

Spread Spectrum Communications

R. Michael Buehrer

Summer 2007

Contents

1	Digital Communications	1
1.1	Introduction	1
1.2	Bandpass Signals	3
1.3	Random Processes	4
1.4	Sampling	6
1.5	Quantization	11
1.6	Digital Modulation	13
1.6.1	Binary Modulation	13
1.6.2	M -ary Modulation	14
1.6.3	Power Spectral Density	14
1.6.4	Pulse Shaping	16
1.7	Optimum Receiver in AWGN	17
1.7.1	The Matched Filter	17
1.7.2	The Correlator Version of the Matched Filter	23
1.7.3	Detection and the Optimal Threshold	23
1.7.4	BER of the Matched Filter with Optimal Detection	28
1.7.5	Performance of General Binary Modulation	34
1.7.6	Non-coherent Demodulation	38
1.7.7	M -ary Modulation	38
1.8	Bandwidth Efficiency and Energy Efficiency	39
1.9	Capacity	41
1.10	Coding	44
1.11	Fading Channels	45
1.12	Conclusion	52
2	Motivation for Spread Spectrum	53
2.1	Introduction	53
2.2	Reasons for Spreading	54
2.2.1	Anti-jamming	55
2.2.2	Interference Randomization	58
2.2.3	LPI/LPD Communications	62
2.2.4	Improved Multiple Access	66
2.2.5	Interference Averaging	68

2.2.6	Resistance to Multipath Fading	71
2.2.7	Ranging	74
2.3	Types of spread spectrum	76
2.4	Conclusions	77

Preface

The following is a (currently) unpublished manuscript on spread spectrum communications. The manuscript is the result of several years of accumulating notes and examples from teaching ECE 5660. I invite you, the student, to provide feedback, corrections and suggestions for ways of improving the text.

Chapter 1

Digital Communications

In this chapter we briefly review digital communications. It is assumed that the reader has a familiarity with digital communication principles. This chapter is intended to simply provide the reader with quick reminders of several key digital communications concepts that will be used in our study of spread spectrum communications. For a more detailed description of general digital communications, the reader is encouraged to refer to one or more of the many fine texts in the area [1, 2, 3, 4, 5].

1.1 Introduction

For the past 25-30 years digital communication systems having been replacing their corresponding analog counterparts. While it has long been known that digital systems offer tremendous advantages over analog systems, only in the last 20 years has technology advanced to the point to allow digital systems to be a cost-effective alternative. Cell-phones, cordless telephones, land-line communications, military communication systems and finally broadcast services have all seen digital technology replace analog. Among the advantages that digital systems offer are

- Increased fidelity through error correction coding
- Mixed information types in a single communication system and even multiplexed in a single data stream
- The opportunity for bandwidth reduction via source coding
- More flexible trade-offs between bandwidth and performance (i.e., fidelity)
- Opportunities to improve performance via advanced signal processing
- The opportunity for encryption

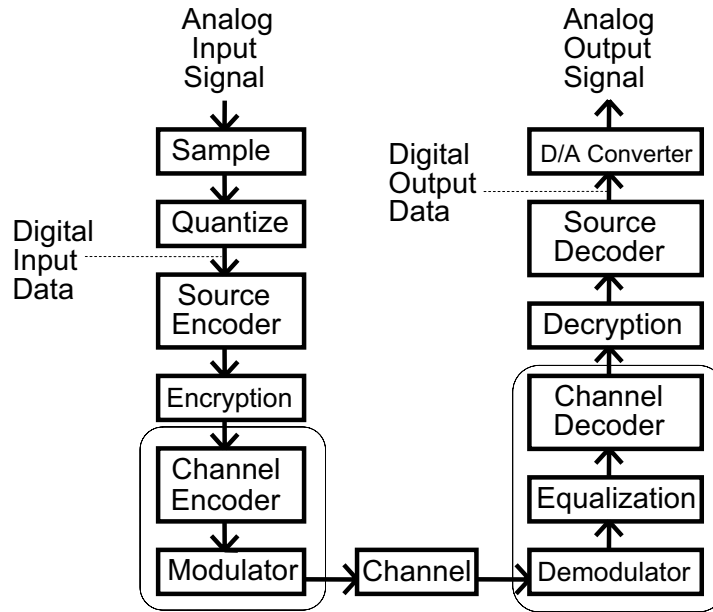


Figure 1.1: Block Diagram of a Typical Digital Communications System

A block diagram of a typical digital communications system is shown in Figure 1.1. Digital Communication systems can transmit either information that is inherently digital (e.g., computer files) or information that is analog but has been converted to a digital format through sampling and quantization. Digital information can also be compressed through *source coding*. Source coding removes redundancy in the information to reduce the requirements for transmission. The source-coded binary data can also be encrypted to provide security. This new information stream can then be further encoded¹ by the *channel coder* using an appropriate forward error correction (FEC) code to provide protection against errors introduced by the *channel*. This operation increases the rate of the binary data by adding redundancy in an intelligent fashion so that it can be exploited to detect and possibly correct bit errors.

The *modulator* accepts the coded bit stream and converts the bits to symbols for transmission using electromagnetic waves. For bandpass transmission (typical for wireless systems) this operation typically involves modulating either the frequency, phase or amplitude of a radio frequency sinusoidal carrier. The resulting signal is launched into the channel which will attenuate and possibly distort the signal. The most benign channel will attenuate the signal based on the distance traveled in a manner which is constant in time and frequency. In many systems, however, the channel may also have a limited bandwidth,

¹The term "coding": is unfortunately over-used in the field of communications. We will typically use the term encryption for security-oriented codes and reserve the term "code" for a forward error correction code. Source codes will generally not be considered in this text.

which can cause frequency distortion. In wireless systems, mobility combined with multipath can induce a time-varying channel which is typically referred to as a *fading channel* since the time-varying attenuation can be quite severe resulting in a "fade" of the received signal. We will discuss this more in Section 1.11. When multipath components have relative arrival times that are separated by more than the symbol duration, the attenuation experienced can also vary across frequency. Typically this "frequency-selective" fading must be combated through equalization techniques, although spread spectrum systems offer a specific advantage over this type of distortion as we will see. Regardless of the attenuation and distortion effects of the channel, the received signal will also be corrupted by an additive signal due to the combined EM radiation of objects which have non-negligible temperatures. The superposition of these radiation sources is typically modeled as a white Gaussian random process. This type of distortion is termed additive white Gaussian noise or AWGN.

At the receiver, the reverse operations must be performed. After filtering, down-conversion and demodulation, the receiver will typically perform equalization, if necessary, followed by decoding. This is in turn followed by decryption, if encryption is employed, source decoding and, if the information is analog, digital-to-analog conversion.

We will discuss several of these steps in detail shortly. However, before we do so, we will review two concepts that will prove to be quite useful in the analysis of communication systems: bandpass signals and random processes. Note that the coverage here is limited and intended to merely reacquaint the reader with these topics.

1.2 Bandpass Signals

Communication signals typically involve carrier modulation and are thus bandpass signals. However, the use of bandpass signals for analysis is somewhat awkward. Thus, typically we will use the complex baseband representation of bandpass signals. We will briefly describe different representations of bandpass signals in this section. Consider a signal $z(t)$ with center frequency f_o which has non-zero frequency components only in a relatively small region $|f - f_o| \leq B$ where $B \ll f_o$. We define B as the signal *bandwidth*. A signal which has this property is considered a *bandpass signal*. This is to be contrasted with a *baseband signal* which has non-zero frequency components only in the region $|f| \leq B$ where again B is termed the bandwidth of the signal.²

A bandpass signal can be represented as

$$z(t) = z_I(t)\cos(2\pi f_o t) - z_Q(t)\sin(2\pi f_o t) \quad (1.1)$$

where $z_I(t)$ and $z_Q(t)$ are real *baseband* signals and are called the in-phase and quadrature components of $z(t)$ respectively. This representation is termed the

²It should be noted that there are many definitions of bandwidth. This is often termed the *absolute bandwidth*.

in-phase and quadrature representation of bandpass signals and is useful for developing transmitter and receiver structures. However, it isn't particularly intuitive when it comes to understanding modulation. Bandpass signals can also be represented as

$$z(t) = A(t) \cos(2\pi f_o t + \theta(t)) \quad (1.2)$$

where $A(t)$ and $\theta(t)$ are real baseband signals that represent the amplitude and phase modulation respectively. It is relatively straightforward to show that

$$A(t) = \sqrt{z_I^2(t) + z_Q^2(t)} \quad (1.3)$$

$$\theta(t) = \tan^{-1} \left(\frac{z_Q(t)}{z_I(t)} \right) \quad (1.4)$$

This representation is often termed the *amplitude/phase representation* and provides more insight into the carrier modulation of the bandpass signal.

Finally, a bandpass signal can be written in *complex baseband (or complex envelope) representation* as

$$z(t) = \text{Re} \{ \tilde{z}(t) e^{j2\pi f_o t} \} \quad (1.5)$$

where $j = \sqrt{-1}$, $\text{Re} \{ x \}$ is the real part of x and $\tilde{z}(t)$ is a complex baseband signal that can be written as

$$\tilde{z}(t) = z_I(t) + jz_Q(t) \quad (1.6)$$

We will find that this last representation is the most useful in terms of analysis, since it provides us a means for eliminating the awkward carrier component.

1.3 Random Processes

The use of random processes is fundamental to the study of communication systems. In this section we briefly review random processes. The reader is referred to [6, 7, 8] for a more detailed discussion. A random process is a collection of time functions where each time function in the set (termed a sample function) has at least one parameter that is the instantiation of a random variable. Each sample function is one realization of the random process and is one member of the ensemble of sample functions. A random process is described by an n -fold joint distribution function:

$$f_{X(t_1), \dots, X(t_n)}(x_1, \dots, x_n; t_1, \dots, t_n) = P \{ X(t_1) \leq x_1; \dots, X(t_n) \leq x_n \} \quad (1.7)$$

where $x_i = X(t_i)$ is a random variable. If the first order density function $f_{X(t_1)}(x_1; t_1)$ is independent of time t_1 , that is

$$f_{X(t_1)}(x_1; t_1) = f_{X(t_1+\tau)}(x_1; t_1 + \tau) \quad (1.8)$$

for any τ then the random process is called *first order stationary*. A result of first order stationarity is that

$$E\{X(t)\} = \mu_X = \text{constant} \quad (1.9)$$

where $E\{\cdot\}$ is the expectation operator and μ_X is termed the mean. In other words, we may measure the mean of the random process at any time t and obtain the same value. A *strictly stationary* process is a random process for which all n , all (t_1, t_2, \dots, t_n) and all τ ,

$$f_{X(t_1), \dots, X(t_n)}(x_1, \dots, x_n) = f_{X(t_1+\tau), \dots, X(t_n+\tau)}(x_1, \dots, x_n) \quad (1.10)$$

Strict stationarity is not typically required for proper analysis. A more relaxed requirement is wide-sense stationarity. A *wide-sense stationary* process is one for which

- $E\{X(t)\} = E\{X(t + \tau)\} = \mu_X$
- $E\{X(t_1)X(t_2)\} = E\{X(t_1)X(t_1 + \tau)\} = R_X(\tau)$

In this book we will typically assume that random processes are wide-sense stationary.

Another important class of random processes is the class of ergodic random processes. An ergodic random process is one for which the time and ensemble averages are interchangeable. For example

$$E\{X(t)\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t) dt \quad (1.11)$$

and

$$E\{X(t)X(t + \tau)\} = \lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T/2}^{T/2} x(t)x(t + \tau) dt \quad (1.12)$$

Note that for strict ergodicity, all ensemble averages must equal their corresponding time averages, not just the first and second order averages. Further, it should be noted that ergodicity requires stationarity and is thus a more strict requirement than strict stationarity. We will commonly assume ergodicity although it is not strictly necessary in many cases.

Another important class of random processes is the class of Gaussian random processes. A Gaussian random process is a process $X(t)$ for which the random variables $\{X(t_i)\}_{i=1}^n$ (i.e., n time samples of the random process) have a joint Gaussian density function. That is

$$f_X(x_1, \dots, x_n; t_1, \dots, t_n) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{C}_X|}} \exp\left\{-(1/2) [\mathbf{x} - \bar{\mathbf{x}}]^T \mathbf{C}_X^{-1} [\mathbf{x} - \bar{\mathbf{x}}]\right\} \quad (1.13)$$

where

$$[\mathbf{x} - \bar{\mathbf{x}}] = \begin{bmatrix} x_1 - \bar{x}_1 \\ x_2 - \bar{x}_2 \\ \dots \\ x_n - \bar{x}_n \end{bmatrix}, \quad (1.14)$$

\bar{x}_i is the mean of x_i ,

$$\mathbf{C}_X = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix}, \quad (1.15)$$

is the covariance matrix with

$$C_{ij} = E \{ (x_i - \bar{x}_i) (x_j - \bar{x}_j) \} \quad (1.16)$$

and $|\mathbf{C}_X|$ is the determinant of \mathbf{C}_X . The most relevant examples of Gaussian random processes are Additive White Gaussian Noise and the Rayleigh fading channel. Note that a *white* process is a process that has a spectral density $S(f)$ which is constant. In other words based on the Wiener-Khintchine Theorem, the autocorrelation function $R_{xx}(\tau) = \mathcal{F} \{ R_{xx}(\tau) \} = \delta(\tau)$ where $\mathcal{F} \{ x(t) \}$ is the Fourier Transform of $x(t)$.

1.4 Sampling

The analog-to-digital process consists primarily of sampling and quantization. These two procedures convert an analog signal to a signal which is discrete in time and amplitude. These values can then be converted directly to data bits or further processed to compress the information (e.g., in vocoders). Sampling is the process of converting an analog signal to one which is discrete in time. If samples are taken at a uniform rate, it was shown by Nyquist [9] that the original signal can be reconstructed from samples that are taken at a rate f_s provided that

$$f_s \geq 2f_{\max} \quad (1.17)$$

where f_{\max} is the maximum frequency component of the signal of interest. For baseband signals f_{\max} is also termed the bandwidth B of the signal. Thus, we can replace f_{\max} with B . We can also view sampling through the interpolation formula which states that a signal can be reconstructed from samples taken at $f_s = 2B$ as

$$x(t) = \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2B}\right) \frac{\sin(2\pi B(t - n/2B))}{2\pi B(t - n/2B)} \quad (1.18)$$

To understand this, recall that the sampling theorem tells us that we can completely represent a band-limited signal $x(t)$ by samples of the signal $x(nT_s)$

taken at an appropriate frequency $f_s = \frac{1}{T_s}$. Formally, if we assume an impulse-sampled signal we can write the sampled signal as

$$\begin{aligned} x_s(t) &= \sum_{n=-\infty}^{\infty} x(nT_s)\delta(t - nT_s) \\ &= x(t) \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \end{aligned} \quad (1.19)$$

where the impulse is defined as

$$\delta(t) = \begin{cases} \text{undefined} & t = 0 \\ 0 & \text{else} \end{cases} \quad (1.20)$$

and

$$\int_{-\infty}^{\infty} \delta(t) dt = 1. \quad (1.21)$$

Now taking the Fourier transform of (1.19) results in

$$\begin{aligned} X_s(f) &= X(f) * \mathcal{F} \left\{ \sum_{n=-\infty}^{\infty} \delta(t - nT_s) \right\} \\ &= X(f) * \frac{1}{T_s} \sum_{n=-\infty}^{\infty} \delta \left(f - \frac{n}{T_s} \right) \\ &= \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X \left(f - \frac{n}{T_s} \right) \end{aligned} \quad (1.22)$$

where $*$ is the convolution operation. Now we see that sampling in the time domain results in replication in the frequency domain. If $f_s > 2f_{\max}$ where f_{\max} is the highest frequency component of the signal, there will be no overlapping in frequency between the replicas of the original spectrum (i.e., no *aliasing*) and thus no distortion. Now if we lowpass filter the signal *i.e.*, we multiply the spectrum by an ideal lowpass filter with response

$$H(f) = \begin{cases} T_s & |f| < B \\ 0 & \text{else} \end{cases} \quad (1.23)$$

we obtain the original spectrum:

$$\begin{aligned} Y(f) &= H(f)X_s(f) \\ &= X(f) \end{aligned} \quad (1.24)$$

Looking at this in the time domain we have

$$\begin{aligned}
 y(t) &= h(t) * x_s(t) \\
 &= 2BT_s \text{sinc}(2Bt) * \sum_{n=-\infty}^{\infty} x(nT_s) \delta(t - nT_s) \\
 &= \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2B}\right) \text{sinc}(2Bt - n) \\
 &= \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2B}\right) \frac{\sin(2\pi Bt - \pi n)}{2\pi Bt - \pi n} \tag{1.25}
 \end{aligned}$$

$$= \sum_{n=-\infty}^{\infty} x\left(\frac{n}{2B}\right) \frac{\sin(2\pi B(t - n/2B))}{2\pi B(t - n/2B)} \tag{1.26}$$

Thus, if we use the *sinc* function for interpolating the signal, we can reconstruct the signal perfectly. Note that $\text{sinc}(x) = 0$ for integer values of x . If the signal is sampled at the Nyquist rate ($f_s = 2f_{\max}$) an ideal brick wall filter is needed. However, such a filter is impossible to build in practice since it requires an infinite (and non-causal) impulse response. Realistic system filters have finite roll-off and will thus introduce some finite amount of aliasing, although it may be negligible.

Example 1.1 Consider a bandpass signal with center frequency $f_c = 2.4\text{GHz}$ and bandwidth $B_T = 300\text{MHz}$ whose magnitude spectrum is shown in Figure 1.2. (a) Plot the magnitude spectrum of the sampled signal when $f_s = 5.6\text{GHz}$ and characterize the ideal filter needed to recover the original spectrum. (b) Repeat for $f_s = 2\text{GHz}$

SOLUTION: (a) When $f_s = 5.6\text{GHz}$ the sampled spectrum is simply the original spectrum replicated at integer multiples of 5.6GHz . The first two replicas are shown in Figure 1.3. The highest frequency component of the original spectrum is $f_{\max} = f_c + B/2 = 2.55\text{GHz}$. We can see that since the sampling frequency is greater than twice the highest frequency of the signal, i.e., $f_s > 2f_{\max}$, we can recover the desired signal using an ideal lowpass filter with a cut-off frequency between 2.55GHz and 3.15GHz . Alternatively, we can recover the original signal using an ideal bandpass filter centered at 2.4GHz with a bandwidth between 300MHz and 1.5GHz . (b) When $f_s = 2.0\text{GHz}$ the sampled spectrum is given in Figure 1.4 (top). We see that although we cannot recover the original signal using a lowpass filter, we can recover the original signal using an ideal bandpass filter with a center frequency of 2.4GHz and a bandwidth between 300MHz and approximately 1.4GHz .

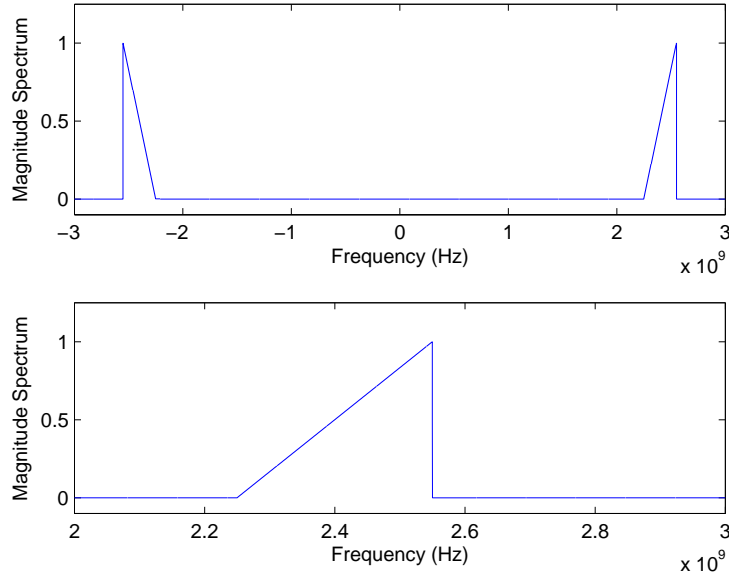


Figure 1.2: Magnitude Spectrum of Example 1.1

In the preceding example, interestingly we found that we could recover the original signal even when the sampling frequency was not higher than the maximum frequency component of the signal. Have we violated Nyquist's Theorem? No. Actually, we found that Nyquist's Theorem, which was originally stated for baseband signals, must be revisited for bandpass signals. In order to understand this better, consider what happens to the sampled spectrum if the sampling rate is slightly increased or decreased. Specifically, consider the spectrum of the sampled signal when the sampling rate is decreased from $2.0GHz$ to $f_s = 1.85GHz$ as shown in Figure 1.4 (middle). Recalling that

$$X_s(f) = \frac{1}{T_s} \sum_{n=-\infty}^{\infty} X(f - nf_s) \quad (1.27)$$

there exists one spectrum replica that corresponds to $n = 0$ whose position does not change with the sampling frequency. However, we can see that the spectrum replica just above the desired (i.e., original) component moves slightly down in frequency. Further, it is clear that as we continue to decrease the sampling frequency this replica will eventually interfere with the original spectrum causing aliasing. Now, if we slightly increase the sampling frequency from $f_s = 2GHz$ to $f_s = 2.15GHz$ the spectrum changes as shown in Figure 1.4 (bottom). Specifically, we can see that the replica slightly below the original spectrum moves up in frequency. Clearly, as we continue to increase the sampling frequency there will be a point at which we will experience aliasing.

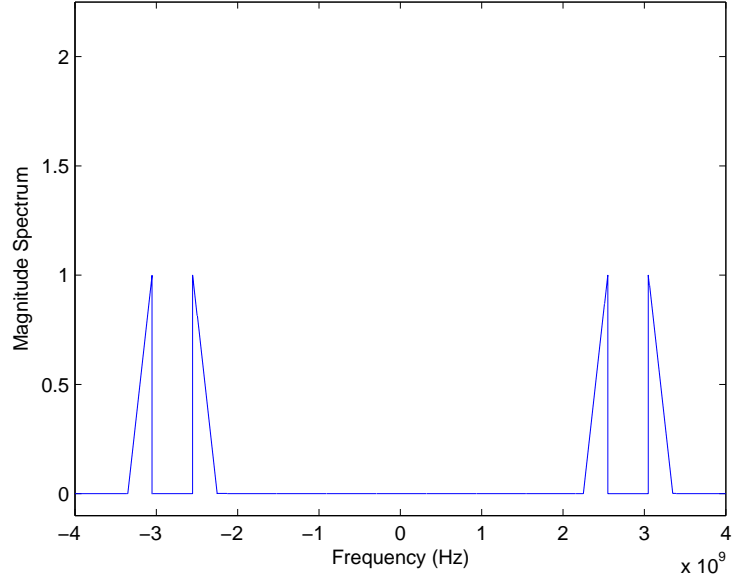


Figure 1.3: Spectrum of the Sampled Signal in Example 1.1 with $f_s = 5.6GHz$

It can be shown that there is a specific range of sampling frequencies for this scenario that will avoid aliasing. Namely

$$\frac{f_c + B/2}{n + 1} \leq f_s \leq \frac{f_c - B/2}{n} \quad (1.28)$$

where n is some integer, in this case $n = 1$. Thus, with $n = 1$ we have a range of sampling frequencies for which no aliasing occurs. However, this does not tell us the absolute minimum sampling frequency since we can always allow more than one replica of the spectrum to appear between the desired spectrum and DC (i.e., we can allow n to be larger values). It can be shown that n cannot exceed $n_{\max} = \frac{f_c - B/2}{2B}$. Thus, the minimum sampling frequency which avoids aliasing can be shown to be

$$f_s = 2B + 2B \left(\frac{f_{\max}/B - \lfloor f_{\max}/B \rfloor}{\lfloor f_{\max}/B \rfloor} \right) \quad (1.29)$$

where $\lfloor x \rfloor$ is the largest integer smaller than x . The sampling frequency as a multiple of B is plotted in Figure 1.5. We can see that as the ratio of the maximum frequency to the bandwidth increases, the minimum sampling frequency approaches $2B$.

Combining this result with the previous result for baseband signals, since the maximum frequency of a baseband signal is also its bandwidth, B , we can

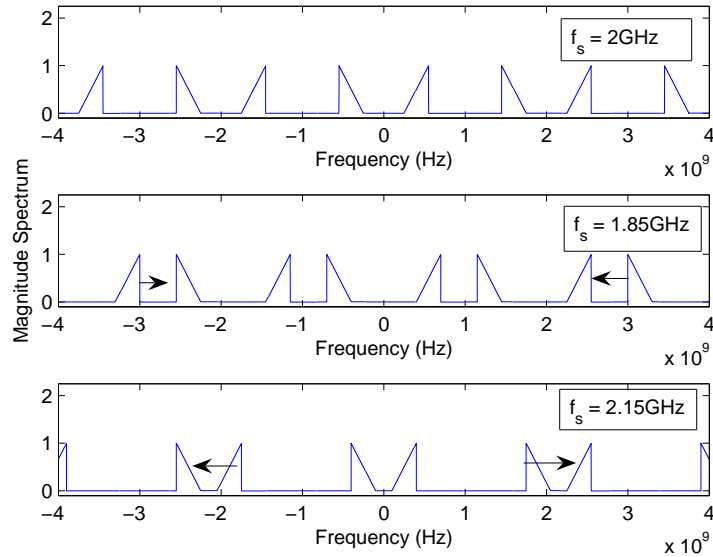


Figure 1.4: Spectrum of Sampled Signal in Example 1.1 when $f_s = 2\text{GHz}$ (top), 1.85GHz (middle) and 2.15GHz (bottom)

state that the minimum sampling rate for any signal, baseband or bandpass, is twice the signal bandwidth. However, for bandpass signals, we must be careful when choosing the exact sampling frequency.

We can also understand the minimum sampling rate of bandpass signals by considering the in-phase and quadrature representation of a bandpass signal given in equation (1.1). Specifically, recall that $z_I(t)$ and $z_Q(t)$ are real baseband signals with absolute bandwidth $B = B_T/2$ where B_T is the bandwidth of the bandpass signal. Now, if we mix the signal down to baseband with both in-phase and quadrature components, we can perfectly represent the in-phase and quadrature components by samples taken at $f_s \geq 2B$. Since the bandpass signal can always be reconstructed from the inphase and quadrature components, we can completely represent the bandpass signal with samples taken at $f_s \geq 4B = 2B_T$.

1.5 Quantization

Quantization is the second step in the analog-to-digital process. Whereas sampling converts a signal which is continuous in time to one which is discrete in time, quantization converts a signal which is continuous in amplitude to one which is discrete in amplitude. This is required simply because we cannot represent an infinite number of values with a finite number of digital words. Unlike

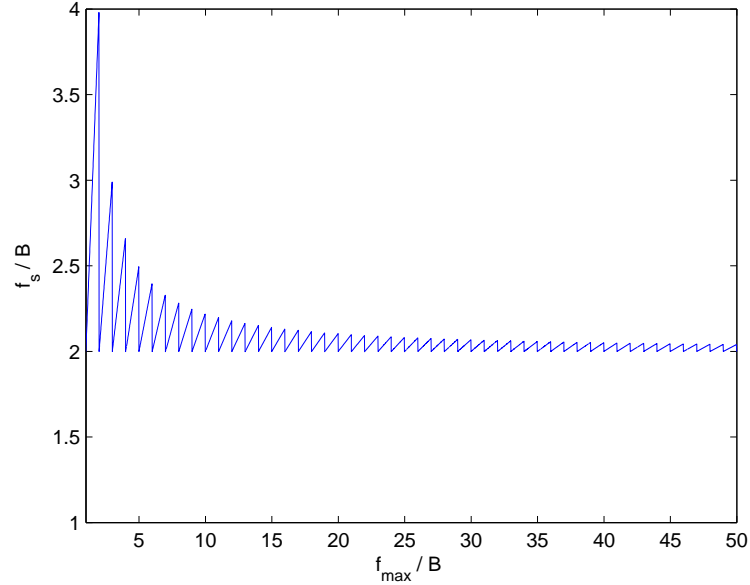


Figure 1.5: Minimum Sampling Frequency (normalized by bandwidth B) to Avoid Aliasing versus f_{max}/B

the sampling process, quantization necessarily introduces distortion into the resulting digital signal. Specifically, at each sampling instance there is some error introduced

$$e_i = x(t_i) - f_q \{x(t_i)\} \quad (1.30)$$

where $f_q \{x\}$ is the quantization function. If the input samples $x(t_i)$ are uniformly distributed (note that in general $x(t)$ is not a deterministic function, but rather a random process) and a uniform quantizer is used, the resulting signal-to-noise ratio can be written as [1]

$$\frac{S}{N} = M^2 = 2^{2n} \quad (1.31)$$

where $M = 2^n$ is the number of quantization levels, n is the number of bits per sample used in the quantization process, $S = E \{x^2(t)\}$ is the average signal power and $N = E \{e_i^2\}$ is the resulting average error power or noise power. In dB, we have

$$\frac{S}{N} (dB) = 10 \log_{10} (2^{2n}) \quad (1.32)$$

$$= 6.02n \quad (1.33)$$

Example 1.2 Consider an analog signal with bandwidth $B=3.5\text{kHz}$. The signal is to be converted to a digital bit stream using a uniform quantizer. Assuming that the analog signal is uniformly distributed, determine the minimum bit rate required to obtain a quantization SNR of 70dB.

SOLUTION: The bit rate can be written as

$$R_b = n f_s$$

where n is the number of quantization bits per sample. To obtain a quantization SNR of 70dB or better we require

$$\begin{aligned} 6.02n &\geq 70 \\ n &\geq \frac{70}{6.02} = 11.6 \end{aligned}$$

Thus, $n_{min} = 12$. To avoid aliasing we require $f_s \geq 2 * 3.5\text{kHz} = 7\text{ksps}$. The minimum bit rate that achieves an SNR of 70dB is

$$\begin{aligned} R_b &= 12\text{bits/sample} * 7\text{ksps} \\ &= 84\text{kbps} \end{aligned}$$

1.6 Digital Modulation

1.6.1 Binary Modulation

Modulation is the process of converting data bits into symbols for transmission. Typically symbols are sinusoids of radio frequency (RF) with a specific set of phases, amplitudes and/or frequencies. For example, one of the simplest modulation schemes is what is referred to as on-off keying (OOK). This can also be called binary amplitude shift keying (BASK). In this scheme we use the mapping

$$\begin{aligned} b = 0 &\rightarrow s(t) = 0 & 0 \leq t \leq T_s \\ b = 1 &\rightarrow s(t) = A \cos(\omega_c t) & 0 \leq t \leq T_s \end{aligned} \quad (1.34)$$

where b is the information bit to be transmitted, $s(t)$ is the transmit symbol, T_s is the symbol duration and ω_c is the carrier frequency in radians per second. In other words, the amplitude of the sinusoid carries the information. An amplitude of 0 represents a binary '0' and an amplitude of A represents a binary '1' and thus this scheme is called Amplitude Shift Keying. Information can also be encoded into the phase or frequency of a sinusoid. When the former is done, it

is called Phase Shift Keying or PSK. In Binary PSK or BPSK, the mapping is

$$\begin{aligned} b = 0 &\rightarrow s(t) = A\cos(\omega_c t) & 0 \leq t \leq T_s \\ b = 1 &\rightarrow s(t) = A\cos(\omega_c t + \pi) & 0 \leq t \leq T_s \end{aligned} \quad (1.35)$$

is used. In Frequency Shift Keying or FSK, the information is encoded in the frequency of the sinusoid. For Binary FSK or BFSK we use the mapping

$$\begin{aligned} b = 0 &\rightarrow s(t) = A\cos(\omega_1 t) & 0 \leq t \leq T_s \\ b = 1 &\rightarrow s(t) = A\cos(\omega_2 t) & 0 \leq t \leq T_s \end{aligned} \quad (1.36)$$

where ω_1 and ω_2 are typically chosen such that the two symbols are orthogonal over a symbol period and are centered around the nominal carrier frequency ω_c .

1.6.2 M -ary Modulation

The previous modulation schemes are known as *binary* modulation schemes since there are only two symbols, one for each bit value. It is also possible to map n bits per symbol. This is called M -ary modulation where there are $M = 2^n$ different symbols³. For example, with M -PSK the modulation is still contained in the phase of the sinusoid, but there are M different possible phase values corresponding to $M = 2^n$ different groups of binary '1's and '0's. The M symbols can be defined as

$$s_i(t) = A\cos(\omega_c t + (i-1)\frac{2\pi}{M}) \quad 0 \leq t \leq T_s \quad (1.37)$$

where $i = 1, 2, \dots, M$. M -ary ASK and M -ary FSK are also possible where each group of $n = \log_2 M$ bits is mapped to one of M symbols each with a different amplitude or frequency respectively.

1.6.3 Power Spectral Density

Of primary importance in digital modulation are the spectral properties of the transmitted signal. In this section we examine the power spectral density (PSD) of digitally modulated signals. The power spectral density of a bandpass signal $x(t)$ can be determined from the PSD of the corresponding complex baseband signal $\tilde{x}(t)$ given in (1.5) as

$$P_{bp}(f) = \frac{1}{4}P_x(-f - f_c) + \frac{1}{4}P_x(f + f_c) \quad (1.38)$$

where $P_x(f)$ is the power spectral density of the complex baseband. Now, the PSD of the baseband equivalent can be written as [1]

$$P_x(f) = \lim_{T \rightarrow \infty} \left(\frac{|X_T(f)|^2}{T} \right) \quad (1.39)$$

³Please note that the number of symbols (and the corresponding number of bits per symbol) is independent of the number of quantization levels (and corresponding number of bits per sample) in the quantization process.

where

$$X_T(f) = \int_{-T/2}^{T/2} \tilde{x}(t) e^{-j2\pi ft} dt \quad (1.40)$$

and \bar{x} is the ensemble average of the random variable x . The baseband equivalent of linearly modulated bandpass signals can be written as

$$\tilde{x}(t) = \sum_{n=-\infty}^{\infty} a_n f(t - nT_s) \quad (1.41)$$

where a_n represents the data modulation (possibly complex) and $f(t)$ is the pulse shape used. Thus, $X_T(f)$ can be written as

$$\begin{aligned} X_T(f) &= \int_{-T/2}^{T/2} \sum_{n=-\infty}^{\infty} a_n f(t - nT_s) e^{-j2\pi ft} dt \\ &= \sum_{n=-N}^N a_n \int_{-\infty}^{\infty} f(t - nT_s) e^{-j2\pi ft} dt \end{aligned} \quad (1.42)$$

where $T = (2N + 1)T_s$. Now the inner integral is simply the Fourier Transform of the delayed pulse shape. Thus,

$$\begin{aligned} X_T(f) &= \sum_{n=-N}^N a_n F(f) e^{-j2\pi f n T_s} \\ &= F(f) \sum_{n=-N}^N a_n e^{-j2\pi f n T_s} \end{aligned} \quad (1.43)$$

Now, returning to our definition of the PSD:

$$\begin{aligned} P_x(f) &= \lim_{N \rightarrow \infty} \left(\frac{\left| F(f) \sum_{n=-N}^N a_n e^{-j2\pi f n T_s} \right|^2}{(2N + 1) T_s} \right) \\ &= |F(f)|^2 \lim_{N \rightarrow \infty} \left(\frac{1}{(2N + 1) T_s} \sum_{n=-N}^N \sum_{m=-N}^N \overline{a_n a_m} e^{-j2\pi f(n-m)T_s} \right) \end{aligned} \quad (1.44)$$

Now, defining the discrete auto-correlation function $R(k) = \overline{a_n a_{n+k}}$, we have

$$\begin{aligned}
P_x(f) &= |F(f)|^2 \lim_{N \rightarrow \infty} \left(\frac{1}{(2N+1)T_s} \sum_{n=-N}^N \sum_{k=-N-n}^{N-n} R(k) e^{-j2\pi f k T_s} \right) \\
&= \frac{|F(f)|^2}{T_s} \lim_{N \rightarrow \infty} \left(\frac{(2N+1)}{(2N+1)} \sum_{k=-N}^N R(k) e^{-j2\pi f k T_s} \right) \\
&= \frac{|F(f)|^2}{T_s} \sum_{k=-\infty}^{\infty} R(k) e^{-j2\pi f k T_s} \tag{1.45}
\end{aligned}$$

Now, assuming that the data is uncorrelated, we have

$$R(k) = \begin{cases} \sigma_a^2 + m_a^2 & k = 0 \\ m_a^2 & k \neq 0 \end{cases} \tag{1.46}$$

where m_a and σ_a^2 are the mean and variance of the data sequence. In this case we can simplify the power spectral density as

$$P_x(f) = \frac{|F(f)|^2}{T_s} \left(\sigma_a^2 + m_a^2 \sum_{k=-\infty}^{\infty} e^{-j2\pi f k T_s} \right) \tag{1.47}$$

It is well known that an infinite sum of complex sinusoids is mathematically equivalent to an infinite impulse train. That is

$$\sum_{k=-\infty}^{\infty} e^{-j2\pi f k T_s} = \frac{1}{T_s} \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{k}{T_s}\right) \tag{1.48}$$

Thus, we have

$$\begin{aligned}
P_x(f) &= \frac{|F(f)|^2}{T_s} \left(\sigma_a^2 + m_a^2 \sum_{k=-\infty}^{\infty} \delta\left(f - \frac{k}{T_s}\right) \right) \\
&= \left(\frac{\sigma_a^2}{T_s} |F(f)|^2 + \frac{m_a^2}{T_s} \sum_{k=-\infty}^{\infty} \left| F\left(\frac{k}{T_s}\right) \right|^2 \delta\left(f - \frac{k}{T_s}\right) \right) \tag{1.49}
\end{aligned}$$

The spectrum clearly has two parts, a continuous portion that depends on the pulse shape and a discrete component that arises due to the mean of the data sequence. A common means of eliminating the discrete terms is to use a zero mean data mapping (e.g., BPSK). The bandpass PSD $P_{bp}(f)$ is then determined from (1.38).

1.6.4 Pulse Shaping

As demonstrated in the previous section, the Fourier Transform of the pulse shape dominates the power spectral density of linearly modulated digital bandpass signals. If the symbols last over a single symbol period, we can think

of each symbol as being a sinusoid with a particular amplitude, phase and frequency multiplied by a square pulse of width T_s and centered at $(i - \frac{1}{2}) T_s$. A square pulse has a spectrum that is related to the sinc function. Because of the slow decay of the sidelobes of the sinc function, other pulse shapes are often used rather than a simple rectangular pulse. The optimal pulse shape is a sinc function, since its spectral properties are related to a square pulse (i.e., it has minimum bandwidth). However, since the sinc function has infinite duration in both the positive and negative time directions, it is impossible to implement in practice. A function which has good bandwidth properties and reasonable implementation complexity is a truncated raised cosine pulse.

1.7 Optimum Receiver in AWGN

1.7.1 The Matched Filter

The received signal in digital communications receiver is corrupted by Additive White Gaussian Noise or AWGN. The PSD of AWGN is a constant of (theoretically) infinite bandwidth. Consider a received signal

$$r(t) = x(t) + n(t) \quad (1.50)$$

where $x(t)$ is the transmit signal and $n(t)$ is AWGN which is modeled as a Gaussian random process with power spectral density

$$P_n(f) = \frac{N_o}{2}$$

Theoretically, the bandwidth (and thus the power) of AWGN is infinite. Although noise does not truly have infinite bandwidth, its bandwidth is sufficiently larger than the bandwidth of the desired signal that the infinite bandwidth model is reasonable. Thus, a crucial operation in any receiver is filtering which limits the received signal bandwidth and thus the power of AWGN. Further, the bandwidth of the filter used in the receiver has a direct impact on the performance of the receiver. Clearly, we would like to filter the signal in such a way that maximizes the desired signal power, while minimizing the noise power. In other words, we wish to maximize the SNR at the output of the filter. To determine the optimal filter (i.e., the filter which maximizes SNR) we will examine the signal in the frequency domain. The desired signal in the frequency domain, after filtering with a filter whose impulse response is $h(t)$ can be written as

$$\tilde{X}(f) = X(f) H(f). \quad (1.51)$$

where $H(f) = \mathcal{F}\{h(t)\}$ is the filter transfer function.

The noise power spectral density at the output of the filter can be written as

$$P_{\tilde{n}}(f) = P_n(f) |H(f)|^2. \quad (1.52)$$

Now, the output at the sampling instant T can be written using the inverse Fourier Transform of (1.51):

$$\tilde{x}(T) = \int_{-\infty}^{\infty} \tilde{X}(f) e^{j2\pi fT} df \quad (1.53)$$

The noise power at the filter output can be calculated from the noise power spectral density as

$$\overline{\tilde{n}^2(t)} = \int_{-\infty}^{\infty} P_n(f) df \quad (1.54)$$

$$= \int_{-\infty}^{\infty} P_n(f) |H(f)|^2 df. \quad (1.55)$$

Note that the noise power does not depend on the sampling instant. The SNR at the filter output at the sampling instant is then

$$\left(\frac{S}{N}\right)_{out} = \frac{|\tilde{x}(T)|^2}{\overline{\tilde{n}^2(t)}} \quad (1.56)$$

Now we wish to find the filter $H(f)$ that maximizes $(S/N)_{out}$. First, let us include the filter in equation (1.56):

$$\left(\frac{S}{N}\right)_{out} = \frac{\left|\int_{-\infty}^{\infty} \tilde{X}(f) e^{j2\pi fT} df\right|^2}{\int_{-\infty}^{\infty} P_n(f) |H(f)|^2 df} \quad (1.57)$$

$$= \frac{\left|\int_{-\infty}^{\infty} X(f) H(f) e^{j2\pi fT} df\right|^2}{\int_{-\infty}^{\infty} P_n(f) |H(f)|^2 df} \quad (1.58)$$

Now, in order to maximize this expression, we apply Schwarz' Inequality which states

$$\left|\int_{-\infty}^{\infty} A(f) B(f) df\right|^2 \leq \int_{-\infty}^{\infty} |A(f)|^2 df \int_{-\infty}^{\infty} |B(f)|^2 df \quad (1.59)$$

where the equality is reached when $A(f) = KB^*(f)$ for an arbitrary real constant K and where $()^*$ is the conjugation operation. Now, using Schwarz' inequality, if we let

$$A(f) = H(f) \sqrt{P_n(f)} \quad (1.60)$$

$$B(f) = \frac{X(f) e^{j2\pi fT}}{\sqrt{P_n(f)}} \quad (1.61)$$

then we can write

$$\left(\frac{S}{N}\right)_{out} \leq \frac{\int_{-\infty}^{\infty} |H(f)|^2 P_n(f) df \int_{-\infty}^{\infty} \frac{|X(f)|^2}{P_n(f)} df}{\int_{-\infty}^{\infty} |H(f)|^2 P_n(f) df} \quad (1.62)$$

$$\leq \int_{-\infty}^{\infty} \frac{|X(f)|^2}{P_n(f)} df \quad (1.63)$$

Thus, we know that the output SNR is bounded by equation (1.63). Further, we know that the maximum is reached when $A(f) = KB^*(f)$. That is the SNR is maximized when

$$H(f) \sqrt{P_n(f)} = K \frac{X^*(f) e^{-j2\pi fT}}{\sqrt{P_n(f)}} \quad (1.64)$$

$$H(f) = K \frac{X^*(f)}{P_n(f)} e^{-j2\pi fT} \quad (1.65)$$

The optimal filter (i.e., the one which maximizes SNR at the sampling instant) is one whose frequency response is directly proportional to the desired signal's spectrum and inversely proportional to the noise power spectral density and includes a delay equal to the symbol duration in order to maximize the signal power at the sampling instant. The constant K is arbitrary and has no effect on the filter's performance as it will equally scale the desired signal and the noise.

For the special and important case of white noise we have

$$\begin{aligned} H(f) &= K \frac{X^*(f)}{N_o/2} e^{-j2\pi fT} \\ &= C \cdot X^*(f) e^{-j2\pi fT} \end{aligned} \quad (1.66)$$

Taking the inverse Fourier Transform

$$h(t) = C \cdot x(T - t) \quad (1.67)$$

Thus, for the case of white noise, the optimal filter is one which is *matched* to the desired signal. In other words, the impulse response of the optimal filter is a time-reversed version of the desired signal. Returning to equation (1.63), the

SNR at the output of the matched filter with white noise is

$$\begin{aligned}
 \left(\frac{S}{N}\right)_{\max} &= \int_{-\infty}^{\infty} \frac{|X(f)|^2}{P_n(f)} df \\
 &= \int_{-\infty}^{\infty} \frac{|X(f)|^2}{N_o/2} df \\
 &= \frac{2}{N_o} \int_{-\infty}^{\infty} |X(f)|^2 df \\
 &= \frac{2}{N_o} \int_{-\infty}^{\infty} |x(t)|^2 dt \\
 &= \frac{2E_s}{N_o}
 \end{aligned} \tag{1.68}$$

where E_s is the energy in the desired signal over one symbol. This is an important result. It says that regardless of the signal, i.e., pulse shape, used provided that a matched filter is employed at the receiver the signal-to-noise ratio at the output is the same and is given by (1.68). Thus, we are free to choose the pulse shape to minimize signal bandwidth since performance is not affected.

Example 1.3 Determine the matched filter for a signal employing a polar non-return-to-zero line code with a square pulse when the noise is white.

SOLUTION: From equation (1.67) we know that the impulse response of the matched filter is

$$h(t) = C \cdot x(T - t) \tag{1.69}$$

where in this case $x(t) = \Pi(t/T)$. An example plot of the received signal before matched filtering (noiseless) is shown in Figure 1.6. The impulse response of the matched filter is a square pulse. In other words the matched filter is a square pulse sliding integrator and the output is shown in Figure 1.7 for the case where we choose $C = 1/T$. The output signal is clearly more sensitive to timing error as can be seen. The reason that it improves performance can be seen in Figure 1.8 which shows the energy spectral densities of an example randomly modulated square pulse train (top) and a sample function of AWGN (bottom). Since the matched filter has a transfer function of

$$H(f) = T \text{sinc}(fT) \tag{1.70}$$

it is matched to the desired signal spectrum and all of the desired signal energy will be captured, while the noise power will be reduced.

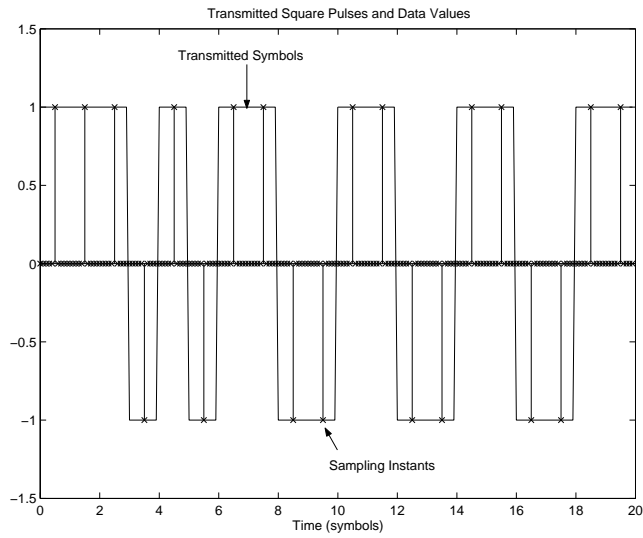


Figure 1.6: Example of Received Signal with Square Pulses and Ideal Sampling Times

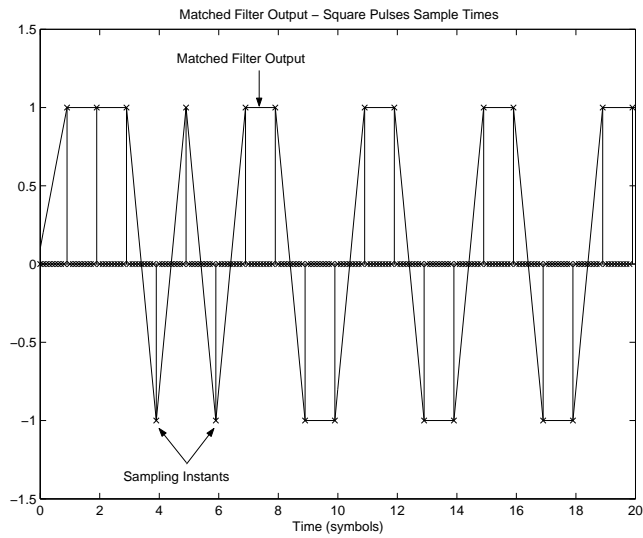


Figure 1.7: Example of Matched Filter Output for Square Pulses from Figure 1.6

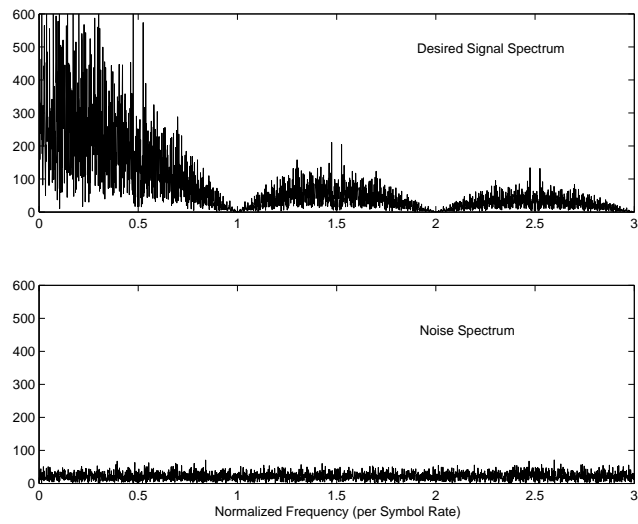


Figure 1.8: Power Spectrum for Example Functions of Random Data Signal (top) and AWGN (bottom)

1.7.2 The Correlator Version of the Matched Filter

The output of the matched filter receiver can be written as

$$y(t) = \int_{-\infty}^t r(\tau) h(t - \tau) d\tau \quad (1.71)$$

where $h(t)$ is the impulse response of the matched filter. Now we have just shown that the impulse response of the matched filter is matched to the pulse shape, $h(t) = x(T - t)$. Thus,

$$y(t) = \int_0^t r(\tau) x(T - (t - \tau)) d\tau \quad (1.72)$$

Now, the optimal sample point is at $t = T$. Thus, decision statistic is

$$\begin{aligned} y(T) &= \int_0^T r(\tau) x(T - (T - \tau)) d\tau \\ &= \int_0^T r(\tau) x(\tau) d\tau \end{aligned} \quad (1.73)$$

The last line shows that the decision statistic of the optimal, i.e., matched, filter is simply the output of a correlator which correlates the received signal with a time synchronous version of the desired signal. That is, the optimal decision metric can be calculated by correlating the received signal with the pulse shape used over the symbol interval.

1.7.3 Detection and the Optimal Threshold

The previous sections detailed the optimal filter, that is the filter which maximizes the signal-to-noise ratio at its output at the sampling time. The optimal detector is a decision device that estimates the transmitted symbol (or bit in the binary case) using the matched filter output such that it minimizes the probability of symbol error. For the binary case, let us define the output of the matched filter due to the desired signal when symbol 1 is sent as s_{O1} and the output when symbol 2 is sent as s_{O2} . Further, in the case of AWGN, the matched filter output $y(T)$ conditioned on symbol 1 being sent is a Gaussian random variable with mean s_{O1} and variance σ^2 . The matched filter output conditioned on symbol 2 being sent is also a Gaussian random variable with mean s_{O2} and variance σ^2 . The probability of error can be written as

$$P_e = \Pr\{e|s_1\} \Pr\{s_1\} + \Pr\{e|s_2\} \Pr\{s_2\} \quad (1.74)$$

where $\Pr\{s_i\}$ is the *a priori* probability of sending symbol s_i and $\Pr\{e|s_i\}$ is the probability of error given that s_i is sent. Assuming that $s_{O2} < s_{O1}$, decisions are made at the detector based on some threshold V_T such that

$$\hat{s}_i = \begin{cases} s_2 & y(T) \leq V_T \\ s_1 & y(T) > V_T \end{cases} \quad (1.75)$$

To determine the optimal threshold, i.e., the threshold which minimizes P_e , we first must write the probability of error in terms of V_T . Now, the probability of error given that s_1 is sent, is simply the probability that $y(T) \leq V_T$ given that s_1 is sent. Similarly, the probability of error given that s_2 is sent, is simply the probability that $y(T) > V_T$ given that s_2 is sent. That is, if we define $Z = y(T)$,

$$P_e = \Pr\{s_1\} \int_{-\infty}^{V_T} f_Z(z|s_1) dz + \Pr\{s_2\} \int_{V_T}^{\infty} f_Z(z|s_2) dz \quad (1.76)$$

where $f_Z(z|s_i)$ is the distribution of z conditioned on s_i being sent. To minimize the probability of error we take the derivative of P_e with respect to V_T and set it equal to zero

$$\begin{aligned} \frac{dP_e}{dV_T} &= 0 \\ \frac{d}{dV_T} \left\{ \Pr\{s_1\} \int_{-\infty}^{V_T} f_Z(z|s_1) dz + \Pr\{s_2\} \int_{V_T}^{\infty} f_Z(z|s_2) dz \right\} &= 0 \end{aligned}$$

Now, the conditional distributions of z are Gaussian with mean values of s_{O1} and s_{O2} when conditioned on s_1 and s_2 respectively. Thus, we have

$$\begin{aligned} \frac{d}{dV_T} \left\{ \Pr\{s_1\} \int_{-\infty}^{V_T} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-s_{O1})^2}{2\sigma^2}} dz \dots \right. \\ \left. + \Pr\{s_2\} \int_{V_T}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-s_{O2})^2}{2\sigma^2}} dz \right\} &= 0 \\ \frac{d}{dV_T} \left\{ \Pr\{s_1\} \int_{-(V_T-s_{O1})/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} d\lambda + \dots \right. \\ \left. \Pr\{s_2\} \int_{(V_T-s_{O2})/\sigma}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} d\lambda dz \right\} &= 0 \end{aligned}$$

where we have made the substitutions $\lambda = -(z - s_{O1})/\sigma$ and $\lambda = (z - s_{O2})/\sigma$ in the first and second terms respectively in the second line. Now, using Leibniz' Rule to evaluate the derivative of the integral (note that the variable which the derivative is being taken with respect to is in the limits of integration):

$$\begin{aligned} \Pr\{s_1\} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(V_T-s_{O1})^2}{2\sigma^2}} - \Pr\{s_2\} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(V_T-s_{O2})^2}{2\sigma^2}} &= 0 \\ \Pr\{s_1\} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(V_T-s_{O1})^2}{2\sigma^2}} &= \Pr\{s_2\} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(V_T-s_{O2})^2}{2\sigma^2}} \end{aligned}$$

Taking the natural log:

$$\begin{aligned}
 \ln(\Pr\{s_1\}/\Pr\{s_2\}) &= -\frac{(V_T - s_{O2})^2}{2\sigma^2} + \frac{(V_T - s_{O1})^2}{2\sigma^2} \\
 \ln(\Pr\{s_1\}/\Pr\{s_2\}) &= \frac{2V_T s_{O2} - s_{O2}^2 - 2V_T s_{O1} + s_{O1}^2}{2\sigma^2} \\
 V_T &= \frac{2\sigma^2 \ln(\Pr\{s_1\}/\Pr\{s_2\}) + s_{O2}^2 - s_{O1}^2}{2(s_{O2} - s_{O1})} \quad (1.77)
 \end{aligned}$$

Now, equation (1.77) provides the optimal threshold in the general case for AWGN signaling. For equally likely symbols, $\Pr\{s_1\} = \Pr\{s_2\}$ and the threshold is

$$\begin{aligned}
 V_T &= \frac{2\sigma^2 \ln(1) + s_{O2}^2 - s_{O1}^2}{2(s_{O2} - s_{O1})} \\
 &= \frac{s_{O2}^2 - s_{O1}^2}{2(s_{O2} - s_{O1})} \\
 &= \frac{s_{O1} + s_{O2}}{2} \quad (1.78)
 \end{aligned}$$

Further, if antipodal signaling is used (along with equally likely symbols), we have $s_{O1} = -s_{O2}$ and

$$\begin{aligned}
 V_T &= \frac{s_{O1} - s_{O1}}{2} \\
 &= 0
 \end{aligned}$$

Example 1.4 Determine the optimal threshold for antipodal signaling when $\Pr\{s_2\} = 4\Pr\{s_1\}$.

SOLUTION: From equation (1.77) we know that the optimal threshold can be written as.

$$V_T = \frac{2\sigma^2 \ln(\Pr\{s_1\}/\Pr\{s_2\}) + s_{O2}^2 - s_{O1}^2}{2(s_{O2} - s_{O1})} \quad (1.79)$$

For antipodal signaling $s_{O2} = -s_{O1}$, thus:

$$\begin{aligned} V_T &= \frac{2\sigma^2 \ln(\Pr\{s_1\}/\Pr\{s_2\}) + s_{O2}^2 - s_{O2}^2}{2(s_{O2} + s_{O2})} \\ &= \frac{\sigma^2 \ln(\Pr\{s_1\}/\Pr\{s_2\})}{2s_{O2}} \\ &= \frac{\sigma^2 \ln(4)}{2s_{O2}} \end{aligned} \quad (1.80)$$

Now for the matched filter we have $\sigma^2 = N_o/2$ and $s_{O2} = -\sqrt{E_b}$. Thus,

$$V_T = -\frac{N_o \ln(4)}{4\sqrt{E_b}} \quad (1.81)$$

If the noise power (relative to the desired signal energy) increases, we can see that the threshold is biased more and more negative (i.e., toward the *a priori* more probable symbol s_2). This is illustrated in Figure 1.9. However, if $\Pr\{s_2\} < \Pr\{s_1\}$, the natural log evaluates to a negative number and the threshold is positive (i.e., biased toward s_1 which is now more probable.). Thus, if the symbols are not equally likely, the optimal decision boundary is biased toward the more probable symbol. This bias is more pronounced for higher noise power since our observations are less reliable. On the other hand, if the noise power is very low, the a priori probabilities are less relevant since the observations are more reliable.

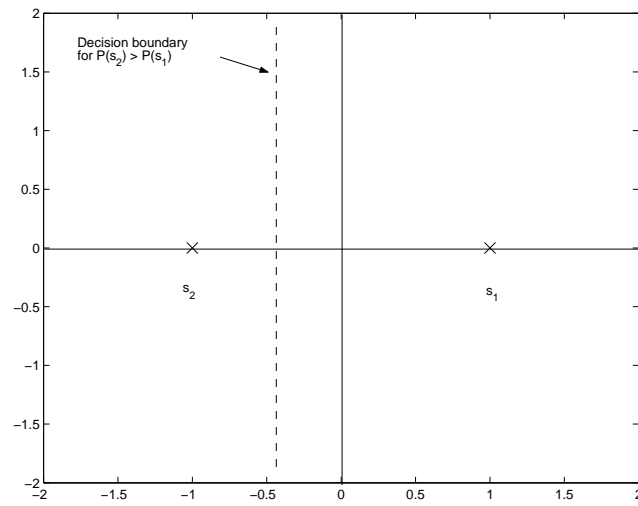


Figure 1.9: Illustration of Decision Boundary for Binary Signaling of Example 1.4

1.7.4 BER of the Matched Filter with Optimal Detection

Now that we have determined the optimal filter which maximizes SNR and the optimal detector, we are in a position to determine the performance of this optimized receiver. Using (1.76) and again assuming AWGN, the probability of error can be written as

$$P_e = \Pr\{s_1\} \int_{-\infty}^{V_T} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-s_{O1})^2}{2\sigma^2}} dz + \Pr\{s_2\} \int_{V_T}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-s_{O2})^2}{2\sigma^2}} dz \quad (1.82)$$

Assuming equally likely symbols and using the optimal threshold, we can express the probability of error as

$$P_e = \frac{1}{2} \int_{-\infty}^{\frac{s_{O1}+s_{O2}}{2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-s_{O1})^2}{2\sigma^2}} dz + \frac{1}{2} \int_{\frac{s_{O1}+s_{O2}}{2}}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(z-s_{O2})^2}{2\sigma^2}} dz \quad (1.83)$$

Making the substitutions $\lambda = -(z - s_{O1})/\sigma$ and $\lambda = (z - s_{O2})/\sigma$ in the first and second terms respectively results in

$$\begin{aligned} P_e &= \frac{1}{2} \int_{\frac{s_{O1}-s_{O2}}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} dz + \frac{1}{2} \int_{\frac{s_{O1}-s_{O2}}{2\sigma}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} dz \\ &= Q\left(\frac{s_{O1} - s_{O2}}{2\sigma}\right) \\ &= Q\left(\sqrt{\frac{(s_{O1} - s_{O2})^2}{4\sigma^2}}\right) \end{aligned} \quad (1.84)$$

Now, examining the numerator inside the radical of (1.84) and assuming a matched filter receiver:

$$\begin{aligned} (s_{O1} - s_{O2})^2 &= \left(\int_0^T s_1(t) x(t) dt - \int_0^T s_2(t) x(t) dt \right)^2 \\ &= \left(\int_0^T (s_1(t) - s_2(t)) x(t) dt \right)^2 \\ &= E_d \end{aligned} \quad (1.85)$$

where E_d is termed the *difference energy* and is simply the desired signal power at the output of a filter matched to the difference between symbols s_1 and s_2 . The term $\frac{(s_{O1}-s_{O2})^2}{\sigma^2}$ is simply the signal to noise ratio at the output of the filter matched to the difference signal $s_1(t) - s_2(t)$. However, from (1.68) we know that the SNR at the output of the matched filter is

$$\frac{(s_{O1} - s_{O2})^2}{\sigma^2} = \frac{2E_d}{N_o} \quad (1.86)$$

Thus, we can rewrite the probability of symbol error as

$$P_e = Q\left(\sqrt{\frac{E_d}{2N_o}}\right) \quad (1.87)$$

Now that we have the probability of bit error for the general binary case, we can derive the performance of specific signaling formats. As a first example, consider unipolar NRZ signaling. In this case

$$\begin{aligned} s_1(t) &= Ax(t) & 0 \leq t \leq T \\ s_2(t) &= 0 & 0 \leq t \leq T \end{aligned} \quad (1.88)$$

Thus, E_d is calculated as

$$\begin{aligned} E_d &= \int_0^T A^2 x^2(t) dt \\ &= A^2 \end{aligned} \quad (1.89)$$

assuming a unit energy pulse. Now the average energy per symbol, E_s , with this format is

$$\begin{aligned} E_s &= \frac{1}{2}(A^2 + 0) \\ &= \frac{A^2}{2} \\ &= \frac{E_d}{2} \end{aligned} \quad (1.90)$$

Thus, we have for unipolar NRZ signaling:

$$P_e = Q\left(\sqrt{\frac{E_s}{N_o}}\right) \quad (1.91)$$

For polar NRZ signaling, the transmit signals are

$$s_1(t) = Ax(t) \quad 0 \leq t \leq T \quad (1.92)$$

$$s_2(t) = -Ax(t) \quad 0 \leq t \leq T \quad (1.93)$$

Thus, the difference energy is calculated as

$$\begin{aligned} E_d &= \int_0^T (Ax(t) - (-Ax(t)))^2 dt \\ &= 4A^2 \end{aligned} \quad (1.94)$$

while $E_s = A^2$. Thus,

$$P_e = Q\left(\sqrt{\frac{2E_s}{N_o}}\right) \quad (1.95)$$

Example 1.5 Compare the performance of three filters for use with a square-pulse NRZ line-code: (a) low-pass filter with a noise equivalent bandwidth of $5R_s$, (b) a low-pass filter with noise-equivalent bandwidth of R_s , and (c) a matched filter.

SOLUTION: The energy spectral density (ESD) of a sample function (without noise) and a sampling rate of $20R_s$ is shown in Figure 1.10 (top-left). The ESD of a received sample function including white noise at an E_b/N_o of 10dB is shown in Figure 1.10 (top-right). The noise ESD is approximately flat, while the desired signal has an ESD that follows a sinc function. A filter with a bandwidth of $5R_s$ was applied to the received signal first. The resulting ESD is shown in Figure 1.10 (bottom-left). It is evident that a majority of the desired signal is maintained along with a large amount of noise. This can also be seen in the time domain in Figure 1.11 which shows the transmit signal (top), desired portion of the received signal (middle), and the entire received signal (bottom). Comparing the top and middle curves, we can see that the large bandwidth of the receive filter allows the desired signal to pass with very little distortion. However, from the bottom plot it is obvious that the filter passes considerable noise power. This can be compared to Figure 1.10 (bottom-right) and Figure 1.12 which plots the frequency and time domain received signal for a low-pass filter with a bandwidth of R_s . From the frequency domain plot in Figure 1.10 (bottom-right) we can see that substantially less noise power is permitted to pass to the detector, but a substantial amount of the desired signal is also lost. The plot in Figure 1.12 (middle) demonstrates that the desired signal is heavily distorted and inter-symbol interference is introduced. However, the reduced noise power can also be seen by comparing Figure 1.12 (middle) to Figure 1.12 (top). The strong correlation between the two plots emphasizes that less noise is permitted as compared to the first filter.

The matched filter is demonstrated in the time domain in Figure 1.13. The filter has a transfer function that is matched to the energy spectral density shown in Figure 1.10 (top-left). The time-domain plots show that (a) there is no inter-symbol interference introduced at the proper sampling instances and (b) relatively little noise is passed. The first point can be seen from Figure 1.13 (middle) where the desired signal at the output of the filter exactly equals the data value (± 1) at the proper sampling instances. Comparing 1.13 (middle) to 1.13 (bottom) we can also see that little noise is present as compared to the previous filters. Finally, the simulated Bit Error Rate (BER) performance is plotted in Figure 1.14 for each of the three filters. The highest bandwidth filter (labeled as 'low distortion' since it passes the desired signal with little distortion) provides the worst performance due to the low resulting SNR. The lower bandwidth filter has improved SNR but also introduces ISI. The two factors result in performance which is roughly 3dB worse than the matched filter which maximizes SNR and avoids ISI.

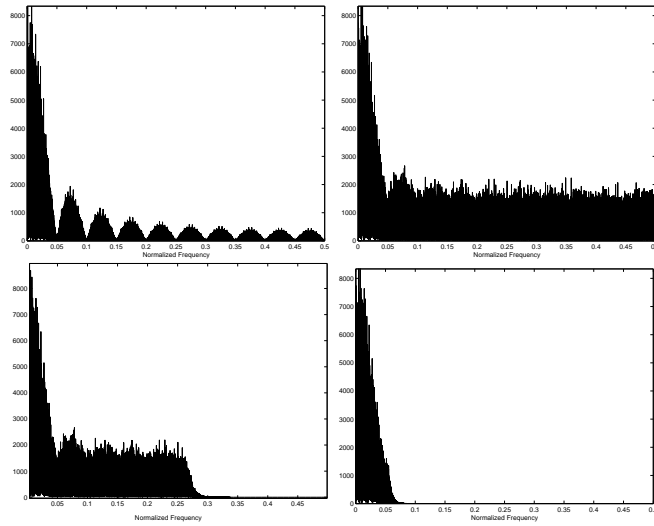


Figure 1.10: Energy Spectral Densities for Sample Waveforms for Example 1.5 Transmit Signal (top-left), Received Signal (top-right), Received Signal after Filtering with First Filter (bottom-left), Received Signal after Filtering with Second Filter (bottom-right)

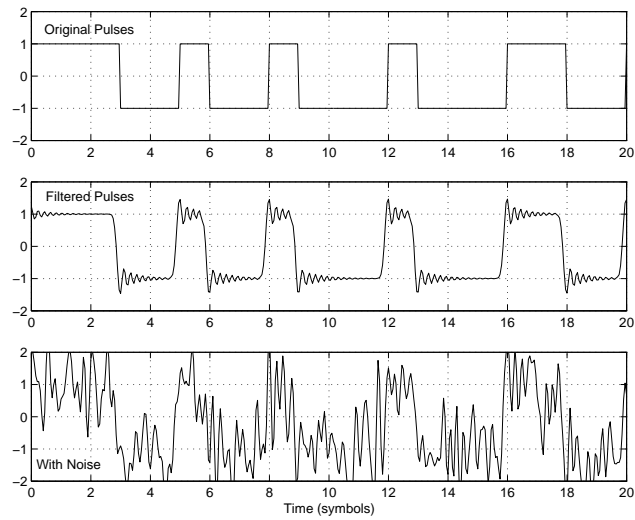


Figure 1.11: Example Time Domain Waveform for Modulated Square Pulse (Polar NRZ Line-Code) assumed in Example 1.5- Top: Original transmitted signal; Middle: Desired portion of the received signal after filtering; Bottom: Entire received signal (including noise) after filtering.

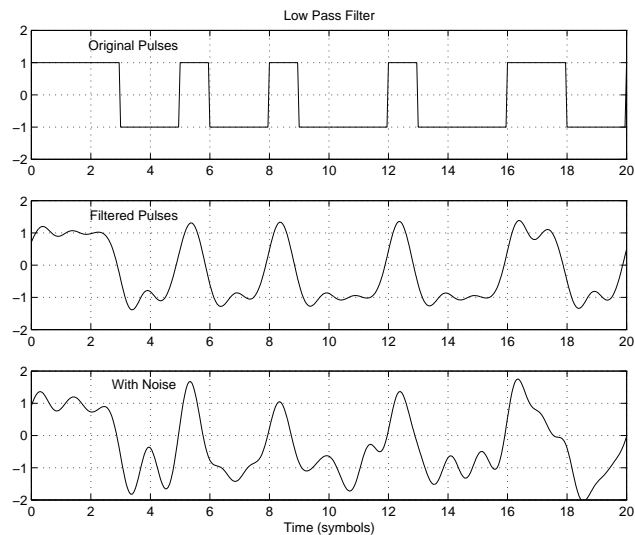


Figure 1.12: Example Time Domain Waveform for Transmitted and Received Square Pulse (Polar NRZ Line-Code, filter $BW = R_s$) from Example 1.5 - Top: Original transmitted signal; Middle: Desired portion of the received signal after filtering; Bottom: Entire received signal (including noise) after filtering

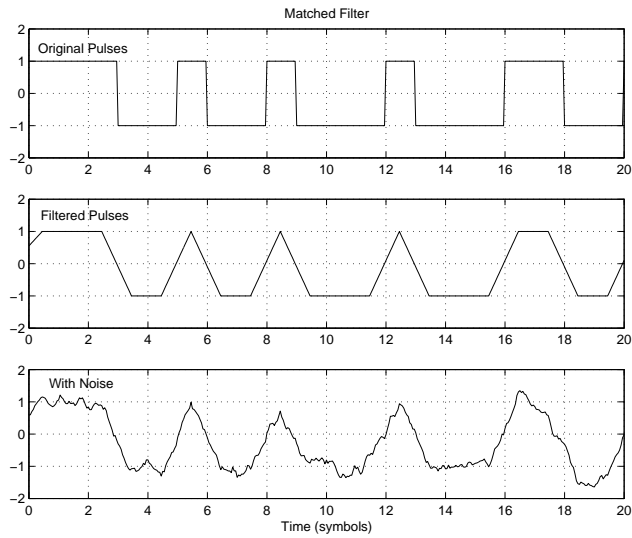


Figure 1.13: Example Time Domain Waveform for Transmitted and Received Square Pulse (Polar NRZ Line-Code, matched filter) from Example 1.5 - Top: Original transmitted signal; Middle: Desired portion of the received signal after filtering; Bottom: Entire received signal (including noise) after filtering.

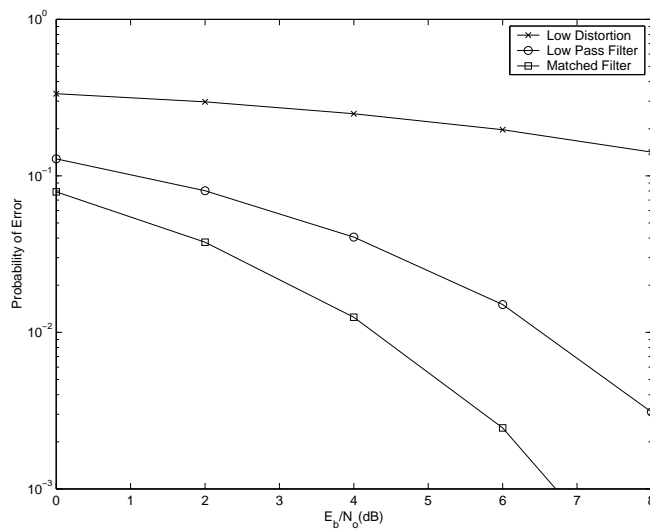


Figure 1.14: Simulated Bit Error Rate Performance for Three Filters with Square Pulse Transmit Signal from Example 1.5

1.7.5 Performance of General Binary Modulation

Let us now consider a generic binary modulation scheme with two symbols $s_1(t)$ and $s_2(t)$. Any M -ary modulation scheme can be represented by $K \leq M$ basis functions [1]. The receiver then consists of a bank of K filters matched to the K basis functions. For general binary modulation, since there are two symbols, we can completely represent the two symbols with at most two basis functions $\phi_1(t)$ and $\phi_2(t)$ where by definition

$$\int_0^{T_s} \phi_1^2(t) dt = \int_0^{T_s} \phi_2^2(t) dt = 1 \quad (1.96)$$

and

$$\int_0^{T_s} \phi_1 \phi_2(t) dt = 0 \quad (1.97)$$

The two signals can be represented as

$$\begin{aligned} s_1(t) &= s_{11} \phi_1(t) + s_{12} \phi_2(t) \\ s_2(t) &= s_{21} \phi_1(t) + s_{22} \phi_2(t) \end{aligned} \quad (1.98)$$

where

$$s_{ij} = \int_0^T s_i(t) \phi_j(t) dt \quad (1.99)$$

Now since we can choose the first basis function arbitrarily, let's choose

$$\phi_1(t) = \frac{1}{\sqrt{E_1}} s_1(t) \quad (1.100)$$

where E_1 is the energy in symbol 1. Once the basis functions are defined, a signal space plot (sometimes termed a constellation diagram) can be drawn which represents the symbols in k -dimensional space (k is the number of basis function which is two in this example). An example constellation diagram is plotted in Figure 1.15 for arbitrary $s_1(t)$ and $s_2(t)$.

Assume that the received signal is $r(t) = s(t) + n(t)$ where $n(t)$ is additive white Gaussian noise with power spectral density $P_N(f) = \frac{N_0}{2}$. We wish to examine the receiver which minimizes the probability of symbol error.

The optimum (maximum SNR) receiver is one which correlates the received signal with the two basis functions (or a matched filter receiver). The outputs of the two correlators can be represented in vector form as

$$\mathbf{z} = \mathbf{s}_i + \mathbf{n} \quad (1.101)$$

where the vectors are of length two corresponding to the outputs of the two correlators. Now the maximum likelihood receiver is the one which maximizes the *a posteriori* probability. That is

$$\hat{\mathbf{s}} = \max_{\mathbf{s}} P(\mathbf{s} | \mathbf{z}) \quad (1.102)$$

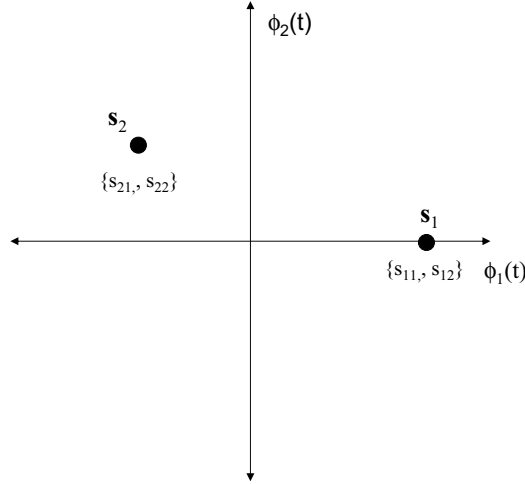


Figure 1.15: Distances for Constellation Diagram for Arbitrary Binary Modulation Scheme

In general, the *a posteriori* probability is difficult to calculate. However, using Bayes Theorem

$$P(\mathbf{s}|\mathbf{z}) = \frac{P(\mathbf{z}|\mathbf{s})P(\mathbf{s})}{P(\mathbf{z})} \quad (1.103)$$

Now, since $P(\mathbf{z})$ is constant regardless of the choice of \mathbf{s} , we can write

$$\hat{\mathbf{s}} = \max_{\mathbf{s}} (P(\mathbf{z}|\mathbf{s})P(\mathbf{s})) \quad (1.104)$$

Further, if the symbols are all equally likely:

$$\hat{\mathbf{s}} = \max_{\mathbf{s}} P(\mathbf{z}|\mathbf{s}) \quad (1.105)$$

Now, since the noise \mathbf{n} is a vector of uncorrelated Gaussian noise samples we can write

$$P(\mathbf{z}|\mathbf{s}_i) = \frac{1}{2\pi\sigma^2} e^{-\frac{(z_1-s_{i1})^2}{2\sigma^2}} e^{-\frac{(z_2-s_{i2})^2}{2\sigma^2}} \quad (1.106)$$

where z_1 and z_2 are the two correlator outputs and σ is the standard deviation of the Gaussian noise in each dimension of the received signal. Further, from the above equation we can see that maximizing $P(\mathbf{z}|\mathbf{s}_i)$ is equivalent to minimizing $(z_1 - s_{i1})^2 + (z_2 - s_{i2})^2$ which is equivalent to choosing the symbol which is closest to the received vector \mathbf{z} .

Now an error will occur if the noise vector projected onto the line connecting the two symbols is greater than one half the distance between the two symbols (see Figure 1.16) or in other words if the received signal is closer in two-dimensional space to the symbol that was not sent. The noise projected onto the line connecting the two symbols is

$$\nu = n_1 \cos(\theta) + n_2 \sin(\theta) \quad (1.107)$$

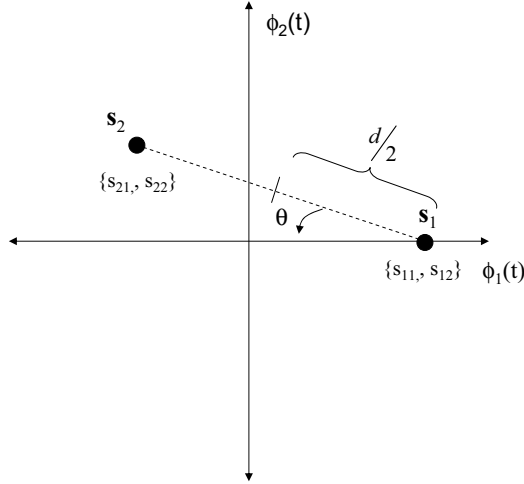


Figure 1.16: Distance Properties for Arbitrary Binary Modulation Scheme

where θ is the angle between the line connecting the two symbols and the x-axis as is shown in Figure 1.16. Further, the distance between the two symbols is

$$d = \sqrt{(s_{11} - s_{21})^2 + (s_{22})^2} \quad (1.108)$$

Now, n_1 and n_2 are zero mean Gaussian random variables (since they are the outputs of linear filters with Gaussian inputs). The variance of the noise terms is

$$\begin{aligned} E\{n_1^2\} &= E\{n_2^2\} = E\left\{\int_0^T \int_0^T n(t)\phi_1(t)n(\tau)\phi_1(\tau)dt d\tau\right\} \\ &= \int_0^T E\{n^2(t)\}\phi_1^2(t)dt \\ &= \frac{N_o}{2} \end{aligned} \quad (1.109)$$

where we have used the fact that the basis functions have unit energy. Thus, we can easily show that

$$E\{\nu^2\} = \frac{N_o}{2} \frac{1}{2} + \frac{N_o}{2} \frac{1}{2} = \frac{N_o}{2} \quad (1.110)$$

Thus, the probability of error can be found as

$$\begin{aligned} P_e &= P\left\{\nu > \frac{d}{2}\right\} \\ &= \int_{d/2}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{N_o/2}} e^{-x^2/N_o} dx \end{aligned} \quad (1.111)$$

Making the substitution $y = \sqrt{2/N_o}x$

$$\begin{aligned} P_e &= \int_{\sqrt{2/N_o}d/2}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= Q\left(\sqrt{\frac{d^2}{2N_o}}\right) \end{aligned} \quad (1.112)$$

where $Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy$ is the standard Q -function. Now, from Figure 1.16 we have

$$\begin{aligned} d^2 &= (s_{11} - s_{21})^2 + (s_{22})^2 \\ &= s_{11}^2 - 2s_{11}s_{21} + s_{21}^2 + s_{22}^2 \\ &= E_1 + E_2 - 2s_{11}s_{21} \\ &= E_1 + E_2 - 2\sqrt{E_1} \int_0^T s_2(t)\phi_1(t)dt \\ &= E_1 + E_2 - 2 \int_0^T s_2(t)s_1(t)dt \\ &= E_1 + E_2 - 2\rho_{12} \end{aligned} \quad (1.113)$$

where ρ_{12} is the correlation between $s_1(t)$ and $s_2(t)$. Substituting we have

$$\begin{aligned} P_e &= Q\left(\sqrt{\frac{E_1 + E_2 - 2\rho_{12}}{2N_o}}\right) \\ &= Q\left(\sqrt{\frac{E_b - \rho_{12}}{N_o}}\right) \end{aligned} \quad (1.114)$$

where E_b is the average energy per bit (which is the same as the average energy per symbol). Now, let's examine a few specific examples of binary modulation. If the modulation scheme is BPSK, there is a single basis function and $s_1(t) = -s_2(t)$. Thus, $E_1 = E_2 = E_b$ and $\rho_{12} = -E_b$. Thus,

$$P_e^{BPSK} = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \quad (1.115)$$

If the modulation scheme is BFSK, the two symbols are orthogonal, $E_1 = E_2 = E_b$ and thus $\rho_{12} = 0$ and

$$P_e^{BFSK} = Q\left(\sqrt{\frac{E_b}{N_o}}\right) \quad (1.116)$$

If the modulation scheme is BASK, $\rho_{12} = 0$ and

$$P_e^{BASK} = Q\left(\sqrt{\frac{E_b}{N_o}}\right) \quad (1.117)$$

From equation (1.114) we can see that the minimum probability of error occurs when the signals are antipodal ($\rho_{12} = -E_b$). Thus, BPSK provides the minimum probability of error for binary modulation.

1.7.6 Non-coherent Demodulation

The performance of the modulation schemes considered so far assume that a coherent reference is available at the receiver. While all of the modulation schemes considered can be demodulated coherently, it is very often more practical to build non-coherent receivers. FSK and ASK easily allow envelope detection which doesn't require a coherent reference. Further, PSK allows for differential encoding and detection. In Differential PSK (DPSK) the change in phase represents the data. A receiver does not need a coherent reference since it must only compare the current signal phase with the previous signal phase. The benefit of such schemes is that they allow for much simpler (and thus inexpensive) receivers. The downside is that they provide inferior performance. It can be shown that the performance of binary DPSK can be approximated by [1]

$$P_b \approx \frac{1}{2} e^{-\frac{E_b}{N_o}} \quad (1.118)$$

Similarly, the performance of non-coherent BASK and BFSK can be approximated by [1]

$$P_b \approx \frac{1}{2} e^{-\frac{1}{2} \frac{E_b}{N_o}} \quad (1.119)$$

A comparison of coherent BPSK and BFSK with non-coherent BFSK and binary DPSK is plotted in Figure 1.17. As we can see, non-coherent demodulation costs the system approximately 1dB in performance at high SNR with larger losses experienced at lower E_b/N_o values.

1.7.7 M -ary Modulation

The probability of error for M -ary modulation schemes can be found in a similar manner to binary modulation by using the Union Bound [1]. The Union Bound says that the probability of the union of several events is less than or equal to the sum of the individual event probabilities. We can use this in determining the probability of symbol error for M -ary modulation symbols by bounding the error probability by the sum of the individual pair-wise error probabilities:

$$\Pr \{ \hat{s} \neq s_i | s = s_i \} \leq \sum_{j \neq i} \Pr \{ \hat{s} = s_j | s = s_i \} \quad (1.120)$$

Using this approximation we can bound the probability of symbol error for MPSK using the error probability given above for binary modulation. Additionally, it can be shown that the above bound is dominated by the pairwise error probability $\Pr \{ \hat{s} = s_j | s = s_i \}$ corresponding to the two nearest symbols.

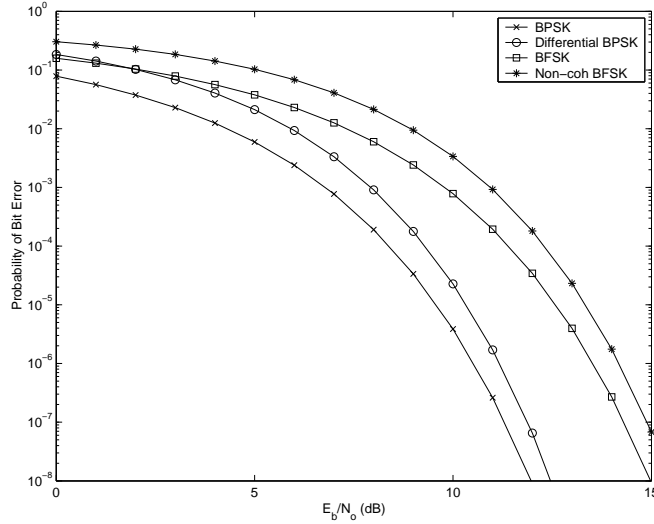


Figure 1.17: Comparison of the Performance of Coherent and Non-coherent (or differentially coherent) Reception

Specifically, by realizing that each symbol has two adjacent neighbors and noting that the distance between symbols is $d = 2\sqrt{E_s}\sin(\pi/M)$, we can write the probability of symbol error using the Union Bound as

$$P_s \leq 2Q\left(\sqrt{\frac{2E_b}{N_o} \log_2 M \left\{\sin\left(\frac{\pi}{M}\right)\right\}}\right) \quad (1.121)$$

Similarly, the probability of error for MFSK can be found (using the fact that the constellation shows $M - 1$ nearest neighbors) as

$$P_s \leq (M - 1)Q\left(\sqrt{\frac{E_b}{N_o} \log_2 M}\right) \quad (1.122)$$

The probability of symbol error for QAM constellations (when the constellation is square) can be shown to be [1]

$$P_s = 1 - \left(1 - 2\left(1 - \frac{1}{\sqrt{M}}\right)Q\left(\sqrt{\frac{3 \log_2 M E_b}{M - 1 N_o}}\right)\right)^2 \quad (1.123)$$

1.8 Bandwidth Efficiency and Energy Efficiency

The two main characteristics of modulation schemes that need to be considered by the communications engineer are bandwidth efficiency and energy efficiency. Bandwidth efficiency can be defined as the bit rate per bandwidth or bits/sec/Hz

for some definition of bandwidth. First let us consider two-dimensional modulation schemes such as PSK, ASK, and QAM. In these modulation schemes, there are only two dimensions (i.e., two basis functions). If square pulses are used, the first null-to-null bandwidth is an appropriate bandwidth measure and is

$$B = 2R_s \quad (1.124)$$

where R_s is the symbol rate and is related to the bit rate by $R_s = \frac{R_b}{\log_2 M}$. Thus, the bandwidth efficiency in terms of null-to-null bandwidth can be written as

$$\begin{aligned} \eta_{bw} &= \frac{R_b}{B} \\ &= \frac{R_b}{2R_s} \\ &= \frac{\log_2(M) R_b}{2R_b} \\ &= \frac{\log_2(M)}{2} \end{aligned} \quad (1.125)$$

Thus, we have

$$\eta_{bw} = \frac{\log_2(M)}{2} \quad \text{MPSK, MQAM, MASK} \quad (1.126)$$

If optimal pulse shaping is used, i.e., sinc pulses, then bandwidth efficiency is (in terms of absolute bandwidth) $\eta_{bw} = \log_2(M)$. Thus, for schemes with a fixed number of dimensions, bandwidth efficiency increases with M . Conversely, for orthogonal modulation schemes (e.g., MFSK) bandwidth efficiency decreases with M . This is due to the fact that orthogonal modulation schemes require an additional dimension for each additional symbol. The null-to-null bandwidth for MFSK depends on whether or not phase synchronous carriers are used for each symbol. In the best case (phase synchronous carriers) the minimum frequency separation is $\Delta f = \frac{1}{2T_s}$ where T_s is the symbol duration. Since there are $M - 1$ intervals and R_s Hz on either end (assuming square pulses), the null-to-null bandwidth is approximately

$$B = \frac{(M - 1) R_s}{2} + R_s \quad (1.127)$$

Since $R_s = \frac{R_b}{\log_2 M}$,

$$\eta_{bw} = \frac{2 \log_2(M)}{M+1} \quad \text{MFSK} \quad (1.128)$$

Thus, while for PSK, QAM and ASK bandwidth efficiency increases with M , the opposite is true for orthogonal (or bi-orthogonal) signals. Note that these trends are true regardless of pulse shape. However, changing the pulse shape for a given modulation scheme can improve the bandwidth efficiency. Since with a matched filter receiver, there is no performance penalty for changing pulse shape, it makes sense to use the most bandwidth efficient pulse possible (limited by implementation complexity).

Energy efficiency is typically defined as the value of $\frac{E_b}{N_p}$ required to obtain a specified probability of bit error. The probability of symbol error for MPSK, QAM and MFSK are given in equations (1.121), (1.123) and (1.122) respectively. For MPSK and QAM the probability of bit error can be found by assuming the use of Gray coding. Gray coding maps bits to symbols such that adjacent symbols differ by one bit only. Since the symbol error probability is dominated by the nearest neighbors, the probability of bit error, P_b , can be approximated as

$$P_b \approx \frac{P_s}{M} \quad (1.129)$$

since one of M bits will be in error for each symbol error. Since in FSK all symbols are nearest neighbors, all bit error patterns are equally likely and thus $P_b = P_s \frac{M}{2(M-1)} \approx \frac{P_s}{2}$. Using these relationships and the probability of symbol error expressions from equations (1.121) and (1.122), we can see that for PSK (and for ASK and QAM) the probability of symbol error increases with M . This is due to the fact that the symbols move closer together as we increase M since the number of dimensions in the symbol space is fixed. FSK on the other hand behaves in an opposite manner. Energy efficiency improves with M which may seem counter-intuitive at first glance. As M increases, the distances between symbols remains constant in terms of the energy per symbol, although the number of nearest neighbors increases. This increases the probability of symbol error directly with M . However, since the number bits per symbol increases, the distance between symbols in terms of E_b increases with $\log_2 M$ and thus the argument inside the Q -function increases with $\log_2 M$. The net result is that probability of bit error improves with M , thus improving energy efficiency. A comparison of MPSK and MFSK is plotted in Figure 1.18 and summarized in Table 1.1 for various values of M . Note that (a) QAM provides better energy efficiency than PSK for the same bandwidth efficiency, and (b) the performance of FSK assumes coherent reception and is also applicable to any M -ary orthogonal modulation scheme. Further MFSK provides a means for improving energy efficiency beyond the performance of BPSK, but requires substantial bandwidth cost.

1.9 Capacity

In the previous section we found that there is a trade-off between energy efficiency and bandwidth efficiency. This raises the question: What is the optimal trade-off between energy and bandwidth? Claude Shannon in his classic paper [10] showed that error-free communication is possible in an AWGN channel provided that the signaling rate does not exceed a certain rate termed the channel *capacity*. The channel capacity is dependent on the channel bandwidth and the

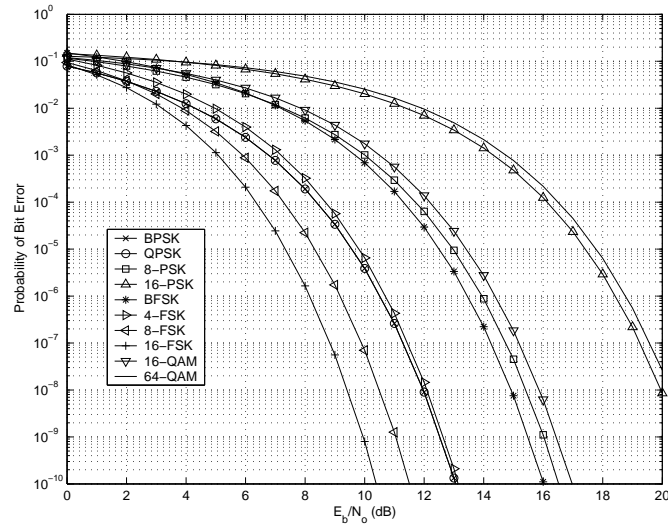


Figure 1.18: Performance of MPSK, QAM, and MFSK in AWGN with Coherent Reception

signal-to-noise ratio:

$$C = B \log_2 \left(1 + \frac{S}{N} \right) \quad (1.130)$$

$$= B \log_2 \left(1 + \frac{S}{N_o B} \right) \quad (1.131)$$

This provides us with the optimal relationship between energy efficiency and bandwidth efficiency. Specifically, the spectral efficiency can be related to the E_b/N_o required to achieve that spectral efficiency:

$$\frac{C}{B} = \log_2 \left(1 + \frac{C E_b}{B N_o} \right) \quad (1.132)$$

This relationship now tells us the best spectral efficiency achievable at a particular E_b/N_o , or equivalently the lowest value of E_b/N_o required for good performance at a specified spectral efficiency. Is there a lower limit on the required E_b/N_o for any spectral efficiency or can we achieve some rate regardless of how low our E_b/N_o is? We will examine this in the next example.

Table 1.1: Comparison of Modulation Schemes

Modulation Scheme	Energy Efficiency (E_b/N_o for $P_b = 10^{-5}$)	BW Efficiency (b/s/Hz)
BPSK	9.5dB	0.5
QPSK	9.5dB	1
8-PSK	13dB	1.5
16-PSK	17.5dB	2
BFSK	12.5dB	0.4
4-FSK	10dB	0.57
8-FSK	8.25dB	0.55
16-FSK	7.25dB	0.42
4-QAM	9.5dB	2
16-QAM	13.5dB	2.5
64-QAM	17.75dB	3

Example 1.6 Determine the minimum value of E_b/N_o that will achieve capacity. At what bandwidth efficiency is this achieved?

SOLUTION: Solving (1.132) for E_b/N_o we have

$$\frac{E_b}{N_o} = \frac{2^{C/B} - 1}{C/B} \quad (1.133)$$

which says that E_b/N_o increases exponentially with C/B . Thus, the minimum value of E_b/N_o can be found by letting C/B approach zero:

$$\left(\frac{E_b}{N_o}\right)_{\min} = \lim_{C/B \rightarrow 0} \frac{2^{C/B} - 1}{C/B} \quad (1.134)$$

$$= \ln(2) \quad (1.135)$$

Thus, the minimum value of $\frac{E_b}{N_o}$ that achieves capacity is $\ln(2) = 0.64$ or -1.6 dB. However, this occurs as the bandwidth efficiency C/B approaches zero.

Example 1.7 We showed previously that MFSK was one means of improving energy efficiency. What is the best energy efficiency achievable by MFSK and at what spectral efficiency is it achieved? How does this compare to capacity?

SOLUTION: The performance of MFSK was stated previously as

$$P_s \leq (M - 1)Q \left(\sqrt{\frac{E_b}{N_o} \log_2 M} \right) \quad (1.136)$$

which can be upper bounded as

$$P_s \leq (M - 1)Q \left(\sqrt{\frac{E_b}{N_o} \log_2 M} \right) < 2e^{-\frac{1}{2} \frac{E_b}{N_o} \log_2 M} = e^{-\log_2 M \left(\sqrt{\frac{E_b}{N_o}} - \sqrt{\ln 2} \right)^2} \quad (1.137)$$

Thus, the probability of error for MFSK will approach zero as $M \rightarrow \infty$ provided that $\frac{E_b}{N_o} > \ln 2 = 0.69$. Thus, we see that MFSK approaches capacity as $M \rightarrow \infty$ but unfortunately, this is at vanishingly small spectral efficiency.

As demonstrated by the previous example, modulation can achieve capacity at extremely low spectral efficiencies. A more complete comparison between modulation and capacity is shown in Figure 1.19. Specifically, the figure plots the spectral efficiency (assuming ideal pulse shaping) versus the E_b/N_o required to achieve a bit error rate of 10^{-4} . From the plot we can see that for spectral efficiencies greater than 1.0 b/s/Hz, modulation alone does not come within 8dB of capacity. However, as spectral efficiency decreases, modulation does approach capacity. In fact, at a spectral efficiency of 0.1 b/s/Hz M -ary orthogonal signaling (e.g., MFSK) comes within 4dB of capacity. The previous example showed that as we let spectral efficiency approach zero, the energy efficiency of MFSK will approach capacity (i.e., -1.6dB).

1.10 Coding

As stated in the previous section modulation approaches capacity only at low values of spectral efficiency. This is due to the fact that modulation does not use the dimensionality of the transmitted signal effectively. Coding is a technique which improves the distance properties of a transmit signal without large increases in the dimensionality (i.e., the bandwidth) of the signal. This is achieved by introducing memory. Specifically, coding is achieved by either (a) mapping k information bits to n coded bits (block codes) or (b) by passing the information sequence through a finite state machine which maps bits to a sequence of coded bits (convolutional codes). In this brief review of digital communications, we will not go into the details of coding techniques. However,

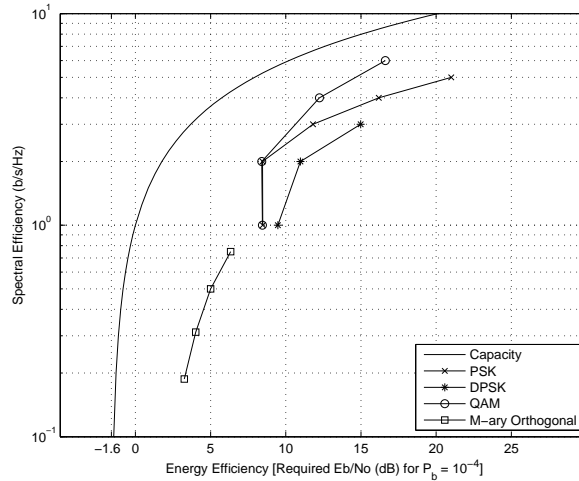


Figure 1.19: Spectral and Energy Efficiency of Modulation Schemes Compared to Capacity

the interested reader is referred to one of many excellent coding texts [1, 11]. Figure 1.20 plots the spectral efficiency (assuming ideal pulse shaping) versus the E_b/N_0 required to achieve a bit error rate of 10^{-4} for various modulation schemes when combined with convolutional codes with different rates. We can see that the use of coding allows communication signals to more closely approach capacity. Specifically, the combination of coding and modulation comes within 4dB of capacity at much better spectral efficiencies than modulation alone (e.g., 1-4 b/s/Hz).

1.11 Fading Channels

The final review topic that we will discuss in this brief review of digital communications is multipath fading or the fading channel. One type of channel that is well suited to spread spectrum is the fading channel. In order to understand the benefits afforded by spreading spectrum in fading channels, we must first review the principles behind multipath channels in general. By modeling the channel as a time-varying linear system, the received signal $r(t)$ can be written as [5][12]

$$r(t) = s(t) \otimes h(t, \tau) + n(t) \quad (1.138)$$

$$= \int_{-\infty}^{\infty} s(t - \tau) h(t, \tau) d\tau + n(t) \quad (1.139)$$

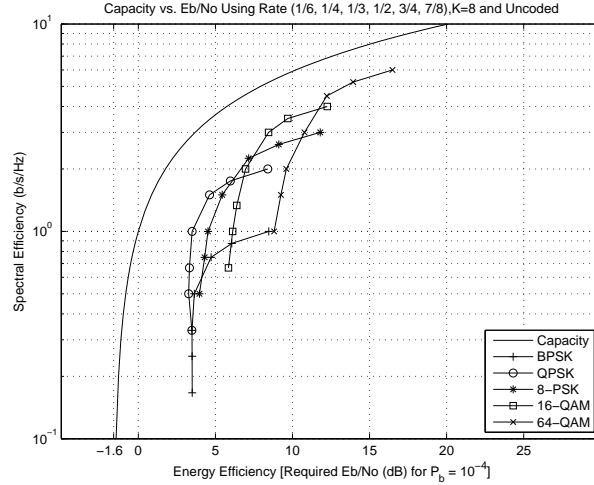


Figure 1.20: Spectral Efficiency versus E_b/N_o Required to Achieve a Bit Error Rate of 10^{-4} for Various Combinations of Modulation and Convolutional Coding Techniques

where $s(t)$ is the transmit waveform, $n(t)$ is AWGN and $h(t, \tau)$ is the time-varying impulse response of the channel. The channel impulse response is typically modeled as a series of impulses as shown in Figure 1.21 where t represents time and τ represents delay.

Multipath fading results when two multipath components arrive with a relative time offset that is much smaller than a symbol interval (i.e., the two paths are *non-resolvable*) but large relative to a single cycle of the carrier frequency. For example, if the symbol rate is 1Mbps and the carrier frequency is 2GHz, a relative time offset between two multipath components of 1.25ns is insignificant relative to the symbol duration of $1\mu s$ but corresponds to a phase difference of 7.9 radians. As a result, when non-resolvable multipath components combine non-coherently, they have the potential of annihilating each other as shown in Figure 1.22. The plot shows the envelope of signal with two multipath components as the delay increases as a fraction of the carrier frequency in radians. When the delay is $\frac{1}{2f_c}$ the two paths are opposite in phase and thus cancel each other resulting in a deep fade. In a mobile scenario, the relative delays will change with time causing the signal to go in and out of fades.

The mobile's movement also causes a second, related phenomenon known as Doppler shift. As shown in Figure 1.23, due to the mobile unit's movement (whether it is the transmitter or the receiver), each multipath component undergoes a (potentially different) phase shift. Specifically, for a mobile unit traveling at velocity v at an angle ϕ relative to the incoming signal (in the azimuthal plane), the phase change over Δt seconds is

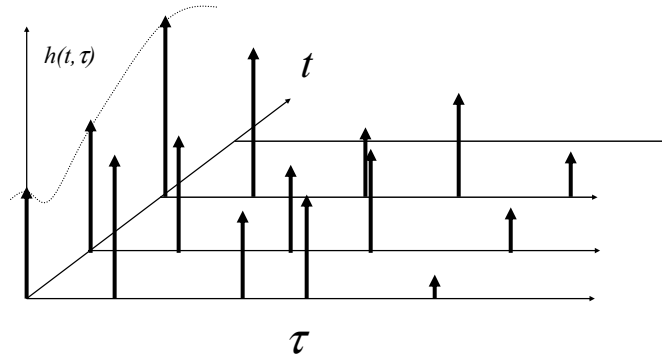


Figure 1.21: Time-Varying Impulse Response of the Multipath Channel

$$\begin{aligned}\Delta\theta &= \frac{2\pi}{\lambda}d\cos(\phi) \\ &= \frac{2\pi}{\lambda}v\Delta t\cos(\phi)\end{aligned}$$

where λ is the carrier wavelength and d is the distance traveled in Δt seconds. Due to this phase shift over Δt seconds, the signal experiences a frequency shift of

$$\begin{aligned}f_d &= \frac{1}{2\pi} \frac{\Delta\theta}{\Delta t} \\ &= \frac{v}{\lambda} \cos(\phi)\end{aligned}$$

and we call this frequency shift a *Doppler shift*. If there is a large number of unresolvable multipath components arriving at the mobile from all directions as shown in Figure 1.24, the signal will experience a large number of Doppler shifts, each corresponding to a different multipath component. This results in a *Doppler spread* which can be seen by examining the Doppler spectrum of the channel. The resulting signal will fade randomly over time as shown in Figure 1.25. The complex baseband channel representation of such a channel can be written as

$$h(t, \tau) = \sqrt{\frac{1}{N}} \sum_{i=1}^N e^{j(2\pi f_i t + \theta_i)} \delta(\tau - \tau_i)$$

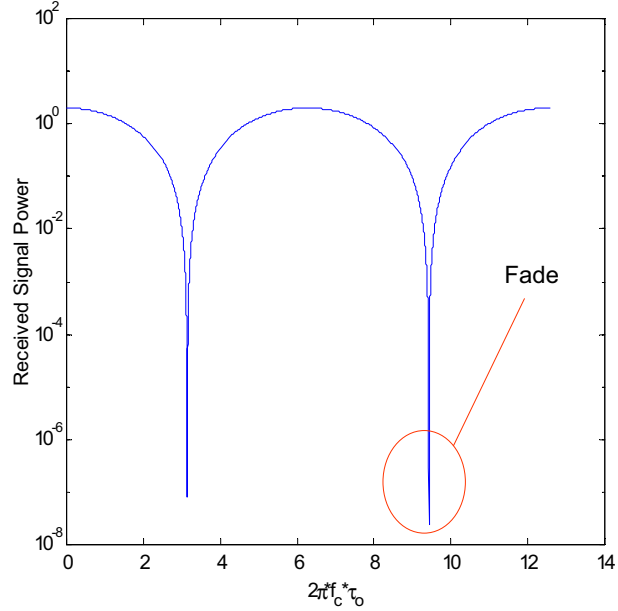


Figure 1.22: Envelope Fading Due to Two Path Channel

where f_i , θ_i and τ_i are the Doppler shift, phase shift, and relative delay associated with the i th multipath. Provided that the differences between the delays are very small as compared to the symbol duration (what is called *flat fading*), the paths all appear to arrive simultaneously and we can model the channel impulse response as

$$h(t, \tau) = \sqrt{\frac{1}{N}} \left\{ \sum_{i=1}^N e^{j(2\pi f_i t + \theta_i)} \right\} \delta(\tau) \quad (1.140)$$

The resulting received complex baseband signal can then be modeled as

$$\begin{aligned} r(t) &= h(t, \tau) \otimes s(t) \\ &= \gamma(t) s(t) \end{aligned}$$

where

$$\gamma(t) = \sqrt{\frac{1}{N}} \left\{ \sum_{i=1}^N e^{j(2\pi f_i t + \theta_i)} \right\} \quad (1.141)$$

is modeled as a complex Gaussian random process (due to the sum of a large number of complex exponentials) and represents multiplicative distortion. Note that the average value of $|\gamma(t)|^2$ is unity, as can be seen from (1.141). An example received signal envelope is shown in 1.25. Assuming that the multipath

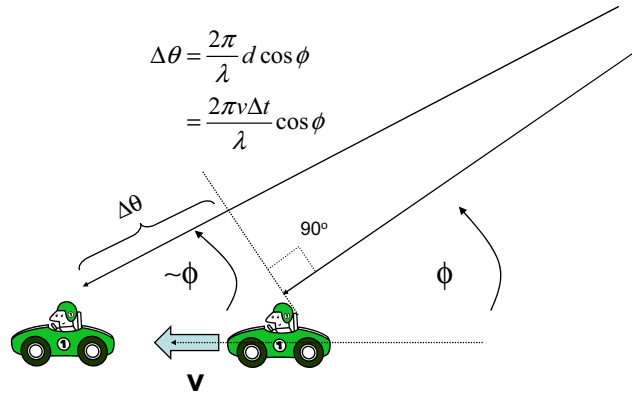


Figure 1.23: Illustration of the Doppler Shift

components are uniformly distributed in angle around the mobile, the autocorrelation function of the channel can be shown to be [13]

$$R(\tau) = J_0(2\pi f_m \tau)$$

where f_m is the maximum Doppler spread and $J_0(x)$ is the zeroth order modified Bessel function. As the Doppler spread increases, the autocorrelation function changes more rapidly thus indicating faster fading. The corresponding power spectral density of the channel is [13]

$$P_C(f) = \frac{1}{\pi f_m \sqrt{1 - \left(\frac{f}{f_m}\right)^2}}$$

Finally, we should note that since $\gamma(t)$ is a complex Gaussian random process, samples of the process are complex Gaussian random variables. The envelope (or magnitude) of a complex Gaussian random variable, follows a Rayleigh distribution. Thus, this channel model is often referred to as a *Rayleigh fading channel*.

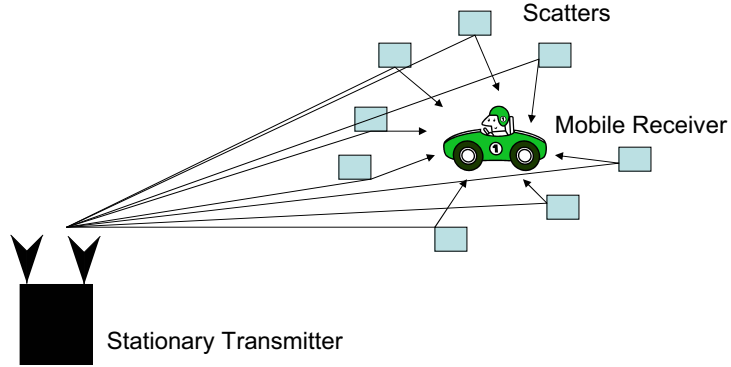


Figure 1.24: Physical Scenario that Provides the Doppler Spectrum

The performance of a spread spectrum signal in Rayleigh fading can be determined in a manner similar to the AWGN case, with the exception that we now must integrate the probability of error over the SNR distribution due to fading. The decision statistic for BPSK modulation can be written as the output of the correlator version of the matched filter:

$$\begin{aligned}
 Z &= \frac{1}{T} \int_0^T r(t) s^*(t) dt \\
 &= \frac{1}{T} \int_0^T (\sqrt{P} s(t) \gamma(t) + n(t)) s^*(t) dt \\
 &= \frac{\sqrt{P} \gamma}{T} \int_0^T s(t) s^*(t) dt + \frac{1}{T} \int_0^T n(t) s^*(t) dt \\
 &= \sqrt{P} \gamma b + N
 \end{aligned}$$

where we have assumed that the channel $\gamma(t)$ remains constant over a symbol interval and is represented by γ . Since the baseband channel is complex, we must remove the channel-induced phase modulation in order to do phase detection. Multiplying the matched filter output by the complex conjugate of the channel gain γ^* , and assuming BPSK modulation, the probability of error conditioned on γ is

$$P_e | \gamma = Q \left(\sqrt{\frac{2E_b}{N_o}} |\gamma|^2 \right)$$

Since γ is a complex Gaussian random variable, $|\gamma|^2$ is a Chi-Square random variable with two degrees of freedom. The average probability of error is conditional probability of error averaged over the distribution of $|\gamma|^2$. The distribution of $\beta = \frac{E_b}{N_o} |\gamma|^2$ is

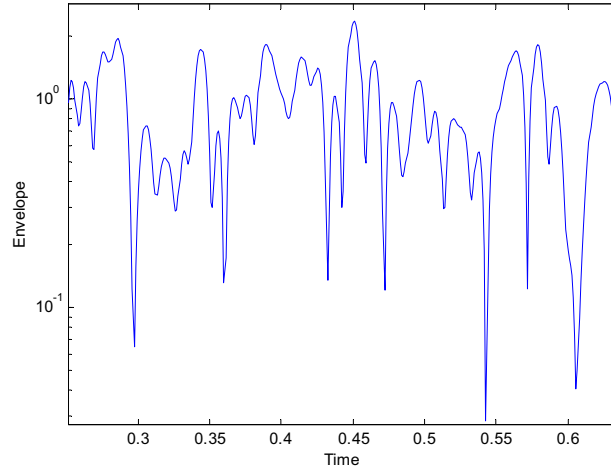


Figure 1.25: Illustration of Received Signal Envelope in Multipath Fading Channel

$$p(\beta) = \frac{1}{\bar{\beta}} e^{-\frac{\beta}{\bar{\beta}}}$$

Since $\overline{|\gamma|^2} = 1$, $\bar{\beta} = \frac{E_b}{N_o}$ which is the *average* energy per bit divided by the noise power spectral density. The average probability of error is then found by integrating over the probability density function of β :

$$\begin{aligned} P_e &= \int_0^{\infty} p(\beta) Q(\sqrt{2\beta}) d\beta \\ &= \int_0^{\infty} \frac{1}{\bar{\beta}} e^{-\frac{\beta}{\bar{\beta}}} Q(\sqrt{2\beta}) d\beta \\ &= \frac{1}{2} \left(1 - \sqrt{\frac{\bar{\beta}}{1 + \bar{\beta}}} \right) \end{aligned} \quad (1.142)$$

where we have used a well-known integration identity [1]. If $\bar{\beta} \gg 1$, we can use the approximation:

$$\begin{aligned} P_e &= \frac{1}{2} \left(1 - \frac{1}{\sqrt{1 + \frac{1}{\bar{\beta}}}} \right) \\ &\cong \frac{1}{2} \left(1 - \left(1 - \frac{1}{2\bar{\beta}} \right) \right) \\ &= \frac{1}{4\bar{\beta}} \end{aligned}$$

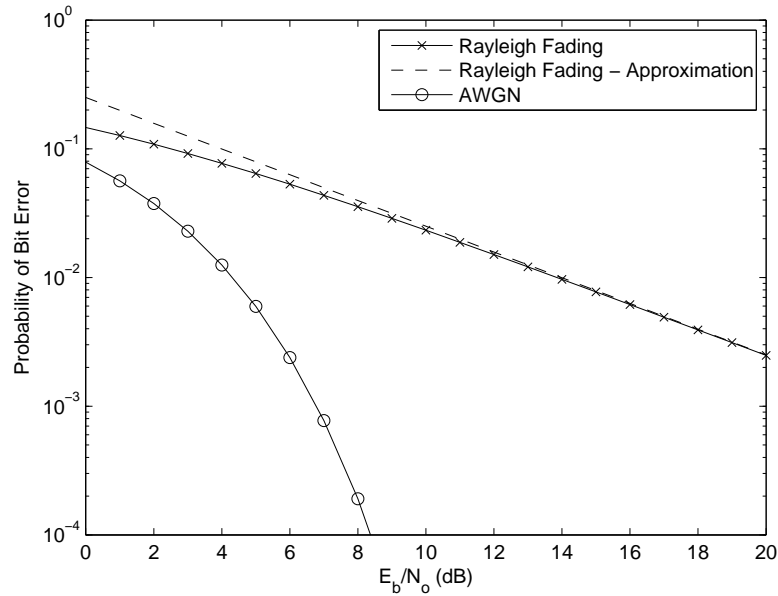


Figure 1.26: Bit Error Rate of BPSK in Flat Rayleigh Fading

The probability of error and the high SNR approximation are both plotted in Figure 1.26. Also plotted is the BER of coherently demodulated BPSK in AWGN from (1.115). As can be seen, the performance in Rayleigh fading is dramatically degraded as compared to the AWGN case. For example, if the desired BER for a link is 10^{-2} , in an AWGN channel the required value of E_b/N_o is approximately 4dB. However, in the presence of Rayleigh fading, we require an E_b/N_o of 14dB, a 10dB increase! As we will find in Chapter 8, spread spectrum will allow us to improve this situation through the use of a Rake receiver and through the use of error correction coding.

1.12 Conclusion

In this chapter we have provided a brief overview of digital communications. Specifically, we have discussed fundamental topics including sampling, quantization, modulation/demodulation, coding, bandwidth efficiency, energy efficiency and multipath fading. Throughout this book, we will assume that the reader has a basic understanding of these and other fundamental aspects of digital communications. For those readers who feel that they need additional review, please refer to the following excellent text books on Digital Communications [1, 2, 5, 3, 4].

Chapter 2

Motivation for Spread Spectrum

2.1 Introduction

The digital communication techniques discussed in the previous chapter were designed for the AWGN channel. When communication must occur in less benign environments, more sophisticated techniques are warranted [14][15][16][17]. Two such environments that are often encountered in communications are the interference-limited environment (often termed the jamming channel in the context of spread spectrum) and the fading environment [5]. One communications technique that improves performance in both of these channel types is spread spectrum. Spread spectrum can be defined as any modulation technique which (a) occupies a bandwidth which is (usually substantially) beyond what is necessary for the data rate being transmitted and (b) uses a pseudorandom¹ signal independent of the data to obtain the increased bandwidth and to recover the spread signal at the receiver. The latter factor distinguishes spread spectrum techniques from standard communication techniques such as FM and high order orthogonal signaling which may also require high bandwidth as compared to the information rate. The pseudorandom spreading is accomplished by adding a spreading block after the modulator as shown in Figure 2.1. Note that spreading can either be done at an intermediate frequency or more commonly at baseband. The pseudorandom nature of the spreading signal means that the receiver recovers the signal by correlating the received signal with a synchronized local version of the spreading waveform (also called despreading) as also shown in Figure 2.1. It should be emphasized that the benefits of spread

¹The reader may find the term "pseudorandom" somewhat strange and thus it deserves a brief discussion. By the term "pseudorandom signal" we simply mean a signal that appears to an outside observer to be random, while in fact being generated by a deterministic algorithm. Such a signal will clearly not be random to a receiver which knows the algorithm, but appears to be random to a receiver which does not know the algorithm.

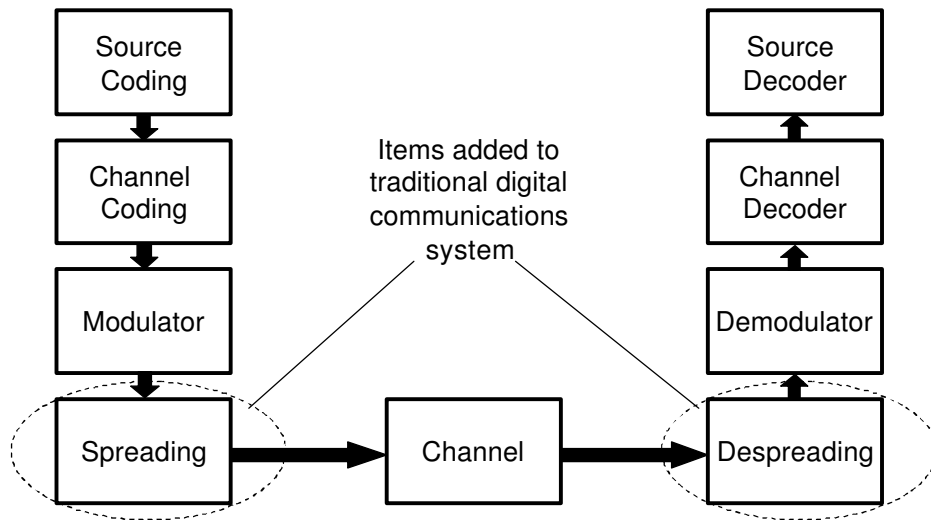


Figure 2.1: Block Diagram of a Digital Communications System with Spreading/Despreading

spectrum techniques are derived from the combination of these two characteristics, i.e., the wide bandwidth (and thus low power spectral density) and the pseudorandom nature of the signal. This will be discussed at a high level in the current chapter and in significantly more detail in the coming chapters.

2.2 Reasons for Spreading

There are several reasons for spreading a signal to a bandwidth well beyond the information rate in a pseudorandom manner. These include resistance to interference (intentional or otherwise), the potential to hide the signal from unintended reception (i.e., low probability of intercept or LPI), resistance to multipath fading, improved multiple access capability, and ranging [18]. The resistance to interference including multiple access interference is related to two more subtle properties of spread spectrum termed *interference averaging* and *interference randomization*. We will discuss these two concepts explicitly. In the following sections we will discuss each of these motivations for spread spectrum. Each of these benefits naturally are derived from the two key characteristics of spread spectrum, its wide spectral occupancy and/or the pseudorandom nature of the signal. By occupying a large bandwidth in a pseudorandom manner, over a long period of time the signal is spread over many dimensions in an unpredictable (to an outside observer) manner. However, during any short time interval (say a few data symbols), the signal occupies only a single dimension. Since individual interference sources are concentrated in one dimension or spread over many dimensions at once, the spread spectrum signal experiences

an improved signal-to-interference ratio (SIR) upon reception. Additionally, this increased dimensionality means that an unintended receiver which does not know the pseudorandom code, must observe all dimensions simultaneously resulting in a very poor signal-to-(thermal)noise ratio (SNR). We will now discuss these benefits of spread spectrum in more detail.

2.2.1 Anti-jamming

Jamming is a term typically used when an RF signal source of some type is used to disrupt the communications of an enemy link. Spread spectrum has long been used in military communications applications as a means of combatting jamming [19]. Although we will describe this benefit in those terms, it should be noted that the anti-jamming capabilities of spread spectrum can also be seen as interference resistance in commercial applications [20][21]. This will be a general theme in our discussion of spread spectrum in this book. Although it originated as a military technology, the same characteristics which make it useful in military applications can also provide powerful benefits in commercial applications.

Jamming is done in various ways including sending a continuous, strong narrowband interferer, wideband noise or a pulsed jamming signal. Each has its advantages and disadvantages. We will examine each of these approaches in Chapter 6, but for now let us consider a pulsed signal with duty cycle ρ where continuous jamming is a special case of pulsed jamming with $\rho = 1$. Now, without jamming (and assuming an AWGN channel), the performance of BPSK modulation (as discussed in Chapter 1) is

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_o}}\right) \quad (2.1)$$

Let us assume that a wideband jammer transmits a noise-like signal which is received with average power J in bandwidth W . While the jammer is transmitting, the performance of a receiver using BPSK modulation is [16]

$$P_b = Q\left(\sqrt{\frac{2E_b}{N_o + \frac{N_j}{\rho}}}\right) \quad (2.2)$$

where $N_j = \frac{J}{W}$ is the one-sided power spectral density of the jammer. Note that the jammer increases its impact while transmitting by transmitting less frequently (i.e., decreasing ρ). This concentrates the jammer's power in time. However, by decreasing ρ the jammer impacts performance less frequently. Thus, there is a trade-off by changing ρ . Combining equations (2.1) and (2.2), the average performance of a BPSK link in the presence of pulsed noise-like jamming is

$$P_b = (1 - \rho) Q\left(\sqrt{\frac{2E_b}{N_o}}\right) + \rho Q\left(\sqrt{\frac{2E_b}{N_o + \frac{N_j}{\rho}}}\right) \quad (2.3)$$

In a typical jamming situation, $\frac{E_b}{N_o} \gg \frac{E_b}{N_j}$, thus

$$P_b \approx \rho Q \left(\sqrt{\frac{2\rho E_b}{N_j}} \right) \quad (2.4)$$

The optimal choice of ρ (in terms of maximizing P_b) can be found by using an upper bound for the Q -function. Namely, it can be shown that [22]

$$Q(x) \leq \frac{1}{x\sqrt{2\pi}} e^{-x^2/2} \quad (2.5)$$

Using this upper bound, we can bound the probability of error as [16]

$$P_b \leq \frac{\rho}{\sqrt{4\pi\rho E_b/N_j}} e^{-\rho E_b/N_j} \quad (2.6)$$

The optimal choice of ρ can be determined by taking the derivative of this expression and setting it equal to zero. In Chapter 8 the solution is shown to be

$$\rho_{opt} = \frac{N_j}{2E_b} \quad (2.7)$$

The resulting probability of error bound is then [16]

$$P_b^{max} \leq \frac{1}{\sqrt{2\pi e}} \frac{N_j}{2E_b} \quad (2.8)$$

In the case of continual jamming the performance from (2.3) is

$$P_b = Q \left(\sqrt{\frac{2E_b}{N_j}} \right) \quad (2.9)$$

As shown in Figure 2.2 optimal pulsing severely degrades the performance of the victim receiver as compared to continuous jamming. For example, at a target error rate of $P_b = 10^{-4}$ optimal jamming degrades performance by over 22dB. It should be noted that optimal jamming is difficult to accomplish in practice since it requires the jammer to know the operating point of the receiver, but it does demonstrate the radical change that jamming can cause to receiver performance. While continuous wideband jamming is identical to AWGN, pulsed jamming results in performance where the BER reduces linearly as E_b/N_j increases, as opposed to exponential improvement. Thus, traditional narrowband modulation and coding techniques which are designed for AWGN channels are no longer appropriate. As we will show in Chapter 8, spread spectrum techniques can return the performance to an exponential improvement with increasing E_b/N_j .

As a simple example of the possible performance improvement with spread spectrum, note that $E_b = PT_b$ where P is the received signal power of the desired signal and T_b is the bit duration and $N_j = \frac{J}{W}$.

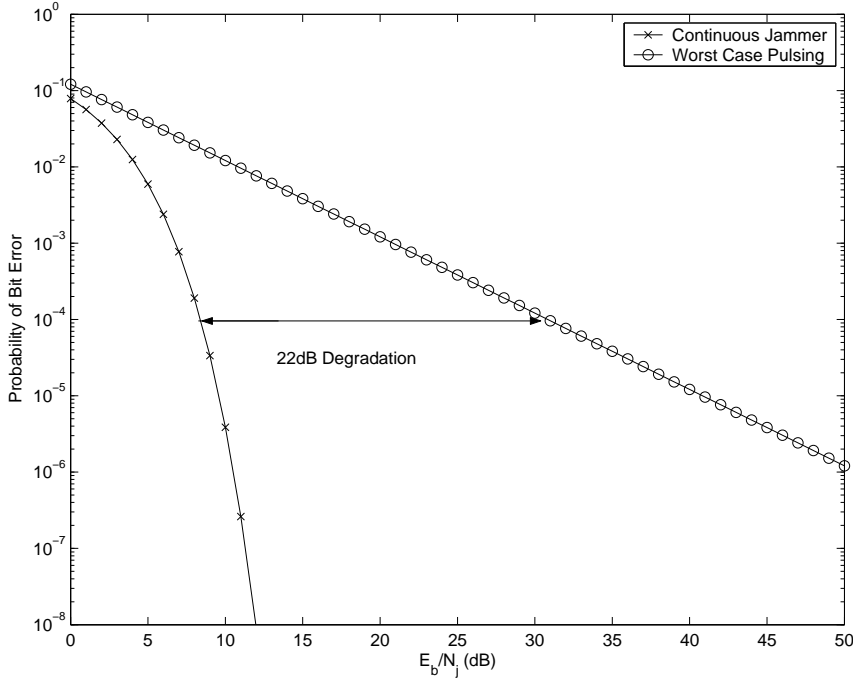


Figure 2.2: Impact of Optimal Pulsing on the Jammer's Effectiveness

Using these definitions we can re-write equation (2.8) as

$$P_b^{\max} = \frac{1}{\sqrt{2\pi e}} \frac{J}{2P} \frac{R_b}{W} \quad (2.10)$$

where $R_b = \frac{1}{T_b}$ is the bit rate, P/J is the signal-to-interference ratio and the ratio $\frac{W}{R_b}$ can be termed the spreading factor since it describes the amount of excess bandwidth used. In other words, in traditional BPSK, $W = kR_b$ where $1 \leq k \leq 2$ depends on the pulse shape used. However, in spread spectrum $W \gg R_b$ and thus, the spreading factor is much greater than one. From equation (2.10) we can see that increasing the spreading factor directly reduces the impact of the jammer as shown in Figure 2.4. Note that the pulsing jammer is still able to impact performance such that the BER reduction is only linear with increasing E_b/N_j , but its effectiveness is reduced in direct proportion to the spreading factor. We will show later (Chapter 8) that even better improvement can be obtained by including error correction coding along with increased bandwidth.

We should note that even without optimal pulsing, jamming can have a strong impact. In Figure 2.3 the performance of a victim receiver using BPSK is plotted versus E_b/N_o with and without continuous jamming. In this case, the jammer is continuous with a resulting $\frac{E_b}{N_j} = 10dB$. The figure demonstrates

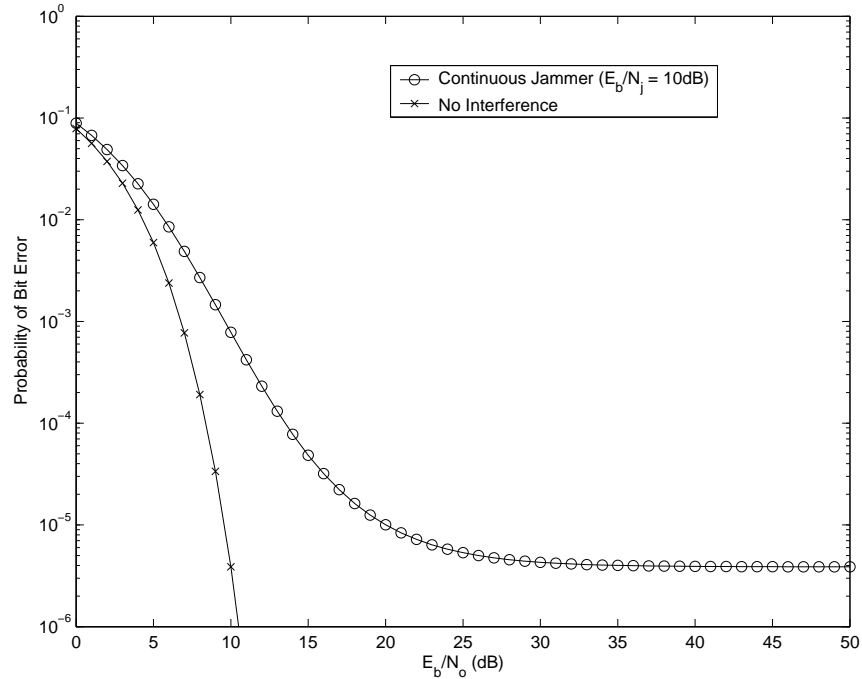


Figure 2.3: Error Floor Due to Continuous Jamming (An illustration of Interference Limited Scenario)

that as $\frac{E_b}{N_o}$ goes to infinity, the BER performance is interference-limited due to the constant $\frac{E_b}{N_j}$ with a resulting error floor. We will show later that spread spectrum can improve this scenario as well.

As a final note in this section, we want to again stress that although this benefit of spread spectrum has been presented in terms of a military application, the benefits apply equally well to commercial systems. Specifically, the anti-jam benefits of spread spectrum allow the possibility of spread spectrum signals being overlaid on existing narrowband systems. The ability to reject interference can then be exploited to allow the spread spectrum system to perform acceptably in the presence of the legacy system which appears as narrowband interference. Additionally, the same properties will be useful in multiple access situations as will be discussed in Section 2.2.4 and Chapter 10.

2.2.2 Interference Randomization

The previous section demonstrated the deleterious effects of jamming and showed that the increased bandwidth of spread spectrum can provide some benefit in terms of performance. As will be explained in detail in later chapters, this benefit is directly related to a reduction in the effective interference power due

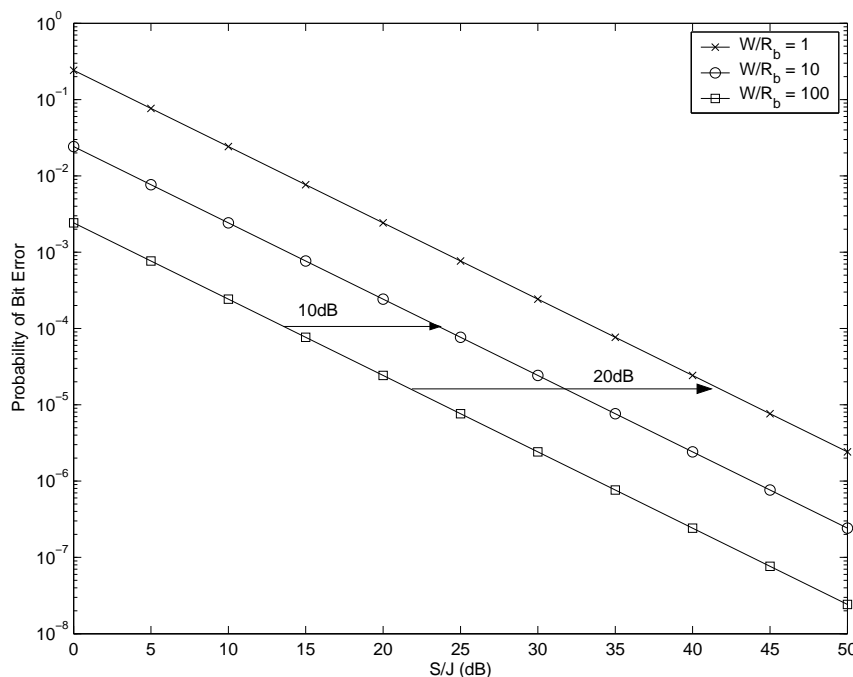


Figure 2.4: Illustration of Spreading Gain on Worst Case Jamming Performance

to spreading/despreading. An additional aspect of spread spectrum demodulation (i.e., the despreading) is randomization of the interference. Specifically, due to the random nature of the spreading waveform, the despreading process has a tendency to randomize the interference. By the term *randomization* we mean that the despreading process tends to change the interference distribution from its original distribution to a Gaussian distribution. This change in the distribution allows for better performance, especially in the presence of error correction coding when using standard receiver structures such as the matched filter [20][23]. The improved performance is due to the fact that communication systems, including modulation and coding, were designed for Gaussian interference, i.e., AWGN. Additionally, performance analysis of systems with Gaussian noise impairments is well-understood and straightforward. Thus, in this sense Gaussian interference can be preferred over other types of interference [24, 25]. The fact that the despreading process of spread spectrum tends to convert interference of any distribution to a Gaussian distribution provides some advantage. However, it should be emphasized that if more sophisticated receiver structures are employed, Gaussian interference may not result in the best performance. Specifically, several interference counter-measures (i.e., interference rejection techniques) that we will discuss in Chapters 12 and 13 can provide excellent performance when the inherent structure of the interference is

exploited.

The interference randomization property of spread spectrum is based on the Central Limit Theorem. The Central Limit Theorem is a fundamental result in statistics which is concerned with the distribution of a sum of random variables. The despreading process in spread spectrum communications essentially results in the summation of a large number of (ideally) independent random variables (i.e., interference terms). Thus, the Central Limit Theorem (CLT) becomes relevant. Specifically, the CLT states

Theorem 1 *If Y is the sum of N independent, identically distributed (iid) random variables $Y = \sum_{i=1}^N X_i$, where X has mean μ and standard deviation σ , in the limit as $N \rightarrow \infty$ the probability density function of $Z = \frac{Y/N - \mu}{\sigma/\sqrt{N}}$ is a unit normal distribution:*

$$p_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (2.11)$$

Again, since the despreading process tends to sum a large number of random variables (not always iid) the central limit theorem tells us that despreading will tend to make interference more like random noise, which can improve performance when using communication techniques designed for AWGN (e.g., the matched filter receiver). We will demonstrate an example of the benefit of Gaussian interference in the following example.

Example 2.1 *Consider the performance impact of two different interference sources on the standard matched filter. After matched filtering, the first interference source has a binary distribution. Specifically, assume that the interference at the filter output takes on the value $+a$ with probability $(1-p)$ and the value $+Xa$ with probability p where $X > 1$. The probability density function (pdf) of the interference source can be written as*

$$f_{I1}(x) = (1-p)\delta(x-a) + p\delta(x-Xa) \quad (2.12)$$

where $\delta(x)$ is standard impulse or delta function defined as

$$\begin{aligned} \delta(x) &= 0 \quad x \neq 0 \\ \int_{-\infty}^{\infty} \delta(x) dx &= 1 \end{aligned} \quad (2.13)$$

The second interference source is Gaussian distributed at the output of the matched filter with pdf

$$f_{I2}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} \quad (2.14)$$

If the k th output of the matched filter with BPSK signaling is

$$r_k = b_k + i_k \quad (2.15)$$

where $b_k \in \{+1, -1\}$ represents the desired signal and i_k represents the interference with pdfs given above, determine the probability of bit error as a function of the Signal-to-Interference Ratio (SIR) assuming a standard BPSK decision rule. Compare the two BERs for 4dB \downarrow SIR \downarrow 10dB when $p = 0.1$ and $X = 10$.

SOLUTION: First let us consider the binary case. Clearly, if $b_k = 1$, the probability of error is zero since the binary interference is always positive. If $b_k = -1$, the probability of error depends on the interference power. Since the desired signal has unit power, the signal-to-noise ratio is

$$SIR = \frac{S}{I} = \frac{1}{E\{i_k^2\}} \quad (2.16)$$

Let us now find the average interference power. This can be found from equation (2.12) as

$$E\{i_k^2\} = (1-p)a^2 + pX^2a^2 \quad (2.17)$$

The probability of error when $b_k = -1$ can be written as

$$\Pr\{e | b_k = -1\} = \Pr\{i_k > 1 | b_k = -1\} \quad (2.18)$$

Now from equations (2.16) and (2.17) we can write:

$$\Pr\{i_k > 1 | b_k = -1\} = \begin{cases} 1 & SIR < \frac{1}{1-p+pX^2} \\ p & \frac{X^2}{1-p+pX^2} > SIR > \frac{1}{1-p+pX^2} \\ 0 & SIR > \frac{X^2}{1-p+pX^2} \end{cases} \quad (2.19)$$

Thus, assuming equally likely data bits the overall BER is

$$P_b = \begin{cases} 1/2 & SIR < \frac{1}{1-p+pX^2} \\ p/2 & \frac{X^2}{1-p+pX^2} > SIR > \frac{1}{1-p+pX^2} \\ 0 & SIR > \frac{X^2}{1-p+pX^2} \end{cases} \quad (2.20)$$

For Gaussian interference we have, the probability of error given that $b_k = -1$ can be found as

$$\begin{aligned} \Pr\{i_{kz} > 1 | b_k = -1\} &= \int_1^\infty \frac{1}{\sqrt{2\pi}\sigma} e^{-x^2/2\sigma^2} dx \\ &= \int_{1/\sigma}^\infty \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy \\ &= Q\left(\frac{1}{\sigma}\right) \\ &= Q\left(\sqrt{SIR}\right) \end{aligned} \quad (2.21)$$

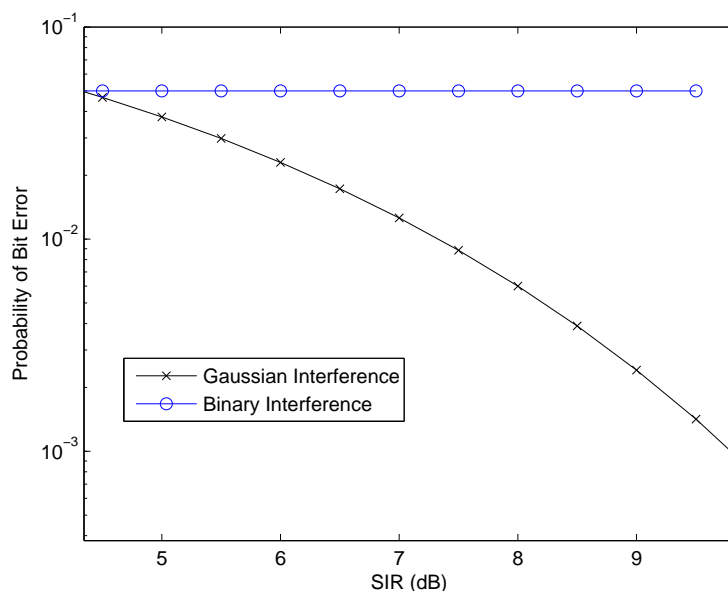


Figure 2.5: Comparison of the Effects of Binary and Gaussian Distributed Interference from Example 2.1

Since the interference is symmetric, the overall probability of error is the same as the probability of error when $b_k = -1$. A comparison of the two BERs (for $p = 0.1$ and $X = 10$) is shown in Figure 2.5. We can see that over the region of interest binary interference is substantially more severe than Gaussian interference. In such a case, randomization of the interference provides a substantial benefit. However, it should be noted that in this example, the receiver used processing (including the decision rule) designed for AWGN. Thus, it is not surprising that performance for Gaussian interference is superior. If the receiver knew that the interference was binary, better performance could have been achieved. However, in practice the exact distribution may be difficult to know. Also, thermal noise (which is Gaussian distributed) is always present. Thus, if despreading can to some degree guarantee Gaussian interference, the job of the communications engineer is made much easier.

2.2.3 LPI/LPD Communications

A traditional application of spread spectrum in military scenarios is Low Probability of Intercept (LPI) or Low Probability of Detection (LPD) communications [16][26]. In many military applications it is desirable (or perhaps necessary) that communications are carried out without knowledge of a third party. The

ability of the third party to know that communications is taking place is defined by the *probability of detection*. We can demonstrate the benefit of spread spectrum in this scenario by examining the performance of a radiometer [27]. A radiometer is a device used to detect RF energy and determine whether or not a signal is present. A typical radiometer block diagram is presented in Figure 2.7 [27]. The incoming signal is first filtered to the bandwidth of the signal of interest W and then amplified using a low power amplifier. The resulting signal is squared or rectified to eliminate phase modulation and consequently integrated over duration T_i . The output of the integration is then compared to a threshold to determine the existence of a signal. The threshold is an important parameter that controls the trade-off between the Probability of Detection (i.e., the probability of detecting that a signal is present when it is in fact present) and the Probability of False Alarm (i.e., the probability of erroneously determining a signal is present when it is not). Lower threshold values lead to a higher probability of detection P_d , but also a higher probability of false alarm P_{fa} . Conversely, increasing the threshold decreases the probability of false alarm, but also decreases the probability of detection. By choosing the threshold to obtain a desired P_{fa} the probability of detection in an AWGN channel can be approximated as [16][26][27]

$$P_d = Q \left\{ Q^{-1}(P_{fa}) - \frac{P}{N_o} \sqrt{\frac{T_i}{W}} \right\} \quad (2.22)$$

We can see that the probability of detection is directly related to the received signal power P and the integration time T_i while it is inversely related to the bandwidth W . Consider the scenario depicted in Figure 2.6 where an unintended receiver (interceptor) is trying to eavesdrop on communications between the transmitter and desired receiver. One measure of performance for the interceptor is the distance between it and the receiver that is required before detection is possible. Specifically, it would like the required distance to be as large as possible. Figure 2.8 plots the probability of detection versus the distance d between the transmitter and eavesdropper relative to the distance d_o between the transmitter and the desired receiver. In this example, the signal of interest has a required $\frac{E_b}{N_o} = 7.25dB$ and a bit rate $R_b = 10kbps$. The interceptor has an integration time $T_i = 10ms$, must maintain $P_{fa} = 2\%$ and it is assumed that the path loss exponent is 4 (i.e., the received power decays with d^4). As the figure shows, by increasing the signal bandwidth the eavesdropper must be closer to the transmitter to obtain the same probability of detection. For example, assuming a data rate of 10kbps and a bandwidth of approximately 10kHz, in order to obtain a probability of detection of 80% the interceptor need only get within $1.55d_o$ of the transmitter. However, if we increase the bandwidth to 10MHz, in order to obtain the same probability of detection the interceptor must get within $0.6d_o$, i.e., it must be closer to the transmitter than the desired receiver. Clearly, increasing the signal bandwidth makes interception more difficult, while (as we will show later) not negatively impacting the performance of the intended receiver.

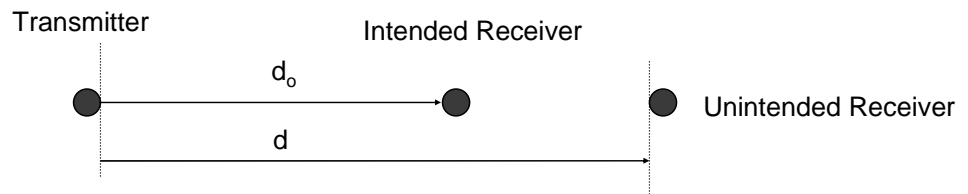


Figure 2.6: Assumed Interceptor Scenario

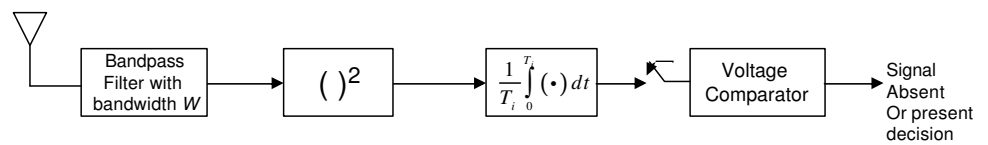


Figure 2.7: Block Diagram of a Simple Radiometer

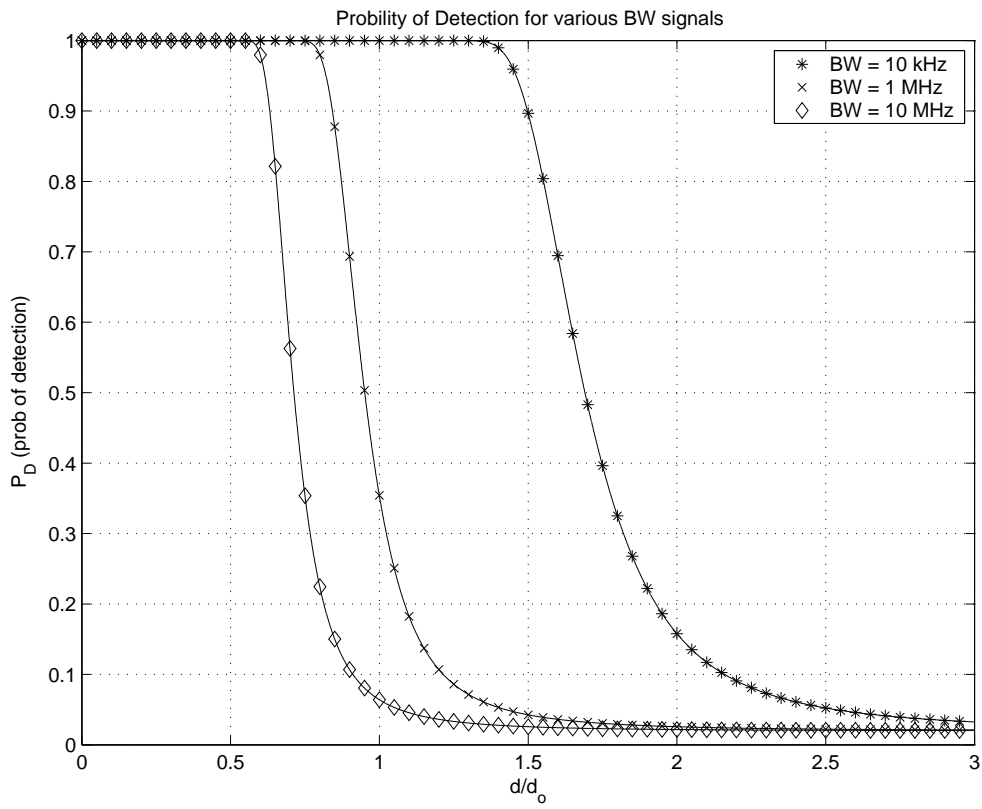


Figure 2.8: Impact of Spreading on Radiometer Performance

2.2.4 Improved Multiple Access

The LPI/LPD and anti-jamming capabilities of spread spectrum have traditionally been exploited in military systems. However, these same characteristics can also be used in commercial systems[20][21]. Specifically, these characteristics allow spread spectrum to be used in overlay systems and for multiple access [28]. In overlay systems, spread spectrum is used where existing legacy systems already occupy the spectrum. The anti-jamming capabilities allow spread spectrum to operate in the existence of narrowband interference from the incumbent systems. The LPI/LPD capabilities mean that spread spectrum signals can be generated that cause minimal interference to the existing systems.

These same qualities can be exploited when spread spectrum signals operate in the presence of other spread spectrum signals within the same system. The way in which transmitters from the same system share the spectrum is referred to as *multiple access* and will be discussed in detail in Chapter 10 and 11. In this section we will demonstrate the usefulness of spread spectrum in commercial systems by making a simplistic comparison of the capacity, in terms of the number of simultaneous channels of traditional narrowband multiple access techniques such as Time Division Multiple Access (TDMA) and Frequency Division Multiple Access (FDMA) with spread spectrum based multiple access (termed Code Division Multiple Access or CDMA). A more detailed discussion of multiple access will be given in Chapters 10-11. Let us first consider an FDMA system and assume that each user transmits at a rate of R_b b/s. Assuming optimal pulse shaping² and PSK, ASK or QAM modulation, the bandwidth per channel is $W = R_s = \frac{R_b}{k}$ where k is the number of bits per symbol. If there is a total system bandwidth of B_T , the number of channels supported by FDMA assuming that $k = 1$ is

$$K_{FDMA} = \frac{B_T}{R_b} \quad (2.23)$$

where we have ignored guard bands. If TDMA is used, all users transmit at a constant symbol rate $R_s = B_T$. If the desired bit rate is R_b , the number of channels that can be supported by the system is

$$\begin{aligned} K_{TDMA} &= \frac{R_s}{R_b} \\ &= \frac{B_T}{R_b} \end{aligned} \quad (2.24)$$

which is the same as FDMA. Now, let us consider a CDMA system. In general, CDMA (unlike TDMA and FDMA) is not an orthogonal multiple access scheme. That is, with proper filtering and timing, TDMA and FDMA channels are orthogonal and signals do not experience interference from other channels. With CDMA, however, this is difficult to maintain, especially for the uplink. Thus, typically a CDMA system is designed to have a controlled amount of interference

²Note that here we are assuming the use of optimal pulse shaping as opposed to square pulses which were assumed in Chapter 1. The null-to-null bandwidth with square pulses is twice that shown here for optimal pulses.

between channels. If we assume that each of the K_{CDMA} channels in the system causes interference to the other channels, the total interference power seen by any one signal is $(K_{CDMA} - 1)P$ where P is the received signal power for each signal (assumed to be the same for simplicity). Assuming that the mutual interference has the same properties as white noise (an assumption we will examine later), interference power spectral density is then

$$I_o = \frac{(K_{CDMA} - 1)P}{B_T} \quad (2.25)$$

where each signal uses the entire bandwidth B_T . The resulting E_b/I_o for each signal is then

$$\begin{aligned} \frac{E_b}{I_o} &= \frac{PT_b}{\frac{(K_{CDMA}-1)P}{B_T}} \\ &= \frac{B_T}{(K_{CDMA} - 1)R_b} \end{aligned} \quad (2.26)$$

Solving for the number of channels K_{CDMA} results in

$$K_{CDMA} \approx \frac{B_T}{R_b} \frac{I_o}{E_b} \quad (2.27)$$

Now, since the required value of $\frac{E_b}{I_o}$ is certainly more than one, CDMA will support fewer users in a *single cell environment*. In fact, a typical value of $\frac{E_b}{I_o}$ is 6dB, meaning that CDMA will support 4 times fewer users than FDMA and TDMA. However, there are other factors that we haven't yet taken into account. Specifically, the value of CDMA comes in a cellular environment. In a cellular environment, the entire system bandwidth must be divided up in a frequency reuse pattern to minimize co-channel interference. If Q is the number of cells with distinct frequency bands, the capacity of a single cell in a TDMA or FDMA cell becomes

$$K_{TDMA} = K_{FDMA} = \frac{B_T}{QR_b} \quad (2.28)$$

However, CDMA uses universal frequency reuse ($Q=1$). Although this increases the interference seen by an individual signal, the net result is a capacity improvement for CDMA as compared to FDMA and TDMA. Specifically, $\frac{E_b}{I_o}$ is now calculated as

$$\frac{E_b}{I_o} = \frac{PT_b}{\frac{(K_{CDMA}-1)P}{B_T}(1+f)} \quad (2.29)$$

where f is the interference increase due to out-of-cell interference as a fraction of in-cell interference and is typically in the range of 0.6. The number of channels supported is then

$$K_{CDMA} \approx \frac{B_T}{(1+f)R_b} \frac{I_o}{E_b} \quad (2.30)$$

Another important factor to consider with CDMA is voice activity. Up to this point, the value of $\frac{E_b}{I_o}$ assumes that all signals are present at all times. However, it is a well established fact that the typical speaker is only speaking $\nu = 3/8$ of the time. Thus, if a transmitter suppresses transmissions during periods of inactivity, the amount of interference is reduced by ν and the number of channels is increased by $\frac{1}{\nu}$. That is

$$K_{CDMA} \approx \frac{B_T}{\nu(1+f)R_b} \frac{I_o}{E_b} \quad (2.31)$$

Finally, we need to consider the impact of sectored antennas. In an FDMA or TDMA system, sectored antennas allow the reuse factor to decrease. Thus, for single sector systems $Q = 12$ is typical. However, for sectored systems $Q = 7$ is possible resulting in an increase in capacity of nearly 2. However, with sectored antennas, CDMA observes an interference reduction which directly improves $\frac{E_b}{I_o}$ and thus directly increases the number of supportable channels. That is

$$K_{CDMA} \approx \frac{GB_T}{\nu(1+f)R_b} \frac{I_o}{E_b} \quad (2.32)$$

where G is the antenna gain of the sector antenna in the azimuthal plane. A typical value for G is 4dB.

Now let us assume the use of three-sector antennas at the base station. A standard reuse pattern for three sector systems is $Q = 7$. The capacity of TDMA and FDMA is then

$$K_{TDMA} = K_{FDMA} = \frac{B_T}{7R_b} \quad (2.33)$$

Now for CDMA, let us assume that $\frac{E_b}{I_o} = 6\text{dB}$, $\nu = 3/8$, $G = 4\text{dB}$, and $f = 0.6$,

$$\begin{aligned} K_{CDMA} &\approx \frac{2.5B_T}{\frac{3}{8}(1.6)R_b} \frac{1}{4} \\ &\approx \frac{B_T}{R_b} \end{aligned} \quad (2.34)$$

which is an improvement of nearly an order of magnitude. Thus, due mainly to universal frequency reuse and the statistical multiplexing capability stemming from voice activity, CDMA can provide substantially better multiple access capability than TDMA or FDMA. A more detailed analysis will be provided in Chapter 11.

2.2.5 Interference Averaging

The previous section discussed the advantage of spread spectrum in multiple access scenarios. This is partly due to the ability of spread spectrum to exploit universal frequency reuse, statistically multiplex bursty signals as well as interference averaging [23]. The dominant reason why narrowband systems

require large frequency reuse patterns (e.g., $Q = 7$), is that co-channel interference can vary dramatically due to the relative positions of co-channel transmitters. Reuse patterns are then designed to account for the worst case scenario [29, 5, 30]. The use of spread spectrum causes interference averaging which reduces the variation of the interference even if it doesn't reduce the average interference power. More specifically, the use of spreading exposes the signal to a larger number of interferers. However, only a fraction of each interferer will be left after matched filtering. Thus, instead of being impacted by the full power of a small number of interferers, spread spectrum signals are impacted by a fraction of the power from a large number of interferers. The benefit of this interference averaging is demonstrated in the following example.

Example 2.2 Consider a cellular system where there is a single co-channel interferer with a normalized received power that follows an exponential distribution with a mean of 1. The desired signal is in outage if the interference power exceeds 3. What is the probability of outage? Now consider a cellular system which has 10 co-channel interferers whose powers follow independent exponential distributions, but now each with a mean power of 0.1. What is the mean interference in this case and how does it compare to the first situation? What is the outage probability in the second case and how does it compare to the first case?

SOLUTION: (a) The interference power follows an exponential distribution. The exponential distribution can be written as [6]

$$f_X(x) = \lambda e^{-\lambda x} \quad x \geq 0 \quad (2.35)$$

where distribution is characterized by the parameter λ which is simply the inverse of the mean

$$\bar{X} = \frac{1}{\lambda} \quad (2.36)$$

The probability of outage can be written as

$$P_{out} = Pr\{x > 3\} \quad (2.37)$$

$$= 1 - F_X(3) \quad (2.38)$$

where $F_X(x)$ is the cumulative distribution function of the interference. In our case, the interference follows an exponential distribution with parameter $\lambda = 1$, thus

$$F_X(x) = 1 - e^{-\lambda x} \quad x \geq 0 \quad (2.39)$$

$$F_X(x) = 1 - e^{-x} \quad x \geq 0 \quad (2.40)$$

$$P_{out} = e^{-x} \quad x \geq 0 \quad (2.41)$$

For our case, $P_{out} = e^{-3} = 0.05$ or 5%.

(b) Now, the interference is a sum of independent exponential distributions each with parameter $\lambda = 10$. It is well-known that the distribution of a sum of independent exponential random variables is a random variable which follows a Gamma distribution [6]. That is, if

$$Y = \sum_{i=1}^n X_i \quad (2.42)$$

where X_i each have a distribution described in equation (2.35), then Y follows a Gamma distribution. The Gamma distribution can be written as [6]

$$f(y; k, \theta) = \frac{y^{k-1} e^{-y/\theta}}{\theta^k \Gamma(k)} \quad y \geq 0 \quad (2.43)$$

where $\Gamma(x)$ is the Gamma function which for integer values is simply $\Gamma(x) = (x-1)!$ Specifically, it can be shown that since Y is the sum of n iid exponential random variables it has the distribution $f_Y(y) = f(y; n, \frac{1}{\lambda})$ which is written as

$$f_Y(y) = \frac{y^{k-1} e^{-\lambda y}}{(1/\lambda)^n \Gamma(n)} \quad y \geq 0 \quad (2.44)$$

where $\lambda = 1/\bar{X} = 1/0.1 = 10$ and $n = 10$. Now the mean of the Gamma distribution is $\bar{Y} = k\theta$. In our case, $\bar{Y} = n/\lambda = 1$. Thus, the mean interference power is identical to the single user case. However, the probability density functions for the interference power for the two cases are plotted in Figure 2.9. We can see that although the mean power is the same in the two cases, the distributions are quite different. Specifically, by interference averaging, the variance is reduced considerably. The outage probability is defined as $P_{out} = 1 - F_Y(3)$. The cumulative distribution function of a Gamma distribution is

$$F_Y(y) = \frac{\gamma(k, y/\theta)}{\Gamma(k)} \quad y \geq 0 \quad (2.45)$$

where $\gamma(k, y/\theta)$ is the incomplete Gamma function. Now, again substituting $k = n = 10$ and $\theta = 1/\lambda = 1/10$ we have

$$P_{out} = 1 - F_Y(3) \quad (2.46)$$

$$= 1 - \frac{\gamma(10, 30)}{\Gamma(10)} \quad (2.47)$$

$$= 7 * 10^{-6} \quad (2.48)$$

Thus, interference averaging causes a dramatic reduction in outage probability even if it doesn't reduce the mean interference power.

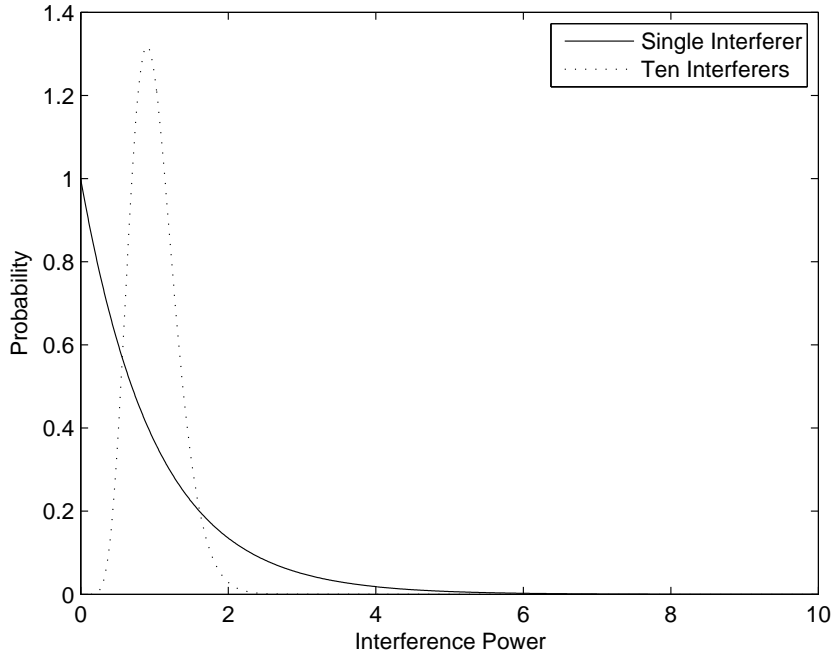


Figure 2.9: The Probability Density Functions for the Case of One and Ten Interferers in Example 2.2

2.2.6 Resistance to Multipath Fading

As we have stated previously, traditional modulation and coding schemes were designed for the relatively benign AWGN channel. A more harsh channel which is common in mobile wireless communications is the multipath fading channel [5]. This channel was discussed in more detail in Chapter 1. We will summarize the key features of the multipath channel, but please refer to Chapter 1 for a more complete description. Due to the existence of a large number of scatterers in the environment, the receiver sees many reflected and delayed versions of the signal. If the relative delays of the signals are small as compared to the symbol rate, the receiver will observe a single time-varying waveform whose amplitude fluctuates with the mobile's movement due to random phase combinations. A common model for such fading channels is what is termed the flat Rayleigh fading model, since the amplitude of the signal follows a Rayleigh distribution [12]. The fading is termed 'flat' fading since the coherence bandwidth of the channel is greater than the signal bandwidth and thus the channel transfer function is flat over the bandwidth of the signal. This causes the all frequency components of the signal to fade simultaneously. As derived in Chapter 1, the probability of error is significantly worse than AWGN in this scenario. As

derived in Chapter 1 and presented in equation (1.142), the probability of bit error for BPSK with coherent reception is [1, 5]

$$P_b = \frac{1}{2} \left(1 - \sqrt{\frac{\frac{E_b}{N_o}}{1 + \frac{E_b}{N_o}}} \right) \quad (2.49)$$

The performance is plotted in Figure 2.10 (labeled 'flat fading') along with the AWGN case. We can see that impact of fading is severe. For example, at a bit error probability of 10^{-3} , Rayleigh fading results in a 17dB degradation in performance.

Again, spread spectrum can be beneficial in such a case [31]. By increasing the signal bandwidth beyond the channel coherence bandwidth, the severe fading can be mitigated. In other words, by decreasing the symbol duration so that it is smaller than the multipath delays, the multipath signals can be resolved and used as a form of diversity. When the signal bandwidth is expanded beyond the coherence bandwidth, the fading is termed 'frequency selective' fading since the channel fades differently over different parts of the signal spectrum. For example, by increasing the symbol rate such that the bandwidth is approximately twice the coherence bandwidth, a diversity of order two can be obtained provided the proper spreading and receiver are used. As shown in Figure 2.10, if the second order diversity ($L=2$) can be harnessed by the receiver an improvement of nearly 10dB can be obtained at an error rate of 10^{-3} . By increasing the bandwidth by greater amounts, greater diversity and thus better performance can potentially be obtained. For example, with $L=4$ order diversity, performance is improved by approximately 14dB and is only 3dB from the AWGN case. Fading and diversity are considered in detail in Chapter 8.

Example 2.3 Assume that a Rayleigh fading channel has a coherence bandwidth of $B_C = B$ where B is bandwidth of a transmitted signal without using spread spectrum. (a) If the average signal-to-noise ratio (SNR) is 10dB and an outage occurs whenever the instantaneous SNR drops below 0dB, determine the probability of outage. (b) Assume that by increasing the bandwidth using spread spectrum to LB_C frequency selective fading is introduced which can be exploited to achieve L -fold diversity with equal power per diversity branch. If the transmit bandwidth is increased to $8B_C$ using spread spectrum, what is the new probability of outage?

SOLUTION: (a) A Rayleigh fading channel follows a Rayleigh distribution in its absolute value. Thus, the received power follows a Chi-Square distribution with two degrees of freedom. The probability density function of the received SNR γ is then [1]

$$f_{\Gamma_1}(\gamma) = \frac{1}{\bar{\gamma}} e^{-\gamma/\bar{\gamma}} \quad \gamma \geq 0 \quad (2.50)$$

where $\bar{\gamma}$ is the average SNR. The outage probability can be written as

$$\begin{aligned} P_{out} &= \Pr\{\gamma < 1\} \\ &= F_{\Gamma_1}(1) \end{aligned}$$

where $F_{\Gamma_{10^0/10=1}}(\gamma)$ is the cumulative distribution function (CDF). For a Chi-Square distribution we have [1]

$$F_{\Gamma_1}(\gamma) = 1 - e^{-\gamma/\bar{\gamma}} \quad \gamma \geq 0 \quad (2.51)$$

Thus, the outage probability is $F_{\Gamma_1}(1) = 0.1$ or 10%.

(b) It is assumed in the problem that increasing the bandwidth by a factor of L over the coherence bandwidth provides L -fold diversity with equal power per channel. In practice this is not always the case, but it is a reasonable approximation. The resulting SNR follows a Chi-Square distribution with $2L$ degrees of freedom [5, 1]

$$f_{\Gamma_L}(\gamma) = \frac{1}{(L-1)!(\bar{\gamma}/L)^L} \gamma^{L-1} e^{-\gamma/(\bar{\gamma}/L)} \quad \gamma \geq 0 \quad (2.52)$$

Note that obtaining diversity in this way does not increase the total average received power, thus the average SNR per channel is $\bar{\gamma}/L$. Other diversity methods (e.g., using multiple receive antennas) can increase average power. Again, the probability of outage is related to the CDF of the SNR which is [5, 1]

$$F_{\Gamma_L}(\gamma) = 1 - e^{-\gamma/(\bar{\gamma}/L)} \sum_{k=0}^{L-1} \frac{1}{k!} \left(\frac{\gamma}{\bar{\gamma}/L} \right)^k \quad \gamma \geq 0 \quad (2.53)$$

Specifically, the new outage probability is $F_{\Gamma_L}(1) = 2 * 10^{-6}$ which is a dramatic reduction. Thus, we find that exploiting diversity in this way can be very powerful. Of course, if there is not equal power per diversity branch, the gains will not be quite as dramatic but still quite substantial.

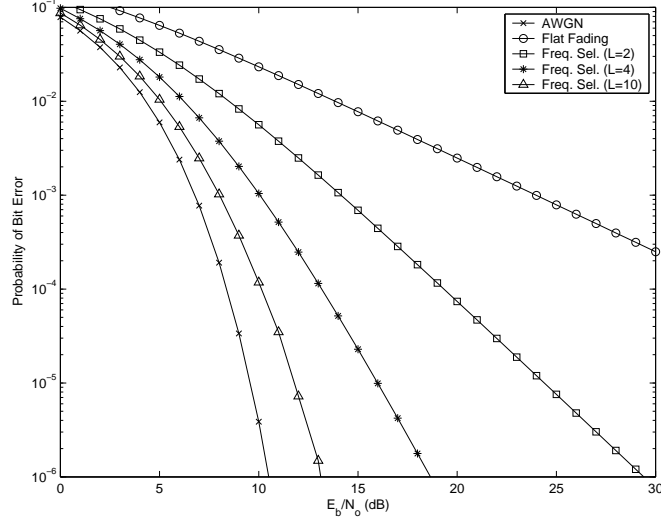


Figure 2.10: Impact of Frequency Selective Rayleigh Fading on the Performance of a Rake Receiver

2.2.7 Ranging

Finally, the last application mentioned here that is useful for spread spectrum is ranging. Ranging is the process of determining the distance a receiver is from a transmitter by measuring the time-of-arrival (TOA) of a signal as compared to a known reference. This is also sometimes referred to as the time-of-flight. The accuracy of this type of measurement is directly related to the symbol duration, with accuracy improving as the symbol duration decreases (or bandwidth increases). Thus, more accurate ranging capabilities are possible with very short symbol duration or very wide bandwidths. Additionally, multipath propagation is typically a limiting factor in ranging applications. Spread spectrum formats (esp. direct sequence spread spectrum) is resistant to multipath as discussed in section 2.2.6. Thus, spread spectrum improves ranging capabilities in two distinct ways.

The improvement due to bandwidth can be shown using the Cramer-Rao lower bound (CRLB) [32]. The CRLB places a lower limit on the variance of an unbiased estimator. In this case we are interested with the estimation of the time-of-arrival τ which is directly related to the distance estimate through the speed of light. Specifically, the variance of the estimate of τ (σ_τ^2) is lower-bounded by the CRLB of τ , $CRLB(\tau)$, which can be shown to be [33]

$$\sigma_\tau^2 \geq CRLB(\tau) = \frac{1}{\gamma} \frac{\int_{-\infty}^{\infty} |W(f)|^2 df}{\int_{-\infty}^{\infty} f^2 |W(f)|^2 df} \quad (2.54)$$

where γ is the signal to noise ratio of the link and $W(f)$ is the Fourier Transform

of the signal being used for ranging. As seen in equation (2.54), as the bandwidth of $W(f)$ increases the denominator increases faster than the numerator and thus the CRLB decreases. We will return to this shortly.

The CRLB for the distance estimate \hat{d} is related to the CRLB for time-of-arrival as $CRLB(d) = c^2 \cdot CRLB(\tau)$ where c is the speed of light. Thus,

$$\sigma_d^2 \geq \frac{c^2 \int_{-\infty}^{\infty} |W(f)|^2 df}{\gamma \int_{-\infty}^{\infty} f^2 |W(f)|^2 df} \quad (2.55)$$

To get a better understanding of the relationship between signal bandwidth and ranging estimation accuracy, let us consider an “ideal” pulse which perfectly uses the bandwidth available. That is the spectrum of the pulse with bandwidth B is

$$W(f) = \Pi\left(\frac{f}{2B}\right) \quad (2.56)$$

where $\Pi(f)$ is a rectangular pulse of unity width. The minimum variance of the distance estimator is then written as

$$\begin{aligned} \sigma_d^2 &\geq \frac{c^2 \int_{-\infty}^{\infty} |W(f)|^2 df}{\gamma \int_{-\infty}^{\infty} f^2 |W(f)|^2 df} \\ &= \frac{c^2 \int_{-B}^B df}{\gamma \int_{-B}^B f^2 df} \\ &= \frac{c^2 \cdot 2B}{\gamma \cdot \frac{2B^3}{3}} \\ &= \frac{c^2 \cdot 3}{\gamma \cdot B^2} \end{aligned} \quad (2.57)$$

Thus, we can clearly see the relationship between the minimum achievable ranging variance and the signal bandwidth. That is, the minimum achievable variance decreases as B^2 increases. We examine this more in the following example.

Example 2.4 Consider a ranging system which uses an ideal pulse, i.e., one which completely uses the bandwidth available. If the received signal-to-noise ratio is 20dB (regardless of bandwidth), plot the minimum achievable standard deviation of a distance estimator versus the signal bandwidths between 1MHz and 10GHz. What is the minimum achievable range estimate standard deviation for a bandwidth of 1GHz?

SOLUTION: From equation (2.57) we know that the minimum achievable variance at an SNR of 100 is

$$\sigma_d^2 \geq \frac{3c^2}{400\pi^2} \frac{1}{B^2} \quad (2.58)$$

This corresponds to a standard deviation bound of

$$\sigma_d \geq \frac{c}{\pi} \sqrt{\frac{3}{400}} \frac{1}{B} \quad (2.59)$$

The resulting minimum achievable standard deviation is plotted in Figure 2.11 for bandwidths between 1MHz and 10GHz. We can see that the standard deviation decreases directly with bandwidth as expected. Further, for a bandwidth of 1GHz, the minimum achievable standard deviation is just under 1cm.

2.3 Types of spread spectrum

The preceding sections have demonstrated that there are many scenarios where increasing the bandwidth of a communications signal provides benefits in performance. However, we have not discussed how this bandwidth expansion (or spreading) is accomplished. There are two main techniques or types of spread spectrum that will be discussed in this book. The first is called direct-sequence spread spectrum (DS/SS). It involves increasing the signal bandwidth by multiplying the information signal by a high rate spreading waveform. The symbols of the spreading waveform are typically pseudorandom and are generated using a pseudorandom noise (PN) sequence. From Fourier Transform theory, the multiplication of two signals in the time-domain results in the convolution of their spectra. Thus, if the spreading waveform has a large bandwidth, the resulting signal will also have a large bandwidth. The ratio of the symbol rate of the spreading waveform to the symbol rate of the information signal is typically called the spreading gain or processing gain and is directly related to the benefits provided by spread spectrum. We will discuss DS/SS in more detail and more rigorously define the processing gain in Chapter 3.

The second main type of spread spectrum is called frequency-hopped spread spectrum (FH/SS). In this form of spread spectrum the original information signal is moved in frequency from one band to another in a pseudorandom fashion. Again, this is due to the use of a pseudorandom noise sequence that

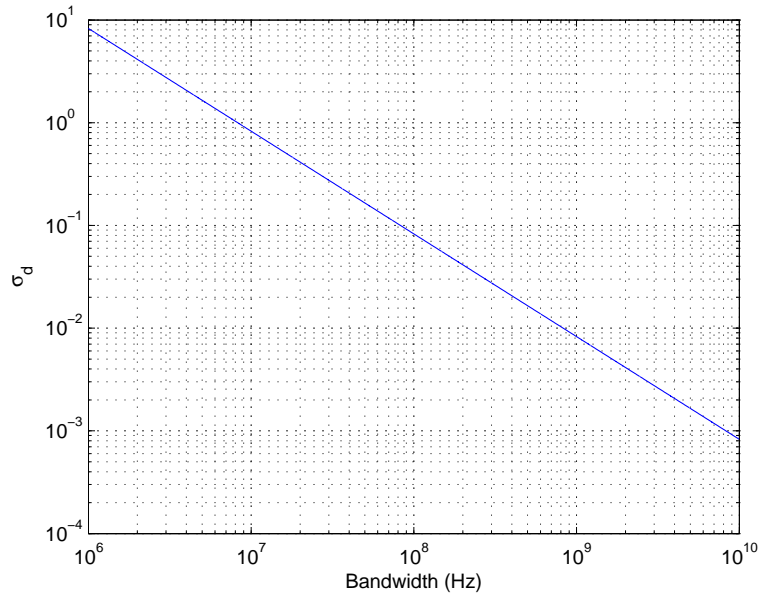


Figure 2.11: The Minimum Achievable Range Estimate Variance versus Bandwidth from Example 2.4

is used to change the center frequency of the carrier to one of N different non-overlapping frequency channels. The number of different hopping frequencies N is also called the spreading gain, although we will discuss this in more detail in Chapter 4.

There are also many combinations of these two techniques as well as variations known as time-hopping and multi-carrier spread spectrum. We will examine each of these schemes in more detail in the coming chapters.

2.4 Conclusions

In this chapter we have discussed briefly the motivations behind the use of spread spectrum communications. These motivations include resistance to jamming (or other interference), low probability of detection/intercept (or the possibility of spectral overlay), the related benefits of interference averaging and randomization, resistance to multipath fading, and improved ranging resolution. In the coming chapters we will discuss these benefits in more detail, especially as they relate to specific spread spectrum formats.

Bibliography

- [1] J. Proakis, *Digital Communications*. New York, NY: McGraw-Hill, fourth ed., 2001.
- [2] M. Pursley, *Introduction to Digital Communications*. Upper Saddle River, NJ: Prentice Hall, 2005.
- [3] B. Sklar, *Digital Communications*. Upper Saddle River, NJ: Prentice Hall, 2000.
- [4] R. Ziemer and R. Peterson, *Introduction to Digital Communications*. Upper Saddle River, NJ: Prentice Hall, second ed., 2001.
- [5] T. S. Rappaport, *Wireless Communications: Principles and Practice*. Upper Saddle River, NJ: Prentice Hall, second ed., 2002.
- [6] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY: McGraw-Hill, third ed., 1991.
- [7] P. Peebles, *Probability, Random Variables, and Random Signal Principles*. New York, NY: McGraw-Hill, second ed., 1987.
- [8] K. Shanmugan and A. Breipohl, *Random Signals: Detection, Estimation, and Data Analysis*. New York, NY: John Wiley and Sons, second ed., 1988.
- [9] H. Nyquist, "Certain topics in telegraph transmission theory," *AIEE Transactions*, pp. 617–644, April 1928.
- [10] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, 1948.
- [11] S. Lin and J. Costello, *Error Control Coding: Fundamentals and Applications*. Englewood Cliffs, NJ: Prentice-Hall, first ed., 1983.
- [12] J. Parsons, *The Mobile Radio Propagation Channel*. New York: John Wiley and Sons, second ed., 2000.
- [13] W. Jakes, *Microwave Mobile Communications*. Wiley-Interscience, 1974.
- [14] R. Scholtz, "The spread spectrum concept," *IEEE Transactions on Communications*, vol. COM-25, pp. 748–755, August 1977.

- [15] D. Torrieri, *Principles of Spread-Spectrum Communication Systems*. New York, NY: Springer, 2005.
- [16] R. Peterson, R. Ziemer, and D. Borth, *Introduction to Spread Spectrum Communications*. Prentice-Hall, 1995.
- [17] M. Simon, J. Omura, R. Scholtz, and B.K.Levitt, *Spread Spectrum Communications Handbook*. New York, NY: McGraw-Hill, electronic ed., 2002.
- [18] R. Pickholtz, D. Schilling, and L. Milstein, "Theory of spread-spectrum communications - a tutorial," *IEEE Transactions on Communications*, vol. COM-30, pp. 855–884, May 1982.
- [19] D. Nicholson, *Spread Spectrum Signal Design: LPE and AJ Systems*. Computer Science Press, 1988.
- [20] A. Viterbi, *CDMA: Principles of Spread Spectrum Communication*. Addison-Wesley, 1995.
- [21] G. Cooper and R. Nettleton, "A spread spectrum technique for high-capacity mobile communications," *IEEE Transactions on Vehicular Technologies*, vol. VT-27, pp. 264–275, November 1978.
- [22] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions*. Dover Publications, 1970.
- [23] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, 2005.
- [24] A. Viterbi, "The evolution of digital wireless technology from space exploration to personal communication," *IEEE Transactions on Vehicular Technologies*, vol. 43, pp. 638–644, August 1994.
- [25] A. Viterbi, "Wireless digital communication: A view based on three lessons learned," *IEEE Personal Communications Magazine*, vol. 29, pp. 33–36, September 1991.
- [26] R. Dillard, "Detectability of spread spectrum signals," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-15, pp. 526–537, 1979.
- [27] R. Mills and G. Prescott, "A comparison of various radiometer detection models," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 467–473, 1996.
- [28] K. Gilhousen and et. al., "On the capacity of a CDMA system," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 303–312, May 1991.
- [29] V. MacDonald, "The cellular concept," *The Bell Systems Technical Journal*, vol. 58, pp. 15–43, January 1979.
- [30] W. Lee, *Mobile Cellular Telecommunications Systems*. McGraw-Hill, 1989.

- [31] J. Holtman and L. Jalloul, "Rayleigh fading effect reduction with wideband DS/CDMA signals," in *Proceedings of the 1991 Global Communications Conference*, pp. 553–557, 1991.
- [32] S. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*. Upper Saddle River, NJ: Prentice-Hall, 1993.
- [33] J. Zhang, R. Kennedy, and T. Abhayapala, "Cramer-Rao lower bounds for the time delay estimation of UWB signals," in *Proceedings of the IEEE International Conference on Communications*, pp. 3424–3428, 20–24 June 2004.