x}} \right)}{\left(d_{\min}^2(\tilde{\mathbf{x}}) / \bar{\mathcal{E}}_{\tilde{\mathbf{x}}} \right)} = \frac{\zeta_{\mathbf{x}}}{\zeta_{\tilde{\mathbf{x}}}} \quad (1.277)$$

where both constellations are used to transmit \bar{b} bits of information per dimension.

A coding gain of $\gamma = 1$ (0dB) implies that the two systems perform equally. A positive gain (in dB) means that the constellation with data symbols \mathbf{x} outperforms the constellation with data symbols $\tilde{\mathbf{x}}$.

As an example, we compare the two constellations in Figures 1.30 and 1.32 and obtain

$$\gamma = \frac{\zeta_{\mathbf{x}}(8\text{AMPM})}{\zeta_{\mathbf{x}}(8\text{PSK})} = \frac{\frac{2}{10}}{\sin^2(\frac{\pi}{8})} \approx 1.37 \text{ (1.4dB)} . \quad (1.278)$$

Signal constellations are often based on N -dimensional structures known as lattices. (A discussion of lattices appears in Chapter 8.) A **lattice** is a set of vectors in N -dimensional space that is closed under vector addition – that is, the sum of any two vectors is another vector in the set. A translation of a lattice produces a **coset** of the lattice. Most good signal constellations are chosen as subsets of cosets of lattices. The **fundamental volume** for a lattice measures the region around a point:

Definition 1.6.3 (Fundamental Volume) *The fundamental volume $\mathcal{V}(\Lambda)$ of a lattice Λ (from which a signal constellation is constructed) is the volume of the decision region for any single point in the lattice. This decision region is also called a **Voronoi Region of the lattice**. The Voronoi Region of a lattice, $\mathcal{V}(\Lambda)$, is to be distinguished from the Voronoi Region of the constellation, $\mathcal{V}_{\mathbf{x}}$ the latter being the union of M of the former.*

For example, an M -QAM constellation as $M \rightarrow \infty$ is a translated subset (coset) of the two-dimensional rectangular lattice Z^2 , so M -QAM is a translation of Z^2 as $M \rightarrow \infty$. Similarly as $M \rightarrow \infty$, the M -PAM constellation becomes a coset of the one dimensional lattice Z .

The coding gain, γ of one constellation based on \mathbf{x} with lattice λ and volume $\mathcal{V}(\Lambda)$ with respect to another constellation with $\tilde{\mathbf{x}}$, $\tilde{\Lambda}$, and $\mathcal{V}(\tilde{\Lambda})$ can be rewritten as

$$\gamma = \frac{\left(\frac{d_{\min}^2(\mathbf{x})}{\mathcal{V}(\Lambda)^{2/N}} \right) \cdot \left(\frac{\mathcal{V}(\tilde{\Lambda})^{2/N}}{\bar{\mathcal{E}}_{\tilde{\mathbf{x}}}} \right)}{\left(\frac{d_{\min}^2(\tilde{\mathbf{x}})}{\mathcal{V}(\tilde{\Lambda})^{2/N}} \right) \cdot \left(\frac{\mathcal{V}(\Lambda)^{2/N}}{\bar{\mathcal{E}}_{\mathbf{x}}} \right)} \quad (1.279)$$

$$= \gamma_f + \gamma_s \text{ (dB)} \quad (1.280)$$

The two quantities on the right in (1.280) are called the **fundamental gain** γ_f and the **shaping gain** γ_s respectively.

Definition 1.6.4 (Fundamental Gain) *The fundamental gain γ_f of a lattice, upon which a signal constellation is based, is*

$$\gamma_f \triangleq \frac{\left(\frac{d_{\min}^2(\mathbf{x})}{\mathcal{V}(\Lambda)^{2/N}} \right)}{\left(\frac{d_{\min}^2(\tilde{\mathbf{x}})}{\mathcal{V}(\tilde{\Lambda})^{2/N}} \right)} . \quad (1.281)$$

The fundamental gain measures the efficiency of the spacing of the points within a particular constellation per unit of fundamental volume surrounding each point.

Definition 1.6.5 (Shaping Gain) *The shaping gain γ_s of a signal constellation is defined as*

$$\gamma_s = \frac{\left(\frac{\mathcal{V}(\Lambda)^{2/N}}{\mathcal{E}_{\mathbf{x}}}\right)}{\left(\frac{\mathcal{V}(\tilde{\Lambda})^{2/N}}{\mathcal{E}_{\tilde{\mathbf{x}}}}\right)} . \quad (1.282)$$

The shaping gain measures the efficiency of the shape of the boundary of a particular constellation in relation to the average energy per dimension required for the constellation.

Using a continuous approximation, the designer can extend shaping gain to constellations with different numbers of points as

$$\gamma_s = \frac{\left(\frac{\mathcal{V}(\Lambda)^{2/N}}{\mathcal{E}_{\mathbf{x}}}\right) \cdot 2^{2\bar{b}(\mathbf{x})}}{\left(\frac{\mathcal{V}(\tilde{\Lambda})^{2/N}}{\mathcal{E}_{\tilde{\mathbf{x}}}}\right) \cdot 2^{2\bar{b}(\tilde{\mathbf{x}})}} . \quad (1.283)$$

Peak-to-Average Power Ratio (PAR)

For practical system design, the peak power of a system may also need to be limited. This constraint can manifest itself in several different ways. For example if the modulator uses a Digital-to-Analog Converter (or Analog-to-Digital Converter for the demodulator) with a finite number of bits (or finite dynamic range), then the signal peaks can not be arbitrarily large. In other systems the channel or modulator/demodulator may include amplifiers or repeaters that saturate at high peak signal voltages. Yet another way is in adjacent channels where crosstalk exists and a high peak on one channel can couple into the other channel, causing an impulsive noise hit and an unexpected error in the adjacent system. Thus, the Peak-to-Average Power Ratio (PAR) is a measure of immunity to these important types of effects.

The peak energy is:

Definition 1.6.6 (Peak Energy) *The N -dimensional peak energy for any signal constellation is \mathcal{E}_{peak} .*

$$\mathcal{E}_{peak} \triangleq \max_i \sum_{n=1}^N x_{in}^2 . \quad (1.284)$$

The peak energy of a constellation should be distinguished from the peak squared energy of a signal $x(t)$, which is $\max_{i,t} |x_i(t)|^2$. This latter quantity is important in analog amplifier design or equivalently in however the filters $\varphi_n(t)$ are implemented.

The peak energy of a constellation concept allows precise definition of the PAR:

Definition 1.6.7 (Peak-to-Average Power Ratio) *The N -dimensional **Peak-to-Average Power Ratio**, $PAR_{\mathbf{x}}$, for N -dimensional Constellation is*

$$PAR_{\mathbf{x}} = \frac{\mathcal{E}_{peak}}{\mathcal{E}_{\mathbf{x}}} \quad (1.285)$$

For example 16SQ QAM has a PAR of 1.8 in two dimensions. For each of the one-dimensional 4-PAM constellations that constitute a 16SQ QAM constellation, the one-dimensional PAR is also 1.8. These two ratios need not be equal, however, in general. For instance, for 32CR, the two-dimensional PAR is $34/20 = 1.7$, while observation of a single dimension when 32CR is used gives a one-dimensional PAR of $25/(\cdot75(5) + \cdot25(25)) = 2.5$. Typically, the peak squared signal energy is inevitably yet higher in QAM constellations and depends on the choice of $\varphi(t)$.

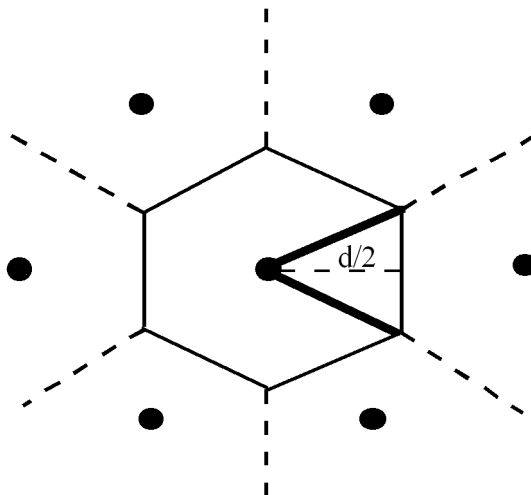


Figure 1.44: Hexagonal lattice.

1.6.4 Hexagonal Signal Constellations in 2 Dimensions

The most dense packing of regularly spaced points in two dimensions is the hexagonal lattice shown in Figure 1.44. The volume (area) of the decision region for each point is

$$\mathcal{V} = 6\left(\frac{1}{2}\right)\left(\frac{d}{2}\right)\left(\frac{d}{\sqrt{3}}\right) = d^2 \frac{\sqrt{3}}{2} \quad . \quad (1.286)$$

If the minimum distance between any two points is d in both constellations, then the fundamental gain of the hexagonal constellation with respect to the QAM constellation is

$$\gamma_f = \frac{d^2}{\frac{\sqrt{3}d^2}{2}} = \frac{2}{\sqrt{3}} = .625\text{dB} \quad . \quad (1.287)$$

The encoder/detector for constellations based on the hexagonal lattice may be more complex than those for QAM.

1.7 Additive Self-Correlated Noise

In practice, additive noise is often Gaussian, but its power spectral density is not flat. Engineers often call this type of noise “self-correlated” or “colored”. The noise remains independent of the message signals but correlates with itself from time instant to time instant. The origins of colored noise are many. Receiver filtering effects, noise generated by other communications systems (“crosstalk”), and electromagnetic interference are all sources of self-correlated noise. A narrow-band transmission of a radio signal that somehow becomes noise for an unintended channel is another common example of self-correlated noise and called “RF” noise (RF is an acronym for “radio frequency”).

Self-correlated Gaussian noise can significantly alter the performance of a receiver designed for white Gaussian noise. This section investigates the optimum detector for colored noise and also considers the performance loss when using a suboptimum detector designed for AWGN case in the presence of Additive Correlated Gaussian Noise (ACGN).

This study is facilitated by first investigating the filtered “one-shot” AWGN channel in Subsection 1.7.1. Subsection 1.7.2 then finds the optimum detector for additive self-correlated Gaussian noise, by adding a whitening filter that transforms the self-correlated noise channel into a filtered AWGN channel.

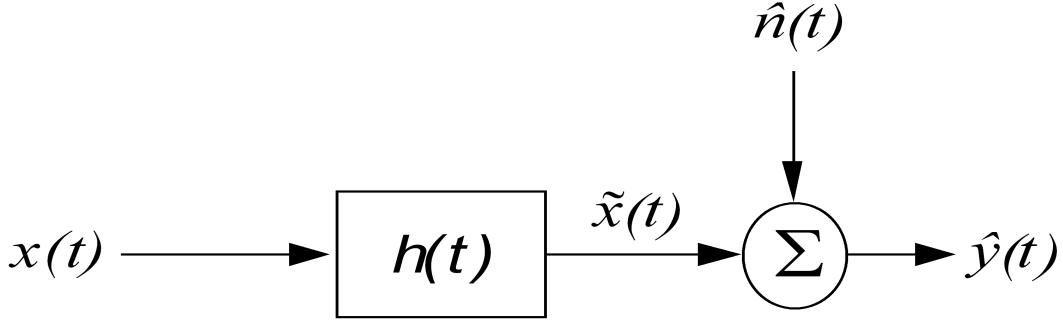


Figure 1.45: Filtered AWGN channel.

Subsection 1.7.3 studies the vector channel, for which (for some unspecified reason) the noise has not been whitened and describes the optimum detector given this vector channel. Finally, Subsection 1.7.4 studies the degradation that occurs when the noise correlation properties are not known for design of the receiver, and an optimum receiver for the AWGN is used instead.

1.7.1 The Filtered (One-Shot) AWGN Channel

The filtered AWGN channel is illustrated in Figure 1.45. The modulated signal $x(t)$ is filtered by $h(t)$ before the addition of the white Gaussian noise. When $h(t) \neq \delta(t)$, the filtered signal set $\{\tilde{x}_i(t)\}$ may differ from the transmitted signal set $\{x_i(t)\}$. This transformation may change the probability of error as well as the structure of the optimal detector. This section still considers only one use of the channel with M possible messages for transmission. Transmission over this type of channel can incur a significant penalty due to intersymbol interference between successively transmitted data symbols. In the “one-shot” case, however, analysis need not consider this intersymbol interference. Intersymbol interference is considered in Chapters 3, 4, and 5.

For any channel input signal $x_i(t)$, the corresponding filtered output equals $\tilde{x}_i(t) = h(t) * x_i(t)$. Decomposing $x_i(t)$ by an orthogonal basis set, $\tilde{x}_i(t)$ becomes

$$\tilde{x}_i(t) = h(t) * x_i(t) \quad (1.288)$$

$$= h(t) * \sum_{n=1}^N x_{in} \varphi_n(t) \quad (1.289)$$

$$= \sum_{n=1}^N x_{in} \{h(t) * \varphi_n(t)\} \quad (1.290)$$

$$= \sum_{n=1}^N x_{in} \phi_n(t) \quad , \quad (1.291)$$

where

$$\phi_n(t) \triangleq h(t) * \varphi_n(t) \quad . \quad (1.292)$$

Note that:

- The set of N functions $\{\phi_n(t)\}_{n=1,\dots,N}$ is not necessarily orthonormal.
- For the channel to convey any and all constellations of M messages for the signal set $\{x_i(t)\}$, the basis set $\{\phi_n(t)\}$ must be linearly independent.

The first observation can be easily proven by finding a counterexample, an exercise for the interested reader. The second observation emphasizes that if some dimensionality is lost by filtering, signals in the original signal set that differed only along the lost dimension(s) would appear identical at the channel output. For example consider the two signals $\tilde{x}_k(t)$ and $\tilde{x}_j(t)$.

$$\tilde{x}_k(t) - \tilde{x}_j(t) = \sum_{n=1}^N (x_{kn} - x_{jn})\phi_n(t) = 0 \quad , \quad (1.293)$$

If the set $\{\phi_n(t)\}$ is linearly independent then the sum in (1.293) must be nonzero: a contradiction to (1.293). If this set of vectors is linearly dependent, then (1.293) can be satisfied, resulting in the possibility of ambiguous transmitted signals. Failure to meet the linear independence condition could mandate a redesign of the modulated signal set or a rate reduction (decrease of M). The dimensionality loss and ensuing redesign of $\{x_i(t)\}_{i=0:M-1}$ is studied in Chapters 4 and 5. This chapter assumes such dimensionality loss does not occur.

If the set $\{\phi_n(t)\}$ is linearly independent, then the Gram-Schmidt procedure in Appendix A generates an orthonormal set of N basis functions $\{\psi_n(t)\}_{n=1,\dots,N}$ from $\{\phi_n(t)\}_{n=1,\dots,N}$. A new signal constellation $\{\tilde{x}_i\}_{i=0:M-1}$ can be computed from the filtered signal set $\{\tilde{x}_i(t)\}$ using the basis set $\{\psi_n(t)\}$.

$$\tilde{x}_{in} = \int_{-\infty}^{\infty} \tilde{x}_i(t)\psi_n(t)dt = \langle \tilde{x}_i(t), \psi_n(t) \rangle . \quad (1.294)$$

Using the previous analysis for AWGN, a tight upper bound on message error probability is still given by

$$P_e \leq N_e Q \left[\frac{d_{\min}}{2\sigma} \right] \quad , \quad (1.295)$$

where d_{\min} is the minimum Euclidean distance between any two points in the filtered signal constellation $\{\tilde{x}_i\}_{i=0:M-1}$. The matched filter implementation of the demodulator/detector does not need to compute $\{\psi_n(t)\}_{n=1,\dots,N}$ for the signal detector as shown in Figure 1.46. (For reference the reader can reexamine the detector for the unfiltered constellation in Figure 1.23).

In the analysis of the filtered AWGN, the transmitted average energy $\mathcal{E}_{\tilde{x}}$ is still measured at the channel input. Thus, while $\mathcal{E}_{\tilde{x}}$ can be computed, its physical significance can differ from that of $\mathcal{E}_{\mathbf{x}}$. If, as is often the case, the energy constraint is at the input to the channel, then the comparison of various signaling alternatives, as performed earlier in this chapter could change depending on the specific filter $h(t)$.

1.7.2 Optimum Detection in the Presence of Self-Correlated Noise

The Additive Self-Correlated Gaussian Noise (ACGN) channel is illustrated in Figure 1.47. The only change with respect to Figure 1.18 is that the autocorrelation function of the additive noise $r_n(\tau)$, need not equal $\frac{N_0}{2}\delta(\tau)$. Simplification of the ensuing development defines and uses a normalized noise autocorrelation function

$$\bar{r}_n(\tau) \triangleq \frac{r_n(\tau)}{\frac{N_0}{2}} \quad . \quad (1.296)$$

The power spectral density of the unnormalized noise is then

$$\mathcal{S}_n(f) = \frac{N_0}{2} \bar{\mathcal{S}}_n(f) \quad , \quad (1.297)$$

where $\bar{\mathcal{S}}_n(f)$ is the Fourier Transform of $\bar{r}_n(\tau)$.

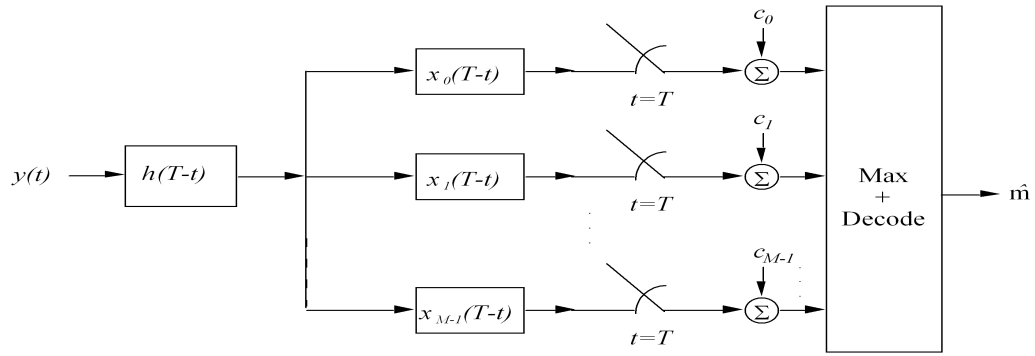


Figure 1.46: Modified signal detector.

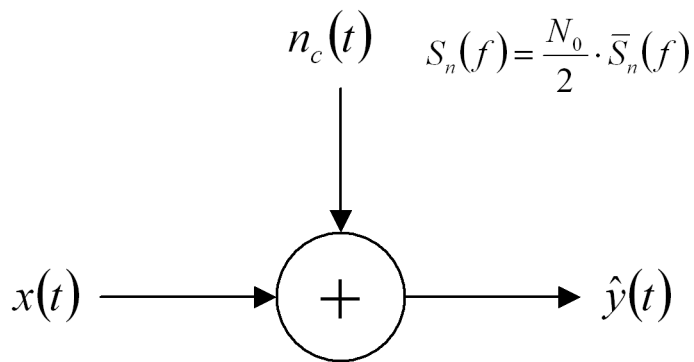


Figure 1.47: ACGN channel.

The Whitening filter

The whitening-filter analysis of the ACGN channel “whitens” the colored noise with a whitening filter $g(t)$, and then uses the results of the previous section for the filtered AWGN channel where the filter $h(t) = g(t)$. To ensure no loss of information when filtering the noisy received signal by $g(t)$, the filter $g(t)$ should be invertible. By the reversibility theorem, the receiver can use an optimal detector for this newly generated filtered AWGN without performance loss. Actually, the condition on invertibility of $g(t)$ is sufficient but not necessary. For a particular signal set, a necessary condition is that the filter be invertible over that signal set. For the filter to be invertible on any possible signal set, $g(t)$ must necessarily be invertible. This subtle point is often overlooked by most works on this subject.

For $g(t)$ to whiten the noise,

$$[\bar{\mathcal{S}}_n(f)]^{-1} = |G(f)|^2 \quad . \quad (1.298)$$

In general many filters $G(f)$, may satisfy Equation (1.298) but only some of the filters shall possess realizable inverses.

To ensure the existence of a realizable inverse $\mathcal{S}_n(f)$ must satisfy the **Paley-Wiener Criterion**.

Theorem 1.7.1 (Paley-Wiener Criterion) *If*

$$\int_{-\infty}^{\infty} \frac{|\ln \mathcal{S}_n(f)|}{1+f^2} df < \infty \quad , \quad (1.299)$$

then there exists a $G(f)$ satisfying (1.298) with a realizable inverse. (Thus the filter $g(t)$ is a 1-to-1 mapping).

If the Paley-Wiener criterion were violated by a noise signal, then it is possible to design transmission systems with infinite data rate (that is when $\mathcal{S}_n(f) = 0$ over a given bandwidth) or to design transmission systems for each band over which Paley-Wiener is satisfied (that is the bands where noise is essentially of finite energy). This subsection’s analysis always assumes Equation (1.299) is satisfied.¹⁴ With a 1-to-1 $g(t)$ that satisfies (1.298), the ACGN channel converts into an equivalent filtered white Gaussian noise channel as shown in Figure 1.45 replacing $h(t)$ with $g(t)$. The performance analysis of ACGN is identical to that derived for the filtered AWGN channel in Subsection 1.7.1.

A further refinement handles the filtered ACGN channel by whitening the noise and then analyzing the filtered AWGN with $h(t)$ replaced by $h(t) * g(t)$.

Analytic continuation of $\bar{\mathcal{S}}_n(s)$ determines an invertible $g(t)$:

$$\bar{\mathcal{S}}_n(s) = \bar{\mathcal{S}}_n \left(f = \frac{s}{2\pi j} \right) \quad , \quad (1.300)$$

where $\bar{\mathcal{S}}_n(s)$ can be canonically (and uniquely) factored into causal (and causally invertible) and anti-causal (and anticausally invertible) parts as

$$\bar{\mathcal{S}}_n(s) = \bar{\mathcal{S}}_n^+(s) \bar{\mathcal{S}}_n^-(s) \quad , \quad (1.301)$$

where

$$\bar{\mathcal{S}}_n^+(s) = \bar{\mathcal{S}}_n^-(-s) \quad . \quad (1.302)$$

If $\bar{\mathcal{S}}_n(s)$ is rational, then $\bar{\mathcal{S}}_n^+(s)$ is “minimum phase,” i.e. all poles and zeros of $\bar{\mathcal{S}}_n^+(s)$ are in the left half plane. The filter $g(t)$ is then given by

$$g(t) = \mathcal{L}^{-1} \left\{ \frac{1}{\bar{\mathcal{S}}_n^+(s)} \right\} \quad (1.303)$$

where \mathcal{L}^{-1} is the inverse Laplace Transform. The matched filter $g(-t)$ is given by $g(-t) = \frac{1}{2\pi j} \oint G(s) e^{-st} ds$, or equivalently by

$$g(-t) = \mathcal{L}^{-1} \left\{ \frac{1}{\bar{\mathcal{S}}_n^-(s)} \right\} \quad . \quad (1.304)$$

¹⁴Chapters 4 and 5 expand to the correct form of transmission that should be used when (1.299) is not satisfied.

$g(-t)$ is anticausal and cannot be realized. Practical receivers instead realize $g(T - t)$, where T is sufficiently large to ensure causality.

In general $g(t)$ may be difficult to implement by this method; however, the next subsection considers a discrete equivalent of whitening that is more straightforward to implement in practice. When the noise is complex (see Chapter 2), Equation (1.302) generalizes to

$$\bar{\mathcal{S}}_n^+(s) = [\bar{\mathcal{S}}_n^-(-s^*)]^* \quad . \quad (1.305)$$

1.7.3 The Vector Self-Correlated Gaussian Noise Channel

This subsection considers a discrete equivalent of the ACGN

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad , \quad (1.306)$$

where the autocorrelation matrix of the noise vector \mathbf{n} is

$$E[\mathbf{n}\mathbf{n}^*] = \mathbf{R}_n = \bar{\mathbf{R}}_n \sigma^2 \quad . \quad (1.307)$$

Both \mathbf{R}_n and $\bar{\mathbf{R}}_n$ are positive definite matrices. This discrete ACGN channel can often be substituted for the continuous ACGN channel. The discrete noise vector can be “whitened”, transforming $\bar{\mathbf{R}}_n$ into an identity matrix. The discrete equivalent to whitening $y(t)$ by $g(t)$ is a matrix multiplication.

The $N \times N$ whitening matrix in the discrete case corresponds to the whitening filter $g(t)$ in the continuous case.

Cholesky factorization determines the invertible whitening transformation according (see Appendix A of Chapter 3):

$$\bar{\mathbf{R}}_n = \bar{\mathbf{R}}^{1/2} \bar{\mathbf{R}}^{*/2} \quad , \quad (1.308)$$

where $\bar{\mathbf{R}}^{1/2}$ is lower triangular and $\bar{\mathbf{R}}^{*/2}$ is upper triangular. These matrices constitute the matrix equivalent of a “square root”, and both matrices are invertible. Noting the definitions,

$$\bar{\mathbf{R}}^{-/2} \triangleq [\bar{\mathbf{R}}^{1/2}]^{-1} \quad , \quad (1.309)$$

and

$$\bar{\mathbf{R}}^{-*/2} \triangleq [\bar{\mathbf{R}}^{*/2}]^{-1} \quad , \quad (1.310)$$

To whiten \mathbf{n} , the receiver passes \mathbf{y} through the matrix multiply $\bar{\mathbf{R}}^{-/2}$,

$$\tilde{\mathbf{y}} \triangleq \bar{\mathbf{R}}^{-/2} \mathbf{y} = \bar{\mathbf{R}}^{-/2} \mathbf{x} + \bar{\mathbf{R}}^{-/2} \mathbf{n} = \tilde{\mathbf{x}} + \tilde{\mathbf{n}} \quad . \quad (1.311)$$

The autocorrelation matrix for $\tilde{\mathbf{n}}$ is

$$E[\tilde{\mathbf{n}}\tilde{\mathbf{n}}^*] = \bar{\mathbf{R}}^{-/2} E[\mathbf{n}\mathbf{n}^*] \bar{\mathbf{R}}^{-*/2} = \bar{\mathbf{R}}^{-/2} (\bar{\mathbf{R}}^{1/2} \bar{\mathbf{R}}^{*/2} \sigma^2) \bar{\mathbf{R}}^{-*/2} = \sigma^2 \mathbf{I} \quad . \quad (1.312)$$

Thus, the covariance matrix of the transformed noise $\tilde{\mathbf{n}}$ is the same as the covariance matrix of the AWGN vector. By the theorem of reversibility, no information is lost in such a transformation.

EXAMPLE 1.7.1 (QPSK with correlated noise) For the example shown in Figure 1.21 suppose that the noise is colored with correlation matrix

$$\mathbf{R}_n = \sigma^2 \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 1 \end{bmatrix} \quad (1.313)$$

Then

$$\bar{\mathbf{R}}^{1/2} = \begin{bmatrix} 1 & 0 \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad (1.314)$$

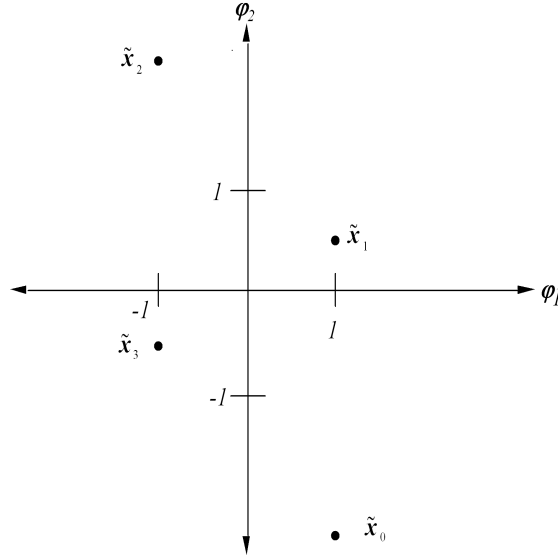


Figure 1.48: Equivalent signal constellation for Example 1.7.1.

and

$$\bar{\mathbf{R}}^{*/2} = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix} . \quad (1.315)$$

From (1.314),

$$\bar{\mathbf{R}}^{-/2} = \begin{bmatrix} 1 & 0 \\ -1 & \sqrt{2} \end{bmatrix} \quad (1.316)$$

and

$$\bar{\mathbf{R}}^{-*/2} = \begin{bmatrix} 1 & -1 \\ 0 & \sqrt{2} \end{bmatrix} . \quad (1.317)$$

The signal constellation after the whitening filter becomes

$$\tilde{\mathbf{x}}_0 = \bar{\mathbf{R}}^{-/2} \mathbf{x}_0 = \begin{bmatrix} 1 & 0 \\ -1 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ (-\sqrt{2} - 1) \end{bmatrix} , \quad (1.318)$$

and similarly $\tilde{\mathbf{x}}_2 = [-1 \ (\sqrt{2} + 1)]'$, $\tilde{\mathbf{x}}_1 = [1 \ (\sqrt{2} - 1)]'$, and $\tilde{\mathbf{x}}_3 = [-1 \ (-\sqrt{2} + 1)]'$. This new constellation forms a parallelogram in two dimensions, where the minimum distance is now along the shorter diagonal (between $\tilde{\mathbf{x}}_1$ and $\tilde{\mathbf{x}}_3$), rather than along the sides and $d_{\min} = 2.164 > 2$. This new constellation appears in Figure 1.48. Thus, the optimum detector for this channel with self-correlated Gaussian noise has larger minimum distance than for the white noise case, illustrating the important fact that having correlated noise is sometimes advantageous.

The example shows that correlated noise may lead to improved performance measured with respect to the same channel and signal constellation with white noise of the same average energy. Nevertheless, the autocorrelation matrix of the noise is often not known in implementation, or it may vary from channel

use to channel use. Then, the detector is designed as if white noise were present anyway, and there is a performance loss with respect to the optimum detector. The next subsection deals with the calculation of this performance loss.

1.7.4 Performance of Suboptimal Detection with Self-Correlated Noise

A detector designed for the AWGN channel is obviously suboptimum for the ACGN channel, but is often used anyway, as the correlation properties of the noise may be hard to know in the design stage. In this case, the detector performance will be reduced with respect to optimum.

Computation of the amount by which performance is reduced uses the error event vectors

$$\boldsymbol{\epsilon}_{ij} \triangleq \frac{\mathbf{x}_i - \mathbf{x}_j}{\|\mathbf{x}_i - \mathbf{x}_j\|} . \quad (1.319)$$

The component of the additive noise vector along an error event vector is $\langle \mathbf{n}, \boldsymbol{\epsilon}_{ij} \rangle$. The variance of the noise along this vector is $\sigma_{ij}^2 \triangleq E \{ \langle \mathbf{n}, \boldsymbol{\epsilon}_{ij} \rangle^2 \}$. Then, the NNUB becomes

$$P_e \leq N_e Q \left[\min_{i \neq j} \left\{ \frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{2\sigma_{ij}} \right\} \right] . \quad (1.320)$$

For Example 1.7.1, the worst case argument of the Q-function in (1.320) is $1/\sigma$, which represents a factor of $(2.164/2)^2 = .7$ dB loss with respect to optimum. This loss varies with rotation of the signal set, but not translation. If the signal constellation in Example 1.7.1 were rotated by 45° , as in Figure 1.26, then the increase in noise variance is $(1 + \sqrt{1/2})/1 = 2.3$ dB, but d_{\min} remains at 2 for this sub-optimum detector, so performance is 3 dB inferior than the optimum detector for the unrotated constellation. However, the optimum receiver for the rotated case would also have changed to have 3 dB worse performance for this rotation, so in this case the optimum rotated and sub-optimum rotated receiver have the same performance.

Chapter 1 Exercises

1.1 Our First Constellation.

a. Show that the following two basis functions are orthonormal. (2 pts)

$$\begin{aligned}\phi_1(t) &= \begin{cases} \sqrt{2} (\cos(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ \phi_2(t) &= \begin{cases} \sqrt{2} (\sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

b. Consider the following modulated waveforms.

$$\begin{aligned}x_0(t) &= \begin{cases} \sqrt{2} (\cos(2\pi t) + \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_1(t) &= \begin{cases} \sqrt{2} (\cos(2\pi t) + 3 \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_2(t) &= \begin{cases} \sqrt{2} (3 \cos(2\pi t) + \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_3(t) &= \begin{cases} \sqrt{2} (3 \cos(2\pi t) + 3 \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_4(t) &= \begin{cases} \sqrt{2} (\cos(2\pi t) - \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_5(t) &= \begin{cases} \sqrt{2} (\cos(2\pi t) - 3 \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_6(t) &= \begin{cases} \sqrt{2} (3 \cos(2\pi t) - \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_7(t) &= \begin{cases} \sqrt{2} (3 \cos(2\pi t) - 3 \sin(2\pi t)) & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_{i+8}(t) &= -x_i(t) \quad i = 0, \dots, 7\end{aligned}$$

Draw the constellation points for these waveforms using the basis functions of (a). (2 pts)

c. Compute $\mathcal{E}_{\mathbf{x}}$ and $\bar{\mathcal{E}}_{\mathbf{x}}$ ($\bar{\mathcal{E}}_{\mathbf{x}} = \mathcal{E}_{\mathbf{x}}/N$) where N is the number of dimensions

(i) for the case where all signals are equally likely. (2 pts)

(ii) for the case where (2 pts)

$$p(x_0) = p(x_4) = p(x_8) = p(x_{12}) = \frac{1}{8}$$

and

$$p(x_i) = \frac{1}{24} \quad i = 1, 2, 3, 5, 6, 7, 9, 10, 11, 13, 14, 15$$

d. Let

$$y_i(t) = x_i(t) + 4\phi_3(t)$$

where

$$\phi_3(t) = \begin{cases} 1 & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases}$$

Compute $\mathcal{E}_{\mathbf{y}}$ for the case where all signals are equally likely. (2 pts)

1.2 Inner Products.

Consider the following signals:

$$\begin{aligned}
 x_0(t) &= \begin{cases} \frac{2}{\sqrt{T}} \cos\left(\frac{2\pi t}{T} + \frac{\pi}{6}\right) & \text{if } t \in [0, T] \\ 0 & \text{otherwise} \end{cases} \\
 x_1(t) &= \begin{cases} \frac{2}{\sqrt{T}} \cos\left(\frac{2\pi t}{T} + \frac{5\pi}{6}\right) & \text{if } t \in [0, T] \\ 0 & \text{otherwise} \end{cases} \\
 x_2(t) &= \begin{cases} \frac{2}{\sqrt{T}} \cos\left(\frac{2\pi t}{T} + \frac{3\pi}{2}\right) & \text{if } t \in [0, T] \\ 0 & \text{otherwise} \end{cases}
 \end{aligned}$$

- Find a set of orthonormal basis functions for this signal set. Show that they are orthonormal. *Hint:* Use the identity for $\cos(a + b) = \cos(a)\cos(b) - \sin(a)\sin(b)$. (4 pts)
- Find the data symbols corresponding to the signals above for the basis functions you found in (a). (3 pts)
- Find the following inner products: (3 pts)
 - $\langle x_0(t), x_0(t) \rangle$
 - $\langle x_0(t), x_1(t) \rangle$
 - $\langle x_0(t), x_2(t) \rangle$

1.3 Multiple sets of basis functions.

Consider the following two orthonormal basis functions:

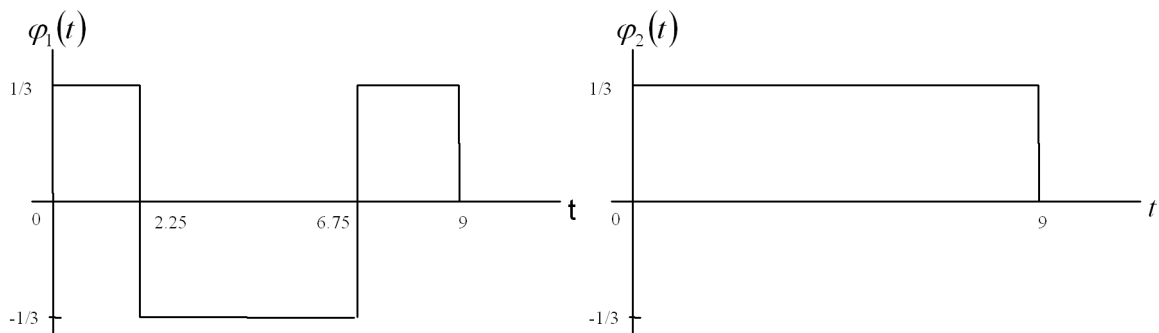


Figure 1.49: Basis functions.

- Use the basis functions given above to find the modulated waveforms $u(t)$ and $v(t)$ given the data symbols $\mathbf{u} = [1 \ 1]$ and $\mathbf{v} = [2 \ 1]$. It is sufficient to draw $u(t)$ and $v(t)$. (2 pts)
- For the same $u(t)$ and $v(t)$, a different set of two orthonormal basis functions is employed for which $\mathbf{u} = [\sqrt{2} \ 0]$ produces $u(t)$. Draw the new basis functions and find the \mathbf{v} which produces $v(t)$. (3 pts)

1.4 Minimal orthonormalization with MATLAB.

Each column of the matrix \mathbf{A} given below is a data symbol that is used to construct its corresponding modulated waveform from the set of orthonormal basis functions $\{\phi_1(t), \phi_2(t), \dots, \phi_6(t)\}$. The set of modulated waveforms described by the columns of \mathbf{A} can be represented with a smaller number of basis functions.

$$A = [\mathbf{a}_0 \ \mathbf{a}_1 \ \dots \ \mathbf{a}_7] \quad (1.321)$$

$$= \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 3 & 5 & 7 & 9 & 11 & 13 & 15 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 2 & 0 & 2 & 0 & 2 & 0 & 2 \\ 2 & 4 & 2 & 4 & 2 & 4 & 2 & 4 \end{bmatrix} \quad (1.322)$$

The transmitted signals $a_i(t)$ are represented (with a superscript of * meaning matrix or vector transpose) as

$$a_i(t) = \mathbf{a}_i^* \begin{bmatrix} \phi_1(t) \\ \phi_2(t) \\ \vdots \\ \phi_6(t) \end{bmatrix} \quad (1.323)$$

$$A(t) = A^* \boldsymbol{\varphi}(t) \quad (1.324)$$

Thus, each row of $A(t)$ is a possible transmitted signal.

- Use MATLAB to find an orthonormal basis for the columns of A. Record the matrix of basis vectors. The MATLAB commands **help** and **orth** will be useful. In particular, if one executes $Q = \text{orth}(A)$ in matlab, a 6×3 orthogonal matrix Q is produced such that $Q^*Q = I$ and $A^* = [A^*Q]Q^*$. The columns of Q can be thought of as a new basis – thus try writing $A(t)$ and interpreting to get a new set of basis functions and description of the 8 possible transmit waveforms. Note that **help orth** will give a summary of the **orth** command. To enter the matrix B in matlab (for example) shown below, simply type $B=[1 \ 2; 3 \ 4]$; (2 pts)

$$B = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad (1.325)$$

- How many basis functions are actually needed to represent our signal set ? What are the new basis functions in terms of $\{\phi_1(t), \phi_2(t), \dots, \phi_6(t)\}$?(2 pts)
- Find the new matrix \hat{A} which gives the data symbol representation for the original modulated waveforms using the smaller set of basis functions found in (b). \hat{A} will have 8 columns, one for each data symbol. The number of rows in \hat{A} will be the number of basis functions you found in (b). (1 pts)

1.5 Decision rules for binary channels.

- The **Binary Symmetric Channel (BSC)** has binary (0 or 1) inputs and outputs. It outputs each bit correctly with probability $1 - p$ and incorrectly with probability p . Assume 0 and 1 are equally likely inputs. State the MAP and ML decision rules for the BSC when $p < \frac{1}{2}$. How are the decision rules different when $p > \frac{1}{2}$? (5 pts)
- The **Binary Erasure Channel (BEC)** has binary inputs as with the BSC. However there are three possible outputs. Given an input of 0, the output is 0 with probability $1 - p_1$ and 2 with probability p_1 . Given an input of 1, the output is 1 with probability $1 - p_2$ and 2 with probability p_2 . Assume 0 and 1 are equally likely inputs. State the MAP and ML decision rules for the BEC when $p_1 < p_2 < \frac{1}{2}$. How are the decision rules different when $p_2 < p_1 < \frac{1}{2}$? (5 pts)

1.6 Minimax [Wesel 1994].

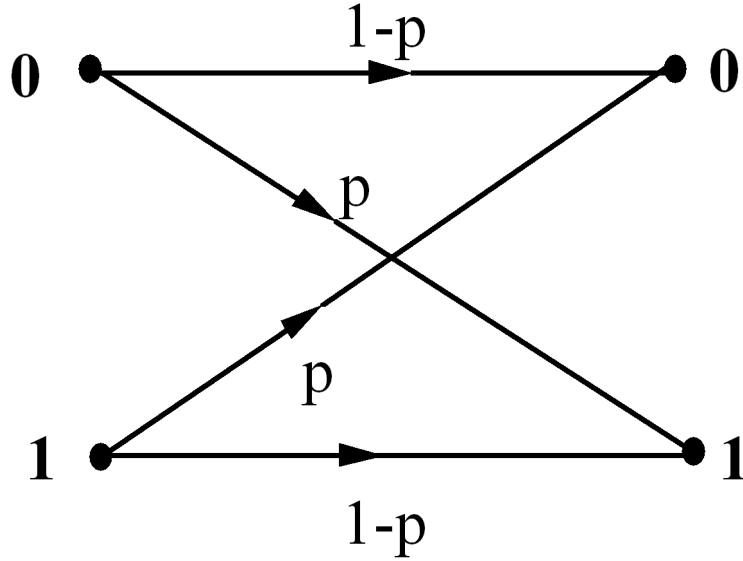


Figure 1.50: Binary Symmetric Channel(BSC).

Consider a 1-dimensional vector channel

$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

where $\mathbf{x} = \pm 1$ and \mathbf{n} is Gaussian noise with $\sigma^2 = 1$. The Maximum-Likelihood (ML) Receiver which is minimax, has decision regions:

$$D_{ML,1} = [0, \infty)$$

and

$$D_{ML,-1} = (-\infty, 0)$$

So if \mathbf{y} is in $D_{ML,1}$ we decode \mathbf{y} as +1; in $D_{ML,-1}$, as -1.

Consider another receiver, R , where the decision regions are:

$$D_{R,1} = \left[\frac{1}{2}, \infty\right)$$

and

$$D_{R,-1} = \left(-\infty, \frac{1}{2}\right)$$

- Find $P_{e,ML}$ and $P_{e,R}$ as a function of $p\mathbf{x}(1) = p$ for values of p in the interval $[0, 1]$. On the same graph, plot $P_{e,ML}$ vs. p and $P_{e,R}$ vs. p . (2 pts)
- Find $\max_p P_{e,ML}$ and $\max_p P_{e,R}$. Are your results consistent with the Minimax Theorem? (2 pts)
- For what value of p is D_R the MAP decision rule? (1 pt)

Note: For this problem you will need to use the $Q(\cdot)$ function discussed in Appendix B. Here are some relevant values of $Q(\cdot)$.

x	Q(x)
0.5	0.3085
1.0	0.1587
1.5	0.0668

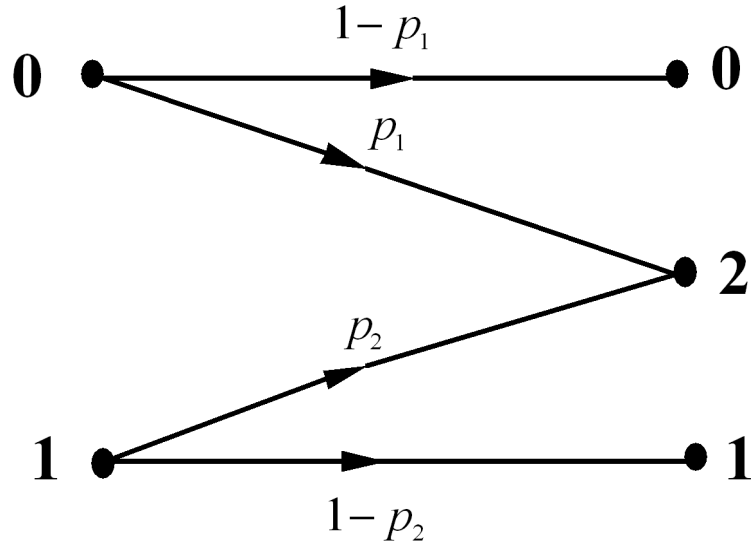
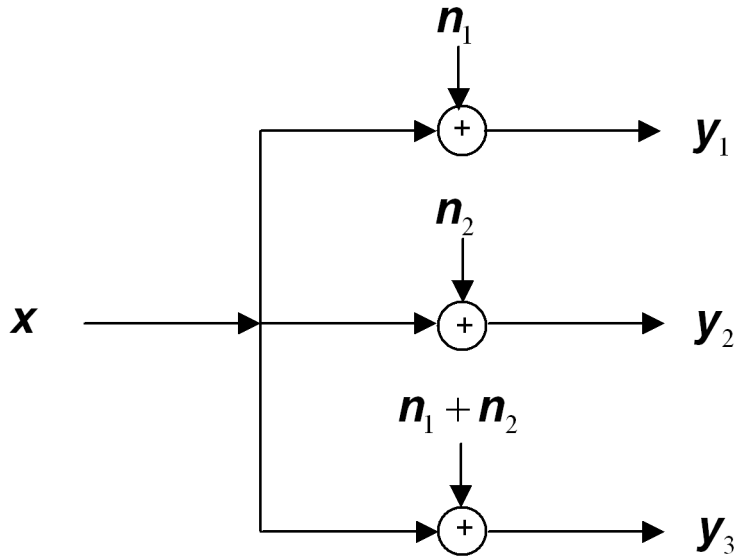


Figure 1.51: Binary Erasure Channel (BEC).

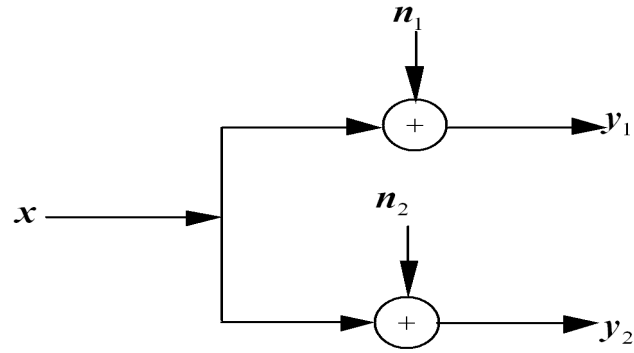
1.7 Irrelevancy/Decision Regions. (From Wozencraft and Jacobs)

- a. Consider the following channel where x , n_1 , and n_2 are independent binary random variables. All the additions shown below are modulo two. (Equivalently, the additions may be considered xor's.)



- (i) Given only y_1 , is y_3 relevant? (1 pt)
- (ii) Given y_1 and y_2 , is y_3 relevant? (1 pt)

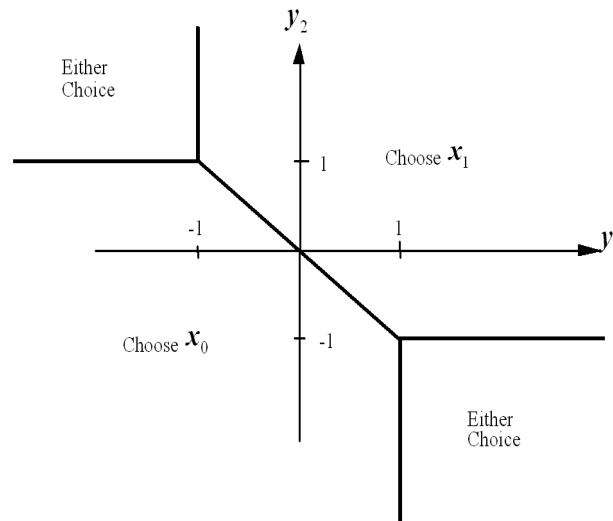
For the rest of the problem, consider the following channel,
 One of the two signals $x_0 = -1$ or $x_1 = 1$ is transmitted over this channel. The noise random



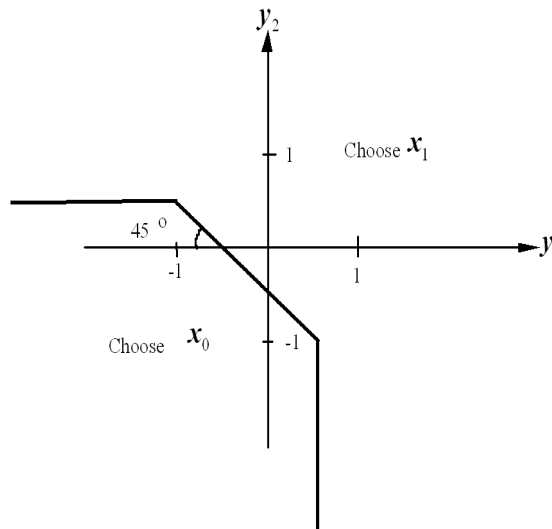
variables n_1 and n_2 are statistically independent of the transmitted signal x and of each other. Their density functions are,

$$p_{n_1}(n) = p_{n_2}(n) = \frac{1}{2}e^{-|n|} \quad (1.326)$$

- b. Given y_1 only, is y_2 relevant ? (1 pt)
- c. Prove that the optimum decision regions for equally likely messages are shown below, (3 pts)



- d. A receiver chooses x_1 if and only if $(y_1 + y_2) > 0$. Is this receiver optimum for equally likely messages ? What is the probability of error ? (Hint: $P_e = P\{y_1 + y_2 > 0/x = -1\}p_x(-1) + P\{y_1 + y_2 < 0/x = 1\}p_x(1)$ and use symmetry. Recall the probability density function of the sum of 2 random variables is the convolution of their individual probability density functions) (4 pts)
- e. Prove that the optimum decision regions are modified as indicated below when $Pr\{X = x_1\} > 1/2$. (2 pts)



1.8 Optimum Receiver. (From Wozencraft and Jacobs)

Suppose one of M equiprobable signals $x_i(t)$, $i = 0, \dots, M - 1$ is to be transmitted during a period of time T over an AWGN channel. Moreover, each signal is identical to all others in the subinterval $[t_1, t_2]$, where $0 < t_1 < t_2 < T$.

- Show that the optimum receiver may ignore the subinterval $[t_1, t_2]$. (2 pts)
- Equivalently, show that if $\mathbf{x}_0, \dots, \mathbf{x}_{M-1}$ all have the same projection in one dimension, then this dimension may be ignored. (2 pts)
- Does this result necessarily hold true if the noise is Gaussian but not white? Explain. (2 pts)

1.9 Receiver Noise (use MATLAB for all necessary calculations - courtesy S. Li, 2005).

Each column of A given below is a data symbol that is used to construct its corresponding modulated waveform from a set of orthonormal basis functions (assume all messages are equally likely):

$$\Phi(t) = [\phi_1(t) \quad \phi_2(t) \quad \phi_3(t) \quad \phi_4(t) \quad \phi_5(t) \quad \phi_6(t)] \quad .$$

The matrix A is given by

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 4 & 6 & 8 & 10 & 12 & 14 & 16 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 3 & 3 & 3 & 3 \\ 5 & 6 & 7 & 8 & 5 & 6 & 7 & 8 \end{bmatrix} \quad (1.327)$$

so that

$$\mathbf{x}(t) = \Phi(t)A = [x_0(t) \quad x_1(t) \quad \dots \quad x_7(t)] \quad . \quad (1.328)$$

A noise vector $\mathbf{n} = \begin{bmatrix} n_1 \\ n_2 \\ n_3 \\ n_4 \\ n_5 \\ n_6 \end{bmatrix}$ is added to the symbol vector \mathbf{x} , such that

$$\mathbf{y}(t) = \Phi(t) \cdot (\mathbf{x} + \mathbf{n})$$

where $n_1 \dots n_6$ are independent, with $n_k = \pm 1$ with equal probability.

The transmitted waveform $y(t)$ is demodulated using the basis detector of Figure ???. This problem examines the signal-to-noise ratio of the demodulated vector $\mathbf{y} = \mathbf{x} + \mathbf{n}$ with $\sigma^2 \triangleq E(n^2)$

- Find $\bar{\mathcal{E}}_{\mathbf{x}}$, σ^2 , and SNR, $\bar{\mathcal{E}}_{\mathbf{x}}/\sigma^2$ if all messages are equally likely. (2 pts)
- Find the minimal number of basis vectors and new matrix \hat{A} as in Problem 1.4, and calculate the new $\bar{\mathcal{E}}_{\mathbf{x}}$, σ^2 , and SNR. (4 points)
- Let the new vector be $\tilde{\mathbf{y}} = \tilde{\mathbf{x}} + \tilde{\mathbf{n}}$, and discuss if the conversion from \mathbf{y} to $\tilde{\mathbf{y}}$ is invariant (namely, if P_e is affected by the conversion matrix). Compare the detectors for parts a and b. (1 points)
- Compare \bar{b} , $\bar{\mathcal{E}}_{\mathbf{x}}$ with the previous system. Is the new system superior? Why or why not? (2 pts)
- The new system now has three unused dimensions, we would like to send 8 more messages by constructing a big matrix \bar{A} , as follows:

$$\bar{A} = \begin{bmatrix} \hat{A} & \mathbf{0} \\ \mathbf{0} & \hat{A} \end{bmatrix}$$

Compare \bar{b} , $\bar{\mathcal{E}}_{\mathbf{x}}$ with the original 6-dimensional system, and the 3-dimensional system in b). (4 pts)

1.10 Tilt. Consider the signal set shown in Figure 1.52 with an AWGN channel and let $\sigma^2 = 0.1$.

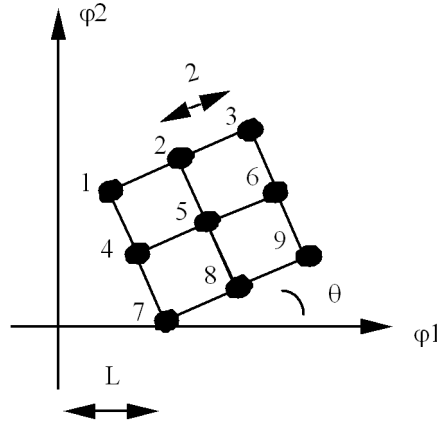


Figure 1.52: A Signal Constellation

- Does P_e depend on L and θ ? (1 pt)
- Find the nearest neighbor union bound on P_e for the ML detector assuming $p_{\mathbf{x}}(i) = \frac{1}{9} \forall i$. (2 pts)
- Find P_e exactly using the assumptions of the previous part. How far off was the NNUB? (5 pts)
- Suppose we have a minimum energy constraint on the signal constellation. How would we change the constellation of this problem without changing the P_e ? How does θ affect the constellation energy? (2 pts)

1.11 Parseval. Consider binary signaling on an AWGN $\sigma^2 = 0.04$ with ML detection for the following signal set. (Hint: consider various ways of computing d_{\min} .)

$$\begin{aligned} x_0(t) &= \text{sinc}^2(t) \\ x_1(t) &= \sqrt{2}\text{sinc}^2(t) \cos(4\pi t) \end{aligned}$$

Determine the exact P_e assuming that the two input signals are equally likely. (5 pts)

1.12 Disk storage channel.

Binary data storage with a thin-film disk can be approximated by an input-dependent additive white Gaussian noise channel where the noise n has a variance dependent on the transmitted (stored) input. The noise has the following input dependent density:

$$p(n) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{n^2}{2\sigma_1^2}} & \text{if } x = 1 \\ \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-\frac{n^2}{2\sigma_0^2}} & \text{if } x = 0 \end{cases}$$

and $\sigma_1^2 = 31\sigma_0^2$. The channel inputs are equally-likely.

- For either input, the output can take on any real value. On the same graph, plot the two possible output probability density functions (pdf's). i.e. Plot the output pdf for $x = 0$ and the output pdf for $x = 1$. Indicate (qualitatively) the decision regions on your graph. (2 pts)

- b. Determine the optimal receiver in terms of σ_1 and σ_0 . (3 pts)
- c. Find σ_0^2 and σ_1^2 if the SNR is 15 dB. SNR is defined as $\frac{\mathcal{E}_x}{\frac{1}{2}(\sigma_0^2 + \sigma_1^2)} = \frac{1}{\sigma_0^2 + \sigma_1^2}$. (1 pt)
- d. Determine P_e when SNR = 15 dB. (3 pts)
- e. What happens as $\frac{\sigma_0^2}{\sigma_1^2} \rightarrow 0$? You may restrict your attention to the physically reasonable case where σ_1 is a fixed finite value and $\sigma_0 \rightarrow 0$. (1 pt)

1.13 *Rotation with correlated noise.*

A two dimensional vector channel $y = x + n$ has *correlated* gaussian noise (that is the noise is not white and so not independent in each dimension) such that $E[n_1] = E[n_2] = 0$, $E[n_1^2] = E[n_2^2] = 0.1$, and $E[n_1 n_2] = 0.05$. n_1 is along the horizontal axis and n_2 is along the vertical axis.

- a. Suppose we use the constellation below with $\theta = 45^\circ$ and $d = \sqrt{2}$. (i.e. $x_1 = (1, 1)$ and $x_2 = (-1, 1)$) Find the mean and mean square values of the noise projected on the line connecting the two constellation points. Note this value more generally is a function of θ when noise is not white. (2 pts)

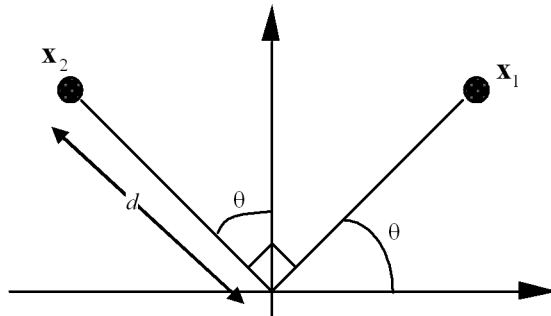


Figure 1.53: Constellation

- b. Note that the noise projected on the line in the previous part is Gaussian. Find P_e for the ML detector. Assume your detector was designed for uncorrelated noise. (2 pts)
- c. Fixing $d = \sqrt{2}$, find θ to minimize the ML detector P_e and give the corresponding P_e . You may continue to assume that the receiver is designed for uncorrelated noise. (2 pts)
- d. Could your detector in part a be improved by taking advantage of the fact that the noise is correlated? (1 pt)

1.14 *Hybrid QAM.* Consider the 64 QAM constellation with $d=2$ (see Figure 1.54.): The 32 hybrid QAM (\times) is obtained by taking one of two points of the constellation. This problem investigates the properties of such a constellation. Assume all points are equally likely and the channel is an AWGN.

- a. Compute the energy \mathcal{E}_x of the 64 QAM and the 32 hybrid QAM constellations. Compare your results. (2 pts)
- b. Find the NNUB for the probability of error for the 64 QAM and 32 hybrid QAM constellations. Which has lower P_e ? Why? (3 pts)
- c. What is d_{min} for a 32 Cross QAM constellation having the same energy? (1 pt)
- d. Find the NNUB for the probability of error for the 32 Cross QAM constellation. Compare with the 32 hybrid QAM constellation. Which one performs better? Why? (2 pts)

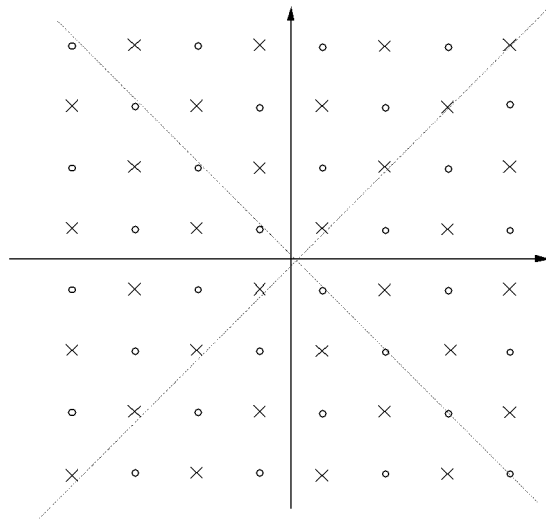


Figure 1.54: 32 SQ embedded in 64 SQ QAM Constellation

- e. Compute the figure of merit for both 32 QAM constellations. Is your result consistent with the one of (d) ? (2 pts)

1.15 Ternary Amplitude Modulation.

Consider the general case of the 3-D TAM constellation for which the data symbols are,

$$(x_l, x_m, x_n) = \left(\frac{d}{2}(2l - 1 - M^{\frac{1}{3}}), \frac{d}{2}(2m - 1 - M^{\frac{1}{3}}), \frac{d}{2}(2n - 1 - M^{\frac{1}{3}}) \right)$$

with $l = 1, 2, \dots, M^{\frac{1}{3}}, m = 1, 2, \dots, M^{\frac{1}{3}}, n = 1, 2, \dots, M^{\frac{1}{3}}$. Assume that $M^{\frac{1}{3}}$ is an even integer.

- a. Show that the energy of this constellation is (2 pts)

$$\mathcal{E}_{\mathbf{x}} = \frac{1}{M} \left[3M^{\frac{2}{3}} \sum_{l=1}^{M^{\frac{1}{3}}} x_l^2 \right] \quad (1.329)$$

- b. Now show that (3 pts)

$$\mathcal{E}_{\mathbf{x}} = \frac{d^2}{4}(M^{\frac{2}{3}} - 1)$$

- c. Assuming an AWGN channel with variance σ^2 , find the NNUB for P_e and \bar{P}_e . (3 pts)
d. Find b and \bar{b} . (1 pt)
e. Find $\bar{\mathcal{E}}_{\mathbf{x}}$ and the energy per bit \mathcal{E}_b . (1 pt)
f. For an equal number of bits per dimension $\bar{b} = \frac{b}{N}$, find the figure of merit for PAM, QAM and TAM constellations with appropriate sizes of M . Compare your results. (2 pts)

1.16 Equivalency of rectangular-lattice constellations.

Consider an AWGN system with a SNR = $\frac{\mathcal{E}_{\mathbf{x}}}{\sigma^2}$ of 22 dB, a target probability of error $P_e = 10^{-6}$, and a symbol rate $\frac{1}{T} = 8$ KHz.

- a. Find the maximum data rate $R = \frac{b}{T}$ that can be transmitted for
- (i) PAM ($\frac{1}{2}$ pt)
 - (ii) QAM ($\frac{1}{2}$ pt)
 - (iii) TAM ($\frac{1}{2}$ pt)
- b. What is the NNUB normalized probability of error \overline{P}_e for the systems used in (a). (1 $\frac{1}{2}$ pts)
- c. For the rest of the problem we will only consider QAM systems. Suppose that the desired data rate is 40 Kbps. What is the transmit power needed to maintain the same probability of error ? The SNR is no longer given as 22 dB. (2 pts)
- d. Suppose now that the SNR was increased to 28 dB. What is the highest data rate that can be reliably sent at the same probability of error 10^{-6} ? (1 pt)

1.17 Frequency separation in FSK. (Adapted from *Wozencraft & Jacobs.*)

Consider the following two signals used in a Frequency Shift Key communications system over an AWGN channel.

$$x_0(t) = \begin{cases} \sqrt{\frac{2\mathcal{E}_x}{T}} \cos(2\pi f_0 t) & \text{if } t \in [0, T] \\ 0 & \text{otherwise} \end{cases}$$

$$x_1(t) = \begin{cases} \sqrt{\frac{2\mathcal{E}_x}{T}} \cos(2\pi(f_0 + \Delta)t) & \text{if } t \in [0, T] \\ 0 & \text{otherwise} \end{cases}$$

$$T = 100\mu s \quad f_0 = 10^5 \text{ Hz} \quad \sigma^2 = 0.01 \quad \mathcal{E}_x = 0.32$$

- a. Find P_e if $\Delta = 10^4$. (2 pts)
- b. Find the smallest $|\Delta|$ such that the same P_e found in part (a) is maintained. What type of constellation is this ? (3 pts)

1.18 Pattern Recognition.

In this problem a simple pattern recognition scheme, based on optimum detectors is investigated. The patterns considered consist of a square divided into four smaller squares, as shown in Figure 1.55,

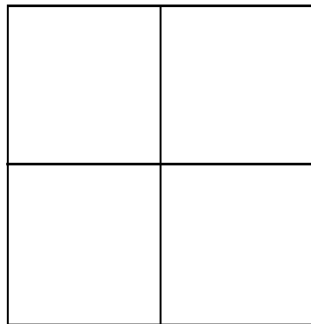


Figure 1.55: Sample pattern

Each square may have two possible intensities, black or white. The class of patterns studied will consist of those having two black squares, and two white squares. For example, some of these patterns are as shown in Figure 1.56,

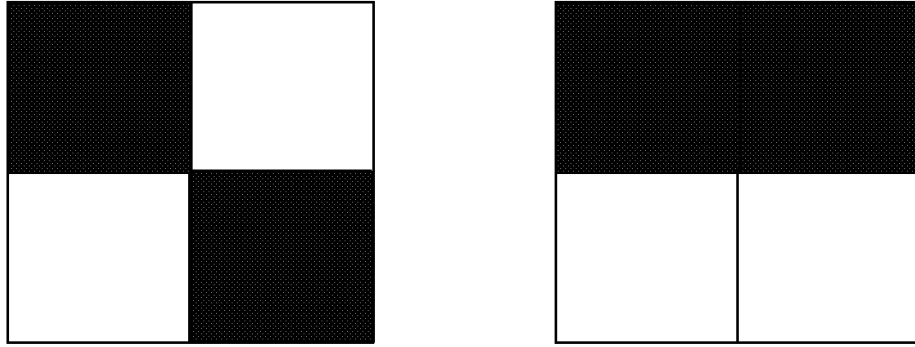


Figure 1.56: Examples of patterns considered

Each pattern can be encoded into a vector $\mathbf{x} = [x_1 \ x_2 \ x_3 \ x_4]$ where each component indicates the ‘intensity’ of a small square according to the following rule,

$$\begin{aligned} \text{Black square} &\Leftrightarrow x_i = 1 \\ \text{White square} &\Leftrightarrow x_i = -1 \end{aligned}$$

For a given pattern, a set of four sensors take measurements at the center of each small square and outputs $\mathbf{y} = [y_1 \ y_2 \ y_3 \ y_4]$,

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{1.330}$$

Where $\mathbf{n} = [n_1 \ n_2 \ n_3 \ n_4]$ is thermal noise (White Gaussian noise) introduced by the sensors. The goal of the problem is to minimize the probability of error for this particular case of pattern recognition.

- What is the total number of possible patterns ? (1 pt)
- Write the optimum decision rule for deciding which pattern is being observed. Draw the corresponding signal detector. Assume each pattern is equally likely. (3 pts)
- Find the union bound for the probability of error P_e . (2 pts)
- Assuming that nearest neighbours are at minimum distance, find the NNUB for the probability of error P_e . (2 pts)

1.19 Shaping Gain.

Find the shaping gain for the following two dimensional voronoi regions (decision regions) relative to the square voronoi region. Do this using the continuous approximation for a continuous uniform distribution of energy through the region.

- equilateral triangle (2 pts)
- regular hexagon (2 pts)
- circle (2 pts)

1.20 (*From Wozencraft and Jacobs*). On an additive white Gaussian noise channel, determine P_e for the following signal set with ML detection. Leave the answer in terms of σ^2 . (Hint: Plot the signals and then the signal vectors.)

$$\begin{aligned} x_1(t) &= \begin{cases} 1 & \text{if } t \in [0, 1] \\ 0 & \text{otherwise} \end{cases} \\ x_2(t) &= \begin{cases} 1 & \text{if } t \in [1, 2] \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

$$\begin{aligned}
x_3(t) &= \begin{cases} 1 & \text{if } t \in [0, 2] \\ 0 & \text{otherwise} \end{cases} \\
x_4(t) &= \begin{cases} 1 & \text{if } t \in [2, 3] \\ 0 & \text{otherwise} \end{cases} \\
x_5(t) &= \begin{cases} 1 & \text{if } t \in [0, 1] \\ 1 & \text{if } t \in [2, 3] \\ 0 & \text{otherwise} \end{cases} \\
x_6(t) &= \begin{cases} 1 & \text{if } t \in [1, 3] \\ 0 & \text{otherwise} \end{cases} \\
x_7(t) &= \begin{cases} 1 & \text{if } t \in [0, 3] \\ 0 & \text{otherwise} \end{cases} \\
x_8(t) &= 0
\end{aligned}$$

1.21 Comparing bounds. Consider the following signal constellation in use on an AWGN channel.

$$\begin{aligned}
\mathbf{x}_0 &= (-1, -1) \\
\mathbf{x}_1 &= (1, -1) \\
\mathbf{x}_2 &= (-1, 1) \\
\mathbf{x}_3 &= (1, 1) \\
\mathbf{x}_4 &= (0, 3)
\end{aligned}$$

Leave answers for parts a and b in terms of σ .

- Find the union bound on P_e for the ML detector on this signal constellation.
- Find the Nearest Neighbor Union Bound on P_e for the ML detector on this signal constellation.
- Let the SNR = 14 dB and determine a numerical value for P_e using the NNUB.

1.22 Basic QAM Design - Midterm 1996 Either square or cross QAM can be used on an AWGN channel with SNR = 30.2 dB and symbol rate $1/T = 10^6$.

- Select a QAM constellation and specify a corresponding integer number of bits per symbol, b , for a modem with the highest data rate such that $P_e < 10^{-6}$.
- Compute the data rate for part a.
- Repeat part a if $P_e < 2 \times 10^{-7}$ is the new probability of error constraint.
- Compute the data rate for part c.

1.23 Basic Detection - One shot or Two? - Final 1996

A 2B1Q signal with $d = 2$ is sent two times in immediate succession through an AWGN channel with transmit filter $p(t)$, which is a scaled version of the basis function. All other symbol times, a symbol value of zero is sent. The symbol period for one of the 2B1Q transmissions is $T = 1$, and the transmit filter is $p(t) = 1$ for $0 < t < 2$ and $p(t) = 0$ elsewhere. At both symbol periods, any one of the 4 messages is equally likely, and the two successive messages are independent. The WGN has power spectral density $\frac{N_0}{2} = .5$.

- Draw an optimum (ML) basis detector and enumerate a signal constellation. (Hint: use basis functions.) (3 pts)
- Find d_{\min} . (2 pts)
- Compute \tilde{N}_e counting only those neighbors that are d_{\min} away. (2pts)
- Approximate P_e for your detector. (3 pts)

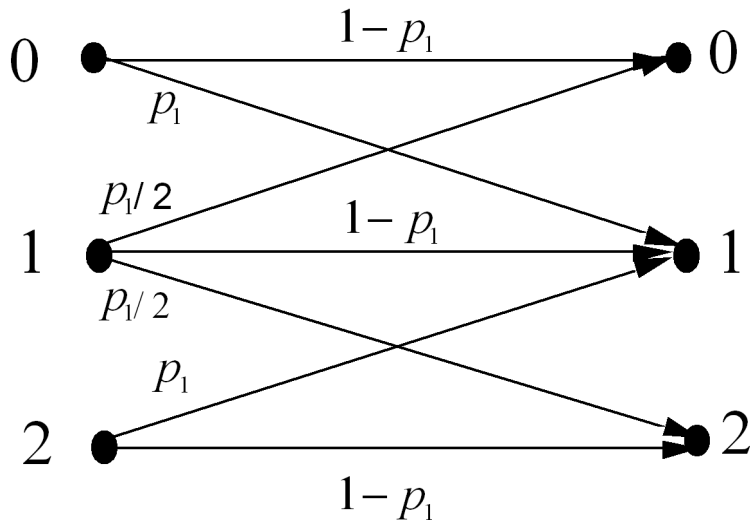


Figure 1.57: Discrete Memoryless Channel

1.24 *Discrete Memoryless Channel - Midterm 1994*

Given a channel with $p_{\mathbf{y}|\mathbf{x}}$ as shown in Figure 1.57: ($y \in \{0, 1, 2\}$ and $x \in \{0, 1, 2\}$) Let $p_1 = .05$

- For $p_{\mathbf{x}}(i) = 1/3$, find the optimum detection rule.
- Find P_e for part a.
- Find P_e for the MAP detector if $p_{\mathbf{x}}(0) = p_{\mathbf{x}}(1) = 1/6$ and $p_{\mathbf{x}}(2) = 2/3$.

1.25 *Detection with Uniform Noise - Midterm 1995*

A one-dimensional additive noise channel, $y = x + n$, has uniform noise distribution

$$p_n(v) = \begin{cases} \frac{1}{L} & |v| \leq \frac{L}{2} \\ 0 & |v| > \frac{L}{2} \end{cases}$$

where $L/2$ is the maximum noise magnitude. The input x has binary antipodal constellation with equally likely input values $x = \pm 1$. The noise is independent of x .

- Design an optimum detector (showing decision regions is sufficient.) (2 pts)
- For what value of L is $P_e < 10^{-6}$? (1 pt)
- Find the SNR (function of L). (2 pts)
- Find the minimum SNR that ensures error-free transmission. (2 pts)
- Repeat part d if 4-level PAM is used instead. (2 pts.)

1.26 *Can you design or just use formulae? - Midterm 1995*

32 CR QAM modulation is used for transmission on an AWGN with $\frac{N_0}{2} = .001$. The symbol rate is $1/T = 400kHz$.

- Find the data rate R .
- What SNR is required for $P_e < 10^{-7}$? (ignore N_e).

- c. In actual transmitter design, the analog filter rarely is normalized and has some gain/attenuation, unlike a basis function. Thus, the average power in the constellation is calibrated to the actual power measured at the analog input to the channel. Suppose $\bar{\mathcal{E}}_x = 1$ corresponds to 0 dBm (1 milliwatt), then what is the power of the signals entering the transmission channel for the 32CR in this problem with $P_e < 10^{-7}$?
- d. *the engineer under stress.* Without increasing transmit power or changing $\frac{N_0}{2} = .001$, design a QAM system that achieves the same P_e at 3.2 Mbps on this same AWGN.

1.27 QAM Design - Midterm 1997

A QAM system with symbol rate $1/T=10$ MHz operates on an AWGN channel. The SNR is 24.5 dB and a $P_e < 10^{-6}$ is desired.

- a. Find the largest constellation with integer b for which $P_e < 10^{-6}$. (2 pts)
- b. What is the data rate for your design in part a? (2 pts)
- c. How much more transmit power is required (with fixed symbol rate at 10 MHz) in dB for the data rate to be increased to 60 Mbps? ($P_e < 10^{-6}$) (2 pts)
- d. With SNR = 24 dB, an reduced-rate alternative mode is enabled to accommodate up to 9 dB margin or temporary increases in the white noise amplitude. What is the data rate in this alternative 9dB-margin mode at the same $P_e < 10^{-6}$? (2 pts)
- e. What is the largest QAM (with integer b) data rate that can be achieved with the same power, \mathcal{E}_x/T , as in part d, but with $1/T$ possibly altered? (2 pts)

1.28 Basic Detection.

A vector equivalent to a channel leads to the one-dimensional real system with $y = x + n$ where n is exponentially distributed with probability density function

$$p_n(u) = \frac{1}{\sigma\sqrt{2}} \cdot e^{-\sqrt{2}\frac{|u|}{\sigma}} \text{ for all } u \quad (1.331)$$

with zero mean and variance σ^2 . This system uses binary antipodal signaling (with equally likely inputs) with distance d between the points. We define a function

$$\tilde{Q}(x) = \begin{cases} \int_x^\infty \frac{1}{\sqrt{2}} e^{-\sqrt{2}u} du = \frac{1}{2} \cdot e^{-\sqrt{2}x} & \text{for } x \geq 0 \\ 1 - \int_{|x|}^\infty \frac{1}{\sqrt{2}} e^{-\sqrt{2}u} du = 1 - \frac{1}{2} \cdot e^{-\sqrt{2}|x|} & \text{for } x \leq 0 \end{cases} \quad (1.332)$$

- a. Find the values $\tilde{Q}(-\infty)$, $\tilde{Q}(0)$, $\tilde{Q}(\infty)$, $\tilde{Q}(\sqrt{10})$. (2 pts)
- b. For what x is $\tilde{Q}(x) = 10^{-6}$? (1 pt)
- c. Find an expression for the probability of symbol error P_e in terms of d , σ , and the function \tilde{Q} . (2 pts)
- d. Defining the SNR as $\text{SNR} = \frac{\bar{\mathcal{E}}_x}{\sigma^2}$, find a new expression for P_e in terms of \tilde{Q} and this SNR. (2 pts)
- e. Find a general expression relating P_e to SNR, M , and \tilde{Q} for PAM transmission. (2 pts)
- f. What SNR is required for transmission at $\bar{b} = 1, 2, \text{ and } 3$ when $P_e = 10^{-6}$? (2 pts)
- g. Would you prefer Gaussian or exponential noise if you had a choice? (1 pt)

1.29 QAM Design - Midterm 1998

QAM transmission is to be used on an AWGN channel with SNR=27.5 dB at a symbol rate of $1/T = 5$ MHz used throughout this problem. You've been hired to design the transmission system. The desired probability of symbol error is $\bar{P}_e \leq 10^{-6}$.

- (2 pts) List two basis functions that you would use for modulation.
- (2 pts) Estimate the highest bit rate, \bar{b} , and data rate, R , that can be achieved with QAM with your design.
- (1 pt) What signal constellation are you using?
- (3 pts) By about how much (in dB) would $\bar{\mathcal{E}}_{\mathbf{x}}$ need to be increased to have 5 Mbps more data rate at the same probability of error? Does your answer change for $\mathcal{E}_{\mathbf{x}}$ or for $P_{\mathbf{x}}$?

1.30 Basic Detection - Midterm 2000 - 10 pts

QAM transmission is used on an AWGN channel with $\frac{N_0}{2} = .01$. The transmitted signal constellation points for the QAM signal are given by $\begin{bmatrix} \pm\frac{\sqrt{3}}{2} \\ \pm\frac{1}{2} \end{bmatrix}$, $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, and $\begin{bmatrix} 0 \\ \pm 1 \end{bmatrix}$, with each constellation point equally likely.

- (1 pt) Find M (message-set size) and $\bar{\mathcal{E}}_{\mathbf{x}}$ (energy per dimension) for this constellation.
- (2 pts) Draw the constellation with decision regions indicated for an ML detector.
- (2 pts) Find N_e and d_{\min} for this constellation.
- (2 pts) Compute a NNUB value for \bar{P}_e for the ML detector of part b.
- (1 pt) Determine \bar{b} for this constellation (value may be non-integer).
- (2 pts) For the same \bar{b} as part e, how much better in decibels is the constellation of this problem than SQ QAM?

1.31 Basic Detection - Midterm 2001 - 7 pts

The QAM “radial” constellation in Figure 1.58 is used for transmission on an AWGN with $\sigma^2 = .05$. All constellation points are equally likely.

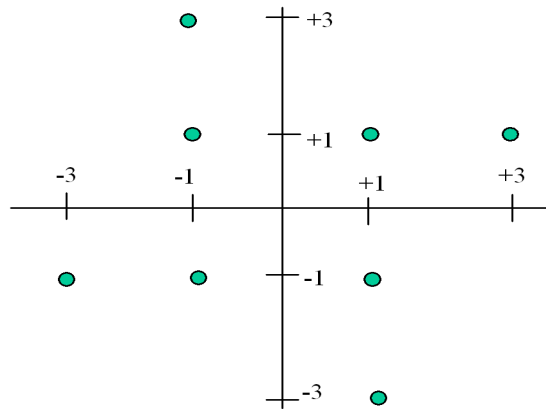


Figure 1.58: Constellation for 8 points.

- (2pts) Find \mathcal{E}_x and $\bar{\mathcal{E}}_{\mathbf{x}}$ for this constellation.
- (3 pts) Find \bar{b} , d_{\min} , and N_e for this constellation.
- (2 pts) Find P_e and \bar{P}_e with the NNUB for an ML detector with this constellation.

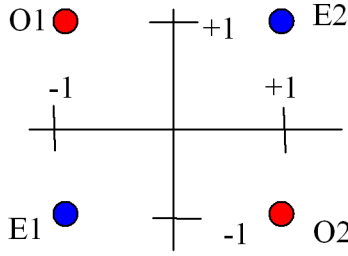


Figure 1.59: Constellation for Problem 1.32.

1.32 *A concatenated QAM Constellation - 2002 Miterm - 15 pts*

A set of 4 orthogonal basis functions $\{\varphi_1(t), \varphi_2(t), \varphi_3(t), \varphi_4(t)\}$ uses the following constellation in both the first 2 dimensions and again in the 2nd two dimensions: The constellation points are restricted in that an E ("even") point may only follow an E point, and an O ("odd") point can only follow an O point. For instance, the 4-dimensional point $[+1 +1 -1 -1]$ is permitted to occur, but the point $[+1 +1 -1 +1]$ cannot occur.

- (2 pts) Enumerate all M points as ordered-4-tuples.
- (3 pts) Find b, \bar{b} , and the number of bits/Hz or bps/Hz.
- (1 pt) Find $\mathcal{E}_{\mathbf{x}}$ and $\bar{\mathcal{E}}_{\mathbf{x}}$ (energy per dimension) for this constellation.
- (2 pts) Find d_{\min} for this constellation.
- (2 pts) Find N_e and \bar{N}_e for this constellation (you may elect to include only points at minimum distance in computing nearest neighbors).
- (2 pts) Find P_e and \bar{P}_e for this constellation using the NNUB if used on an AWGN with $\sigma^2 = 0.1$.
- (3 pts) Compare this 4-dimensional constellation fairly (which requires increasing the number of points in the constellation to 6 to get the same data rate). 4QAM.

1.33 *Detection Fundamentals - 2003 Miterm - 15 pts*

A random variable x_1 takes the 2 values ± 1 with equal probability independently of a second random variable x_2 that takes the values ± 2 also with equal probability. The two random variables are summed to $x = x_1 + x_2$, and x can only be observed after zero-mean Gaussian noise of variance $\sigma^2 = .1$ is added, that is $y = x + n$ is observed where n is the noise.

- (1 pt) What are the values that the discrete random variable x takes, and what are their probabilities? (1 pt)
- (1 pt) What are the means and variances of x and y ?
- (2 pts) What is the lowest probability of error in detecting x given only an observation of y ? Draw corresponding decision regions.
- (1 pt) Relate the value of x with a table to the values of x_1 and x_2 . Explain why this is called a "noisy DAC" channel.
- (1 pt) What is the (approximate) lowest probability of error in detecting x_1 given only an observation of y ?
- (1 pt) What is the (approximate) lowest probability of error in detecting x_2 given only an observation of y ?

- g. Suppose additional binary independent random variables are added so that the two bipolar values for x_u are $\pm 2^{u-1}$, $u = 1, \dots, U$. Which x_u has lowest probability of error for any AWG noise, and what is that P_e ? (1 pt)
- h. For $U = 2$, what is the lowest probability of error in detecting x_1 given an observation of y and a correct observation of x_2 ? (1 pt)
- i. For $U = 2$, what is the lowest probability of error in detecting x_2 given an observation of y and a correct observation of x_1 ? (1 pt)
- j. What is the lowest probability of error in any of parts e through i if $\sigma^2 = 0$? What does this mean in terms of the DAC? (1 pt)
- k. Derive a general expression for the probability of error for all bits $u = 1, \dots, U$ where $x = x_1 + x_2 + \dots + x_U$ in AWGN with variance σ^2 for part g? (2 pts)

1.34 Honey Comb QAM - 2005 Miterm - 15 pts

The QAM constellation below is used for transmission on an AWGN with symbol rate 10MHz and a carrier frequency of 100 MHz.

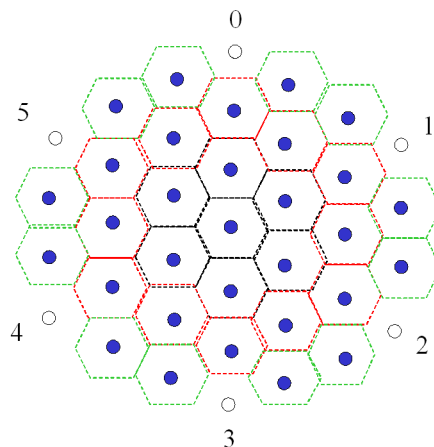


Figure 1.60: Constellation for Problem 1.34.

Each of the solid constellation symbol possibilities is at the center of a perfect hexagon (all sides are equal) and the distance to any of the closest sides of the hexagon is $\frac{d}{2}$. The 6 empty points represent a possible message also, but each is used only every 6 symbol instants, so that for instance, the point labelled 0 is a potential message only on symbol instants that are integer multiples of 6. The 1 point can only be transmitted on symbol instants that are integer multiples of 6 plus one, the 2 point only on symbol instants that are integer multiples of 6 plus two, and so on. At any symbol instant, any of the points possible on that symbol are equally likely.

- a. What is the number of messages that can be possibly transmitted on any single symbol? What are b and \bar{b} ? (3 pts)
- b. What is the data rate? (1 pt)
- c. Draw the decision boundaries for time 0 of a ML receiver. (2 pts)
- d. What is d_{\min} ? (1 pt)

- e. What are $\mathcal{E}_{\mathbf{x}}$ and $\bar{\mathcal{E}}_{\mathbf{x}}$ for this constellation in terms of d ? (3 pts)
- f. What is the average number nearest neighbors? (1 pt)
- g. Determine the NNUB expression that tightly upper bounds \bar{P}_e for this constellation in terms of SNR. (2 pts)
- h. Compare this constellation fairly to Cross QAM transmission. (1 pt)
- i. Describe an equivalent ML receiver that uses time-invariant decision boundaries and a constant decision device with a simple preprocessor to the decision device. (1 pt).

Appendix A

Gram-Schmidt Orthonormalization Procedure

This appendix illustrates the construction of a set of orthonormal basis functions $\varphi_n(t)$ from a set of modulated waveforms $\{x_i(t), i = 0, \dots, M-1\}$. The process for doing so, and achieving minimal dimensionality is called **Gram-Schmidt Orthonormalization**.

Step 1:

Find a signal in the set of modulated waveforms with nonzero energy and call it $x_0(t)$. Let

$$\varphi_1(t) \triangleq \frac{x_0(t)}{\sqrt{\mathcal{E}_{x_0}}} \quad , \quad (\text{A.1})$$

where $\mathcal{E}_{x_0} = \int_{-\infty}^{\infty} [x_0(t)]^2 dt$. Then $\mathbf{x}_0 = [\sqrt{\mathcal{E}_{x_0}} \ 0 \ \dots \ 0]$.

Step i for $i = 2, \dots, M$:

- Compute $x_{i-1,n}$ for $n = 1, \dots, i-1$ ($x_{i-1,n} \triangleq \int_{-\infty}^{\infty} x_{i-1}(t)\varphi_n(t)dt$).
- Compute

$$\theta_i(t) \triangleq x_{i-1}(t) - \sum_{n=1}^{i-1} x_{i-1,n}\varphi_n(t) \quad (\text{A.2})$$

- if $\theta_i(t) = 0$, then $\varphi_i(t) = 0$, skip to step $i+1$.
- If $\theta_i(t) \neq 0$, compute

$$\varphi_i(t) = \frac{\theta_i(t)}{\sqrt{\mathcal{E}_{\theta_i}}} \quad , \quad (\text{A.3})$$

where $\mathcal{E}_{\theta_i} = \int_{-\infty}^{\infty} [\theta_i(t)]^2 dt$. Then $\mathbf{x}_{i-1} = [x_{i-1,1} \ \dots \ x_{i-1,i-1} \ \sqrt{\mathcal{E}_{\theta_i}} \ 0 \ \dots \ 0]'$.

Final Step:

Delete all components, n , for which $\varphi_n(t) = 0$ to achieve minimum dimensional basis function set, and reorder indices appropriately.

Appendix B

The Q Function

The **Q Function** is used to evaluate probability of error in digital communication - It is the integral of a zero-mean unit-variance Gaussian random variable from some specified argument to ∞ :

Definition B.0.1 (Q Function)

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du \quad (\text{B.1})$$

The integral cannot be evaluated in closed form for arbitrary x . Instead, see Figures B.1 and B.2 for a graph of the function that can be used to get numerical values. Note the argument is in dB ($20 \log_{10}(x)$).

Note $Q(-x) = 1 - Q(x)$, so we need only plot $Q(x)$ for positive arguments.

We state without proof the following bounds

$$\left(1 - \frac{1}{x^2}\right) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi x^2}} \leq Q(x) \leq \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi x^2}} \quad (\text{B.2})$$

The upper bound in (B.2) is easily seen to be a very close approximation for $x \geq 3$.

Computation of the probability that a Gaussian random variable u with mean m and variance σ^2 exceeds some value d then uses the Q-function as follows:

$$P\{u \geq d\} = Q\left(\frac{d-m}{\sigma}\right) \quad (\text{B.3})$$

The Q-function appears in Figures B.3, B.1, and B.2 for very low SNR (-10 to 0 dB), low SNR (0 to 10 dB), and high SNR (10 to 16 dB) using a very accurate approximation (less than 1% error) formula from the recent book by Leon-Garcia:

$$Q(x) \approx \left[\frac{1}{\frac{\pi-1}{\pi}x + \frac{1}{\pi}\sqrt{x^2 + 2\pi}} \right] \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \quad (\text{B.4})$$

For the mathematician at heart, $Q(x) = .5 \cdot \text{erfc}(x/\sqrt{2})$, where erfc is known as the complimentary error function by mathematicians.

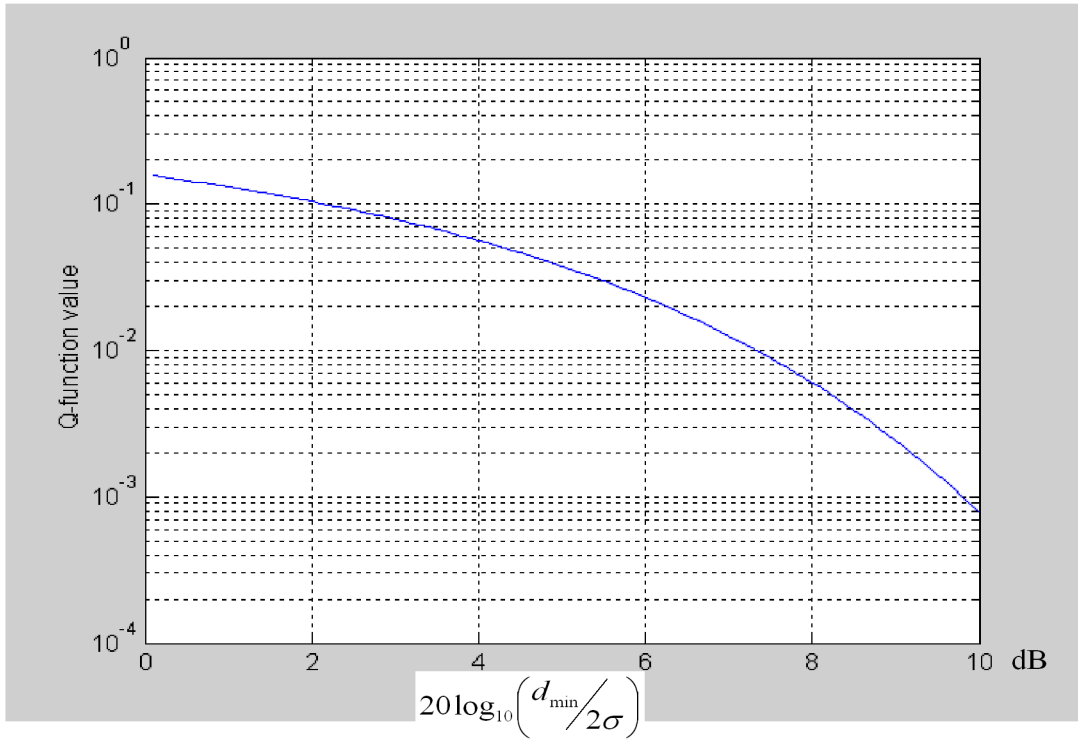


Figure B.1: Low SNR Q-Function Values

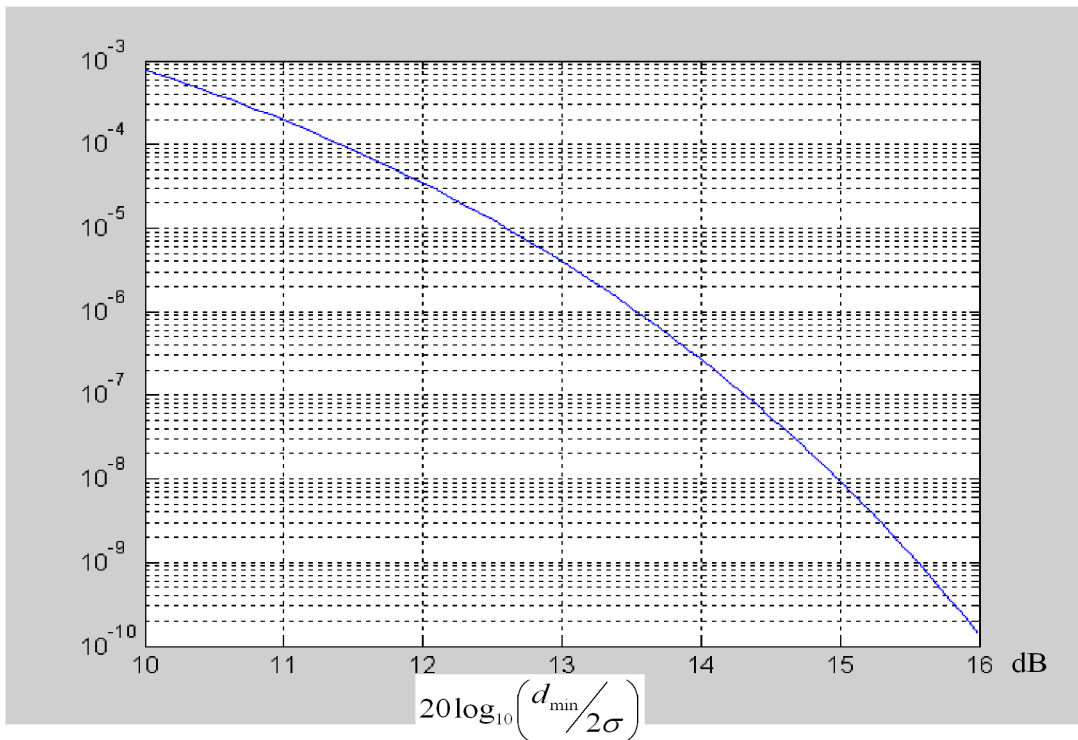


Figure B.2: High SNR Q-Function Values

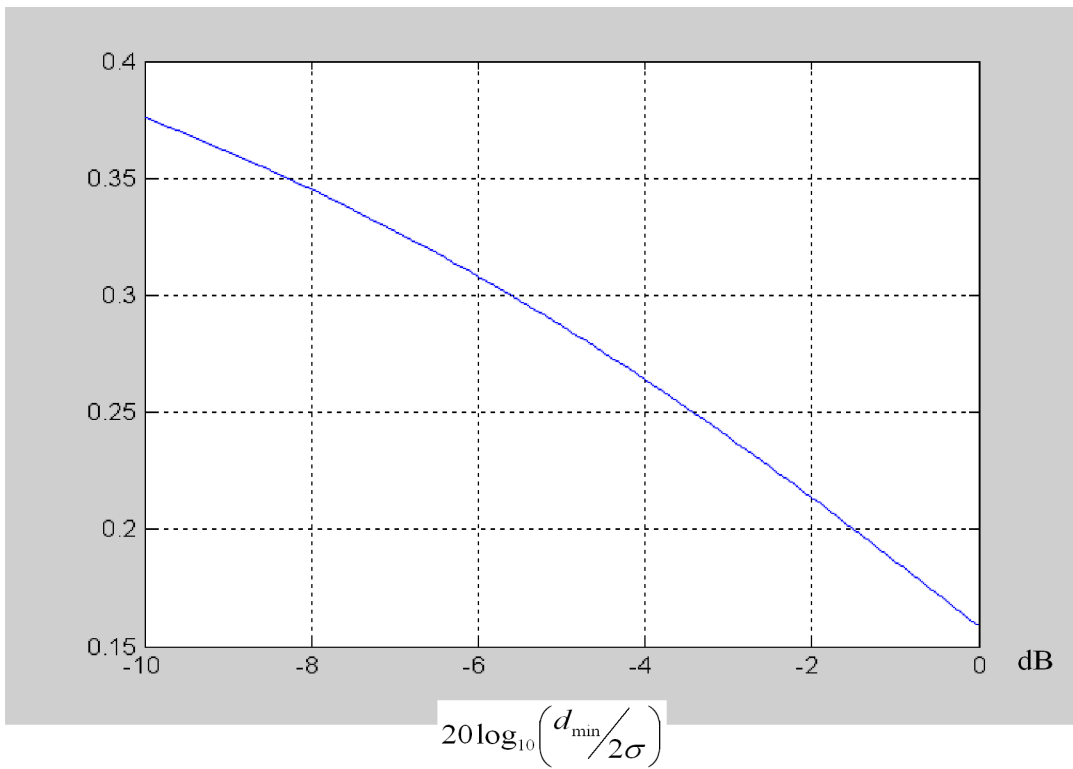


Figure B.3: Very Low SNR Q-Function Values

Bibliography

- [1] J.M. Wozencraft and I.M. Jacobs. *“Principles of Communication Engineering”*. Wiley, New York, 1965.
- [2] B.W. Lindgren. *“Statistical Theory, Third Edition”*. Macmillan, New York, 1968.