Phase Noise in Low-Power Radio Communications

by

Donald C. Wilcoxson

B.S.E.E., Purdue University
(1988)
M.S.E.E., Purdue University
(1989)

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degrees of

Electrical Engineer

and

Master of Science in Electrical Engineering and Computer Science

at the

Massachusetts Institute of Technology
September, 1996

Signature of Author_____
Department of Electrical Engineering and Computer Science
September, 1996

Certified by_____
Assistant Professor Mitchell D. Trott
Thesis Supervisor

Accepted by_____
Professor Frederic R. Morgenthaler
Chair, Department Committee on Graduate Students

Phase Noise in Low-Power Radio Communications

by

Donald C. Wilcoxson

Submitted to the
Department of Electrical Engineering and Computer Science on August 9, 1996
in partial fulfillment of the requirements for the Degrees of Electrical Engineer and
Master of Science in Electrical Engineering and Computer Science

Abstract

Current radio communication design places a high priority on minimizing power consumption, and as such places tight constraints on the design of various components, including oscillators used in the system. As power consumption requirements become more stringent, the ability to minimize unwanted phase noise in the oscillator diminishes, thus characterization of system performance with phase noisy oscillators is warranted. We analyze the effects of a phase noisy oscillator on two different modulation schemes: quadrature amplitude modulation (QAM), which is currently used in mobile radio, and digital pulse-position modulation (PPM), which is a new proposal for low-power operation. In both cases, the main result of our analysis is the characterization of the symbol error probabilities for each scheme in the presence of phase noisy oscillators and additive white Gaussian noise in the channel.

We use Quadrature phase shift keying (QPSK) as a simple representative example of QAM, where we show that phase noise destroys the orthogonality between the in-phase and quadrature channels. Although orthogonality is lost, we show that the impact on the symbol error probabilities is negligible if perfect carrier recovery (complete knowledge of the phase noise process) is possible. With imperfect knowledge of the phase noise at the receiver it is shown that the symbol error probability is degraded significantly.

With the digital PPM scheme, we note that phase noise in the reference oscillator of the transmitter manifests itself as timing jitter in the transmitter's clock. We show that for a Brownian motion phase noise process that under certain reasonable assumptions the timing jitter variable is approximately Gaussian. Several different receiver structures, including both optimal and suboptimal, for digital PPM are studied and it is shown that the effect of this timing jitter on the symbol error probability is significant when perfect timing recovery is not possible.

Thesis Supervisor: Mitchell D. Trott

Title: Assistant Professor of Electrical Engineering and Computer Science

Acknowledgments

This thesis represents the end of a tumultuous time in my life, and is probably the most difficult thing I have ever done. It has changed my outlook on my life and career, and has helped me redefine my goals. I have many people to thank for helping me through this ordeal, not the least of which is my advisor, Mitchell Trott. Without his patience and guidance I would not have been able to make it through these last few years. I would also like to thank the other professors and instructors who've given me invaluable insight and advice throughout the last few years, particularly Professor Robert Gallager and Dr. G. David Forney. I would also like to thank the support staff in the Laboratory for Information and Decision Systems for their friendliness and support throughout my time in the laboratory.

I would like to thank my friends on campus for all their support and time in working on this research, for without them I would have never gotten through it. Particularly I would like to personally thank Mark Ibanez, Patrick Mauer, Brett Schein, James Sarvis, Aradhana Narula, and even Kevin Yu (although he defected to Stanford). I would also like to thank my friends at Lincoln Laboratory, among them Steven Bernstein, Steve Kolek, and Dr. Eytan Modiano with whom I've had many discussions about careers in addition to my research. Most of what I learned has come from interaction with the above people, and without them I would have not made it through the last several years.

Finally, I'd like to thank the most important person in my life, my wife Diana. She has been the rock upon which I've had to lean through the last several years. Her dedication to keeping up my spirits and listening to my complaining have given me strength to muddle through and overcome. This thesis is dedicated to her.

Contents

List of Figures

1. Introduction

The unprecedented popularity of cellular telephony has created a huge amount of interest in the area of radio communications in the last decade [1],[2],[3],[4]. Because there is an intensely profitable market for cellular radio, there has been an explosion of radio communications research in both industry and academia. The majority of the research has been focused in the broad areas of device research, antenna design, modulation design, and multiple-access information theory. Most, if not all, of the device research has been focused into two main areas: low-cost and low-power design techniques. The goal, as with any commercial enterprise, is to make products that are cheaper and last longer than the competition.

In the area of antenna design the effort has been to design antennas that increase the performance of the system in such a way that either more users can be accommodated or that performance is improved in the mobile (i.e., fading) environment. For example, the current use of three antennas on each cellular base station, which allows for three distinct "sectors" per cell, increases available capacity by a factor of three. In addition, there has been research into antenna arrays which allow improved reception even when the user's signal is fading.

As for modulation design, the major accomplishment has been the implementation of digital modulation schemes for second-generation cellular radio, which has been

occurring over the last few years. Previous to about 1990 all existing cellular services in the U.S. and worldwide were analog (typically frequency-modulated). Today there are numerous digital standards and proposals that are both being fielded and tested for the 900 MHz cellular band and the newly developed 2 GHz Personal Communications Services (PCS) band. These include time-division multiple-access (TDMA) and spread-spectrum multiple-access (SSMA) methods as well as hybrid systems combining features from both. The advantages of these digital schemes include the possibility of voice compression (which reduces the bandwidth of the signal), voice activity detection (which allows the transmitter to turn off when no words are being spoken), and more predictable performance in harsh electromagnetic environments (for example, during fading). In addition, digital signal processing is possible which can be performed with lower power consumption than with comparable analog processing.

As far as multiple-access information theory is concerned, most of the research has been in the academic arena, focusing on finding capacity regions in fading environments [5], joint-detection and estimation methods, and evaluation of certain coding schemes in those environments. Although the results from the these analyses are important to understanding the performance of cellular or other multiple-access radio systems, it is sometimes difficult to draw connections from the abstract results to practical systems. Thus, at this time, little of the information-theoretic research has found direct application in commercial systems.

Motivated by three of the major areas of research noted above, namely device research, modulation design, and information theory, the research in this thesis focuses on a specific problem associated with low-cost, low-power communications system design. Specifically, the problem of operating a communication system with oscillators having poor phase noise characteristics is explored in terms of modulation design and performance analysis, with some exploration of related information-theoretic results. Until now the concept of phase noise impairment has only been explored in depth in optical communications systems, because low-cost semiconductor laser sources tend to be quite poor in terms of phase noise performance [6],[7]. Phase noise has been of secondary concern in radio communication systems (although there has been some

research with regard to radar [8] and satellite systems [9]) since thermal noise dominated in most systems. However, since competition in the commercial radio market has become intense, equipment manufacturers are striving for more ways to produce systems of ever-decreasing cost and lower power consumption. It is exactly these goals of low-cost and low-power design that necessitate consideration of the effects of poor quality oscillators on communications systems performance. Because antenna design does not directly impact the following discussion of phase noise, it will not be addressed in this thesis.

1.1 Definition of a "Low-Power" Radio Communications System

The research in this thesis was motivated by a project in which many sensors were connected by radio to a central base station. The data rate is variable and bursty, ranging from 1 bit/sec to over 1 Mb/s. For illustration purposes, the particular sensor application that seems to meet this criteria is a digital camera, whose data is compressed and motion-encoded, such that the data stream is low-rate when the motion detected by the camera is small and is high-rate when the change from frame-to-frame is high. These sensors transmit their data back to a central base station via a radio link. Figure 1.1 illustrates the wireless sensors that motivated this study.

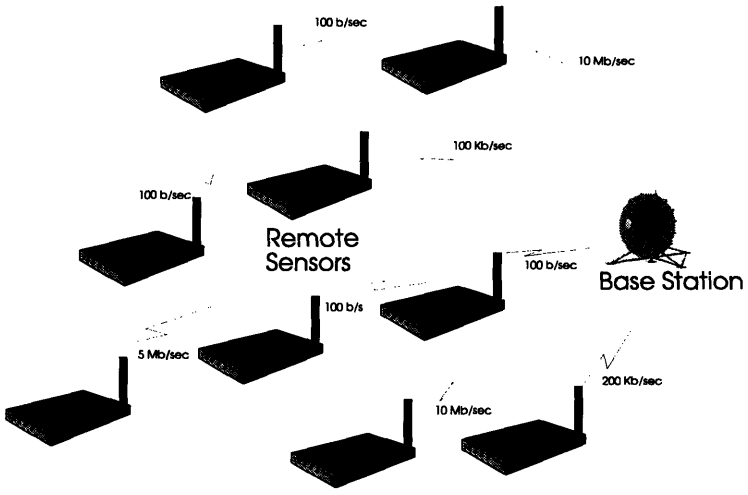

Figure 1.1: Wireless sensors that motivated the research

The research goal was to construct "low-power" sensors, however, the term "low-power" is somewhat nebulous. Two definitions of how "low-power" might be defined come to mind. "Low-power" might correspond to the definition that the transmitted power (or effective isotropic radiated power, EIRP) is low when compared to some standard. A second definition comes from the device aspect, i.e., that the power consumed by the sensor unit is low when compared to some standard, not necessarily implying anything about the transmitted power. It is the latter definition that we mean when we say "low-power," i.e., our definition is that the sensor's total power consumption (including all power consumed in the sensor electronics, analog/digital conversion, compression, encoding, other digital signal processing, and the radio portion including the transmitter/receiver front-end, oscillator, amplifiers etc.) is low compared with some standard. A consequence of designing for low-power consumption is that the constraints on the design of the oscillator used for radio transmission become tight [10].

1.2 Phase Noise

Anyone who has taken a modern digital communications course is familiar with the concept of transmitting a signal through an additive white Gaussian noise (AWGN) channel. This is usually the only degradation considered when initially evaluating a particular modulation scheme. However, under certain circumstances, many other degradations might be equally or more important. For example, those degradations could include antenna efficiency, polarization mismatch, galactic and terrestrial noise sources, fading, intermodulation products, co-channel interference, and phase noise [11]. In particular we are concerned with the effect of phase noise on communications system performance. The simplest (and most general) definition of a phase noisy oscillator is given in (1.1),

$$s(t) = A\cos(2\pi f_o t + \theta(t) + \phi) \tag{1.1}$$

where $\theta(t)$ is a random process that represents the phase noise process, ϕ represents an arbitrary phase offset (a constant), and f_o is the nominal frequency of the oscillator wave

s(t). How θ(t) affects the sinusoid depends on the exact statistics of the phase noise process. Figure 1.2 illustrates what s(t) might look like with a nonzero θ(t).

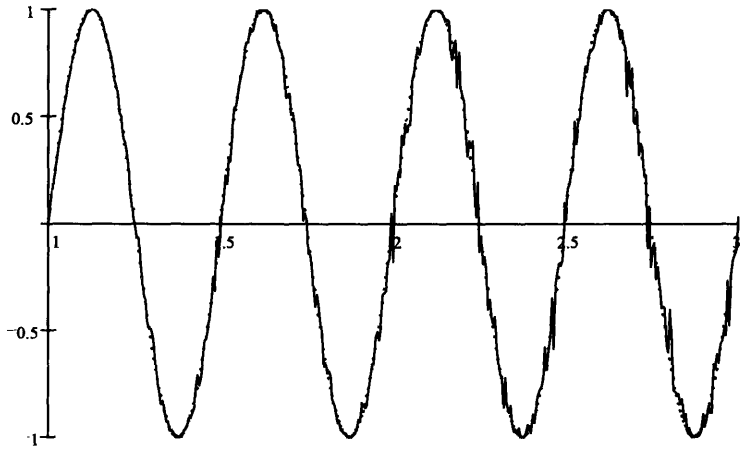

Figure 1.2: Phase noisy sinusoid (A=1)

The preponderance of literature concerned with the effect of phase noise has been in the realm of fiber-optic communications, and in this case the most accepted model for θ(t) has been as a Brownian motion process. This model was first derived empirically, then later shown to be accurate analytically. Since the underlying physical phenomena is not important in this thesis the reader is referred to Salz [7] or Henry [16] for more information. In addition to the Brownian motion model, there have been several more complex models proposed describing the phase noise process in both fiber-optic and radio oscillators (for example, see [12], [13],[14], and [15]). Because of the relative simplicity of the Brownian motion model, however, we will not incorporate the more complex models of phase noise into this thesis. Thus, we can express θ(t) as

$$\theta(t) = 2\pi \int_0^t \mu(\tau) d\tau \quad (\text{for } t > 0) \tag{1.2}$$

where μ(t) is a zero-mean white Gaussian noise process. If we define the (two-sided) height of the μ(t) process's power spectral density as N_1, then θ(t) is a zero-mean Gaussian process with variance

$$Var[\theta(t)] = (2\pi)^2 N_1 t \tag{1.3}$$

That is, the variance of θ(t) grows linearly with time. Jumping a little ahead for the moment it is clear that as t grows large, the phase of s(t) is effectively a random variable uniformly distributed over [0,2π) (reduced modulo 2π). It would seem intuitive that straightforward application using s(t) as a carrier for a coherent[1] phase modulation scheme might be imprudent, since it would be difficult to estimate the actual carrier (i.e., s(t)) at the receiver. In Chapter 2 we will show that good estimation of s(t) is essential to the operation of any phase-modulation scheme.

At this point, one might ask, what does the power spectrum of s(t) look like? Ideally, the spectrum of a pure cosine wave is given by the Fourier transform pair

$$\cos(2\pi f_o t) \overset{FT}{\Longleftrightarrow} \tfrac{1}{2}[\delta(f - f_o) + \delta(f + f_o)] \tag{1.4}$$

However, with the addition of the noise process θ(t) in the argument of the cosine, s(t) is a noise process in itself. Thus we can calculate the power spectrum as the Fourier transform of the autocorrelation function, $R_{ss}(t, t+\tau) = E[s(t)s(t+\tau)]$. One should note that if ϕ is a fixed, *known* constant then s(t) is *nonstationary*. However, given the nature of the problem, it is more natural to assume that ϕ is random variable (fixed for all t). Choosing ϕ to be uniformly distributed over [0,2π), we can calculate the autocorrelation function in a straightforward manner since this choice makes the s(t) process stationary. Skipping the details, we find that $R_{ss}(t, t+\tau) = R_{ss}(\tau)$ is given by [7]

$$R_{ss}(\tau) = \frac{A^2}{2}\cos(2\pi f_o \tau)\, e^{-2\pi^2 N_1 |\tau|} \tag{1.5}$$

The power spectral density of s(t), $S_{ss}(f)$, is therefore given by the Fourier transform of $R_{ss}(\tau)$:

$$S_{ss}(f) = FT\{R_{ss}(\tau)\} = \frac{A^2 N_1}{(2\pi)^2}\left[\frac{1}{1+\left(f-f_o/\pi N_1\right)^2} + \frac{1}{1+\left(f+f_o/\pi N_1\right)^2}\right] \tag{1.6}$$

[1] Typical definitions of "coherent modulation" assume the carrier is known perfectly at the receiver (at least for calculation of probability of error)

Figure 1.3 shows this power spectral density, which is commonly referred to as Lorentzian.

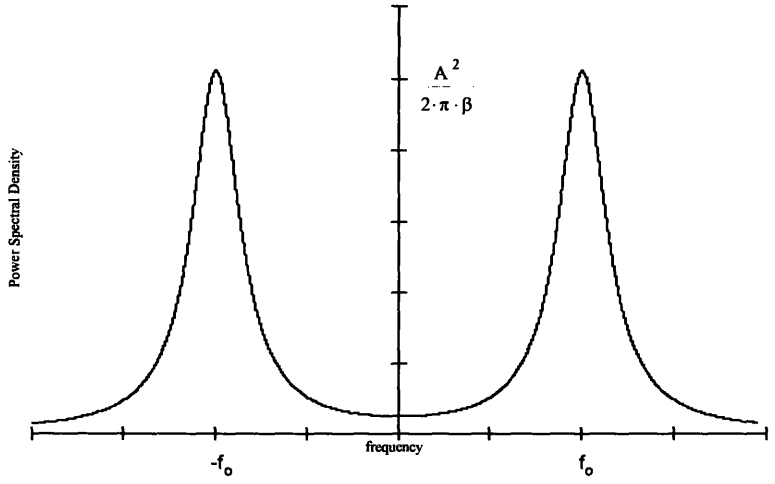

Figure 1.3: Lorentzian power spectral density

From inspection of (1.6) it is clear that the half-power bandwidth of the phase noisy sinusoid is equal to $2\pi N_1$. In fiber-optic systems, this is referred to as the *linewidth* of the signal source. To keep with this common definition, we will also refer to this quantity as the linewidth, and denote it by the symbol β:

$$\beta \equiv 2\pi N_1 \tag{1.7}$$

Having defined β it is important to note the differences between typical values for semiconductor laser sources and radio frequency sources. For laser sources of the type considered in references [6] or [7] for example, a typical value of β would be in the range 5-100 MHz. Considering the carrier frequency (e.g., 1μm system → 300,000 GHz), however, this corresponds to a linewidth/carrier frequency ratio of about $1.7\text{x}10^{-8}$ to $2.2\text{x}10^{-7}$. For a comparable RF system operating nominally around 2 GHz, this ratio would correspond to a linewidth of about 33-667 Hz. Reasonably high-quality signal sources available at this carrier frequency (2 GHz) have linewidths more likely in the range of 1-10 Hz, thus the phase noise present in fiber-optic systems is more severe than in typical radio systems. However, in the realm of low-power componentry, some of this stability is sacrificed to lower the overall power-consumption of the signal source [10].

Thus, because of our low-power consumption goal we will be dealing with significantly more phase noise, and thus the need for this research.

1.3 Overview of Thesis

Because assessing the impact of phase noise on radio communications systems is a much broader topic than could be contained in a thesis of this type, two examples of digital modulation schemes used in radio communications are explored. In Chapter 2 we explore the effect of phase noise on quadrature amplitude modulation (QAM) systems, specifically quadrature phase shift keying (QPSK). QPSK has become a popular choice for current digital radio implementations and is representative of the whole class of QAM systems.

In Chapter 3 we explore the effect of phase noise on a digital pulse-position modulation (PPM) scheme. The main difference with PPM is that phase noise in the reference oscillators in the transmitter manifests itself in timing jitter of the pulse locations, i.e., the "clock" has errors. While digital PPM is not nearly as common as QAM schemes in current commercial systems, it seems to hold promise as a ultra-low power scheme [26].

2. Detection of Quadrature Amplitude Modulation Signals

In fiber-optic communications the available bandwidth is abundant, thus binary modulation formats are almost universally used [6]. By contrast, radio communications has a usable bandwidth (for most applications) of only a fraction of that available with fiber-optics. Until relatively recently, cellular radio and most private commercial radio used predominantly analog modulation schemes such as frequency modulation (FM), but most new radio communications systems are switching to digital modulation. Because of the scarcity of bandwidth, the use of non-binary modulation schemes has become necessary in order to send higher data rates within the same bandwidth. As mentioned in the introduction, quadratrure amplitude modulation (QAM), and in particular quadriphase shift keying (QPSK), has become a popular choice for mobile radio. In fact, in certain circumstances, QPSK has been shown to be preferable to other modulation schemes, for example in a fading environment [11]. The inclusion of phase noisy signal sources, as dictated by the low-power constraints, is the basis for the following analysis.

In the following sections we will present an analysis of QPSK where the quadrature carriers are impaired by phase noise. Analysis of QPSK rather than a more general QAM signal structure was chosen because of the popularity of QPSK in current mobile radio implementations, and since QPSK represents the simplest form of QAM. For the purpose of presenting a common framework from which to base the subsequent

analysis, we will first present a short tutorial on QPSK modulation. Next, the performance of the several possible receiver structures are considered, including a fully coherent receiver, where perfect side information in the form of the phase noise waveform ($\theta(t)$) is available at the receiver. We will show that the phase noise present in the quadrature carriers destroys their orthogonality, which is implicit in standard QAM-type modulation. In later sections we will relax our assumption of perfect knowledge of $\theta(t)$ and consider performance when the receiver has knowledge of samples of the phase noise process, rather than complete knowledge for all time. Although the analysis in this chapter is specific to QPSK, it can be easily extended to higher-order QAM modulations, since these signal sets lie in the same two-dimensional space as QPSK.

2.1 QPSK Background

The structure of the QPSK signal extends simply from binary PSK, although there are many ways to express the resulting waveform. For more detail than is provided below, the reader is referred to any standard digital communications textbook, for example [12],[13],[14],[15],[16].

The data stream $\mathbf{m} = \{m_i, i \in ..-2,1,0,1,2...\}$ to be transmitted is a (possibly infinite) sequence of bits, represented by +1 and -1. Thus, m_i=+1 or -1 for each element of the sequence $\mathbf{m}$. For QPSK two data bits are combined at a time to form a 4-ary complex-valued symbol stream $\mathbf{m}^{(q)}$. The two data bits are transformed into coordinates in the complex-plane via the mapping:

$$m_i^{(q)} = \begin{cases} 1 & \text{if } m_{2i-1} = 1, m_{2i} = 1 \\ j & \text{if } m_{2i-1} = -1, m_{2i} = 1 \\ -1 & \text{if } m_{2i-1} = -1, m_{2i} = -1 \\ -j & \text{if } m_{2i-1} = 1, m_{2i} = -1 \end{cases} \tag{2.1}$$

This mapping may be described graphically as a signal set as in Figure 2.1:

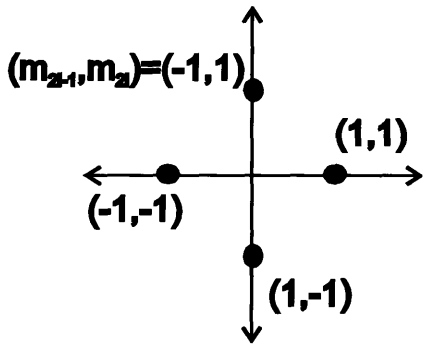

Figure 2.1: Graphical representation of the signal set for $m_i^{(q)}$

Any mapping that maintains the same geometry between the symbols (i.e., such that all inner products between the symbols are unchanged) is essentially equivalent to the one given in (2.1). The symbol stream $\mathbf{m}^{(q)}$ is then converted to an impulse stream $m^{(q)}(t)$ where the "area" of the i^{th} impulse is $m_i^{(q)}$, and the impulses are spaced T seconds apart. This impulse stream is then fed as an input to a pulse shaper, where the resulting output x(t) is a stream of pulses at baseband. The baseband signal is upconverted to a radio frequency suitable for transmission by multiplying x(t) by a complex sinusoid $s(t) = Ae^{j(2\pi f_o t+\theta(t)+\phi)}$. Since the resulting waveform is complex, the last step is to strip off the imaginary part and send only the real part of the signal. This sequence of steps is shown in block diagram form in Figure 2.2.

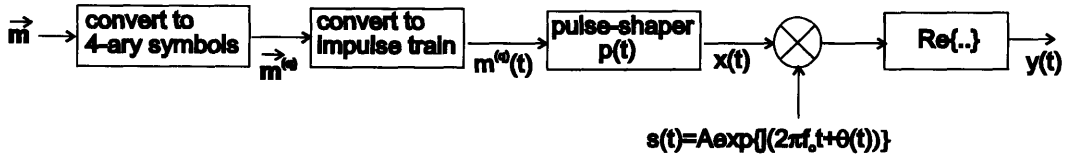

Figure 2.2: QPSK transmitter, implementation 1

In the ideal case the phase noise θ(t) is identically equal to zero. In addition, we will assume that φ (a static phase offset) equals zero since it does not affect the analysis of the system if this quantity is known at the receiver. Given the mapping in equation (2.1) we see that for a each pulse of duration T, if we choose p(t) to be a rectangular pulse

$$p(t) = \begin{cases} 1 & \text{for } 0 \le t < T \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

then the output y(t) on the interval $iT \le t < (i+1)T$ is a cosine wave with a constant phase equal to the angle of $m_i^{(q)}$, i.e., a multiple of $\pi/2$. Specifically,

$$y(t) = A\cos\left(2\pi f_o t + \theta(t) + \angle m_i^{(q)}\right) \qquad \text{for } iT \le t < (i+1)T \tag{2.3}$$

Any pulse shape p(t) can be used, however, because of our low-power constraints we are interested in only those pulse shapes that allow y(t) to be constant envelope. Thus, pulse shapes such as the family of raised cosine pulses are not appropriate. On the other hand, a Haar wavelet or Walsh function (see [22]) will do, since their absolute values are constant. In order to keep our analysis simple, we will not consider these other pulse shapes here.

An equivalent representation of a QPSK transmitter using all real signals is given in Figure 2.3.

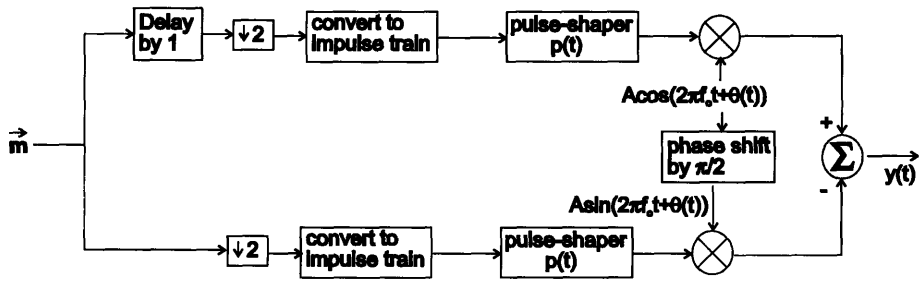

Figure 2.3: Equivalent QPSK transmitter using only real signals

Note that the signal set corresponding to Figure 2.3 is rotated by $\pi/4$ and scaled by $\sqrt{2}$ relative to that of Figure 2.2. To see that y(t) in Figure 2.3 has the form of equation (2.3), rewrite (2.3) as

$$\begin{aligned} y(t) = A\cos\left(2\pi f_o t + \theta(t) + \angle m_i^{(q)}\right) &= A\cos\left(2\pi f_o t + \theta(t) + \tfrac{n\pi}{2}\right) && \text{(for n = -1,0,1,2)} \\ = A\cos\left(\tfrac{n\pi}{2}\right)\cos\left(2\pi f_o t + \theta(t) + \tfrac{\pi}{4}\right) &- A\sin\left(\tfrac{n\pi}{2}\right)\sin\left(2\pi f_o t + \theta(t) + \tfrac{\pi}{4}\right) && \text{for } iT \le t < (i+1)T \\ & && \text{and} \quad n = -1,0,1,2 \end{aligned} \tag{2.4}$$

The terms $\cos\left(\frac{n\pi}{2}\right)$ and $\sin\left(\frac{n\pi}{2}\right)$ have values +1, 0, and -1 as n ranges over -1,0,1 and 2. In contrast, y(t) in Figure 2.3 is just a manipulation of (2.4), with the $\pi/4$ term moved from the carrier terms to the coefficient terms, i.e., the coefficients are $\sqrt{2}\cos\left(\frac{n\pi}{2}+\frac{\pi}{4}\right)$ and $\sqrt{2}\sin\left(\frac{n\pi}{2}+\frac{\pi}{4}\right)$. As an aside, if the delay block in Figure 2.3 is changed to $1/2$ instead of 1, the resulting system is commonly referred to as offset QPSK (OQPSK).

Ideal coherent receivers for y(t) transmitted across an additive white Gaussian noise (AWGN) channel can easily be defined in the case where θ(t) is *identically zero*. Figure 2.4 and Figure 2.5 show equivalent receivers.

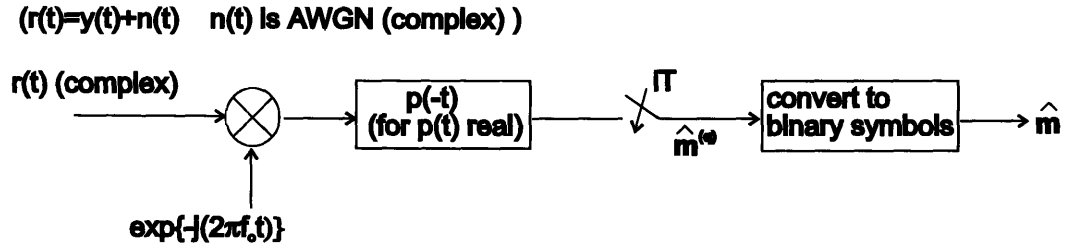

Figure 2.4: QPSK receiver, implementation 1

In Figure 2.4, we've assumed that y(t) is complex, i.e., we've left off the Re{·} block in Figure 2.2. Demodulating the real version of y(t) with a real cosine is completely equivalent.

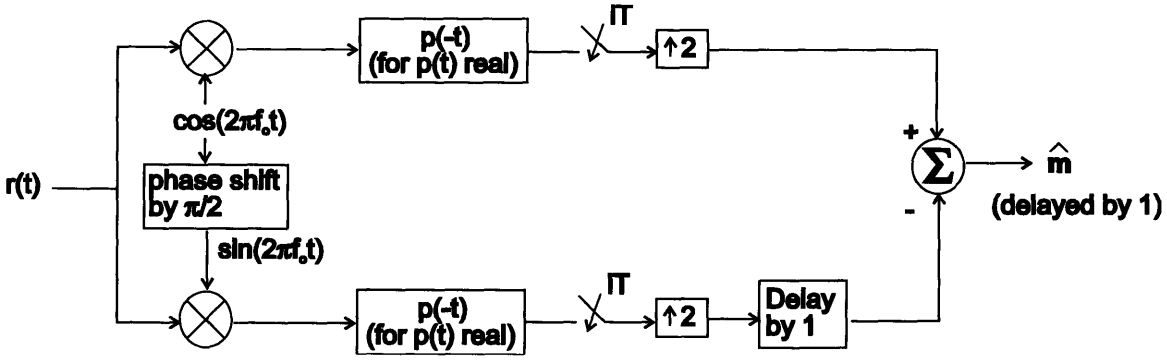

Figure 2.5: QPSK receiver, implementation 2

In Figure 2.5 , y(t) is real, and thus could be used with either transmitter exactly as shown.

There is a significant difference between the real and complex transmitters, however, when θ(t) is not identically zero, even when θ(t) is known perfectly at the receiver. If the y(t) transmitted were somehow complex, Figure 2.4 would indeed be the ideal receiver in additive white Gaussian noise - in fact it is the matched filter. However, if y(t) is real, the addition of a nonzero θ(t) causes the in-phase and quadrature channels to *not* be orthogonal over the symbol period, so that there is *crosstalk* between the in-phase and quadrature channels. Since the transmitter-receiver pair of Figure 2.3 and Figure 2.5 is the only one that is physically realizable, this is a real problem. This will be discussed in more detail in the subsequent section.

Ideally, assuming no phase noise and that the symbol period T is a multiple of $1/f_o$ (i.e., an integral number of cycles are transmitted for each symbol - a standard assumption) then the inner product between the in-phase and quadrature channels is zero. That is,

$$\rho = \left\langle \cos(2\pi f_o t), \sin(2\pi f_o t) \right\rangle = \int_0^T \cos(2\pi f_o t)\sin(2\pi f_o t)dt = 0 \tag{2.5}$$

The orthogonality of sine and cosine shows that QPSK is equivalent to two orthogonal binary PSK signals. On the other hand, when the phase noise is nonzero over [0,T), ρ is not zero in general:

$$\begin{aligned} \rho &= \int_0^T \cos(2\pi f_o t + \theta(t))\sin(2\pi f_o t + \theta(t))dt = \tfrac{1}{2}\int_0^T \sin\left(2(2\pi f_o t + \theta(t))\right)dt \\ &= \tfrac{1}{2}\int_0^T \sin\left(2(2\pi f_o t)\right)\cos(2\theta(t))dt + \tfrac{1}{2}\int_0^T \cos\left(2(2\pi f_o t)\right)\sin(2\theta(t))dt \end{aligned} \tag{2.6}$$

As an example, consider the deterministic case where $\theta(t) = -2\pi f_o t + C$, where C is an arbitrary constant. In this extreme case $\cos(2\pi f_o t + \theta(t)) = \cos(C)$ and $\sin(2\pi f_o t + \theta(t)) = \sin(C)$ and the inner product is not zero. The effect of the crosstalk is that the data bits sent "in quadrature" are subject to *dependent* errors, thus making the error probability higher than it otherwise would be. As we will see in the next section, the four possible points in the two-dimensional signal set lie on an ellipse, rather than a circle as in the ideal QPSK case.

2.2 Performance with an Ideal Phase-Tracking Receiver

Consider the case where the receiver *does* know f_o and θ(t) (for all t) at the receiver, i.e., fully coherent reception is possible. This is the case when a perfect reference signal (or pilot tone) is available at the receiver. In the literature, this type of receiver is often called a "genie" receiver. Although possibly unrealistic, this assumption does provide an upper bound to the performance possible with *any* receiver using suboptimal carrier recovery.

The receiver we derive in this section is the optimal coherent receiver, where the optimality criteria is in terms of minimizing the probability of symbol (detection) error. In the subsequent section we will show that when the phase noise and carrier frequency have values representative of radio communications the optimal receiver has performance virtually identical to the ideal receiver in the AWGN-only case, *when* f_o and θ(t) *are known perfectly at the receiver.*

2.2.1 Definitions

Consider the second transmitter-receiver pair (Figure 2.3 and Figure 2.5). The transmitted signal y(t) is

$$y(t) = m_{2i-1} A \cos\left(2\pi f_o t + \theta(t)\right) - m_{2i} A \sin\left(2\pi f_o t + \theta(t)\right) \qquad \text{for } iT \le t < (i+1)T \tag{2.7}$$

Assuming for the moment no noise in the channel, the estimate after the sampler at time t=(i+1)T in Figure 2.5 is given by

$$\begin{aligned}
\hat{m}_{2i-1} &= \int_{iT}^{(i+1)T} y(t)\cos\left(2\pi f_o t + \theta(t)\right)dt \\
&= \int_{iT}^{(i+1)T} \left(m_{i-2} A \cos\left(2\pi f_o t + \theta(t)\right) - m_{i-1} A \sin\left(2\pi f_o t + \theta(t)\right)\right)\cos\left(2\pi f_o t + \theta(t)\right)dt \\
&= \frac{m_{2i-1} AT}{2} + \frac{m_{2i-1} A}{2} \int_{iT}^{(i+1)T} \cos\left(2\left(2\pi f_o t + \theta(t)\right)\right)dt \\
&\quad - m_{2i} A \int_{iT}^{(i+1)T} \sin\left(2\pi f_o t + \theta(t)\right)\cos\left(2\pi f_o t + \theta(t)\right)dt.
\end{aligned} \tag{2.8}$$

The estimate $\hat{m}_{2i}$ is given by a similar equation. Note that the last integral on the right is the same integral as in equation (2.6), which we'll denote by ρ_i. We will see shortly that the other integral will appear often in the subsequent analysis, thus we'll define it as ε_i, that is,

$$\varepsilon_i = \frac{1}{2}\int_{iT}^{(i+1)T} \cos\left(2\left(2\pi f_o t + \theta(t)\right)\right)dt \tag{2.9}$$

Applying the same definitions to $\hat{m}_{2i}$ (incorporating the minus sign from the later sum), we see that

$$\begin{aligned} \hat{m}_{2i-1} &= \frac{m_{2i-1}AT}{2} + m_{2i-1}A\varepsilon_i - m_{2i}A\rho_i \\ \hat{m}_{2i} &= \frac{m_{2i}AT}{2} - m_{2i}A\varepsilon_i - m_{2i-1}A\rho_i \end{aligned} \tag{2.10}$$

In the absence of phase noise, ρ_i=0 and ε_i=0, but they are in general non-zero. At first glance it seems that this would not be too difficult to deal with, since the *same* ρ_i and ε_i appear in both equations. These equations demonstrate the *crosstalk* previously implied, and in fact, some *self-interference* as well (due to the ε_i terms). Because the phase noise is a random process, ρ_i and ε_i are random variables. In addition, since ρ_i and ε_i are both derived from the same random process, $\theta(t)$, they are in general not independent. For the moment we will not comment further on the structure of their probability distributions.

Another problem becomes apparent if we consider the channel to add additive white Gaussian noise to y(t). The received signal, denoted by r(t), now is given by y(t)+n(t) where n(t) is a zero-mean Gaussian random process, with two-sided spectral density $\frac{N_o}{2}$. When there is no phase noise, the additive noise variables at each sampler, denoted by n_{2i-1} and n_{2i}, are independent, since cosine and sine are orthogonal (over a multiple of $1/f_o$, the period of one cycle of the carrier). However, the addition of phase noise creates correlation between the demodulating phase noisy sinusoids, thus $E[n_{2i-1}n_{2i}] \neq 0$. In fact, a quick calculation shows that $E[n_{2i-1}n_{2i}] = \rho_i$.

A lower bound on the probability of error (either bit ($P_{err,bit}$) or symbol ($P_{err,sym}$)) for any system of this type can be found by assuming fully coherent reception, i.e., that

$\theta(t)$, f_o, and ϕ are known exactly at the receiver. This is completely analogous to the well-defined case (in textbooks) with $\theta(t)=0$. In the textbook case with $\theta(t)=0$, the probability of symbol error, $P_{err,sym}$, is given exactly by (see [9],[12],[13],[14], or [16])

$$P_{err,sym} = 2Q\left(\sqrt{\frac{E_s}{N_o}}\right) - \left[Q\left(\sqrt{\frac{E_s}{N_o}}\right)\right]^2 \tag{2.11}$$

where Q(x) is the complementary distribution function of a zero-mean, unit variance Gaussian random variable evaluated at x, E_s is the energy per symbol ($E_s = \frac{A^2T}{2}$ in our derivation with no phase noise), and $\frac{N_o}{2}$ is the (two-sided) additive white Gaussian noise density in Watts/Hz. For high energy/symbol to noise spectral density ratios, i.e., for $\frac{E_s}{N_o} >> 1$, the second term is essentially negligible and the probability of symbol error is well approximated by

$$P_{err,sym} \cong 2Q\left(\sqrt{\frac{E_s}{N_o}}\right) \tag{2.12}$$

and the bit error probability, $P_{err,bit}$, is approximated by

$$P_{err,bit} \cong \frac{P_{err,sym}}{\log_2(4)} \cong Q\left(\sqrt{\frac{E_s}{N_o}}\right) \tag{2.13}$$

These will be the standards against which we compare our receivers in the presence of both phase noise and AWGN.

2.2.2 Analysis

There are two equivalent optimal receiver configurations, both easy to visualize. The first is just the logical extension of Figure 2.5 where the phase noise process is included in the in-phase and quadrature branches. In this case we are matched filtering the signal, however, since the signal set is no longer bi-orthogonal due to the phase noise, the noise components after sampling are no longer orthogonal (as shown before). Thus we have to perform some additional processing before the decision device to make a decision. This approach is shown in Figure 2.6 below.

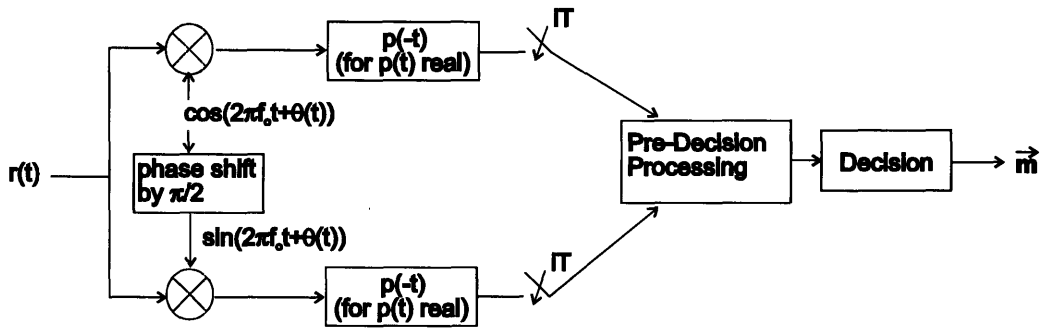

Figure 2.6: QPSK receiver, matched filter with phase noise known at receiver

An equivalent and conceptually simpler receiver maps the received signal onto an *orthogonal* basis that spans the same space as $\cos(2\pi f_o t + \theta(t))$ and $\sin(2\pi f_o t + \theta(t))$. This is similar to the receiver in Figure 2.5 that maps the received signal onto the space spanned by $\cos(2\pi f_o t)$ and $\sin(2\pi f_o t)$ (in the case with no phase noise). The advantage of this receiver over the receiver in Figure 2.6 is that the additive noise variables at the samplers are now *independent*, rather than correlated as in the previous receiver. This orthogonal basis is easily determined using the Gram-Schmidt orthogonalization procedure. Since the phase noise, θ(t), is not the same during each symbol, *the orthogonal basis will be different for each symbol.* In accordance with the G-S procedure we pick one of the signals as the first orthogonal basis signal,

$$\Psi_{1,i}(t) = \cos(2\pi f_o t + \theta(t)) \qquad \text{for } iT \le t < (i+1)T \tag{2.14}$$

Then, the other orthogonal basis signal, $\Psi_{2,i}(t)$, is given by the G-S orthogonalization procedure as

$$\Psi_{2,i}(t) = \sin(2\pi f_o t + \theta(t)) - \frac{\langle \sin(2\pi f_o t + \theta(t)), \Psi_{1,i}(t) \rangle}{\|\Psi_{1,i}(t)\|^2} \Psi_{1,i}(t) \qquad \text{for } iT \le t < (i+1)T \tag{2.15}$$

where $\langle a, b \rangle$ denotes the usual inner product. The need for a different G-S-derived basis for each symbol makes this is a physically unrealizable receiver (at least in real-time), but is one which provides the upper bound on $P_{err,sym}$. The inner product $\langle \sin(2\pi f_o t + \theta(t)), \Psi_{1,i}(t) \rangle$ was already calculated in (2.6) (denoted by ρ), then defined as ρ_i

in (2.8). The squared norm in the denominator is the energy per symbol (in the cosine direction), i.e., $\|\Psi_{1,i}(t)\|^2 = \frac{T}{2} + \varepsilon_i$, where ε_i is given in (2.9). Thus, our basis signals are random processes generated by the single random process $\theta(t)$:

$$\begin{aligned} \Psi_{1,i}(t) &= \cos\left(2\pi f_o t + \theta(t)\right) \\ \Psi_{2,i}(t) &= \sin\left(2\pi f_o t + \theta(t)\right) - \frac{\rho_i}{\frac{T}{2}+\varepsilon_i}\Psi_{1,i}(t) \qquad \text{for } iT \le t < (i+1)T \end{aligned} \tag{2.16}$$

By construction, $\Psi_{1,i}(t)$ and $\Psi_{2,i}(t)$ are orthogonal for each i, and span the two-dimensional subspace that contains all the signal energy (including phase noise). The additive white Gaussian noise introduced by the channel leads to independent noise variables at the decision device, thus allowing minimum distance decoding. A "genie" receiver is of the type in Figure 2.7.

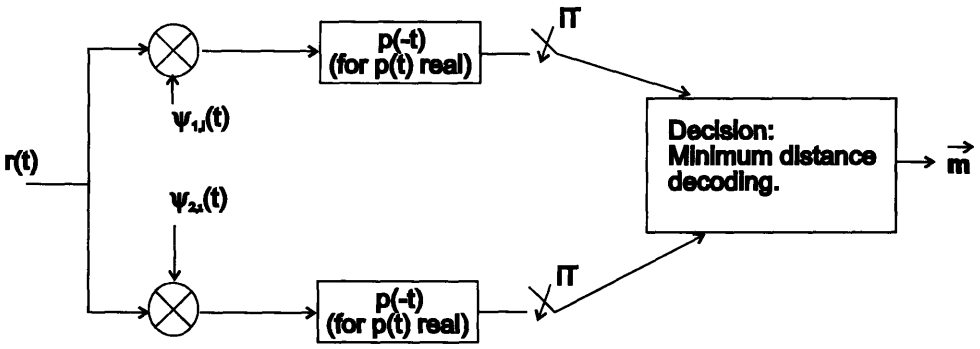

Figure 2.7: "Genie" QPSK receiver, orthogonal projection in the presence of phase noise

In Figure 2.7, we have replaced $\Psi_{1,i}(t)$ and $\Psi_{2,i}(t)$ with ortho*normal* basis signals $\psi_{1,i}(t)$ and $\psi_{2,i}(t)$. The norm of $\Psi_{2,i}(t)$ is

$$\|\Psi_{2,i}(t)\|^2 = \frac{T}{2} - \varepsilon_i - \frac{\rho_i^2}{\frac{T}{2}+\varepsilon_i} = \frac{\left(\frac{T}{2}\right)^2 - \left(\varepsilon_i + \rho_i\right)^2}{\frac{T}{2}+\varepsilon_i} \tag{2.17}$$

Thus, the orthonormal basis signals, $\psi_{1,i}(t)$ and $\psi_{2,i}(t)$ are

$$\psi_{1,i}(t) = \frac{\Psi_{1,i}(t)}{\left\|\Psi_{1,i}(t)\right\|} = \frac{\cos\left(2\pi f_o t + \theta(t)\right)}{\sqrt{\frac{T}{2} + \varepsilon_i}}$$

$$\psi_{2,i}(t) = \frac{\Psi_{2,i}(t)}{\left\|\Psi_{2,i}(t)\right\|} = \frac{\left[\sqrt{\frac{T}{2} + \varepsilon_i}\,\sin\left(2\pi f_o t + \theta(t)\right) - \rho_i \psi_{1,i}(t)\right]}{\sqrt{\left(\frac{T}{2}\right)^2 - \left(\varepsilon_i^2 + \rho_i^2\right)}} \qquad \text{for } iT \le t < (i+1)T \tag{2.18}$$

The statistical properties of the orthonormal basis signals in (2.18) are hard to characterize, since ε_i and ρ_i are, in general, dependent random variables that are themselves hard to characterize.

As indicated before, the additive noise components entering the decision device from each branch are independent, so finding the $P_{err,sym}$ is would seem to be straightforward. To do this, however, requires characterizing ε_i and ρ_i fully, for example with their joint probability density function. As we've alluded to before this proves to be a difficult task. We shall see shortly that we only need the first and second moments of ε_i and ρ_i to evaluate the receiver's performance in the radio communications. For now, we will proceed with specifying the optimal receiver structure without commenting on the density functions of ε_i and ρ_i (or $\psi_{1,i}(t)$ and $\psi_{2,i}(t)$ for that matter).

Using the basis functions $\psi_{1,i}(t)$ and $\psi_{2,i}(t)$ as axes we can find the locations of the received signal points in the absence of AWGN. Two points lie on the $\psi_{1,i}(t)$ axis, at $\pm A\left\|\Psi_{1,i}(t)\right\| = \pm A\sqrt{\frac{T}{2} + \varepsilon_i}$, while the other two points (corresponding to signals $\pm A\sin\left(2\pi f_o t + \theta(t)\right)$) are *not* on the $\psi_{2,i}(t)$ axis, as the signal directions are not orthogonal due to the phase noise process. The locations ,(x,y) and (-x,-y), of these signal can be found by taking the inner product of $\pm A\sin\left(2\pi f_o t + \theta(t)\right)$ and $\psi_{1,i}(t)$ and $\psi_{2,i}(t)$.

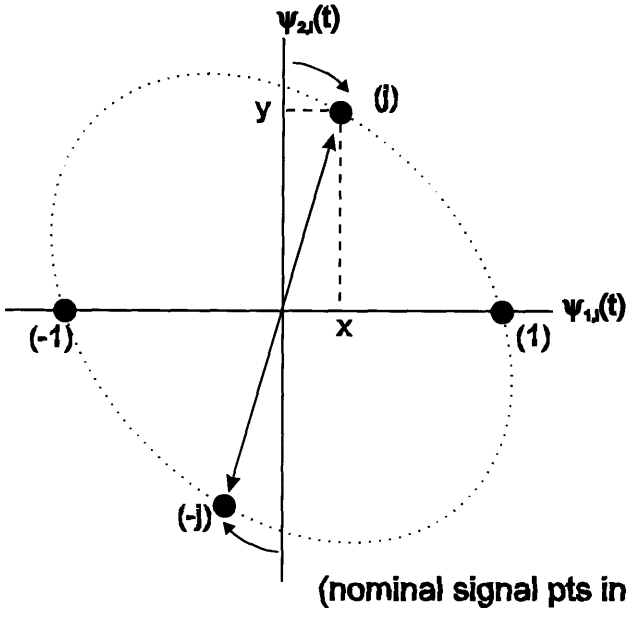

Figure 2.8: Signal space representation of signal points for "genie" receiver

The signal point locations are summarized in Table 2.1 and a representative plot of those points is given in Figure 2.8.

signal point	Cartesian coordinates in direction of: $\psi_{1,i}(t)$	$\psi_{2,i}(t)$
$A\cos(2\pi f_o t+\theta(t))$ (1)	$A\sqrt{\frac{T}{2}+\varepsilon_i}$	0
$-A\cos(2\pi f_o t+\theta(t))$ (-1)	$-A\sqrt{\frac{T}{2}+\varepsilon_i}$	0
$A\sin(2\pi f_o t+\theta(t))$ (j)	$A\frac{\rho_i}{\sqrt{\frac{T}{2}+\varepsilon_i}}$	$A\frac{\sqrt{\left(\frac{T}{2}\right)^2-\left(\varepsilon_i^2+\rho_i^2\right)}}{\sqrt{\frac{T}{2}+\varepsilon_i}}$
$-A\sin(2\pi f_o t+\theta(t))$ (-j)	$-A\frac{\rho_i}{\sqrt{\frac{T}{2}+\varepsilon_i}}$	$-A\frac{\sqrt{\left(\frac{T}{2}\right)^2-\left(\varepsilon_i^2+\rho_i^2\right)}}{\sqrt{\frac{T}{2}+\varepsilon_i}}$

Table 2.1: Signal space representation of signal points in ($\psi_{1,i}(t)$, $\psi_{2,i}(t)$) space

2.2.2.1 *Mean and Variance of ε_i and ρ_i*

Having established the location of the signal points, we could calculate or bound the $P_{err,sym}$, if we could characterize the joint distribution of the random variables, ε_i and ρ_i. Their definitions are repeated here for convenience:

$$\begin{aligned}\rho_i &= \tfrac{1}{2}\int_{iT}^{(i+1)T}\sin\left(2\left(2\pi f_o t+\theta(t)\right)\right)dt\\ \varepsilon_i &= \tfrac{1}{2}\int_{iT}^{(i+1)T}\cos\left(2\left(2\pi f_o t+\theta(t)\right)\right)dt\end{aligned} \tag{2.19}$$

What we are really concerned about is operation in "steady-state," i.e., for t=iT>>0. In this regime the random variables ε_i and ρ_i have the same marginal probability density functions, since the difference between sine and cosine dissipates as $\theta(t)|_{t=iT}$ (modulo 2π) becomes uniformly distributed on $[0,2\pi)$. We therefore focus on ρ_i. Without loss of generality we can focus on the first symbol only, and drop the dependence on i. That is, we can write ρ_i as merely ρ, and ε_i as ε:

$$\begin{aligned}\rho &= \tfrac{1}{2}\int_0^T\sin\left(4\pi f_o t+\zeta+2\tilde{\theta}(t)\right)dt\\ \varepsilon &= \tfrac{1}{2}\int_0^T\cos\left(4\pi f_o t+\zeta+2\tilde{\theta}(t)\right)dt\end{aligned} \tag{2.20}$$

where ζ is uniform on $[0,2\pi)$, (represents the initial state of the Brownian motion process at the beginning of the symbol period) and $\tilde{\theta}(t)$ represents the Brownian motion process since beginning of the symbol period. The means of ε and ρ are both zero in steady-state (as we should expect due to the symmetry of the problem) since $E[\sin(\zeta)]$ and $E[\cos(\zeta)]$ both equal zero. Explicitly,

$$\begin{aligned}E[\rho] &= E\left[\tfrac{1}{2}\int_0^T\sin\left(4\pi f_o t+\zeta+2\tilde{\theta}(t)\right)dt\right]\\ &= \tfrac{1}{2}\int_0^T E\left[\sin\left(4\pi f_o t+\zeta+2\tilde{\theta}(t)\right)\right]dt\\ &= \tfrac{1}{2}\int_0^T E[\sin(\zeta)]E\left[\cos\left(4\pi f_o t+2\tilde{\theta}(t)\right)\right]+E[\cos(\zeta)]E\left[\sin\left(4\pi f_o t+2\tilde{\theta}(t)\right)\right]dt\\ &= 0\end{aligned} \tag{2.21}$$

with a similar equation for the mean of ε. Now, consider the variances of ε and ρ in steady-state. In detail:

$$\begin{aligned} E\left[\rho^2\right] &= \frac{1}{4}E\left[\int_0^T \sin\left(2\pi(2f_o)t+\zeta+2\tilde{\theta}(t)\right)dt \int_0^T \sin\left(2\pi(2f_o)s+\zeta+2\tilde{\theta}(s)\right)ds\right] \\ &= \frac{1}{4}\int_0^T\int_0^T E\left[\sin\left(2\pi(2f_o)t+\zeta+2\tilde{\theta}(t)\right)\sin\left(2\pi(2f_o)s+\zeta+2\tilde{\theta}(s)\right)\right]dtds \\ &= \frac{1}{8}\int_0^T\int_0^T \left(\begin{array}{l} E\left[\cos\left(2\pi(2f_o)(t-s)+2\left(\tilde{\theta}(t)-\tilde{\theta}(s)\right)\right)\right] \\ +E\left[\cos\left(2\pi(2f_o)(t+s)+2\zeta+2\left(\tilde{\theta}(t)+\tilde{\theta}(s)\right)\right)\right] \end{array}\right)dtds \end{aligned} \tag{2.22}$$

Note that the second expectation is zero since 2ζ (modulo 2π) is uniform on $[0,2\pi)$. Thus,

$$\begin{aligned} E\left[\rho^2\right] &= \frac{1}{8}\int_0^T\int_0^T E\left[\cos\left(2\pi(2f_o)(t-s)+2\left(\tilde{\theta}(t)-\tilde{\theta}(s)\right)\right)\right]dtds \\ &= \frac{1}{8}\int_0^T\int_0^T E\left[\cos\left(2\pi(2f_o)(t-s)+2\left(\theta(t)-\theta(s)\right)\right)\right]dtd \end{aligned} \tag{2.23}$$

Now, because of the integrals over both t and s, we need to divide the analysis to consider both the integrals over the region where t>s and s>t. For t>s, we can replace $\theta(t)-\theta(s)$ by $\theta^*(t-s)$, where $\theta^*(t)$ is Brownian motion with $\theta^*(0)=0$. For s>t we get similar expressions, and note that we can split the double integral in (2.23) into two sets of double integrals, one for t>s and the other for s>t. Without getting lost in the details we find that (2.23) can be expressed as

$$E\left[\rho^2\right] = \frac{1}{4}\int_0^T\int_0^t E\left[\cos\left(2\pi(2f_o)(t-s)+2\theta^*(t-s)\right)\right]dsdt \tag{2.24}$$

Leaving the details to the reader, the expectation in the integrand is

$$E\left[\cos\left(2\pi(2f_o)(t-s)+2\theta^*(t-s)\right)\right] = \cos\left(2\pi(2f_o)(t-s)\right)e^{-2(2\pi)^2N_1(t-s)} \tag{2.25}$$

Substituting (2.25) into(2.24),

$$E\left[\rho^2\right] = \frac{\left[\begin{array}{l} \left(3(2\pi N_1 f_o)^2-4f_o^4+(2\pi N_1)^4\right)e^{-2(2\pi)^2N_1T}+160N_1^3\pi^4f_o^2T \\ -3(2\pi N_1f_o)^2-(2\pi N_1)^4+128N_1^5\pi^6T+32N_1\pi^2f_o^4T+4f_o^4 \end{array}\right]}{256\pi^2\left(16\pi^6N_1^6+24\pi^4N_1^4f_o^2+9\pi^2N_1^2f_o^4+f_o^6\right)} = Var[\rho] \tag{2.26}$$

since ρ is zero-mean. In addition, remember that Var[ε]=Var[ρ]. Recalling that in the introduction we defined the *linewidth* of the signal source to be $\beta = 2\pi N_1$, and further defining a "phase noise SNR", γ, as the variance of the phase noise process at time t=T, i.e.

$$\gamma = 2\pi\beta T \tag{2.27}$$

then we can rewrite the previous equations in a more compact and useful form. Specifically the variance ε and ρ can be rewritten by combining (2.27) into (2.26),

$$Var[\rho] = Var[\varepsilon] = \frac{\left[\begin{array}{l}\left(3(\beta f_o)^2 + \beta^4 - 4f_o^4\right)e^{-2\gamma} + \left(2\beta^4 + 10\beta^2 f_o^2 + 8f_o^4\right)\gamma \\ -3(\beta f_o)^2 - \beta^4 + 4f_o^4\end{array}\right]}{64\pi^2\left(\beta^6 + 6\beta^4 f_o^2 + 9\beta^2 f_o^4 + 4f_o^6\right)} \tag{2.28}$$

The expression in (2.28) is still complicated, but for the radio communications problem, we are really only interested in the values for certain values of β, γ, and f_o, which we explore below.

First, as a sanity check, note that in the case of *no phase noise* (θ(t)=0), γ=0 and thus the variance of ρ (and thus ε) is zero, as we would expect. Secondly, for conventional radio communications, the carrier should resemble a sinusoid, or collection of sinusoids at frequencies around f_o (for adequate propagation through the atmosphere), which is possible only if $f_o >> \beta$. In this case we can accurately approximate the variance as

$$\begin{aligned} Var[\rho] = Var[\varepsilon] &\cong \frac{\left[-4f_o^4 e^{-2\gamma} + 8f_o^4\gamma + 4f_o^4\right]}{64\pi^2 f_o^6} \\ &= \frac{\left[1 - e^{-2\gamma} + 2\gamma\right]}{16\pi^2 f_o^2} \end{aligned} \tag{2.29}$$

By noting that $e^{-x} > 1 - x$ we can upper bound the variance (treating the approximately equals sign as a true equality) as

$$Var[\rho] < \frac{\gamma}{4\pi^2 f_o^2} \tag{2.30}$$

Likewise, for very large γ, the first two terms in (2.29) become negligible compared to 2γ, thus we can lower bound the variance as $\frac{\gamma}{8\pi^2 f_o^2}$. Combining the previous results the variance is restricted to be

$$\frac{\gamma}{8\pi^2 f_o^2} < Var[\rho] < \frac{2\gamma}{8\pi^2 f_o^2} \tag{2.31}$$

so long as the $f_o >> \beta$ condition is met. Thus, to within a factor of two, the variance of ρ (and ε) is linear in γ.

At this point we have dealt with the means and variances of ε and ρ but we haven't dealt with their marginal or joint probability density functions, which seem quite difficult to specify. From the mean and variance discussion however, we have some insight into their behavior as a function of f_o, β, and T. In the next section we shall see that it will not be necessary to compute $f_{\varepsilon,\rho}(\varepsilon,\rho)$ to analyze the performance of the "genie" receiver set forth in this section, at least for the radio communications problem that we are interested in.

2.2.2.2 Error Probabilities for an Ideal Phase-Tracking Receiver

Given the characterization of the signal point locations and the random variables ε and ρ in the previous section, we can easily determine the optimal decision regions for the receiver of Figure 2.7. It was previously noted that the four signal points of the phase noisy QPSK signal set lie in a two-dimensional plane on an ellipse (refer to Figure 2.8). Furthermore, because the signal points lie on an ellipse, the optimal decision regions are not as simple as no phase noise case. A particularly egregious example of the possible irregularity of the decision regions is given in Figure 2.9.

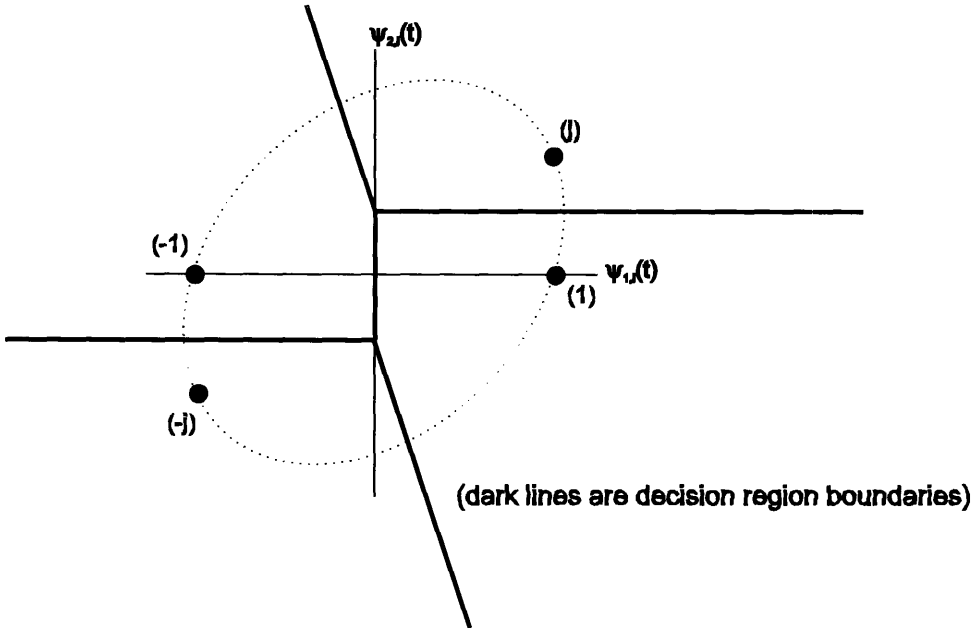

Figure 2.9: Optimal decision regions (exaggerated case)

The decision regions displayed in Figure 2.9 change from symbol to symbol, as the "genie" receiver makes use of it's knowledge of θ(t) to recalculate $\psi_{1,i}(t)$ and $\psi_{2,i}(t)$ for each symbol.

Since we have projected the signal onto an subspace with two orthonormal basis signals $\psi_{1,i}(t)$ and $\psi_{2,i}(t)$, the probability of symbol error, conditional on θ(t) and the signal point sent $m_i^{(q)}$, can be found as follows:

$$\begin{aligned} P_{err,sym}\left(\theta(t), m_i^{(q)}\right) &= 1 - \Pr(\text{signal} + \text{noise is in correct decision region} | \theta(t), m_i^{(q)}) \\ &= 1 - P_{correct}\left(\theta(t), m_i^{(q)}\right) \end{aligned} \tag{2.32}$$

where, in general, $P_{correct}\left(\theta(t), m_i^{(q)}\right)$ is obtained by a double integral over the decision region, i.e.,

$$P_{correct}\left(\theta(t), m_i^{(q)}\right) = \iint_{\substack{decision \\ region\ for\ m_i^{(q)}}} \frac{1}{\pi N_o} e^{-\frac{(x-x_1)^2}{N_o}} e^{-\frac{(y-y_2)^2}{N_o}} dxdy \tag{2.33}$$

and where x_1 and y_2 represent the coordinates of the $m_i^{(q)}$ signal point in the $(\psi_{1,i}(t), \psi_{2,i}(t))$-plane (as listed in Table 2.1) in the absence of AWGN, conditional on a particular θ(t) process. The integral in (2.33) is in general difficult to calculate, since it is in essence a two-dimensional Q-function. In addition, the unconditional probability of a correct decision requires averaging over all possible $m_i^{(q)}$ and θ(t) - itself a formidable task. Explicitly, the unconditional probability of symbol error is given by

$$P_{err,sym} = E_{\theta(t),m_i^{(q)}}\left[P_{err,sym}\left(\theta(t), m_i^{(q)}\right)\right] \tag{2.34}$$

Despite all the apparent difficulties, however, the physical reality of our radio communication problem simplifies our analysis immensely. Namely, the transmission of electromagnetic waves into free space is only possible for relatively large f_o, on the order of at least several megahertz to sustain reasonable propagation and data rates of at least a 1 kilobit/sec. In addition, the assumption of an extremely phase noisy oscillator in physical terms equates to a β to f_o ratio of only perhaps 1/10000 or less. For example, at 1 KHz this would be a linewidth of 0.1 Hz; at 1 GHz this would be 100 KHz. In these cases we are in the $f_o >> \beta$ region and thus the variances of ε and ρ in steady state are given by (2.29), which is repeated here for convenience:

$$Var[\rho] = Var[\varepsilon] \cong \frac{\left[1 - e^{-2\gamma} + 2\gamma\right]}{16\pi^2 f_o^2} \tag{2.29}$$

In the last section we also bounded these variances (equation (2.31)) to within a factor of 2 from the lower to upper bound, and showed that for large γ, the variances can be well approximated by

$$Var[\rho] = Var[\varepsilon] \cong \frac{\gamma}{8\pi^2 f_o^2} \quad \text{for large } \gamma \tag{2.35}$$

Now, we note from the signal point coordinates in Table 2.1 that for the received signal point locations to be perturbed significantly from nominal (no AWGN case), the *standard deviation* of ε and ρ would need to be a significant fraction of T/2. In the case of $f_o >> \beta$, however, the standard deviation of ε and ρ *can not* be a significant fraction of

T/2 (for them to be so would violate the $f_o >> \beta$ assumption). This can be shown by contradiction. Suppose that the standard deviation of ε *is* on the same order of magnitude as T. That is,

$$\sqrt{\frac{\gamma}{8\pi^2 f_o^2}} \approx order\ of\ T \Rightarrow \sqrt{\frac{(2\pi)^2 N_1 T}{8\pi^2 f_o^2}} \approx order\ of\ T \Rightarrow$$
$$\sqrt{\frac{N_1}{f_o^2}} \approx order\ of\ \sqrt{\mathrm{T}} \Rightarrow \frac{N_1}{T} \approx order\ of\ f_o^2 \Rightarrow \tag{2.36}$$
$$\frac{\beta}{T} \approx order\ of\ f_o^2$$

Since $f_o >> \beta$ by assumption, the last line above implies that $\frac{1}{T} >> f_o$, i.e., the symbol rate has to be much greater than the carrier frequency. It should be clear that no *conventional* radio system employing a carrier could operate in this fashion and in any event our analysis is valid only when T is an integer multiple of $1/f_o$. Thus the only way the standard deviation of ε (or ρ) could be large enough to perturb the signal locations significantly is for β to be much larger, which violates our $f_o >> \beta$ assumption.

Thus, for all practical purposes the variables ε and ρ are almost indistinguishable from zero, and thus the probability of symbol error for the "genie" receiver is *virtually identical* to that of the ideal receiver in AWGN alone. Namely the probability of error is as given in (2.11), i.e.

$$P_{err,sym} = 2Q\left(\sqrt{\frac{E_s}{N_o}}\right) - \left[Q\left(\sqrt{\frac{E_s}{N_o}}\right)\right]^2 \tag{2.11}$$

which is derived from the decision regions as in Figure 2.10.

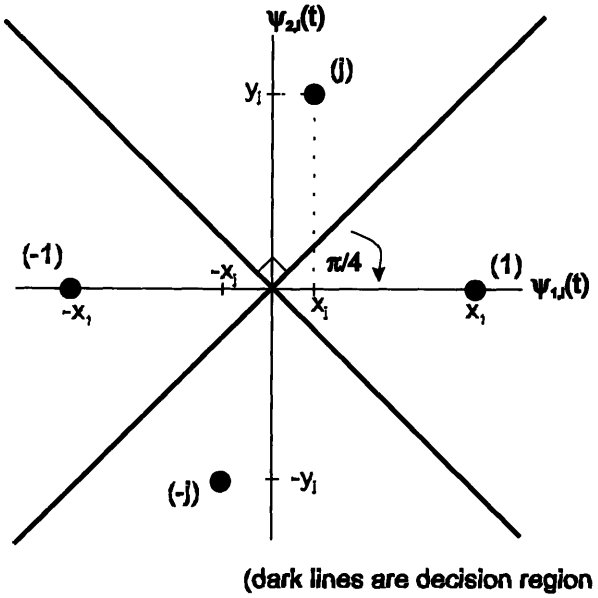

Figure 2.10: "Almost" optimal decision regions in phase noisy case

In addition, since any higher-order QAM receiver would the same front-end structure as Figure 2.7 the results extend similarly for M-QAM signal structures. The main result is that, although phase noise in the carrier genuinely destroys the orthogonality of the generator signals (the cosine and sine), in the *radio* communications problem this deviation is so small as to be imperceptible if perfect carrier recovery is possible. On the other hand, the preceding analysis would suggest that a baseband (or near baseband) system, not subject to electromagnetic propagation constraints, might be severely affected by an extremely phase noisy oscillator.

2.3 Performance with Partial Phase Tracking

In the preceding section we analyzed the performance of QPSK with a "genie" receiver in which *perfect* coherent detection was possible. Most importantly we found that under the constraints of f_o and β suitable for radio communication, perfect knowledge of f_o and $\theta(t)$ at the receiver allows performance virtually *identical* to the ideal receiver operating with AWGN-only, even though the "in-phase" and "quadrature" channels are not completely orthogonal as in the no phase noise case. The availability of such perfect side information at the receiver is, of course, not realistic in most cases, therefore in this section we will relax our assumption of perfect side information and consider how performance is degraded if the receiver has only *partial* information about $\theta(t)$. We will first consider performance when only the initial phase at the beginning of each symbol is known, i.e., only the samples $\theta(t)|_{t=iT}$ are known. We will then extend this analysis to a simple receiver that has N (N>1) samples of $\theta(t)$ available per symbol. As we would expect, performance degrades significantly when very little information is known about $\theta(t)$, and improves has N increases.

2.3.1 Phase Known only at Sampling Instants

In a manner somewhat analogous to the discrete updating nature of decision-directed carrier recovery, in this section we will consider the performance of the receiver in Figure 2.6 when f_o is perfectly known, but $\theta(t)$ is known *only* at the sampling instants, i.e., the samples $\{\theta(iT)\}$ of the $\theta(t)$ process are all that is known about $\theta(t)$ at the receiver. This "semi-coherent" receiver has a structure given in Figure 2.11.

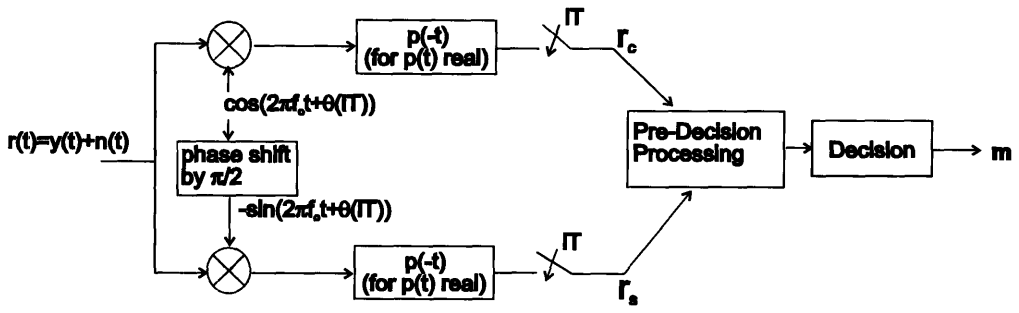

Figure 2.11: "Semi-coherent" QPSK receiver with partial phase noise information

In this receiver configuration the demodulating carriers, $\cos(2\pi f_o t + \theta(iT))$ and $\sin(2\pi f_o t + \theta(iT))$ *are* orthogonal, unlike the receiver studied in Figure 2.6. Thus our additive noise samples after the samplers are *independent* and we can perform minimum-distance decoding in the decision device without any further processing. The received signal r(t) can be written as

$$r(t) = A\cos\left(2\pi f_o t + \theta(iT) + \tilde{\theta}(t-iT) + \angle m_i^{(q)}\right) + n(t) \tag{2.37}$$

where we've expressed the phase noise θ(t) as $\theta(iT) + \tilde{\theta}(t-iT)$, where $\tilde{\theta}(t-iT)$ represents the innovations in the phase noise process from time iT to time t, and where n(t) is zero-mean AWGN with two-sided spectral density $\frac{N_o}{2}$. Since we consider only single-symbol detection, we can set i=0 without loss of generality, thus define the initial phase $\theta(iT)|_{i=0}$ as a random variable ζ, uniformly distributed over [0,2π) (modulo 2π), and we can write r(t) as

$$r(t) = A\cos\left(2\pi f_o t + \zeta + \tilde{\theta}(t) + \angle m_i^{(q)}\right) + n(t) \tag{2.38}$$

Thus, the random variables r_c and r_s after the samplers can be expressed as

$$\begin{aligned} r_c &= \int_0^T r(t)\cos(2\pi f_o t + \zeta)dt \\ r_s &= \int_0^T r(t)\sin(2\pi f_o t + \zeta)dt \end{aligned} \tag{2.39}$$

The integrands in (2.39) can be expanded to

$$
\begin{aligned}
r(t)\cos(2\pi f_o t+\zeta) &= A\cos\left(2\pi f_o t+\zeta+\tilde{\theta}(t)+\angle m_i^{(q)}\right)\cos(2\pi f_o t+\zeta) \\
&= \frac{A}{2}\cos\left(\tilde{\theta}(t)\right)\cos\left(\angle m_i^{(q)}\right)-\frac{A}{2}\sin\left(\tilde{\theta}(t)\right)\sin\left(\angle m_i^{(q)}\right)+n(t)\cos(2\pi f_o t+\zeta) \\
&+\frac{A}{2}\cos\left(2\pi(2f_o)t+2\zeta+\tilde{\theta}(t)+\angle m_i^{(q)}\right)
\end{aligned}
\tag{2.40}
$$

and

$$
\begin{aligned}
r(t)\sin(2\pi f_o t+\zeta) &= \frac{A}{2}\sin\left(\tilde{\theta}(t)\right)\cos\left(\angle m_i^{(q)}\right)-\frac{A}{2}\cos\left(\tilde{\theta}(t)\right)\sin\left(\angle m_i^{(q)}\right)+n(t)\sin(2\pi f_o t+\zeta) \\
&+\frac{A}{2}\sin\left(2\pi(2f_o)t+2\zeta+\tilde{\theta}(t)+\angle m_i^{(q)}\right)
\end{aligned}
\tag{2.41}
$$

In Section 2.2 we showed that the integrals from 0 to T of the $2f_o$ terms (i.e., the last terms in both (2.40) and (2.41), which are essentially the same as ε and ρ) are essentially negligible for radio communications $(f_o >> \beta)$, thus *for our problem* we can write

$$
\begin{aligned}
r_c &\cong \int_0^T \frac{A}{2}\cos\left(\tilde{\theta}(t)\right)\cos\left(\angle m_i^{(q)}\right)-\frac{A}{2}\sin\left(\tilde{\theta}(t)\right)\sin\left(\angle m_i^{(q)}\right)+n(t)\cos(2\pi f_o t+\zeta)dt \\
&= \frac{A}{2}\cos\left(\angle m_i^{(q)}\right)\int_0^T\cos\left(\tilde{\theta}(t)\right)dt-\frac{A}{2}\sin\left(\angle m_i^{(q)}\right)\int_0^T\sin\left(\tilde{\theta}(t)\right)dt+n_c
\end{aligned}
\tag{2.42}
$$

and

$$
r_s \cong \frac{A}{2}\cos\left(\angle m_i^{(q)}\right)\int_0^T\sin\left(\tilde{\theta}(t)\right)dt+\frac{A}{2}\sin\left(\angle m_i^{(q)}\right)\int_0^T\cos\left(\tilde{\theta}(t)\right)dt+n_s \tag{2.43}
$$

where n_c and n_s are independent, zero-mean Gaussian random variables with variance $\frac{N_o}{4}T$. The remaining integrals are random variables in their own right, both derived from the random process $\tilde{\theta}(t)$, thus they are not, in general, independent. Since they are central to our analysis we will define them as

$$
\begin{aligned}
\mu_c &= \int_0^T\cos\left(\tilde{\theta}(t)\right)dt \\
\mu_s &= \int_0^T\sin\left(\tilde{\theta}(t)\right)dt
\end{aligned}
\tag{2.44}
$$

Later in this section we shall see that simplifications in the notation and analysis can be made by normalizing these random variables by T, thus we also define

$$\mu_c' = \frac{\mu_c}{T} = \frac{1}{T}\int_0^T \cos\left(\tilde{\theta}(t)\right)dt$$
$$\mu_s' = \frac{\mu_s}{T} = \frac{1}{T}\int_0^T \sin\left(\tilde{\theta}(t)\right)dt \tag{2.45}$$

Thus we can express r_c and r_s compactly as (treating the $\cong$ as a strict equality for convenience)

$$r_c = \frac{AT}{2}\left[\cos\left(\angle m_i^{(q)}\right)\mu_c' - \sin\left(\angle m_i^{(q)}\right)\mu_s'\right] + n_c$$
$$r_s = \frac{AT}{2}\left[\cos\left(\angle m_i^{(q)}\right)\mu_s' + \sin\left(\angle m_i^{(q)}\right)\mu_c'\right] + n_s \tag{2.46}$$

As we saw with the variables ε and ρ in the last section, the probability density functions for the random variables μ_c' and μ_s' seem to be difficult to evaluate explicitly. Thus we will concentrate on evaluating their first and second moments to give us insight into the problem at hand.

As a sanity check, note that in the case of no phase noise ($\theta(t)=0$) that $\mu_c' = 1$ and $\mu_s' = 0$, which reduces (2.46) to the expected result of the ideal receiver in AWGN-only. Then, if we consider $\tilde{\theta}(t)$ to be a Gaussian random variable with zero mean and variance of $2\pi\beta t$ at t, we can evaluate the expected value of μ_c' as

$$E[\mu_c'] = E\left[\frac{1}{T}\int_0^T \cos\left(\tilde{\theta}(t)\right)dt\right] = \frac{1}{T}\int_0^T E\left[\cos\left(\tilde{\theta}(t)\right)\right]dt \tag{2.47}$$

where the inner expectation is given by

$$E\left[\cos\left(\tilde{\theta}(t)\right)\right] = \int_{-\infty}^{\infty} \cos\left(\tilde{\theta}(\tau)\right)\frac{1}{\sqrt{(2\pi)^2\beta t}}e^{-\frac{\tilde{\theta}(\tau)}{4\pi\beta t}}d\left(\tilde{\theta}(\tau)\right) \tag{2.48}$$
$$= e^{-\pi\beta t}$$

Thus,

$$E[\mu_c'] = \frac{1}{T}\int_0^T e^{-\pi\beta t}dt = \frac{1}{\pi\beta T}\left(1 - e^{-\pi\beta T}\right) = \frac{2}{\gamma}\left(1 - e^{-\frac{\gamma}{2}}\right) \tag{2.49}$$

Note that in the limit of small T, $E[\mu_c'] \approx 1$, which is as we would expect. As T→∞, $E[\mu_c'] \to \frac{1}{\pi\beta T} = \frac{2}{\gamma}$ which is not as we might expect. In an analogous manner, the expected value of μ_s' is given by

$$E[\mu_s'] = E\left[\frac{1}{T}\int_0^T \sin\left(\tilde{\theta}(t)\right)dt\right] = \frac{1}{T}\int_0^T E\left[\sin\left(\tilde{\theta}(t)\right)\right]dt = 0 \tag{2.50}$$

since $E[\sin(X)] = 0$ for any random variable X whose probability density function is symmetric about X=0.

The second moments (thus the variances) are quite a bit more involved. In detail:

$$\begin{aligned} E\left[\mu_s'^2\right] &= E\left[\frac{1}{T^2}\int_0^T \sin\left(\tilde{\theta}(t)\right)dt \int_0^T \sin\left(\tilde{\theta}(s)\right)ds\right] \\ &= \frac{1}{T^2}\int_0^T \int_0^T E\left[\sin\left(\tilde{\theta}(t)\right)\sin\left(\tilde{\theta}(s)\right)\right]dtds \end{aligned} \tag{2.51}$$

In a matter completely analogous to (2.22)-(2.24), we see that

$$\begin{aligned} E\left[\mu_s'^2\right] &= \frac{1}{T^2}\int_0^T \int_0^t e^{-2\pi^2 N_1(t-s)}\left(1 - e^{-2(2\pi)^2 N_1 s}\right)dtds \\ &= \frac{1}{48\pi^4 N_1^2 T^2}\left[24\pi^2 N_1 T - e^{-2(2\pi)^2 N_1 T} + 16e^{-2\pi^2 N_1 T} - 15\right] \\ &= \frac{1}{3\gamma^2}\left[6\gamma - 15 - e^{-2\gamma} + 16e^{-\frac{\gamma}{2}}\right] \end{aligned} \tag{2.52}$$

Likewise,

$$\begin{aligned} E\left[\mu_c'^2\right] &= \frac{1}{48\pi^4 N_1^2 T^2}\left[24\pi^2 N_1 T + e^{-2(2\pi)^2 N_1 T} + 8e^{-2\pi^2 N_1 T} - 9\right] \\ &= \frac{1}{3\gamma^2}\left[6\gamma - 9 + e^{-2\gamma} + 8e^{-\frac{\gamma}{2}}\right] \end{aligned} \tag{2.53}$$

Exploring the limits on γ, as γ gets very small

$$E\left[\mu_s'^2\right] \approx \frac{\gamma}{3} \quad \text{for } \gamma \text{ small} \tag{2.54}$$

And since the expected value of μ_s' is zero, this is also the variance of μ_s'. This serves as a sanity check, since as the phase noise disappears, so does Var[μ_s']. On the other hand,

in the limit of large γ, $E\left[\mu_s'^2\right] \approx \frac{2}{\gamma}$, which goes to zero with increasing γ. Before considering this rather odd result, we shall first consider the same limits for μ_c'. As γ gets very small

$$E\left[\mu_c'^2\right] \approx 1 \quad \text{and} \quad Var\left[\mu_c'\right] \approx 0 \quad \text{for } \gamma \text{ small} \tag{2.55}$$

as we would expect. In the limit of very large γ, $E\left[\mu_c'^2\right]$ and $Var\left[\mu_c'\right] \approx \frac{2}{\gamma}$ which is also expected, since for large phase noise the differences between the cosine and sine in definitions of μ_c' and μ_s' dissipate. Now, we've shown that the variances for μ_c' and μ_s' approach $\frac{2}{\gamma}$ in the limit of increasing γ, which at first glance seems odd. Upon more careful inspection, however, we notice that the mean of μ_c' also approaches $\frac{2}{\gamma}$ in the limit of increasing γ (the mean of μ_s' is zero, regardless of γ). Thus, for large γ the signal point locations are very close to origin and with non-negligible variances this means that the decision device basically has to guess, i.e., $P_{err,bit}=½$.

At this point one should notice that the means and variances of μ_c' and μ_s' do not depend on f_o, unlike similar expressions for ε and ρ in Section 2.2. Thus, we will *not* be able to ignore the affects of these random variables on our problem as we could in the last section. Since an explicit formulation of the probability density functions (including the joint density function) still seems difficult, a MATLAB* simulation was constructed to estimate the marginal density functions for μ_c' and μ_s' in order to get a better idea of their behavior for various γ regimes. Figures 2.12-2.17 show MATLAB-generated histograms estimating $f(\mu_c')$ and $f(\mu_s')$ for small, moderate, and large values of γ. In all cases, the plots were generated using a Monte-Carlo algorithm of 5000 runs each.

For large γ, both $f(\mu_c')$ and $f(\mu_s')$ approach Gaussian densities, although they will never be exactly Gaussian, even for very large γ, since μ_c' and μ_s' are limited to be between -1 and +1 (they are "truncated" Gaussians). In addition, we should again emphasize that the variances for μ_c' and μ_s' actually decrease in the limit of large γ, thus

* MATLAB is a digital signal processing software package.

it is entirely possible for $f(\mu_s')$ to be "narrower" for very large γ than for small γ. Thus, the comparison between Figure 2.12 and Figure 2.14 is entirely accurate. Later, we will show that μ_c' and μ_s' are uncorrelated, thus in the large γ regime, they are approximately independent since they are approximately Gaussian. For small γ, we see that μ_s' is also approximately Gaussian. In this same regime, μ_c' is seen to have a density function that is highly concentrated near +1, a fact we will later use to lower bound the error probability of our receiver. For moderate γ, the densities are quite different from the other two extremes. The density of μ_s' seems to be approximately uniform over a given range, while the density of μ_c' is something altogether different.

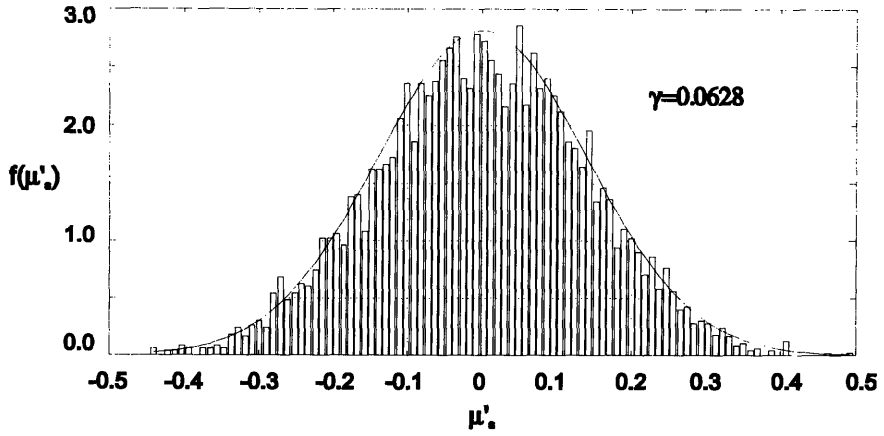

Figure 2.12: Estimated probability density function for μ_s' (small γ)

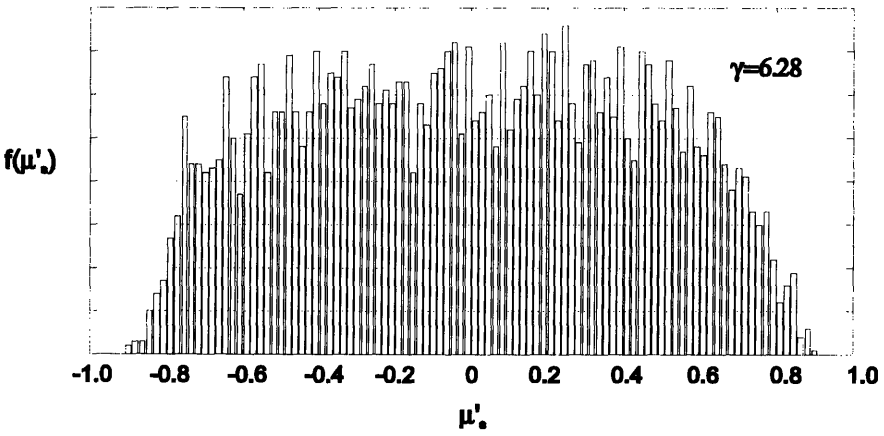

Figure 2.13: Estimated probability density function for μ_s' (moderate γ)

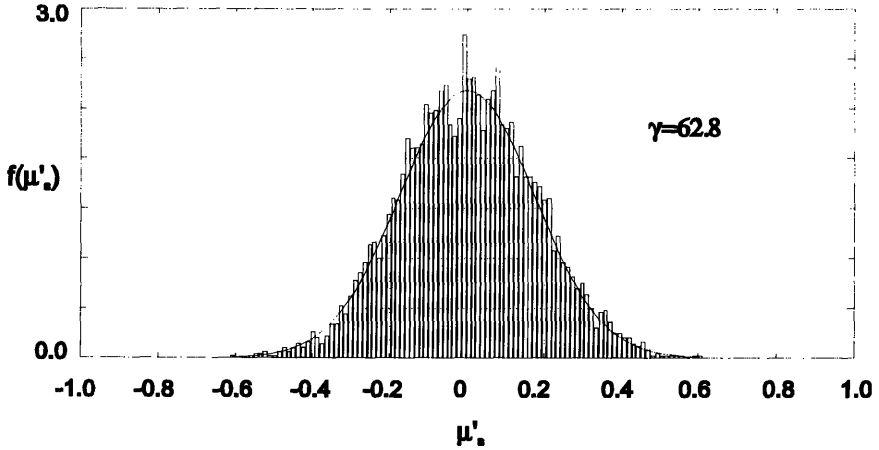

Figure 2.14: Estimated probability density function for μ_s' (large γ)

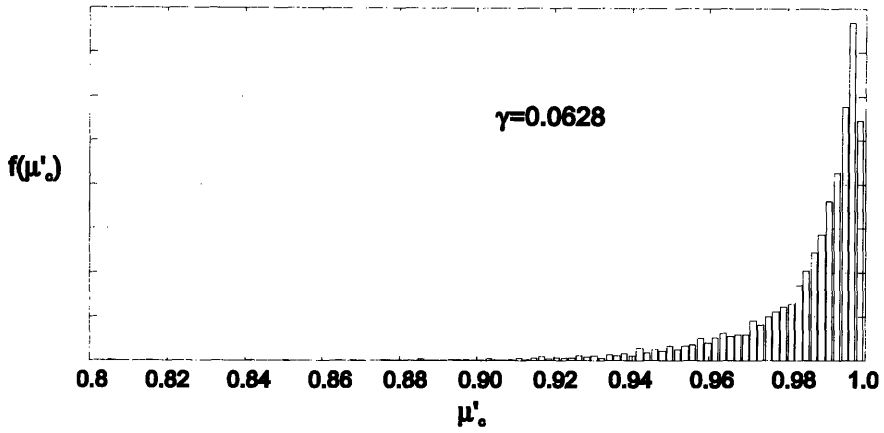

Figure 2.15: Estimated probability density function for μ_c' (small γ)

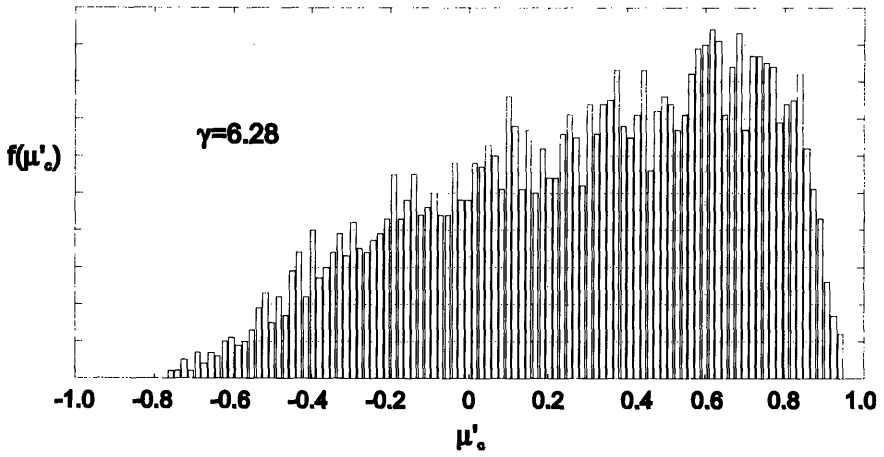

Figure 2.16: Estimated probability density function for μ_c' (moderate γ)

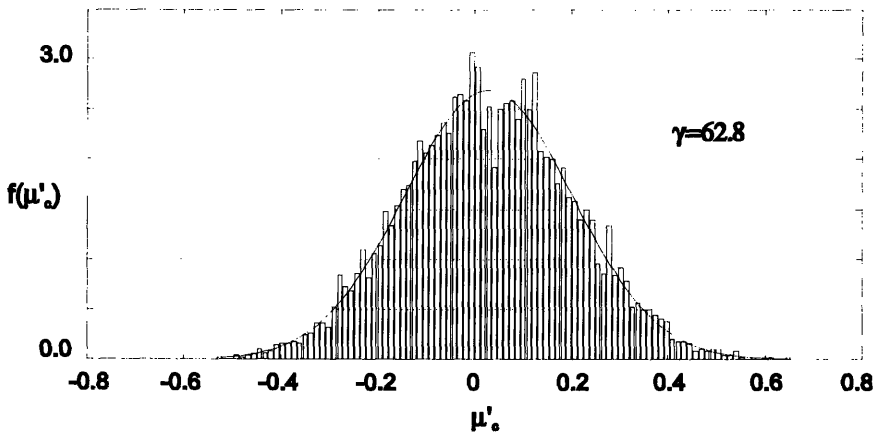

Figure 2.17: Estimated probability density function for μ_c' (large γ)

In addition to the qualitative "feel" given by the MATLAB-generated histograms, the first and second order statistics of μ_c' and μ_s' as functions of γ, are graphed in Figure 2.18 and Figure 2.19.

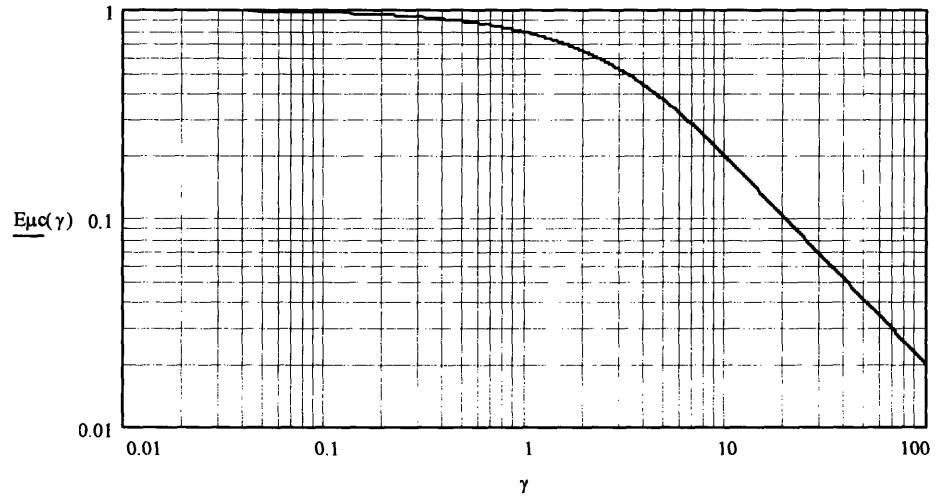

Figure 2.18: $E[\mu_c']$ as a function of the "phase noise SNR," γ

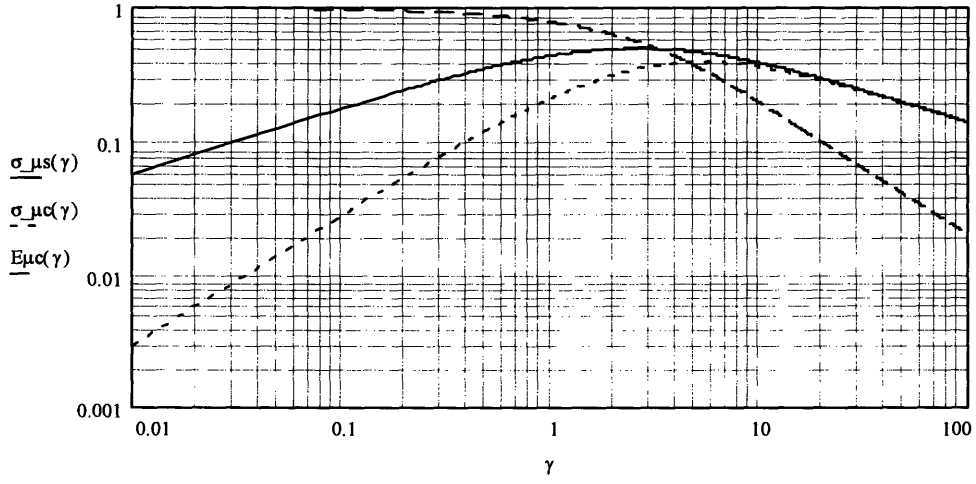

Figure 2.19: Standard deviation of μ_c' and μ_s', along with $E[\mu_c']$

Another important quantity to consider is the standard deviation of μ_c' and μ_s' normalized by the mean of μ_c', graphed in Figure 2.20. Note that as γ increases, the ratio approaches, then exceeds 1. This is the quantitative reason why the variances decreasing

for increasing γ does not bother us (i.e., the probability of error goes to ½ for large phase noise, as expected). Thus in this region, as we will see explicitly shortly, it will be virtually impossible to communicate reliably.

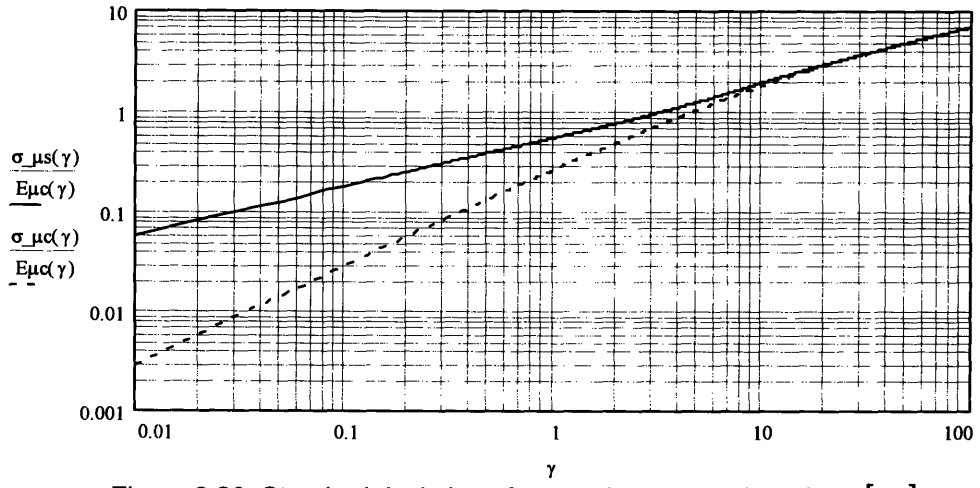

Figure 2.20: Standard deviation of μ_c' and μ_s' normalized by $E[\mu_c']$

Up until this point we've paid extraordinary attention to the random variables μ_c' and μ_s', without paying any attention to the signal locations r_c and r_s expressed in (2.46). However, since r_c and r_s depend linearly on μ_c' and μ_s', we've done this indirectly. The expectations for r_c and r_s (conditional on $\angle m_i^{(q)}$) are

$$\begin{aligned} E\left[r_c \mid \angle m_i^{(q)}\right] &= \frac{AT}{2}\cos\left(\angle m_i^{(q)}\right)E[\mu_c'] \\ E\left[r_s \mid \angle m_i^{(q)}\right] &= \frac{AT}{2}\sin\left(\angle m_i^{(q)}\right)E[\mu_c'] \end{aligned} \tag{2.56}$$

The second moments (once again, conditional on $\angle m_i^{(q)}$) are

$$\begin{aligned} E\left[r_c^2 \mid \angle m_i^{(q)}\right] &= \frac{(AT)^2}{4}\left[\cos\left(\angle m_i^{(q)}\right)\right]^2 E\left[\mu_c'^2\right] + \frac{(AT)^2}{4}\left[\sin\left(\angle m_i^{(q)}\right)\right]^2 E\left[\mu_s'^2\right] + E\left[n_c^2\right] \\ &\quad - \frac{(AT)^2}{2}\cos\left(\angle m_i^{(q)}\right)\sin\left(\angle m_i^{(q)}\right)E[\mu_c'\mu_s'] + AT\cos\left(\angle m_i^{(q)}\right)E[\mu_c']E[n_c] \\ &\quad - AT\sin\left(\angle m_i^{(q)}\right)E[\mu_s']E[n_c] \\ &= \frac{(AT)^2}{4}\left[\cos\left(\angle m_i^{(q)}\right)\right]^2 E\left[\mu_c'^2\right] + \frac{(AT)^2}{4}\left[\sin\left(\angle m_i^{(q)}\right)\right]^2 E\left[\mu_s'^2\right] + E\left[n_c^2\right] \end{aligned} \tag{2.57}$$

since $E[\mu_c'\mu_s']$ is given by

$$E[\mu_c'\mu_s'] = \frac{1}{T^2}\int_0^T\int_0^T E\left[\cos\left(\tilde{\theta}(t)\right)\sin\left(\tilde{\theta}(t)\right)\right]dtds$$
$$= \frac{1}{2T^2}\int_0^T\int_0^T E\left[\sin\left(\tilde{\theta}(t)+\tilde{\theta}(s)\right)\right] - E\left[\sin\left(\tilde{\theta}(t)-\tilde{\theta}(s)\right)\right]dtds \quad (2.58)$$
$$= 0$$

That is, the two random variables, μ_c' and μ_s' are *uncorrelated* since $E[\mu_s'] = 0$.

Likewise, $E\left[r_s^2 | \angle m_i^{(q)}\right]$ is given similarly to (2.57) as

$$E\left[r_s^2 | \angle m_i^{(q)}\right] = \frac{(AT)^2}{4}\left[\cos\left(\angle m_i^{(q)}\right)\right]^2 E\left[\mu_s'^2\right] + \frac{(AT)^2}{4}\left[\sin\left(\angle m_i^{(q)}\right)\right]^2 E\left[\mu_c'^2\right] + E\left[n_s^2\right] \quad (2.59)$$

Thus the variances are

$$Var\left[r_c | \angle m_i^{(q)} = 0, \pi\right] = Var\left[r_s | \angle m_i^{(q)} = \pm\tfrac{\pi}{2}\right] = \frac{A^2T^2}{4}Var[\mu_c'] + \frac{N_oT}{4}$$
$$Var\left[r_c | \angle m_i^{(q)} = \pm\tfrac{\pi}{2}\right] = Var\left[r_s | \angle m_i^{(q)} = 0, \pi\right] = \frac{A^2T^2}{4}Var[\mu_s'] + \frac{N_oT}{4} \quad (2.60)$$

Qualitatively, for small phase noise, the signal point projections r_c and r_s are close to the their nominal (no phase noise) locations. Their sample values are in a region relatively close to the nominal points, however, always of less energy (closer to the origin). As the phase noise increases from zero, the expected position of the signal points move toward the origin according to (2.56) and the variances as described in (2.60) , (2.54), and (2.55) increase. The nature of the projection, however, maintains the bi-orthogonality of the signal set; that is, for any given phase noise sample path and no AWGN, the two pairs of signal points would lie on orthogonal axes (although without *complete* knowledge of the phase noise the receiver does not know the exact orientation of those axes). Clearly since the means of r_c and r_s are decreasing monotonically with the phase noise strength (γ) and the variances are increasing (for $\gamma<10$ anyway), the performance of the system should become progressively worse with higher γ. We showed earlier through MATLAB simulation results that the cumulative density functions of μ_c' and μ_s' approach "truncated" Gaussian densities as γ increases, and since μ_c' and μ_s' are uncorrelated they become approximately independent for large γ. In addition, from Figure 2.19 it is clear that for $\gamma>10$ the variances of μ_c' and μ_s' become the same, thus the probability of error

in this case can be found directly (although we also know from the same figure that it's going to be very poor). Using the same derivation as was used for (2.11) from Section 2.2,

$$P_{err,sym} = \left[1 - Q\left(\sqrt{\frac{\left(\frac{A}{2}E\left[\mu_c|\angle m_i^{(q)}\right]\right)^2}{\left(\frac{A}{2}\right)^2 Var\left[\mu_c|\angle m_i^{(q)}\right] + \frac{N_o T}{2}}}\right)\right]^2$$
$$= \left[1 - Q\left(\sqrt{\frac{\left(E\left[\mu_c'|\angle m_i^{(q)}\right]\right)^2}{Var\left[\mu_c'|\angle m_i^{(q)}\right] + \frac{N_o}{E_s}}}\right)\right]^2 \qquad (2.61)$$

Now, the above equation was derived for the region of $\gamma>10$, thus using Figure 2.19 we can approximate (2.61). The second term in the denominator is a signal-to-noise ratio type term; even if $E_s/N_0 \to \infty$ the probability of error would be *at least* $\left[1 - Q\left(\frac{1}{4}\right)\right]^2 = 0.36$, and more for larger γ (using Figure2.20). As we expected this is *very* poor, thus in the next section we will consider more reasonable values of γ.

2.3.1.1 Error Probabilities when Phase Known only at Sampling Instants

Given the difficulties in finding the probability density functions for μ_c' and μ_s', we are restricted in our ability to find exact expressions for the probability of symbol error. However, as with the example of very large phase noise just shown, we can also bound the probability of error for small amounts of phase noise. For $\gamma<0.1$ for example, it is clear from our MATLAB simulations that μ_s' is approximately a Gaussian random variable, zero-mean, with the variance given in equation (2.52). Note that for $\gamma<0.1$, the standard deviation is less than 0.2, and thus the approximation with a Gaussian density is quite good since the amount of probability in the tails is very small. In addition, we can note that in this case, the probability density function for μ_c' is heavily concentrated near one, and although we do not propose any particular approximation for the probability

density function, we can calculate the probability of error given μ_c', and then average over all possible μ_c'. To do this, we need to make the approximation that μ_c' and μ_s' are independent. This seems reasonable since for small γ the variance of μ_c' is very small, that is, the error in our approximation will be small since μ_c' is with high probability very close to one. In addition, because of the symmetry of the problem, we can assume without loss of generality that the $\angle m_i^{(q)} = 0$ symbol is sent. Specifically, we can write

$$P_{err,sym} = \Pr(|r_s| > r_c) = 2\Pr(r_s > r_c) \tag{2.62}$$

where the second equality is possible because of the symmetry of the probability density function of r_s about zero (and also from the fact that we've assumed $r_c>0$). Writing in terms of μ_c', μ_s', n_c, and n_s, this becomes

$$\begin{aligned} P_{err,sym} &= 2\Pr\left(\frac{AT}{2}\mu_s' + n_s > \frac{AT}{2}\mu_c' + n_c\right) \\ &= 2\Pr\left(\frac{AT}{2}\mu_s' > \frac{AT}{2}\mu_c' + n_c - n_s\right) \\ &= 2\Pr\left(\frac{AT}{2}\mu_s' > \frac{AT}{2}\mu_c' + n_1\right) \end{aligned} \tag{2.63}$$

where $n_1 \sim N\left(0, \frac{N_o T}{2}\right)$ and is independent from μ_c' and μ_s'. Using our assumption of μ_s' being approximately Gaussian, given a particular value of μ_c',

$$\begin{aligned} P_{err,sym}(\mu_c') &= 2\Pr\left(\frac{AT}{2}\mu_s' - n_1 > \frac{AT}{2}\mu_c'\right) \\ &= 2Q\left(\frac{\frac{AT}{2}\mu_c'}{\sqrt{\left(\frac{AT}{2}\right)^2 E\left[\mu_s'^2\right] + \frac{N_o T}{2}}}\right) = 2Q\left(\sqrt{\frac{\mu_c'^2}{E\left[\mu_s'^2\right] + \frac{2N_o}{A^2 T}}}\right) \\ &= 2Q\left(\sqrt{\frac{\mu_c'^2}{E\left[\mu_s'^2\right] + \frac{N_o}{E_s}}}\right) \end{aligned} \tag{2.64}$$

where we've defined E_s in the usual way, i.e., $E_s = \frac{A^2 T}{2}$. Noting that the Q(x) function is convex for positive arguments, i.e., the second derivative is positive for all x>0, we can use Jensen's inequality [18], to lower bound the probability of error. Jensen's inequality says that if f(x) is a convex function, and X is a random variable with an expectation then

$$E[f(X)] \geq f(E[X]) \tag{2.65}$$

Since the probability of error in (2.64) is a Q-function, and we've already assumed that $\mu_c' > 0$, then the probability of error is a convex function of μ_c' and thus

$$P_{err,sym} = E_{\mu_c'}\left[P_{err,sym}(\mu_c')\right] \geq 2Q\left(\sqrt{\frac{E[\mu_c']^2}{E\left[\mu_s'^2\right] + \frac{N_o}{E_s}}}\right) \tag{2.66}$$

which is valid for $\gamma<0.1$ as mentioned previously. Substituting in the expressions for $E[\mu_c']$ and $E\left[\mu_s'^2\right]$, (2.66) becomes

$$P_{err,sym} \geq 2Q\left(\sqrt{\frac{\left[\frac{2}{\gamma}\left(1-e^{-\frac{\gamma}{2}}\right)\right]^2}{\left[\frac{1}{3\gamma^2}\left(6\gamma-15-e^{-2\gamma}+16e^{-\frac{\gamma}{2}}\right)\right] + \frac{N_o}{E_s}}}\right) \tag{2.67}$$

Making the small γ assumption for the exponential terms, i.e., replacing them with $e^x \cong 1+x+\frac{x^2}{2}+\frac{x^3}{6}$, the expression above reduces to

$$P_{err,sym} \geq 2Q\left(\sqrt{\frac{6-3\gamma}{\frac{6N_o}{E_s}+2\gamma}}\right) \tag{2.68}$$

Note that for $\gamma \to 0$, the above expression reduces to that of the no phase noise case (equation (2.12)) as we would expect. Figure 2.21 shows the probability of symbol error as a function of $\frac{E_s}{N_o}$ for various values of γ.

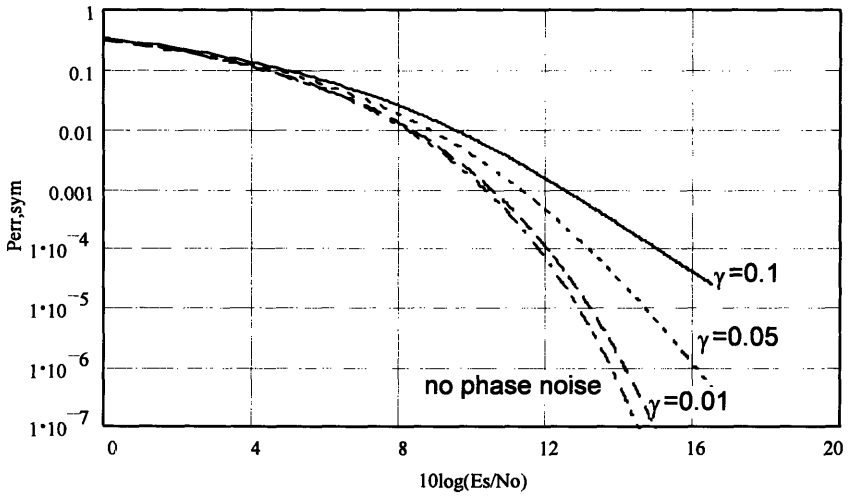

Figure 2.21: Lower bound on probability of symbol error for $\gamma<0.1$

As Figure 2.21 shows, even for relatively small amounts of phase noise, the probability of symbol error increases significantly. In addition, we should remark that these are lower bounds under relatively loose assumptions, thus the probability of error might be significantly higher. As an example, consider the $\gamma=0.1$ case. The "phase noise penalty," that is, the amount of increased signal energy needed for the same error performance as the AWGN-only case, is greater than 5.5 dB at a $P_{err,sym} = 10^{-5}$. Just to illustrate what kind of linewidths (β) our γ's imply, $\gamma=0.1$ corresponds to a linewidth of 16 Hz when the symbol rate is 1000 symbols/sec, 160 Hz when the symbol rate is 10,000 symbols/sec, etc. Thus, for high data rate communications, this means we can withstand a higher linewidth. Of course, implicit in this statement is a better estimation of our phase noisy carrier, since we've assumed perfect estimation of the phase at *one* time during the symbol (the beginning), thus with more symbols/sec, our estimation is better. In the next section we'll address this estimation issue, by considering performance when the receiver has multiple phase estimates per symbol.

2.3.2 Phase Known at N points per Symbol

In the last section we considered the performance of a receiver with perfect phase estimation at one point -- the beginning of each symbol. In this section we will consider how much performance can be improved if we have available perfect phase estimation at N equally spaced points within each symbol, i.e., $\theta(iT)$, $\theta(iT+T/N)$, ... $\theta((i+1)T-T/N)$ are known perfectly at the receiver. This is somewhat unrealistic in the sense of a buildable receiver, but the performance of this receiver provides a bound to the performance of any receiver that estimates the phase of the carrier at multiple points during each symbol. A continuous example of this type of carrier recovery circuit is a phase-locked loop (PLL), which provides phase estimates *continuously* through each symbol. At the other end of the spectrum is a decision-directed carrier recovery loop which is representative of the kind of estimation performance analyzed in the last section.

Our receiver structure is basically be the same as that in Figure 2.11, with the exception that we change the phase in the correlators N times per symbol, the pulse shape p(t) is a rectangular pulse of length T/N, and the sampling occurs N times per symbol. In addition, the processing in the decision box is different, since we now have 2N samples per symbol (as opposed to 2 samples per symbol as in Section 2.3.1). Without making any claims to optimality, we analyze perhaps the simplest decision device possible, namely one that simply adds the samples to form an "average" estimate of the signal point transmitted (this is quite similar to the "double filtering" described in [6]).

Assuming without loss of generality that $\angle m_i^{(q)} = 0$, the samples at the input of the decision device are

$$\begin{aligned} r_{c,i} &= \frac{A}{2}\mu_{c,i} + n_{c,i} \\ r_{s,i} &= \frac{A}{2}\mu_{s,i} + n_{s,i} \qquad \text{for i} = 1,2..\text{N} \end{aligned} \tag{2.69}$$

where $n_{c,i}$ and $n_{s,i} \sim N\left(0, \frac{N_o T}{4N}\right)$ and are independent. The random variables $\mu_{c,i}$ and $\mu_{s,i}$ are similarly defined by replacing T with T/N in (2.44). Note that some care must be taken in properly normalizing the variables $\mu_{c,i}$ and $\mu_{s,i}$ correctly, so we will deal with

the unnormalized variables here and take care of the normalization later. Thus, the mean of $\mu_{s,i}$ is still zero, and the mean of $\mu_{c,i}$ is given by

$$E[\mu_{c,i}] = \frac{1}{\pi\beta}\left(1 - e^{-\pi\beta\frac{T}{N}}\right) = \frac{1}{\pi\beta}\left(1 - e^{-\frac{\gamma}{2N}}\right) \tag{2.70}$$

Likewise, the variances are given by

$$Var[\mu_{s,i}] = E[\mu_{s,i}^2] = \frac{1}{12\pi^2\beta^2}\left[\frac{6\gamma}{N} - 15 - e^{-\frac{2\gamma}{N}} + 16e^{-\frac{\gamma}{2N}}\right] \tag{2.71}$$

and

$$Var[\mu_{c,i}] = \frac{1}{12\pi^2\beta^2}\left[\frac{6\gamma}{N} - 21 + e^{-\frac{2\gamma}{N}} - 12e^{-\frac{\gamma}{N}} + 32e^{-\frac{\gamma}{2N}}\right] \tag{2.72}$$

Denoting r_c and r_s to be the sum of $r_{c,i}$ and $r_{s,i}$ over each symbol respectively, we see that

$$\begin{aligned} E[r_c] &= \sum_{i=1}^{N} E[r_{c,i}] = \sum_{i=1}^{N} \frac{A}{2} E[\mu_{c,i}] \\ &= \frac{AN}{2}\frac{1}{\pi\beta}\left(1 - e^{-\frac{\gamma}{2N}}\right) \\ &= \frac{AT}{2}\left(\frac{2N}{\gamma}\right)\left(1 - e^{-\frac{\gamma}{2N}}\right) \end{aligned} \tag{2.73}$$

while $E[r_s] = 0$. The variances are given similarly,

$$\begin{aligned} Var[r_s] &= \sum_{i=1}^{N} Var[r_{s,i}] = \sum_{i=1}^{N} \left(\frac{A}{2}\right)^2 Var[\mu_{s,i}] + \frac{N_o T}{4N} \\ &= N\left(\frac{A}{2}\right)^2 \frac{1}{12\pi^2\beta^2}\left[\frac{6\gamma}{N} - 15 - e^{-\frac{2\gamma}{N}} + 16e^{-\frac{\gamma}{2N}}\right] + \frac{N_o T}{4} \\ &= \left(\frac{AT}{2}\right)^2 \left(\frac{N}{3\gamma^2}\right)\left[\frac{6\gamma}{N} - 15 - e^{-\frac{2\gamma}{N}} + 16e^{-\frac{\gamma}{2N}}\right] + \frac{N_o T}{4} \end{aligned} \tag{2.74}$$

and

$$\begin{aligned} Var[r_c] &= \sum_{i=1}^{N} Var[r_{c,i}] = \sum_{i=1}^{N} \left(\frac{A}{2}\right)^2 Var[\mu_{c,i}] + \frac{N_o T}{4N} \\ &= \left(\frac{AT}{2}\right)^2 \left(\frac{N}{3\gamma^2}\right)\left[\frac{6\gamma}{N} - 21 + e^{-\frac{2\gamma}{N}} - 12e^{-\frac{\gamma}{N}} + 32e^{-\frac{\gamma}{2N}}\right] + \frac{N_o T}{4} \end{aligned} \tag{2.75}$$

Now, we can make similar definitions for normalized random variables, $\mu'_{c,i}$ and $\mu'_{s,i}$, that is, normalizing $\mu_{c,i}$ and $\mu_{s,i}$ by T (note: *not* T/N), then

$$\begin{aligned} E\left[\mu'_{c,i}\right] &= \left(\frac{2}{\gamma}\right)\left(1-e^{-\frac{\gamma}{2N}}\right) \\ Var\left[\mu'_{c,i}\right] &= \left(\frac{1}{3\gamma^2}\right)\left[\frac{6\gamma}{N}-21+e^{-\frac{2\gamma}{N}}-12e^{-\frac{\gamma}{N}}+32e^{-\frac{\gamma}{2N}}\right] \\ Var\left[\mu'_{s,i}\right] &= \left(\frac{1}{3\gamma^2}\right)\left[\frac{6\gamma}{N}-15-e^{-\frac{2\gamma}{N}}+16e^{-\frac{\gamma}{2N}}\right] \end{aligned} \tag{2.76}$$

and we can rewrite equations (2.73)-(2.75) as

$$\begin{aligned} E[r_c] &= \frac{AT}{2}NE\left[\mu'_{c,i}\right] \\ Var[r_c] &= \left(\frac{AT}{2}\right)^2 NVar\left[\mu'_{c,i}\right]+\frac{N_oT}{4} \\ Var[r_s] &= \left(\frac{AT}{2}\right)^2 NVar\left[\mu'_{s,i}\right]+\frac{N_oT}{4} \end{aligned} \tag{2.77}$$

Now, since the variables $\mu'_{s,i}$ are independent (since $\theta(t)$ is an independent increments process) and can be approximated as Gaussian for small γ/N, then their sum is also Gaussian, with the variance given in (2.76). Under these assumptions, we can bound the probability error in a manner completely analogous to that in Section 2.3.1.1.

2.3.2.1 Error Probabilities When Phase Known at N points per Symbol

Under the assumption that we have phase estimates often enough so that the phase does not wander too much between any successive estimates (that is, $\gamma/N < 0.1$) then we can lower bound the error probability in the same manner as Section 2.3.1.1. Thus,

$$P_{err,sym} = \Pr\left(|r_s| > r_c\right) = 2\Pr\left(r_s > r_c\right) \tag{2.78}$$

where r_c and r_s are as defined as in section 2.3.2 as the sum of $r_{c,i}$'s and $r_{s,i}$'s, respectively. Thus, the probability of error, given $\sum_i \mu'_{c,i}$, is

$$P_{err,sym}\left(\sum_{i=1}^{N}\mu'_{c,i}\right)=2\Pr\left(\frac{AT}{2}\sum_{i=1}^{N}\mu'_{s,i}-n_1>\frac{AT}{2}\sum_{i=1}^{N}\mu'_{c,i}\right)$$

$$=2Q\left(\frac{\frac{AT}{2}\sum_{i=1}^{N}\mu'_{c,i}}{\sqrt{\left(\frac{AT}{2}\right)^2\sum_{i=1}^{N}Var\left[\mu'_{s,i}\right]+\frac{N_oT}{2}}}\right)=2Q\left(\sqrt{\frac{\left(\sum_{i=1}^{N}\mu'_{c,i}\right)^2}{NVar\left[\mu'_{s,i}\right]+\frac{N_o}{E_s}}}\right) \tag{2.79}$$

Then by the convexity of Q(x) for x>0 and Jensen's inequality, we can lower bound the error probability as

$$P_{err,sym}\geq 2Q\left(\sqrt{\frac{\left(NE\left[\mu'_{c,i}\right]\right)^2}{NVar\left[\mu'_{s,i}\right]+\frac{N_o}{E_s}}}\right)$$

$$=2Q\left(\sqrt{\frac{\left(\frac{2N}{\gamma}\right)^2\left(1-e^{-\frac{\gamma}{2N}}\right)^2}{\left(\frac{N}{3\gamma^2}\right)\left[\frac{6\gamma}{N}-15-e^{-\frac{2\gamma}{N}}+16e^{-\frac{\gamma}{2N}}\right]+\frac{N_o}{E_s}}}\right) \tag{2.80}$$

which by approximating the exponentials with $e^x\cong 1+x+\frac{x^2}{2}+\frac{x^3}{6}$, can be simplified to

$$P_{err,sym}\geq 2Q\left(\sqrt{\frac{\left(6-3\frac{\gamma}{N}\right)N^2}{6N^2\left(\frac{N_o}{E_s}\right)+2\gamma}}\right) \tag{2.81}$$

which, again, is valid for $\gamma/N<0.1$. If we keep γ/N fixed, and let $N\to\infty$, the lower bound on the probability of symbol error approaches

$$\lim_{\frac{\gamma}{N}\text{ fixed},\,N\to\infty}P_{err,sym}=2Q\left(\sqrt{\frac{2-\frac{\gamma}{N}}{2\left(\frac{N_o}{E_s}\right)}}\right) \tag{2.82}$$

which, perhaps not surprisingly, doesn't quite approach the phase noise free case (equation (2.12)), except approximately if $\gamma/N<<2$. However, since this limit was derived under the assumption that $\gamma/N<0.1$, it is very close. Figure 2.22 shows the dependence on N for fixed γ/N ratio. For higher values of N, the receiver takes advantage of the

increased averaging, thus decreasing the variability of the signal point locations due to phase noise.

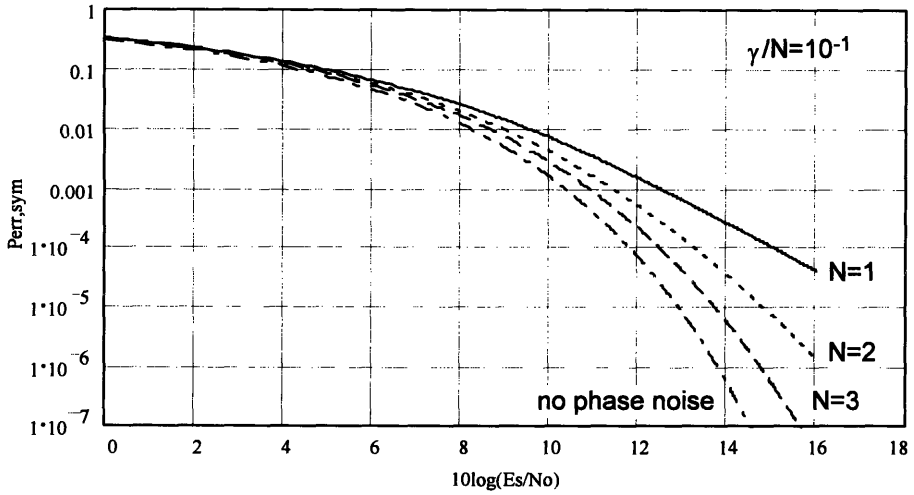

Figure 2.22: Probability of symbol error for fixed $\frac{\gamma}{N}$

Figure 2.23 shows the performance for fixed γ, also taking advantage of the averaging by having more estimates of the phase at points within the symbol.

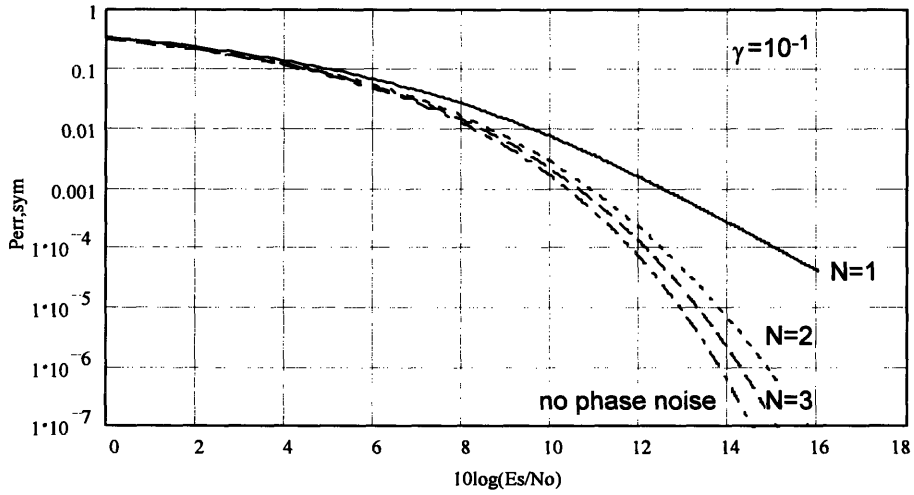

Figure 2.23: Probability of symbol error for fixed γ

At this point is should be clear that an improvement can be made to the receiver of Section 2.3.1 by using the knowledge of the phase noise at the beginning of the next symbol and correlating backward in time to the middle of the symbol. Thus, with a

change in processing (and perhaps necessitating off-line (non-real-time) processing) the performance of the receiver with only 1 sample of $\theta(t)$ per symbol can be made equivalent to that of the N=2 line shown in Figure 2.23. A similar improvement of the receiver in this section can be made if off-line processing is allowed.

Note that the error performance in Figure 2.23 gets better with larger N at a faster rate than in Figure 2.22, due to the "phase noise SNR" (γ) being constant rather than the γ/N ratio. This is due to the fact that holding γ constant as N increases means that the phase is allowed to wander less and less with increasing N, as opposed to when γ/N is constant. In this case as $N \rightarrow \infty$, the probability of symbol error approaches that of the phase noise free case.

2.4 Results and Discussion

In this chapter we have shown, through the representative QPSK case, that QAM-type signaling is not appreciably degraded by phase noise *if* the phase noise process is known perfectly at the receiver (at least in the radio case; a baseband system would behave differently). This result is true even though the presence of phase noise in the transmitter removes the orthogonality between the in-phase and quadrature carriers. On the other hand, we have also shown that limited knowledge of the phase noise process at the receiver can result in a fairly severe increase in the SNR required to maintain the same probability of error as in the AWGN-only case. This *phase noise penalty* was shown to be greater than 5 dB at a $P_{err,sym}$ of 10^{-5} for γ as small as 0.1, even when perfect knowledge of the phase noise process at the beginning of each symbol was known at the receiver.

These results indicate that if severe phase noise is unavoidable in the transmitter, accurate carrier recovery (i.e., good estimation of the phase noise process) is essential at the receiver. One way of accomplishing this would be to send along a reference signal (a pilot tone) with the standard QAM signal solely for the purpose of allowing better estimation of $\theta(t)$. This was proposed in [6] for optical communications systems and has been implemented in several radio systems, for example Qualcomm's IS-95 cellular radio

system [19]. The main disadvantage of such a scheme is the additional energy needed to transmit such a signal, and the fact that the reference signal is corrupted by AWGN, thus carrier recovery can not be ideal. In [6] it was shown that the increase in energy need not be 3 dB to add such a signal (i.e., the reference signal need not be as strong as the modulated signal), thus if the phase noise penalty is greater than 3 dB without the reference signal, this might be a reasonable alternative. In analytic terms, such a scheme has many similarities to a differential scheme, since the information is essentially contained in the difference between the reference signal and the modulated signal. In many cases, especially if transmitter power is at a premium, sending a reference signal might not be an alternative. In this case, the analysis in this section shows that accurate *intra-symbol* carrier recovery is necessary for good performance. Thus, decision-directed schemes that update the phase of the local oscillator in the receiver once per symbol would not be appropriate.

It should be noted that the estimation of symbol error probabilities in this chapter was only possible in several limited cases. In order to characterize the symbol error probability more fully, the joint and marginal probability density functions of ε, ρ, μ_c', and μ_s' would be needed. In particular the density functions of μ_c', and μ_s' for moderate values of γ (0.1 to 10) would be needed. In addition, if we were analyzing a system where f_o is *not* much greater than β, then the density functions for ε and ρ would also be needed. Since these variables involve integrals of non-linear functions of $\theta(t)$, this seems to be a formidable task. In fact, the closest analogy to show the difficulty in deriving theses distributions can be seen in [6], where an *approximation* to the density of $\sqrt{\mu_c^2 + \mu_s^2}$ was derived more than 20 pages.

As a general conclusion we should point out that the results in this chapter and the preceding discussion do not really apply to an information-theoretic point of view. The real problem in lowering the probability of symbol error is to accurately estimate the phase noise process, which one can look upon as additional "information" added to the signal at the transmitter. Thus, if the phase noise process is sufficiently noisy, the amount of information in the phase noise process may be greater than the capacity of the AWGN

channel. Thus, an information theorist would then conclude that accurate carrier recovery is beyond the ability of *any* receiver, and even with the most sophisticated codes it may not be possible to send the *intended* data at rates even *much* less than the capacity of the channel. This concept is beyond the scope of this chapter, but is pointed out to show the limitations of this type of analysis.

3. Detection of Pulse-Position Modulation Signals

As stated in the introduction the goal of this thesis is to consider the effects of a phase noisy oscillator on the low-power radio communications problem. In this chapter we consider the effect of such an oscillator on a communications system proposed by Pulson Communications [26]. The Pulson system approaches the problem of low-power differently than other modulation schemes (for example those in reference [17]). Instead of using a continuously transmitting modulation scheme, Pulson proposes a digital pulse-position modulation (PPM) system with pulse repetition intervals on the order of microseconds and pulse widths on the order of nanoseconds. Thus, low-power operation is attained in the average sense, since even with high peak power during each pulse transmission, the vast majority of time the transmitter is not transmitting a pulse. Modulation of this type leads to a system more akin to radar than a traditional communications system, and as such is not as widely studied in the communications literature.

Simple analysis [26] shows that the power spectrum of such a signal has a significant bandwidth on the order of hundreds to thousands of megahertz, thus is termed an ultrawideband system. In addition, since pulse-position modulation carries information in the time position of the pulse, timing accuracy is of prime importance in considering the performance of the system. Since timing in this type of system is likely derived from an phase noisy oscillator, there will be errors caused by the phase noise

process manifesting itself through inaccuracies in the transmitter's clock. It is this facet that will interest us.

Section 3.1 discusses in detail how a phase noisy reference oscillator can cause timing jitter in the system. Section 3.2 discusses optimal (in the sense of minimizing probability of error) receiver structures and their performance when only one sample is taken during for each possible pulse location. That is, a projection is made through linear filtering for each possible location and the decision device uses the resulting scalar projection variables to render a decision. Representative examples using both rectangular and sin(x)/x-type pulses are used as illustrations. For the rectangular pulse in particular, suboptimal receiver structures, suggested by the structure of the optimal receiver, are studied. Section 3.3 discusses optimal (in maximum likelihood fashion) receiver structures, when estimation of the pulse location is attempted. Detailed analysis is limited because of the difficulties in accurately characterizing the probability density function of the pulse location estimate. However, an approximation to the probability of error is derived under several basic assumptions.

3.1 Phase Noise Manifestation as Timing Jitter

Pulse-position modulation does not use a continuous carrier to transmit, hence it does not make sense to consider phase noise of the type considered in Chapters 1 and 2. Instead, phase noise affects the system through a reference oscillator which influences directly the timing accuracy of the system "clock." For simplicity we assume the clock is derived by counting the number of zero-crossings of a reference oscillator. This is a reasonable assumption since quartz oscillators at 5 MHz and 10 MHz are incredibly common and given than there are two zero crossings per cycle of a sinusoidal waveform, we have timing accurate (at a minimum) to the tenths of a microsecond (more sophisticated timing circuits based on this reference oscillator can attain sub-nanosecond accuracy).

Consider a sinusoidal waveform with phase noise, i.e.,

$$s(t) = \cos\left(2\pi f_T t + \theta(t)\right) \tag{3.1}$$

If the nominal (not accounting for the addition of data to the pulse location) interval between pulses is T_{PRI}, then a zero-crossing timer will cue the transmission of the next pulse when $M = {}^{T_{PRI}}\!/_{\frac{1}{2f_T}} = 2f_T T_{PRI}$ zero crossings are counted. In the phase noisy case the time of the pulse's transmission is determined by the sample path of the phase noise process since the last pulse. That is, the phase noise present in the reference signal s(t) will manifest itself in the modulation as a timing jitter of the pulse location. It is our goal in this section to quantify this timing jitter.

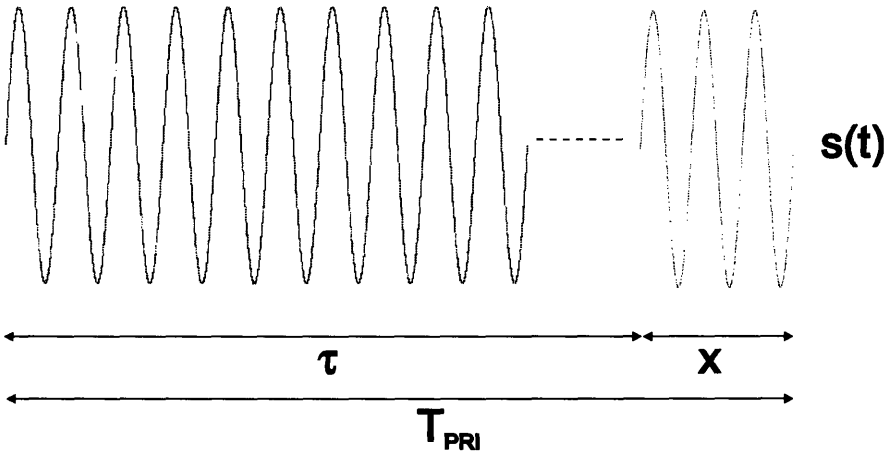

Figure 3.1: Illustration of timing jitter resulting from a phase noisy reference signal

Let τ be the actual time of the pulse transmission (as determined by M zero-crossings in the phase noisy s(t)), and let the variable x in Figure 3.1 be the timing jitter variable we wish to characterize ($T_{PRI} = \tau + x$). If we assume $\theta(t)$ is the same type of Brownian motion process considered in Chapters 1 and 2, then we can determine the probability density function $f_X(x)$ of x. We assume that the zero-crossings are actually crossings of a threshold above and below zero (to eliminate the many zero-crossings that might occur due to phase noise when the s(t) is near zero due to $\theta(t)$). In addition, we assume the phase noise process is sufficiently weak so that the waveform s(t) still roughly resembles a sinusoid, i.e., it's power spectrum is still fairly concentrated around f_T. For small Δx,

$$\Pr(x \le X < x + \Delta x) \cong f_X(x)\Delta x \tag{3.2}$$

A timing error of x seconds corresponds to a phase error of $+2\pi f_T x$ radians, hence the left side of the above equation can be expressed as

$$\Pr(x \le X < x+\Delta x) = \Pr(2\pi f_T x \le \theta(\tau) < 2\pi f_T (x+\Delta x)) \tag{3.3}$$

Thus,

$$\begin{aligned}\Pr(2\pi f_T x \le \theta(T_{PRI}+x) < 2\pi f_T x + (2\pi f_T \Delta x)) &\cong f_{\theta(T_{PRI}+x)}(2\pi f_T x)[2\pi f_T x] \\ &= \frac{f_T}{\sqrt{2\pi N_1 (T_{PRI}+x)}} e^{-\frac{f_T^2 x^2}{2N_1(T_{PRI}+x)}} \Delta x\end{aligned} \tag{3.4}$$

Since $T_{PRI} >> x$ for any reasonable value of N_1, in (3.4) we can approximate $T_{PRI}+x \approx T_{PRI}$. In other words, we expect no appreciable phase change between $T_{PRI}+x$ and T_{PRI}. Thus, x can be reasonably approximated as a Gaussian random variable with zero mean and variance $\frac{N_1 T_{PRI}}{f_T^2}$:

$$f_X(x) \cong \frac{1}{\sqrt{2\pi\left(N_1 T_{PRI}/f_T^2\right)}} e^{-\frac{x^2}{2\left(N_1 T_{PRI}/f_T^2\right)}} \tag{3.5}$$

Some of the typical values for the system we are considering are $T_{PRI} \cong 10^{-6}$ sec, $\beta = 2\pi N_1 \cong 1$-1000 Hz, and $f_T = 5$ MHz. Using these values the standard deviation of x is on the order of

$$8x10^{-11} \le \sigma_x \le 8x10^{-9} \tag{3.6}$$

Since our pulses are on the order of 1 nanosecond (10^{-9}) it is clear that timing jitter with these statistics will cause an increase in the probability of error above what one would expect in with AWGN-only.

Summarizing, the key result is that under some reasonable approximations the probability density function of the timing jitter, x, for a zero-crossing derived clock is Gaussian with zero-mean and variance $\frac{N_1 T_{PRI}}{f_T^2}$. As we would expect, the variance of the timing jitter increases with increasing linewidth ($\beta = 2\pi N_1$) and pulse repetition interval (T_{PRI}). In addition, the variance decreases with increasing f_T, since increasing f_T

improves the timing precision of the clock by allowing more zero-crossings within a given time period.

3.2 Performance of Single Sample Receivers

It is a common to assume in introductory analysis of digital communications systems that both carrier recovery and timing recovery are perfect (for example, in [18],[19],[20], and [22]). In the PPM system under consideration in this chapter there is no carrier in the normal sense, thus carrier recovery is not an issue. Timing recovery, on the other hand, is at the very heart of the system we are considering. If timing recovery is perfect, then the effects of jitter are completely removed from the communications system and the analysis reduces to the classic case. As background we will study such a system first, in Section 3.2.1, and note the resulting optimal receiver is a matched filter plus a sampler. Without perfect timing recovery, however, the optimal receiver is not as easy to derive. Under the restriction of allowing only one sample per possible pulse location (single-sample), we are constraining the subspaces onto which we are allowed to project the incoming signal. This is a logical extension, however, since the optimal receiver with perfect timing recovery only *needs* one sample per possible pulse location. Under the single-sample constraint we find that a "modified" matched filter, which incorporates the probability density function $f_X(x)$, followed by a single sample for each possible pulse location, is the optimal receiver.

3.2.1 Optimal PPM Receiver with no Timing Jitter

As a baseline from which to compare our later analysis, we present a short tutorial on binary PPM and the optimal receiver structure when AWGN is the only impairment. First note that with proper pulse shapes and locations for the two possible symbols, PPM is an orthogonal modulation scheme. For example, this can be accomplished with rectangular pulses such as in Figure 3.2.

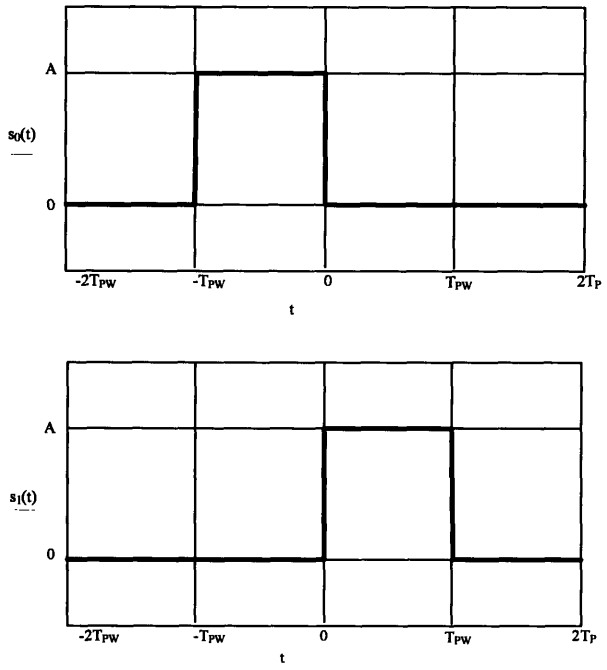

Figure 3.2: Possible pulses for binary PPM

Defining a pulse shape p(t) as

$$p(t) = \begin{cases} 1 & \text{for } 0 \le t < T_{PW} \\ 0 & \text{for all other t} \end{cases} \tag{3.7}$$

then it is clear that the pulses in Figure 3.2 can be written as

$$\begin{aligned} s_0(t) &= Ap(t + T_{PW}) \\ s_1(t) &= Ap(t) \end{aligned} \tag{3.8}$$

These pulses are clearly orthogonal since they do not overlap in time. One should also note that the pulses could be separated by a guard time (i.e., one pulse could extend from $-T_{PW}-\delta$ to $-\delta$ and the other from δ to $T_{PW}+\delta$) and they would still be orthogonal, and the performance, as we shall see shortly (in the absence of timing jitter), is the same. The optimal receiver for this signal set is well known (for example, see [28]) and amounts to filtering the received signal with a time-reversed rectangular pulse, h(t)=p(-t), and sampling at the appropriate times. This is illustrated in Figure 3.3.

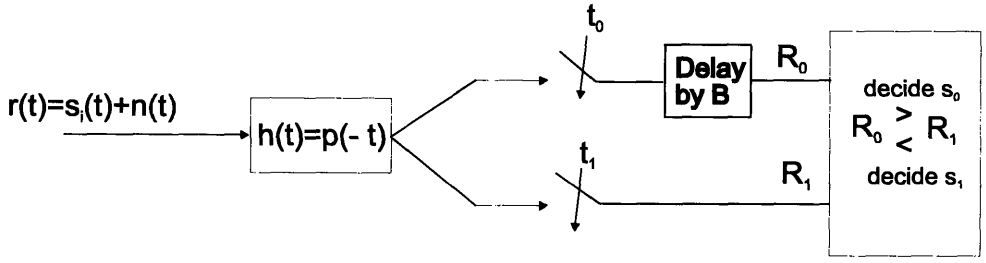

Figure 3.3: Matched filter receiver for binary PPM

The sampling times t_0 and t_1 are determined by the actual pulses $s_0(t)$ and $s_1(t)$. For the example of the pulses in Figure 3.2, $t_0=-T_{PW}$, $t_1=0$, and the delay variable B is T_{PW}. If there was a guard time in between the pulses, it would be reflected in the sampling times and the delay variable B. For M-ary PPM modulation in AWGN-only, the only modification necessary is to add more sampling time branches to Figure 3.3 and to pick the largest of the R_i in the decision device. This is, of course, for an uncoded system, since *hard* decisions are made as to what symbol was received. A *soft* decision receiver would instead pass all the R_i along to the decoder and a decision would be made at a (possible distant) future time as to the most likely symbol sent, given any coding added to the symbol stream. In the following discussion we will focus on hard decision receivers.

For binary PPM in AWGN-only, assuming equiprobable bits, the probability of (bit) error can be found exactly [18]:

$$\begin{aligned} P_{err} &= \Pr(R_0 > R_1 | s_1(t) \text{ sent}) \\ &= \Pr(n_0 > s_1 + n_1 | s_1(t) \text{ sent}) \\ &= \Pr(n > s_1 | s_1(t) \text{ sent}) \end{aligned} \tag{3.9}$$

where $n \sim N(0, N_o T_{PW})$ and $s_1 = AT_{PW}$. Thus,

$$P_{err} = Q\left(\frac{s_1}{\sqrt{N_o T_{PW}}}\right) = Q\left(\frac{AT_{PW}}{\sqrt{N_o T_{PW}}}\right) = Q\left(\sqrt{\frac{E_s}{N_o}}\right) \tag{3.10}$$

where $E_s = A^2 T_{PW}$. This result is valid for *any* binary PPM system, *with or without guard times* between the pulses in each symbol epoch, or even for pulses with guard times between symbol epochs. (By symbol epoch we mean the time period around each

nominal transmission time where pulses are allowed to be transmitted, and to some extent the additional time the receiver may choose to look over to accommodate timing jitter added to the pulse locations.)

In the case where there are M possible pulse locations per symbol epoch (M-ary PPM), we still have a M-ary orthogonal signal set. Typically the error probability in this case is not found exactly, but rather is upper bounded using the union bound. Using the union bound, the probability of symbol error is [18]

$$P_{err,sym} \leq (M-1)Q\left(\sqrt{\frac{E_s}{N_o}}\right) \tag{3.11}$$

Again, this probability of symbol error is unchanged by the insertion or lack of guard times between symbols within each symbol epoch, or between symbol epochs. In addition, although we used rectangular pulses for illustration, any orthogonal signal set would yield the same result, including sin(x)/x-type pulses (ideal bandlimited pulses) that are discussed in later sections.

3.2.2 Optimal Single-Sample PPM Receiver with Timing Jitter and AWGN

The addition of a second noise source, namely timing jitter, significantly changes the reception and signal set design problem. Whereas the previous section's receiver could be viewed as having perfect timing recovery (thus eliminating all timing jitter from the signal) the receiver in this section can be viewed as one with imperfect timing recovery. In addition, the previous section's analysis noted that the presence or lack of guard times did not change the probability of error. With the inclusion of timing jitter and imperfect timing recovery information, however, the insertion of guard times between pulses in the signal set changes the probability of error.

The receiver described in this section is the logical extension of the one described in Figure 3.3, i.e., the optimal receiver in the presence of AWGN-only. Namely, we take the incoming signal, project it onto one of M possible signal directions and declare the direction with the largest projection as the one corresponding to the most likely symbol

sent by the transmitter. Thus, we will derive the optimal linear filter for such a receiver when timing jitter and AWGN are present, such that only M samples (thus the term "single-sample," since only one sample is taken for each possible pulse (symbol) location) of the output of the filter are used in the decision device.

Foreshadowing the results that follow, it would seem intuitive that for a rectangular pulse p(t), an improvement in the probability of error could be made by widening h(t) in Figure 3.3, which is equivalent to integrating the incoming signal over a longer time. Then, small amounts of timing jitter would not cause a decrease in the signal component (s_1) at the sampler, while the resulting increase in the AWGN components (n_0 and n_1) might still be acceptable. Clearly this approach is ad hoc, but we will see that under certain the circumstances the optimal receiver has a very similar structure. This suboptimal receiver structure will be studied in more detail in Section 3.2.3.

The derivation of the optimal linear filter with single-sampling in the presence of timing jitter is similar to pulse amplitude estimation in the presence of random time delay in the channel, which is a problem that has been examined in [28]. In [28] it is shown that minimizing the mean-square error in the estimation of the pulse amplitude can be achieved via a "matched filter," where the matched filter (h(t)) is the time-reversed pulse shape appropriately modified by the probability density function of the timing jitter.

To modify the analysis in [28] to meet our needs, we only need only change a few details. Namely, we recognize that a PPM signal set is basically equivalent to a degenerate sequence of pulse amplitude modulation (PAM) symbols. That is, for each possible pulse location, the pulse is either there or not, thus detection of that pulse corresponds to the detection of the degenerate PAM case of on-off keying (OOK). Thus, the filter we derive minimizes the mean-squared error of the estimation of the pulse amplitude, and hence whether the pulse is present at each possible pulse location. A block diagram of the receiver is shown in Figure 3.4.

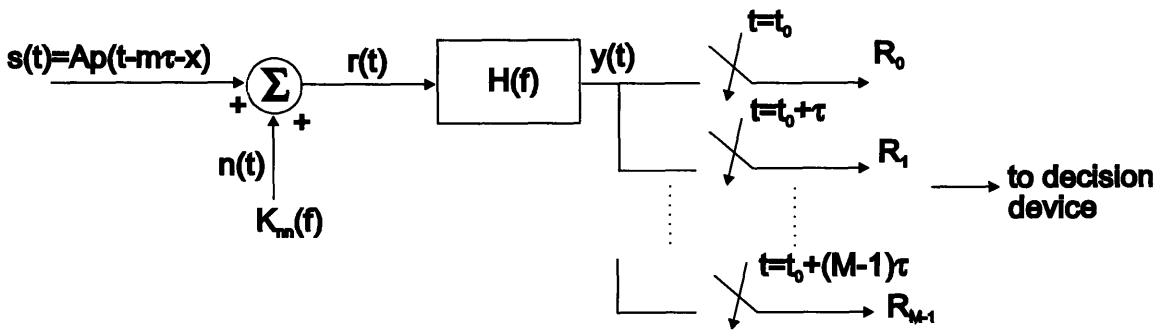

Figure 3.4: Optimal single-sample matched filtering in timing jitter and AWGN

Summarizing the derivation of the optimal filter h(t), the signal r(t) is projected on the signal directions $\{h(t_0-k\tau-t)$ for $0\le k\le M-1\}$ where k represents the possible pulse location index, and $t_0-k\tau$ represents the sampling time. The variables after sampling, i.e., $\{R_k\}$, are the estimates of the pulse amplitude during each possible pulse location. If h(t) is the optimal linear filter then

$$(R_k - a) \perp r(t) \quad \text{for all t, k} \tag{3.12}$$

that is, this represents the minimum mean-squared error criterion. The random variable a represents the true amplitude of the pulse in location k, and is equal to A when pulse location k is sent and zero when any other location is sent. Thus, assuming equiprobable inputs, $\Pr(a = A) = 1/M$ and $\Pr(a = 0) = (M-1)/M$. Without loss of generality we will focus on only the top branch of the receiver in Figure 3.4 (further assuming $t_0=0$). Thus, the orthogonality of r(t) to the error (R_k-a) can be expressed as

$$E\left[r(t)\left(a - \int r(\tau)h(\tau)d\tau\right)\right] = 0 \quad \text{for all t} \tag{3.13}$$

where the expectation is over a and the jitter variable x. Equation (3.13) can be rewritten as

$$\begin{aligned} E[r(t)a] &= E\left[\int r(t)r(\tau)h(\tau)d\tau\right] \\ &= \int E[r(t)r(\tau)]h(\tau)d\tau \end{aligned} \tag{3.14}$$

The left side of (3.14) can be simplified,

$$E[r(t)a] = E\left[a^2 p(t-x)\right] + E[a]E[n(t)]$$
$$= E\left[a^2\right]\int p(t-x) f_X(x)dx \tag{3.15}$$

The expectation inside the integral in (3.14) can be rewritten as

$$E[r(t)r(\tau)] = E\left[a^2 p(t-x)p(\tau-x)\right] + k_{nn}(t-\tau)$$
$$= E\left[a^2\right]\int p(t-x)p(\tau-x) f_X(x)dx + k_{nn}(t-\tau) \tag{3.16}$$

where $k_{nn}(t-\tau)$ is the autocorrelation function of the additive noise, n(t) (which is assumed to be stationary and zero-mean). Substituting (3.16) and (3.15) into (3.14),

$$E\left[a^2\right]\int p(t-x) f_X(x)dx = \int E\left[a^2\right]\int p(t-x)p(\tau-x) f_X(x)h(\tau)dx + k_{nn}(t-\tau)\, h(\tau)d\tau$$
$$= E\left[a^2\right]\int\int p(t-x)p(\tau-x) f_X(x)h(\tau)dx d\tau + \int k_{nn}(t-\tau)h(\tau)d\tau \tag{3.17}$$

Rewriting (3.17) as series of convolutions,

$$(p * f_X)(t) = \left[\left((h * p_B)\cdot f_X\right) * p\right](t) + \frac{1}{E\left[a^2\right]}(k_{nn} * h)(t) \tag{3.18}$$

where $(x * y)(t)$ is the convolution of x and y, with the resulting signal a function of t. In addition, we've defined $p_B(t)=p(-t)$. Taking the Fourier transform of both sides of (3.18), we get

$$P(f)F_X(f) = \left(H(f)P^*(f) * F_X(f)\right)P(f) + \frac{K_{nn}(f)H(f)}{E\left[a^2\right]} \tag{3.19}$$

If we choose our pulse, p(t) to be an ideal bandlimited pulse, i.e.,

$$p(t) = \frac{\sin\left(\pi T_{PW}^{-1} t\right)}{\pi T_{PW}^{-1} t} \tag{3.20}$$

then $P(f) = T_{PW}$ for $|f| \le \frac{1}{2}T_{PW}^{-1}$ and is zero for all other f. From (3.19), this forces H(f) to also be bandlimited to frequencies below $\frac{1}{2}T_{PW}^{-1}$. If we further assume AWGN, $K_{nn}(f) = {}^{N_o}\!/_{2}$, then (3.19) simplifies to

$$F_X(f) = H(f) * \left(T_{PW}F_X(f) + \frac{\delta(f)N_o M}{2A^2 T_{PW}}\right) \quad \text{for } |f| \le \frac{1}{2T_{PW}} \tag{3.21}$$

where we have used the fact that $E[a^2] = A^2/M$. Even more simplification is possible if we assume that the "bandwidth" of $F_X(f)$ is much smaller than that of P(f), i.e., that the typical jitter is much more than T_{PW}. Then (3.21) is approximately valid for $|f| > \frac{1}{2} T_{PW}^{-1}$, and we can take the inverse Fourier transform of (3.21) to solve for h(t). That is, h(t) is given by

$$h(t) = \frac{1}{T_{PW}} \left(\frac{f_X(t)}{\alpha^{-1} + f_X(t)} \right) \tag{3.22}$$

where α is a signal-to-additive noise type term, and is defined as

$$\alpha = \frac{2A^2 T_{PW}^2}{MN_o} = \frac{2T_{PW}}{M} \frac{E_s}{N_o} \tag{3.23}$$

In Section 3.1 we showed that the probability density function of x, $f_X(t)$ is zero-mean Gaussian, with a variance of $\frac{N_1 T_{PRI}}{f_T^2}$. A quick analysis of (3.22) shows that for small α, h(t) will be approximately $\frac{\alpha}{T_{PW}} f_X(t)$ and that for large α, h(t) is approximately $1/T_{PW}$ over the range where $\alpha f_X(t) >> 1$. The impulse response h(t) is shown for various values of α in Figure 3.5.

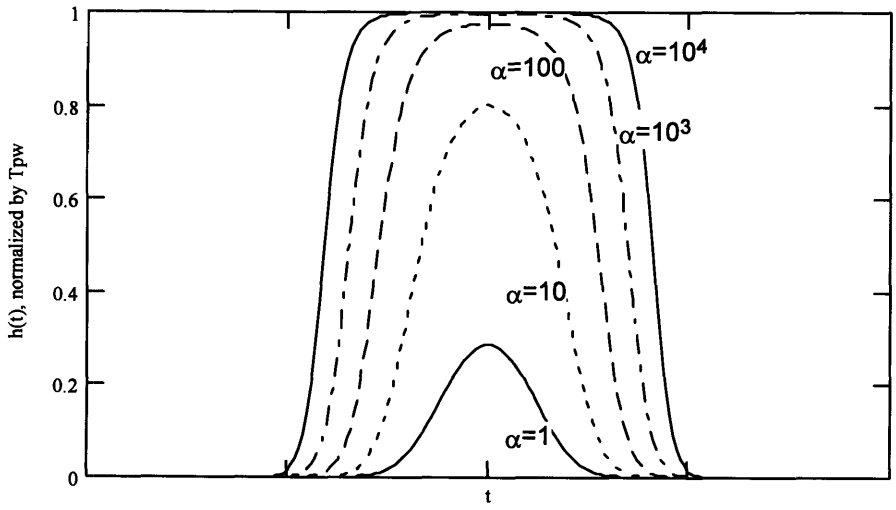

Figure 3.5: Optimal filter impulse response, h(t), for various values of α (sin(x)/x pulse shape and Gaussian timing jitter)

What this means is that for small signal-to-noise ratios (SNRs), the optimal receiver basically tries to "look" only where the pulse is most likely to be in order to decrease the amount of AWGN entering the decision device. Because of the low SNR, integrating over a large effective time, while increasing the ability to find a time jittered pulse, would et in too much AWGN to offset the gain in detection of the jitter.

For large signal-to-noise ratios, the optimal receiver basically filters with something that looks like a unit-amplitude rectangular pulse, where the width of the pulse increases with higher SNR. Since filtering with a rectangular pulse is equivalent to integration over the time-duration of the pulse, this says that the receiver basically tries to extract as much energy from the (jittered) pulse as it can. Since the SNR is already high by assumption, the lengthening of the integration does not substantially increase the AWGN components at the decision device. In the limit of infinitely high SNR, the ideal receiver would thus integrate over the entire time axis to estimate the pulse amplitude. In the PPM case, however, this lengthening of the integration time is *limited* since otherwise the receiver would wind up detecting other pulse locations and incorrectly outputting their estimated amplitude in the wrong pulse location bin. The effect of increasing the integration time is essentially what is discussed in the suboptimal receiver section, Section 3.2.3.

Calculating the probability of (symbol) error for this receiver is a formidable task. It should be clear from the preceding discussion that if the pulses are very close together, although orthogonal, filtering by h(t) might cause the noise components n_i at the decision device to become dependent. This is caused by the fact that the implementation in Figure 3.4 with one filter and a bank of samplers is equivalent to a bank of time delayed replicas of h(t) as filters followed by sampling all at the *same* time. Thus, unless the time-delayed replicas h(t) are orthogonal, the n_i will not be independent. In the preceding example, the n_i are never strictly independent, since the support of h(t), is in actuality, infinite. However, since the values of h(t) for t very far from its peak (the "middle") are very small, an approximation to the ideal filter h(t) could be made that allows the n_i to be independent. That is, if the variance of $f_X(t)$ is small enough, or in case of large α, if the

region where $f_X(t) >> \alpha^{-1}$ is small enough (this is the part of h(t) approaching one, i.e., the "flat-topped" part) we could assume the n_i are approximately independent.

Even with the approximation that the additive noise components are independent, however, the task of calculating the probability of symbol error is still difficult. This is evident from the fact that the probability density function for the signal component in the branch(s) containing the signal has not been found. In fact this probability density is rather hard to find, as is shown in [28], and in general only the variance of the signal component is available. Thus, any probability of error analysis relying solely on the variance would most likely be limited to use of the Chebyshev Inequality which is too weak to be meaningful. Therefore, we will concentrate our attention to evaluating the performance of the receiver suggested by the optimal h(t) found in the high SNR case, namely an integrator followed by a bank of samplers.

3.2.3 Suboptimal Receiver Performance

The results from the last section show that in the region of high E_s/N_o (SNR), the optimal single-sample receiver is essentially an integrator. Thus, in this section we will consider the performance of such an integrator receiver over all E_s/N_o ranges.

For our analysis in this section we will assume a rectangular pulse (p(t), as defined in (3.7)) for convenience, although the results should be similar for a sin(x)/x-type pulse. As noted before, in the presence of substantial timing jitter, putting the pulses "back-to-back" (as in Figure 3.2) makes them vulnerable to errors. Thus, with no other users sharing the medium, it would be advantageous to move the pulses apart by a distance such it would be unlikely that timing jitter would cause a "cross-over" error. In the multiple-user case, however, the solution is not so clear since separating the pulses makes less time available for other users to transmit, but this is beyond the scope of our analysis. We will proceed with systems based on the "back-to-back" signaling scheme, and will later consider what happens if the pulses are separated by a guard time.

The integrator receiver is essentially the same as that in Figure 3.3, however, we will show the structure from a different point-of-view to emphasize the integration aspect. We will first consider the performance of the receiver when the AWGN is negligible, then add in the effects of the AWGN later. This receiver will be composed of a series of "gates" matched to the signals $\{s_m(t)\}$ (=$Ap(t-mT_{PW})$). That is, each "gate" will integrate the incoming signal over non-overlapping time periods of T_{PW} and compare the resulting variables $\{R_i\}$, deciding that the branch with the largest R_i represents the symbol that was sent. If we denote the received signal with timing jitter as r(t), then the receiver is given as in Figure 3.6.

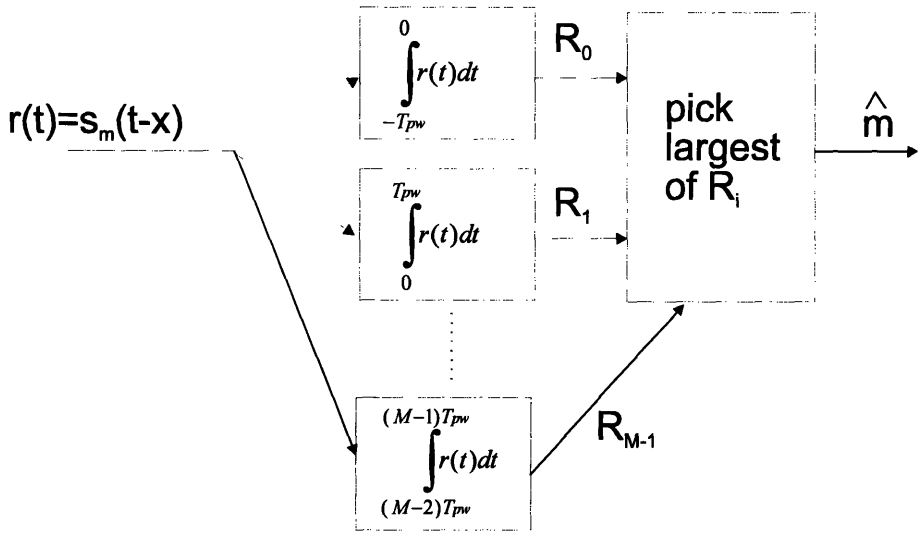

Figure 3.6: Integration receiver for M-ary PPM signal set

The probability of error for M=2 (binary case) is given as

$$P_{err} = \Pr(error|s_0(t) \text{ sent}) = \Pr\left(X \geq \frac{T_{PW}}{2}\right) + \frac{1}{2}\left[\Pr(X \leq -T_{PW}) - \Pr(X \geq 2T_{PW})\right] \qquad (3.24)$$

where the second terms are due to "guesses" made if the pulse does not land in either of the two gates. Using the fact that we know x is approximately Gaussian (zero-mean, variance of $\frac{N_1 T_{PRI}}{f_T^2}$) as given in Section 3.1, the probability error is

$$P_{err} = Q\left(\frac{T_{PW} f_T}{2\sqrt{N_1 T_{PRI}}}\right) + \frac{1}{2}\left[Q\left(\frac{T_{PW} f_T}{\sqrt{N_1 T_{PRI}}}\right) - Q\left(\frac{2T_{PW} f_T}{\sqrt{N_1 T_{PRI}}}\right)\right] \qquad (3.25)$$

For M>2 the probability of error is somewhat more complicated, taking into account the differences between pulses in the "middle" and pulses at the "edges" of the symbol epoch. In addition, the terms relating to "guesses" made if the pulse is jittered out of the symbol epoch entirely become quite complicated for large M.

Now, as mentioned previously, if we separate the pulses by a guard time we may improve probability of error performance at the expense of slightly altering the power spectrum of the signal. Consider the binary case as an example. By expanding the gate width to T_B, and centering the pulses $s_0(t)$ and $s_1(t)$ such that they are nominally in the middle of the gates (i.e., the center of the pulses at $\pm T_B/2$ instead of $\pm T_{PW}/2$) then the probability of error becomes

$$P_{err} = Q\left(\frac{T_B f_T}{2\sqrt{N_1 T_{PRI}}}\right) + \frac{1}{2}\left[Q\left(\frac{(T_B + T_{PW})f_T}{2\sqrt{N_1 T_{PRI}}}\right) - Q\left(\frac{(3T_B + T_{PW})f_T}{2\sqrt{N_1 T_{PRI}}}\right)\right] \tag{3.26}$$

which of course, reduces to the original formula if $T_B=T_{PW}$. If we fix the width of the time the receiver "looks" for pulses, i.e., the symbol epoch (the duration of which is $2T_B$) to be a constant, then we can arrange up to $\left\lfloor \frac{2T_B}{T_{PW}} \right\rfloor$ orthogonal pulses within that epoch (where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x). If $M < \frac{2T_B}{T_{PW}}$ then there are guard times between the pulses. Thus, the symbol bin boundaries are of width $\frac{2T_B}{M}$ and *ignoring edge effects* (that is, treating all pulses locations with the same statistics as the middle pulses), the probability of error is

$$P_{err,sym} \cong Q\left(\frac{T_B f_T}{M\sqrt{N_1 T_{PRI}}}\right) \tag{3.27}$$

If we do not ignore the edge effects, the probability of error expression can be quite complicated, since the probability of error is different for each pulse location. However, if we assume either M is large or if we assume the jitter variance is small enough that the probability of jitter moving the pulse more than one location is negligible then (3.27) is a good estimate of the probability of error.

3.2.3.1 Performance with Timing Jitter and AWGN

Although in the preceding section is useful, our real goal is to consider what happens when the received waveform has significant amounts of both timing jitter and AWGN. In this case, the received signal (around the nominal reception time) is

$$r(t) = s_i(t-x) + n(t) \tag{3.28}$$

where n(t) is AWGN and x represents the timing jitter. We will consider the same type receiver as before, i.e., as described in Figure 3.6. Thus, if we can calculate the probability density functions for R_i conditional on particular $s_i(t)$'s, then we can calculate the probability of error. As we will see, however, we do not actually need the densities for the R_i to calculate the probability of error. We see that

$$\begin{aligned} R_i|s_j &= \int_{(i-1)T_B/M}^{iT_B/M} s_j(t-x)dt + \int_{(i-1)T_B/M}^{iT_B/M} n(t)dt \\ &= S_i|s_j + N_i \end{aligned} \tag{3.29}$$

where N_i is a Gaussian random variable with mean zero and variance N_0T_B/M (for n(t) a Gaussian process of zero mean and variance $\frac{N_0}{2}$ for any time t), and $S_i|s_j$ is a as yet undefined random variable representing the signal component. Note that because of the receiver structure, the $\{N_i\}$ are independent Gaussian random variables. However, the $\{S_i|s_j\}$ are not independent as we will show. With equiprobable inputs, we can assume without loss of generality that $s_0(t)$ is sent, then, and assuming hard decisions the probability of error (for the M=2 case) is given as

$$\begin{aligned} P_{err} &= \Pr(R_1 > R_0|s_0(t) \text{ sent}) \\ &= \Pr(S_1 + N_1 > S_0 + N_0|s_0(t) \text{ sent}) \\ &= \Pr(S_1 - S_0 > N_0 - N_1|s_0(t) \text{ sent}) \\ &= \Pr(S > N|s_0(t) \text{ sent}) \end{aligned} \tag{3.30}$$

where $S=S_1-S_0$ and $N=N_0-N_1$. Since N_0 and N_1 are independent Gaussian random variables, N is a Gaussian random variable also, with zero mean and variance N_0T_B. The probability density function of S is not quite as simple to determine, in fact it is a mixed continuous/discrete density, i.e., there are points with a probability mass at S=0 and

$\pm AT_{PW}$. Again, we consider the set of signals where $s_0(t)$ and $s_1(t)$ are similar to those in Figure 3.2, but are spaced into the middle of the two symbol bins, that is, centered at $\pm T_B/2$. Considering the probability mass points separately, we see that the probability that S=0 (i.e., the pulse doesn't land in either gate) is given by

$$\begin{aligned}\Pr(S=0|s_0(t)\text{ sent}) &= \Pr\left(X \le -\tfrac{T_B+T_{PW}}{2}\right) + \Pr\left(X > \tfrac{3T_B+T_{PW}}{2}\right) \\ &= Q\left(\frac{(T_B+T_{PW})f_T}{2\sqrt{N_1 T_{PRI}}}\right) + Q\left(\frac{(3T_B+T_{PW})f_T}{2\sqrt{N_1 T_{PRI}}}\right)\end{aligned} \tag{3.31}$$

Similarly, the probability that $S=+AT_{PW}$ (which is the probability that the pulse lands entirely in the wrong gate) is given by

$$\begin{aligned}\Pr(S=AT_{PW}|s_0(t)\text{ sent}) &= \Pr\left(\tfrac{T_B+T_{PW}}{2} < X < \tfrac{3T_B+T_{PW}}{2}\right) \\ &= Q\left(\frac{(T_B+T_{PW})f_T}{2\sqrt{N_1 T_{PRI}}}\right) - Q\left(\frac{(3T_B+T_{PW})f_T}{2\sqrt{N_1 T_{PRI}}}\right)\end{aligned} \tag{3.32}$$

Lastly, the probability that $S=-AT_{PW}$ (which is the probability that the pulse lands entirely in the correct gate) is given by

$$\begin{aligned}\Pr(S=-AT_{PW}|s_0(t)\text{ sent}) &= \Pr\left(-\tfrac{T_B+T_{PW}}{2} < X < \tfrac{T_B+T_{PW}}{2}\right) \\ &= 1-2Q\left(\frac{(T_B+T_{PW})f_T}{2\sqrt{N_1 T_{PRI}}}\right)\end{aligned} \tag{3.33}$$

The more complicated section of the probability density function is the region where $|S|<AT_{PW}$, which we will consider in two parts. First, consider the section where S is negative and note that

$$\begin{aligned}\Pr\left(s < S < s+\Delta s|s_o(t)\text{ sent}\right) = \Pr&\left(s < -\tfrac{A}{2}(T_B+T_{PW}+2x) < s+\Delta s|s_o(t)\text{ sent}\right) \\ &+ \Pr\left(s < -A(T_B-2x) < s+\Delta s|s_o(t)\text{ sent}\right)\end{aligned} \tag{3.34}$$

and noting that

$$\begin{aligned}\Pr\left(s < -\tfrac{A}{2}(T_B+T_{PW}+2x) < s+\Delta s|s_o(t)\right) &= \Pr\left(\tfrac{2s}{A} < -(T_B+T_{PW}+2x) < 2\tfrac{s+\Delta s}{A}|s_o(t)\text{t}\right) \\ &= \Pr\left(-\tfrac{s}{A}-\tfrac{T_B}{2}-\tfrac{T_{PW}}{2} > x > -\tfrac{s+\Delta s}{A}-\tfrac{T_B}{2}-\tfrac{T_{PW}}{2}|s_o(t)\right) \\ &\cong \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(2s+AT_B+AT_{PW})^2}{2(2A)^2\sigma^2}} \frac{\Delta s}{A}\end{aligned} \tag{3.35}$$

where $\sigma^2 = \frac{N_1 T_{PRI}}{f_T^2}$ is the variance of the timing jitter variable. Likewise,

$$\Pr\left(s < -A(T_B - 2x) < s + \Delta s \,|\, s_o(t) \text{ sent}\right) \cong \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(s+AT_B)^2}{2(2A)^2\sigma^2}} \frac{\Delta s}{2A} \tag{3.36}$$

Thus, as we let $\Delta s \to 0$, we obtain the probability density function in the region of negative S, i.e.,

$$f_S(s|s_0(t)) = \frac{1}{\sqrt{2\pi(2A)^2\sigma^2}} \left[2e^{-\frac{(2s+AT_B+AT_{PW})^2}{2(2A)^2\sigma^2}} + e^{-\frac{(s+AT_B)^2}{2(2A)^2\sigma^2}} \right] \quad \text{for } -AT_{PW} < s < 0 \tag{3.37}$$

Skipping the details, we can obtain the density function for positive s in a similar manner, i.e.,

$$f_S(s|s_0(t)) = \frac{1}{\sqrt{2\pi(2A)^2\sigma^2}} \left[2e^{-\frac{(2s-AT_B-AT_{PW})^2}{2(2A)^2\sigma^2}} + e^{-\frac{(s+AT_B)^2}{2(2A)^2\sigma^2}} \right] \quad \text{for } 0 < s < AT_{PW} \tag{3.38}$$

Now that we have the probability density function for S (given that $s_0(t)$ is sent) and N, we can evaluate the probability of error in a manner similar to a likelihood ratio test (i.e., as if S and N were the variables received by the decision device).

The points of probability mass make this calculation somewhat harder than it would be without them, but with careful attention, it can be done correctly. For instance, for the case of $S=-AT_{PW}$, an erroneous decision would only be made if $N<-AT_{PW}$, thus we just need to integrate that part of the Gaussian density for N and multiply it times the probability that $S=-AT_{PW}$. For S=0 and $S=AT_{PW}$ we do the same, conditional on the probability that S is equal each of those values, respectively. For the continuous part of the density of S, we need to perform a double integral over the S and N. That is, the probability of error is the probability that S>N, thus conditional on $S \neq 0, \pm AT_{PW}$,

$$P_{err}(S \neq 0, \pm AT_{PW}) = \int_{-AT_{PW}}^{AT_{PW}} \int_{-\infty}^{s} f_{S,N}(s,n)\,dn\,ds \tag{3.39}$$

Since S and N are independent, the joint probability density function is just the product of the marginal densities, thus using (3.37) and (3.38)

$$P_{err}(S \neq 0, \pm AT_{PW}) = \int_{-AT_{PW}}^{AT_{PW}} f_S(s) \int_{-\infty}^{s} f_N(n) dn ds = \int_{-AT_{PW}}^{AT_{PW}} f_S(s) \left[1 - Q\left(\frac{s}{\sqrt{N_o T_B}} \right) \right] ds \quad (3.40)$$

which at seems to be inexpressible in closed form. However, we can still write an expression for the unconditional probability of error,

$$\begin{aligned} P_{err} &= P_{err}(S \neq 0, \pm AT_{PW}) + \frac{1}{4} \Pr(S = 0) + \\ &\Pr(S = AT_{PW}) \left[1 - Q\left(\frac{AT_{PW}}{\sqrt{N_o T_B}} \right) \right] + \Pr(S = -AT_{PW}) Q\left(\frac{AT_{PW}}{\sqrt{N_o T_B}} \right) \end{aligned} \quad (3.41)$$

Plugging in the expressions found previously for the probability masses, we have the following expression for the probability of error,

$$\begin{aligned} P_{err} = & \int_{-AT_{PW}}^{AT_{PW}} f_S(s) \left[1 - Q\left(\frac{s}{\sqrt{N_o T_B}} \right) \right] ds \\ & + \frac{1}{4} \left[Q\left(\frac{(T_B + T_{PW})}{2\sigma} \right) + Q\left(\frac{(3T_B + T_{PW})}{2\sigma} \right) \right] \\ & + \left[Q\left(\frac{(T_B + T_{PW})}{2\sigma} \right) + Q\left(\frac{(3T_B + T_{PW})}{2\sigma} \right) \right] \left[1 - Q\left(\frac{AT_{PW}}{\sqrt{N_o T_B}} \right) \right] \\ & + Q\left(\frac{AT_{PW}}{\sqrt{N_o T_B}} \right) \left[1 - 2Q\left(\frac{(T_B + T_{PW})}{2\sigma} \right) \right] \end{aligned} \quad (3.42)$$

which reduces to

$$\begin{aligned} P_{err} = & \int_{-AT_{PW}}^{AT_{PW}} f_S(s) \left[1 - Q\left(\frac{s}{\sqrt{N_o T_B}} \right) \right] ds + \frac{5}{4} \left[Q\left(\frac{(T_B + T_{PW})}{2\sigma} \right) + Q\left(\frac{(3T_B + T_{PW})}{2\sigma} \right) \right] \\ & + Q\left(\frac{AT_{PW}}{\sqrt{N_o T_B}} \right) \left[1 - Q\left(\frac{(T_B + T_{PW})}{2\sigma} \right) + Q\left(\frac{(3T_B + T_{PW})}{2\sigma} \right) \right] \end{aligned} \quad (3.43)$$

Noting that A^2T_{PW} is the signal energy, defining $\alpha = T_B / T_{PW}$, and $\chi = T_{PW} / \sigma$, we can further express the probability of error as

$$\begin{aligned} P_{err} = & \int_{-AT_{PW}}^{AT_{PW}} f_S(s) \left[1 - Q\left(\frac{s}{\sqrt{N_o T_B}} \right) \right] ds + \frac{5}{4} \left[Q\left(\frac{\alpha + 1}{2} \chi \right) + Q\left(\frac{3\alpha + 1}{2} \chi \right) \right] \\ & + Q\left(\sqrt{\frac{E_s}{N_o}} \right) \left[1 - Q\left(\frac{\alpha + 1}{2} \chi \right) + Q\left(\frac{3\alpha + 1}{2} \chi \right) \right] \end{aligned} \quad (3.44)$$

Now, since we cannot find a closed for expression for the integral in (3.44), but we know it necessarily has a positive value, we can lower bound the probability of error by ignoring the integral. Thus, a lower bound on the probability of error is

$$P_{err} > \frac{5}{4}\left[Q\left(\frac{\alpha+1}{2}\chi\right)+Q\left(\frac{3\alpha+1}{2}\chi\right)\right]+Q\left(\sqrt{\frac{E_s}{N_o}}\right)\left[1-Q\left(\frac{\alpha+1}{2}\chi\right)+Q\left(\frac{3\alpha+1}{2}\chi\right)\right] \qquad (3.45)$$

There are several important observations to be made from (3.45). First, note that if the signal to noise term, E_s/N_o, is fixed, then the pulse-width to timing jitter ratio, χ, can be increased only up to a certain point before the probability of error stops decreasing. The same is true if χ is fixed, and the SNR is increased, i.e., a error floor is encountered. In addition, since α represents the integration time to pulse width ratio, we see that performance improves with increasing α only up until a certain point, at which time an error floor is encountered. These relations are shown graphically in Figures 3.7-3.10.

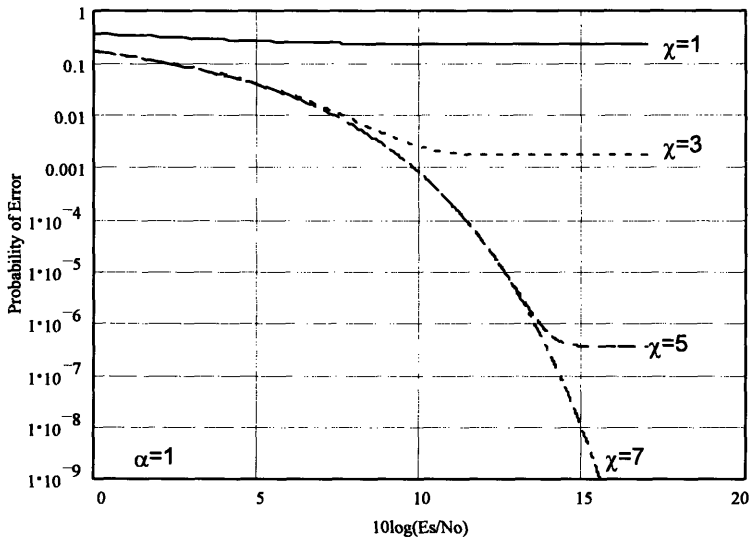

Figure 3.7: Lower bound on P_{err} with α=1 for various values of χ

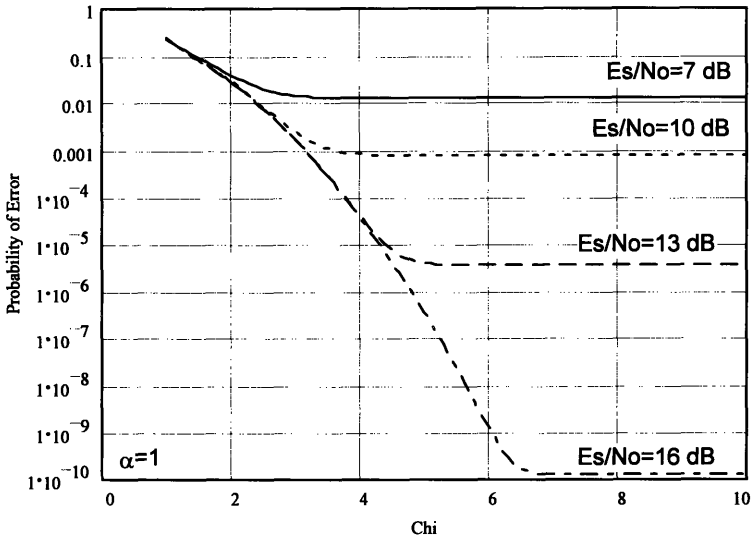

Figure 3.8: Lower bound on P_{err} with α=1 for various values of E_s/N_o

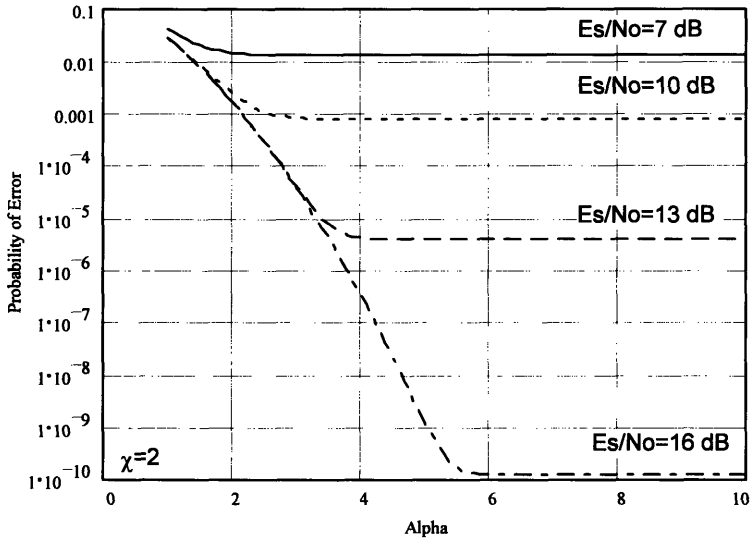

Figure 3.9: Lower bound on P_{err} with χ=2 for various values of E_s/N_o

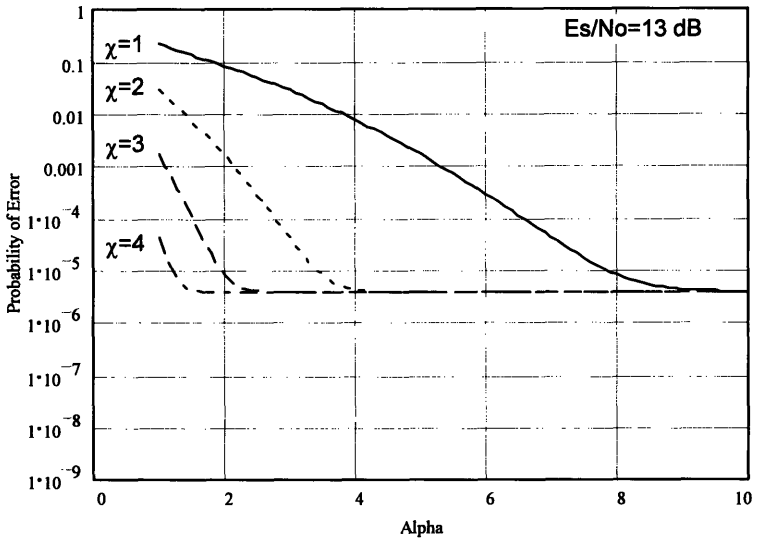

Figure 3.10: Lower bound on P_{err} with E_s/N_o=13 dB for various values of χ

It should be clear that extending this analysis for M>2 becomes a formidable task, and since the binary case exemplifies the error floor that we expect, it is sufficient to show the limitations of this suboptimal receiver.

3.3 Performance of Pulse Location Estimation Receivers

The previous discussion of Section 3.2 shows several limitations that should be obvious to the reader. First, the restriction of only one sample per possible pulse location, while optimal in the case of no timing jitter, is clearly inferior when timing jitter is present. The optimal "matched" filter for the single-sample receiver becomes just an integrator at high SNR, and as such lets in much more noise than is needed to accurately detect the pulse. This is so because the "forcing" of the receiver to make an estimation of whether the pulse is there or not for each pulse location *bin* is inferior to actually trying to estimate the location of the pulse. In this section, we will treat the PPM scheme as an *analog* signal and derive the maximum likelihood pulse *location* estimation receiver. This estimate of the pulse location is inherently a soft-decision detector, thus it can be fed into a decision device using coding or to a hypothesis-tester (hard decision device) equally well.

3.3.1 Optimal PPM Receiver with no Timing Jitter

The basic analysis of an optimal (maximum-likelihood) analog PPM system can be found in [27], the major portions of which will be presented here as a framework for the analysis that follows. The basic idea is to filter the incoming signal with a matched filter (i.e., h(t)=p(-t)) and instead of sampling at only one time, allow the entire analog waveform (y(t)) to go to a peak detection device. The peak of y(t), over some observation time ($-T_B$ to $+T_B$ by our definitions before), is declared as the estimate of the pulse location. A block diagram of this receiver is shown in Figure 3.11.

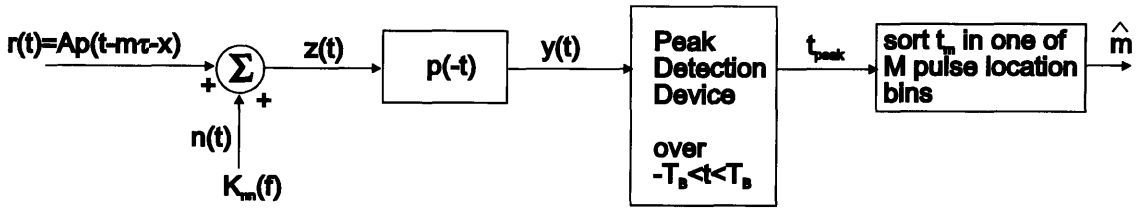

Figure 3.11: Maximum likelihood pulse location estimation and pulse detection

The nomenclature "maximum likelihood" refers to the term t_{peak}, where this is the estimation of the location of the pulse. Thus, this receiver maximizes the likelihood that t_{peak} is the pulse location that was sent by the transmitter, and in turn that m is the most likely symbol that was sent. The term t_{peak} is actually a random variable, with mean $m\tau$, however it's density is difficult to determine exactly. What we do know is that it is symmetric about it's mean, and that we can lower bound it's variance through a Cramer-Rao method. The derivation of the bound is somewhat tedious, so the reader is referred to [27] for more details.

For fairly large E_s/N_o the error variance is dominated by non-anomalous types of errors. "Anomalous" errors are due entirely to AWGN creating false peaks in pulse location bins other than the intended bin. A good estimate of the variance of t_{peak} in the absence of anomalous errors is shown to be [27]

$$Var\left[t_{peak}\right] \cong \frac{12}{\pi^2}\left(\frac{T_{PW}}{2T_B}\right)^2 \frac{N_o}{2E_s} \tag{3.46}$$

for the sin(x)/x-type pulse p(t) as in (3.20). Note that the variance is proportional to $(T_{PW})^2$ as one would expect, since smaller T_{PW} means the sin(x)/x-typepulse is "narrower" and "taller" for the same E_s, thus errors in the estimation of the location of the pulse should be smaller.

For smaller E_s/N_0 ratios, or for larger values of T_B, the chance of an anomalous reception rises. What we mean here is that there is a chance that AWGN alone will create a peak in y(t) that is greater than that created by the signal component (with AWGN added to it) near the actual pulse location, $m\tau$. The probability of an anomalous error can be approximated by the probability of (symbol) error with an orthogonal signal set [27], and is given by

$$\Pr(anomaly) \le (2T_B T_{PW} - 1)Q\left(\sqrt{\frac{E_s}{N_o}}\right) \tag{3.47}$$

In fact this bound is very tight, thus the inequality can be replaced by an approximate equality.

In terms of evaluating the probability of error in deciding which of M pulses was actually sent, we can assume that if an anomalous pulse position estimation error is made, that an error in the (digital) pulse position location is made. In fact, this makes up the entire probability of symbol error, since for sin(x)/x-type pulses small errors of the type described by the variance in (3.46) are *by definition* closer to the intended pulse's center than other pulses' centers. Thus in the last box of Figure 3.11 no error will be made, i.e., $\hat{m} = m$. Thus, the probability of error for this type of detector (with no timing jitter) is for all practical purposes identical to that described in Section 3.2.1, and is given by

$$P_{err,sym} \le (M-1)Q\left(\sqrt{\frac{E_s}{N_o}}\right) \tag{3.48}$$

for $M \le 2T_B T_{PW}^{-1}$. The position of the pulses in the time period observed ($-T_B$ to T_B) by the peak detector is not important when there is no timing jitter, but in the next section we

will show that in the presence of timing jitter the actual position of the pulses is very important.

3.3.2 Optimal PPM Receiver with Timing Jitter and AWGN

The optimal PPM receiver does not change from that described in the previous section when timing jitter is included. Rather, the estimation of the pulse location just includes the timing jitter variable, x. Thus, the probability of symbol error can only be estimated since no probability density function for t_{peak} has been derived. We will now define the pulse location estimate as t'_{peak}, where

$$t'_{peak} = t_{peak} + x \tag{3.49}$$

To estimate the probability of error, we will make a few assumptions and approximations:

1. We will assume T_B is fixed; thus, there is a limit to how many orthogonal pulses can be fit into this time span (we have previously defined this time span as the *symbol epoch*) before T_{PW} must be shortened. In the limit of large numbers of pulses we would have to adjust the height of the pulses to keep the energy per pulse constant.
2. We will ignore edge effects of the pulses abutting the T_B boundaries, thus, we are effectively adding a guard time to the outside pulses. In the limit of large M, this is an equivalent approximation to that made in evaluating higher order M-QAM constellations, i.e., treating all points as if they are interior points in the constellation.
3. We will treat the local randomness (small errors in t_{peak} due to AWGN, not timing jitter) as Gaussian, even though we have not shown that they are Gaussian. Thus, since the timing jitter, x, is Gaussian (see Section 3.1) the location estimation t_{peak} will be approximated as Gaussian.
4. The error induced by anomalous peak detections will be treated as independent from the error induced by the timing jitter and local randomness about the true signal peak. Thus, the probability of error will be the sum of the conditional probabilities of error given whether or not an anomaly occurs.

Taking these assumptions into account, estimating the probability of error is relatively straightforward. The symbol epoch is broken up into M equally sized bins, and without loss of generality we can assume a particular pulse $s_i(t)$ is sent. There are two types of errors: anomalous errors, which could cause an erroneous detection in any of the M-1 other bins, and "jitter+AWGN" errors, which will cause the incorrect symbol detection in one of the neighboring bins, with decreasing probability for bins farther away from the correct bin location. The first probability is unrelated to jitter and was given in (3.48). The second probability of error, under the assumptions made above, is easily seen to be

$$\begin{aligned}\Pr(\text{jitter causes detection in neighboring bins}) &= \Pr\left(\left|t'_{peak}\right| > \frac{T_B}{M}\right) \\ &= 2Q\left(\frac{T_B/M}{\sqrt{Var[x]+Var[t_{peak}]}}\right)\end{aligned} \tag{3.50}$$

Substituting in the variance of t_{peak} found in (3.46) we can estimate the probability of error in (3.50) as

$$\begin{aligned}\Pr(\text{jitter causes detection in neighboring bins}) &\cong 2Q\left(\frac{T_B/M}{\sqrt{\frac{N_1 T_{PRI}}{f_T^2}+\frac{12}{\pi^2}\left(\frac{T_{PW}}{2T_B}\right)^2\frac{N_o}{2E_s}}}\right) \\ &= 2Q\left(\frac{2\pi T_B}{M}\sqrt{\frac{1}{\frac{\gamma}{f_T^2}+\frac{6}{\alpha^2}\frac{N_o}{E_s}}}\right)\end{aligned} \tag{3.51}$$

where $\gamma = (2\pi)^2 N_1 T_{PRI}$ is the "phase noise SNR" described in Chapter 2, and $\alpha = T_B / T_{PW}$. Thus, the overall probability of symbol error can be approximated as

$$P_{err,sym} \cong 2Q\left(\frac{2\pi T_B}{M}\sqrt{\frac{1}{\frac{\gamma}{f_T^2}+\frac{6}{\alpha^2}\frac{N_o}{E_s}}}\right)+(M-1)Q\left(\sqrt{\frac{E_s}{N_o}}\right) \tag{3.52}$$

Substituting in representative values for the variables in (3.52) it becomes clear that the variance of the local randomness is much greater than that introduced by the timing jitter unless the observation period, $2T_B$, is very large when compared with T_{PW} (i.e., α must be large) *or* E_s/N_o is very large. Using representative values of $T_{PRI}=10^{-6}$ sec., $f_T=5$ MHz, and $T_{PW}=10^{-9}$ sec. for the Pulson ultrawideband PPM system, we can graph the probability of error as a function of T_B for M=2. This is shown in Figure 3.12 for E_s/N_o of 10 and 20 dB.

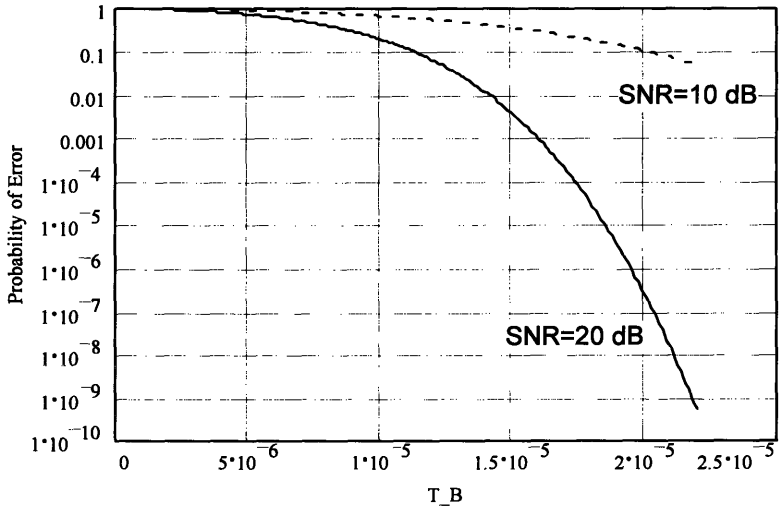

Figure 3.12: Probability of symbol error versus T_B

Since it is clear that T_B as graphed in Figure 3.12 is not feasible (i.e., $T_B>T_{PRI}$ for a reasonable P_{err}, which by the definition of the modulation scheme cannot happen) then E_s/N_o must be much larger than 20 dB. This is not really a restriction, however, since the E_s/N_o we are discussing actually represents the *peak* SNR, rather than the *average* SNR we've been implicitly dealing with since Chapter 2. For an ultrawideband system such as Pulson's proposal, the duty cycle for pulse transmission is on the order of 1/1000 (i.e., $T_{PW}/T_{PRI}=10^{-3}$), thus, a more representative value for E_s/N_o is 40-50 dB (or even higher with smaller duty cycles). Given this observation, the Pulson system appears much more reasonable.

3.4 Results and Discussion

In this chapter we have examined the effect of timing jitter on a digital pulse-position modulation scheme. Since our principle interest in this thesis has been to examine the effects of phase noisy oscillators on the radio communications problem, we first showed how phase noise in a reference oscillator manifests itself in timing jitter in the transmitter clock. Using a simple example, we showed that under certain circumstances the resulting timing jitter can be approximated as a Gaussian random variable.

We have examined several receivers, including both optimal and suboptimal structures, for communicating over AWGN channels when the transmitted signal is also corrupted by timing jitter. The results indicate that knowledge of the probability density function of the jitter improves the performance of such a system, and in the limit of high signal to additive noise ratios, the optimal receiver is essentially an integrator. The exact performance of this receiver was shown to be difficult to analyze exactly, primarily because of the structure of the filter h(t). In particular, the optimal h(t) generally causes the noise components at the decision device to become dependent. The (suboptimal) receiver structure suggested by the optimal receiver in the high SNR case was examined for the binary signaling case, and it was found that there is an error floor, if either the SNR or the pulse width to jitter deviation ratio (χ) is fixed, even in the limit of infinite χ or SNR, respectively.

In addition to the single-sample receivers structures, a maximum-likelihood receiver structure, based on estimating the actual location of the pulse (in an analog fashion) was studied. Under certain assumptions, we derived an estimate of the probability of error for this receiver, and noted that for the ultrawideband Pulson-type modulation scheme, there may be some (possibly minor) limitations. Namely, the variance of the pulse location estimate in AWGN (without regard to the additional timing jitter introduced at the transmitter) can be large enough to severely degrade operation of a receiver without either a large observation interval (T_B) or large *peak* SNR.

As a general conclusion we point out that analysis of a digital PPM scheme of the type shown in this chapter is quite complex. Namely, the nonlinear nature of the

modulation and the complex interaction between the timing jitter and additive noise sources create formidable obstacles for the system designer. Although we did not propose or study any nonlinear receiver schemes, it is possible that such a receiver might have significantly better performance.

4. Conclusions and Future Work

4.1 Conclusions

In this thesis we have studied the effects of a phase noisy oscillator on the radio communications problem. The relatively recent emphasis on minimizing power consumption in these systems has spurred interest in these effects. Previously, phase noise was considered a secondary impairment and thus was neglected in much of the modulation design. It is hoped that better understanding of how phase noise degradations manifest themselves in various modulation formats will lead to better modulation design. Better modulation design would, hopefully, allow relaxation of phase noise requirements on the device designers, thus allowing lower power consumption designs with poorer phase noise characteristics. In this regard, we have concentrated our efforts on the symbol error performance of two different modulation schemes that are currently used or proposed for low-power radio communications.

In the second chapter we examined the currently popular quadrature phase shift keying format, which is essentially a simple case of more general quadrature amplitude modulation. In fiber-optic communications, where phase noise has been studied more extensively, it was noted that binary modulation formats are used almost exclusively due to the large available bandwidth. By contrast, in radio communications the available

bandwidth is highly constrained, thus higher-order modulations such as QAM allow for more data to be transmitted at the expense of more complicated signaling. In Chapter 2 we showed that phase noise in the transmitter's oscillator destroys the orthogonality between the in-phase and quadrature channels. Since QAM is designed to take advantage of this orthogonality it would seem that this would be a significant degradation in the performance of such systems. While this is true in general, we showed that for the type of carrier frequencies and linewidths likely to be encountered in *radio* communications systems this non-orthogonality is negligible in terms of affecting symbol error performance. We showed that even in the presence of significant phase noise, if perfect carrier recovery is possible that the symbol error performance of QPSK, and in general QAM, is virtually identical to the no phase noise case. However, we also showed that with imperfect carrier recovery, specifically if only samples of the phase of the carrier are known at the receiver, that symbol error performance *is* significantly degraded. Although no information-theoretic results were obtained, it was noted that estimation of the phase of the carrier is equivalent to trying receive even *more* information from the transmitter (beyond the actual data) across the channel. We noted that this might not be information theoretically possible, given the level of additive noise in the channel.

In Chapter 3 we examined an ultrawideband digital pulse-position modulation system, the examination of which was spurred by the proposal of such a system by a commercial company. This system is essentially a baseband-type PPM system, with no defined carrier, thus the analysis of the system focused on the effect of a phase noisy reference oscillator from which a clock was generated. It was shown that phase noise in the reference oscillator manifests itself as timing jitter in the clock, and that under some relatively reasonable assumptions the timing jitter has a density that is approximately Gaussian (when the phase noise is a Brownian motion process). Similarly to Chapter 2, we focused on perfect and imperfect *timing* recovery (which is completely analogous to carrier recovery in the QPSK system) and their subsequent effects on symbol error performance. It was shown that perfect timing recovery removes all timing jitter from the system, and thus performance of the digital PPM system is identical to any orthogonal modulation scheme. With imperfect timing recovery, specifically where the sampling

time is a random variable with a known probability density function, we found that an optimal single-sample receiver can be derived, but that its performance is difficult to characterize. This formulation, however, did suggest a suboptimal receiver structure which approximates the optimal formulation in a high signal-to-additive noise (SNR) environment. The symbol error performance of the suboptimal receiver was analyzed for both binary and M-ary formats, and it was shown that there exists an error floor for a fixed SNR or fixed pulse width-to-jitter standard deviation ratio ($\chi=T_{PW}/\sigma$). We further showed that a superior receiver can be constructed that estimates the location of the pulse (i.e., an analog quantity) in a maximum likelihood fashion. Once again, it was difficult to characterize the symbol error performance of such a system, however, an approximation of the error performance was obtained.

The main results of the analysis in this thesis can be summarized in two main statements. First, the availability or accurate estimation of the carrier waveform in QAM modulation is of paramount importance. In some instances, the transmission of a reference waveform, devoid of modulation, might improve performance enough to offset the increase in transmission power necessary for the second signal. And in any event, the lack of accurate estimation of the carrier waveform degrades the symbol error probability of QAM signaling significantly. It was also pointed out that accurate estimation of the carrier waveform is equivalent to trying to send *more* information across the channel, and if the carrier waveform is sufficiently phase noisy, then the total amount of information (estimation and data) might be greater than the capacity of the AWGN channel. Secondly, timing recovery is as essential to digital PPM as is carrier recovery to QAM (in a sense, they are the same thing, relative to their respective systems). However, the transmission of a reference "clock" does not seem to make any sense, thus the performance results stand alone and do not suggest any alternative technique. Rather the major result in dealing with PPM seems to be that there is a definite tradeoff between additive noise and timing jitter (phase noise) in designing the system. In other words, we can be more tolerant to timing jitter if we can allow more additive noise, and vice-versa. This tradeoff is seen for all receivers studied in Chapter 3.

In conclusion, we note that this thesis has been a first attempt to understand the impact of phase noisy oscillators on radio communications, in particular those with low-power consumption as a high priority design goal. We feel that the significance of the thesis is not so much in the quantitative results, such as the symbol error probabilities, but more in the increased awareness of the importance of phase noise in the design of such systems. Therefore it is hoped that some of the ideas of this thesis might spur further research into more phase noise tolerant modulation schemes for radio communications.

4.2 Future Work

As with any technical work, this thesis does not end the problems related to its topic. In fact, there are many open problems that remain and that need to be solved before we can ignore the effect of phase noise in low-power radio communications systems design. We will briefly describe some of these problems.

4.2.1 Statistics of Phase Noisy Random Variables

We have seen throughout this thesis certain random variables that are uniquely determined by the phase noise process. Because of the apparent difficulty in evaluating their marginal and joint probability density functions, we were often forced to employ approximations based on their means and variances. Some of these variables' densities were estimated through simulation, but no closed form expressions were found. It would be invaluable to have a compact statistical description of such random variables, for example through a closed-form marginal or joint probability density function. Only then could we more accurately predict the performance of the systems studied.

4.2.2 Performance of Other Modulation Schemes

Because of the constraints imposed by a thesis of this type, there were many prevalent modulation techniques currently being used in radio communications that were not studied. For example, spread spectrum modulation schemes, such a direct sequence or frequency-hopped spread spectrum were not studied. Since these modulation schemes have many advantages in fighting other degradations such as fading, and in handling multiple-users simultaneously, it would be important to consider the effects of phase noisy oscillators on their performance.

4.2.3 Effect of Multiple Users on System Performance

Throughout the thesis it was implicit that we were discussing point-to-point communications links, with a single user (or pair of users). With other users either frequency or time division multiplexed into the available radio spectrum several complications occur. For example, in the introduction we showed that phase noisy oscillators widen the spectrum of the carrier, thus in FDM systems, wider guard frequencies would be needed to prevent adjacent channel interference. In the digital PPM case, the timing jitter would necessitate wider guard times in a similar manner. Also, it was shown that the pulse location estimation receiver for an ultrawideband system might not be feasible if the observation time is not sufficient (and since long observations time directly influence the number of users that could be serviced, this affects the multiple-user performance of such a system). Thus, more study is warranted to see if this is a limitation of the analysis or truly a limitation of the modulation scheme. Since most future commercial radio systems will be classified as multiple-access and will incorporate various low-power design techniques, an analysis of the degradations and design constraints imposed by phase noisy oscillators is needed to fully predict their performance.

4.2.4 Comparison of Phase Noise to Doppler Effects

Although we have specifically considered the phase noise on the carrier as introduced by the oscillator in the transmitter, we can easily characterize such phase noise as a channel impairment. In that way, there are similarities between the effects of Doppler on a carrier and a phase noisy carrier. Since there is a multitude of research on how to best handle the effects of Doppler in the mobile channel, it would seem natural to try to exploit this research by fully understanding the similarities and differences between phase noise and Doppler effects.

Bibliography

[1] D.C. Cox, "A Radio System Proposal for Widespread Low-Power Tetherless Communications," *IEEE Transactions on Communications*, Vol. 39, No. 2, pp. 324-335, February 1991.

[2] D.C. Cox, "Wireless Network Access for Personal Communications," *IEEE Communications Magazine*, pp. 96-115, December 1992.

[3] N. Abramson, "Multiple-access in Wireless Digital Networks," *Proceedings of the IEEE*, Vol. 82, No. 9, pp. 1360-1370, September 1994.

[4] S. Sheng, A. Chandrakasan, R.W. Broderson, "A Portable Multimedia Terminal," *IEEE Communications Magazine*, pp. 64-75, December 1992.

[5] M. Medard, "The Capacity of Time Varying Multiple User Channels in Wireless Communications," Sc.D. Thesis, Massachusetts Institute of Technology, 1995.

[6] M. Azizoglu, "Phase Noise in Coherent Optical Communications," Ph.D. Thesis, Massachusetts Institute of Technology, 1991.

[7] J. Salz, "Coherent Lightwave Communications," *AT&T Technical Journal*, Vol. 64, No. 10, pp. 2153-2209, December 1985.

[8] S.J. Goldman, *Phase Noise Analysis in Radar Systems Using Personal Computers*, New York: John Wiley & Sons, 1989.

[9] W.C. Lindsey and M.K. Simon, *Telecommunication Systems Engineering*, Englewood Cliffs N.J.: Prentice-Hall, 1973.

[10] M.H. Perrott, "Frequency Synthesizer Proposal," Ph.D. Thesis proposal, Massachusetts Institute of Technology, 1995.

[11] B. Sklar, *Digital Communications: Fundamentals and Applications*, Englewood Cliffs N.J.: Prentice-Hall, 1988.

[12] J.R. Vig, *Quartz Crystal Resonators and Oscillators for Frequency Control and Timing Applications*, Army Research Laboratory, SLCET-TR-88-1 (Rev 6.2), 1994.

[13] D.B. Sullivan, D.W. Allan, D.A. Howe, and F.L. Walls, Eds., "Characterization of Clocks and Oscillators," *NIST Technical Note 1337*, March 1990.

[14] J. Rutman, F.L. Walls, "Characterization of Frequency Stability In Precision Frequency Sources," *Proceedings of the IEEE*, Vol. 79, No. 6, pp. 952-960, June 1991.

[15] J. Rutman, "Characterization of Phase and Frequency Instabilities in Precision Frequency Sources: Fifteen Years of Progress," *Proceedings of the IEEE*, Vol. 66, No. 9, pp.1048-1075, September 1978.

[16] C.H. Henry, "Theory of Linewidth of Semiconductor Lasers," *IEEE Journal of Quantum Electronics*, vol. QE-18, pp. 259-264, February1982.

[17] W.T. Webb, "Modulation Methods for PCNs," *IEEE Communications Magazine*, pp. 90-95, December 1992.

[18] J.G. Proakis, *Digital Communicaitons*, New York: McGraw-Hill, 3rd Ed., 1995

[19] E.A. Lee and D.G. Messerschmitt, *Digital Communicaion*, Boston: Kluwer Academic Publishers, 2nd Ed., 1994.

[20] R.E. Zeimer and W.H. Tranter, *Principles of Communications, Systems, Modulation, and Noise*, Boston: Houghton Mifflin, 2nd Ed., 1985.

[21] R. Steele, Ed., *Mobile Radio Communications*, New York: IEEE Press, 1992.

[22] G.R. Cooper and C.D. McGillem, *Modern Communications and Spread Spectrum*, New York: McGraw-Hill, 1986.

[28] G. Strang, T. Nguyen, *Wavelets and Filter Banks*, Wellesley MA, Wellesley-Cambridge Press, 1996.

[24] R.G. Gallager, *Discrete Stochastic Processes*, Boston, Kluwer Academic Publishers, 1996.

[25] Qualcomm, Inc., "An Overview of the Application of Code Division Multiple Access to Digital Cellular Systems and Personal Communications Networks," a submission to the Cellular Telecommunications Industry Association "CDMA Technology Investigation Subcommittee," 1991.

[26] Pulson Communications Corporation, "Impulse Radio Basics," 1994.

[27] J.M. Wozencraft, I.M. Jacobs, *Principles of Communication Engineering*, New York, John Wiley & Sons, 1965.

[28] L.E. Franks, *Signal Theory*, Stroudsburg, PA, Dowden & Culver, 1981.