Capacity of and Coding for Multiple-Aperture, Wireless, Optical Communications

by

Shane M. Haas

B.S. Mathematics, University of Kansas (1998)
B.S. Electrical Engineering, University of Kansas (1998)
M.S. Electrical Engineering, University of Kansas (1999)
M.A. Mathematics, University of Kansas (1999)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2003

Author .....
Department of Electrical Engineering and Computer Science
May 14, 2003

Certified by .....
Jeffrey H. Shapiro
Julius A. Stratton Professor of Electrical Engineering
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Capacity of and Coding for Multiple-Aperture, Wireless, Optical Communications

by

Shane M. Haas

Submitted to the Department of Electrical Engineering and Computer Science
on May 14, 2003, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Abstract

Refractive index turbulence causes random power fluctuations in optical communication systems, making communication through the atmosphere difficult. This same phenomenon makes the stars twinkle at night, and pavement shimmer on a hot summer day. True to the old adage, "don't put all your eggs in one basket," we examine laser communication systems that use multiple transmit and receive apertures. These apertures provide redundant replicas of the transmitted message to the receiver, each corrupted separately by the atmosphere. Reliable communication occurs when not all of these paths are deeply faded. We quantify the maximum rate of reliable communication, or capacity, and study space-time coding techniques for both direct- and coherent-detection receivers. We also experimentally verify the performance of some simple techniques for optically-preamplified, direct-detection receivers.

Thesis Supervisor: Jeffrey H. Shapiro
Title: Julius A. Stratton Professor of Electrical Engineering

Dedicated to my father, who instilled in me the desire to learn...

Acknowledgments

I would like to foremost thank my research advisor Jeff Shapiro. This thesis would not have been possible without his guidance and support. I am continually amazed at how he juggles being R.L.E. director, teaching classes, and still meeting every week individually with all of his graduate students. I have learned so much from him during these past four years.

I would like to thank my thesis committee member and academic advisor Dave Forney. He really helped me with my transition to M.I.T. Every semester, he would talk with me, and give me great advice on classes, career paths, and life. He also was a mentor to so many students, including myself, during the Area I Graduate Seminars.

I also want to thank Vahid Tarokh. Our discussions on space-time codes and information theory greatly influenced the development of this thesis. He also served as a member on my thesis committee.

For his help with the experimental aspect of this thesis, I would like to thank Franco Wong. In addition to his laboratory expertise, I really appreciated our conversations about investing. Someday, I hope to convince him that derivatives are not evil. I also thank Baris Erkmen, Etty Shin, and Mohsen Razavi for helping with the experimental setup.

I would like to thank my officemate Chris Kuklewicz for our many enjoyable conversations. I am also honored to have introduced Chris to the fine beverage of Guinness.

My family and friends have always given me great support and encouragement. I want to thank my mother Judy, and my brothers Brett and Ty for being there for me during all of these years. I also want to thank Lillian for the time that we have spent together, and for the more to come.

Finally, I thank the Defense Advanced Research Projects Agency for supporting this research in part through grant MDA972-00-1-0012.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

An estimated 95 percent of United States buildings are within 1.5 km of fiber-optic communication infrastructure, but currently unable to access it. One factor contributing to this inability is the high cost of optical fiber installation, approximately $100,000–$200,000 per kilometer in metropolitan areas, with trenching costs responsible for 85 percent of the total. Point-to-point optical communication through the atmosphere (i.e., wireless optical communication) has the potential to provide gigabit per second data rates at roughly one-fifth the price of ground-based, fiber-optic technologies [65].

Communicating optically through the atmosphere, however, poses many inherent challenges. Bad weather (e.g., fog, snow, rain, etc.) and atmospheric molecular constituents (e.g., carbon dioxide and oxygen molecules) cause absorption and scattering that degrade the performance of optical communication systems. Furthermore, the temporal and spatial evolution of thermal inhomogeneities in the troposphere under clear weather conditions cause random fluctuations in the refractive index at optical wavelengths [62]. These refractive-index perturbations—usually referred to as atmospheric turbulence—lead to amplitude and phase fluctuations on light beams propagating through the atmosphere [62], [31]. These fluctuations, in turn, have profound effects on the performance of laser communication systems operating over turbulent paths [49].

One strategy to combat these deleterious effects is to make available to the re-

ceiver multiple copies of the transmitted message, each corrupted separately by the atmosphere. In this thesis, we will develop such methods to establish reliable communication through the turbulent atmospheric channel.

1.1 Optical Detection

We will examine two methods of converting an optical field into an electrical signal. Direct detection refers to receivers that respond only to optical power, i.e., the magnitude squared of the optical complex field. If the inherent randomness in photon arrivals is much greater than thermal noise, we can count the individual photons and make decisions based on photon arrival times. More realistically, we can use a conditional Gaussian approximation to examine the influence of all the noise sources that arise in a practical communication system.

In contrast to direct detection, coherent detection receivers mix the incoming optical field with a spatially and temporally coherent local oscillator. This heterodyne structure essentially yields a traditional additive, white Gaussian noise channel.

Our main focus will be exploring spatial and temporal diversity using multiple transmit and receive apertures. We will derive the information-theoretic capacity of communication, and study coding techniques for two direct detection and one coherent detection atmospheric channels. The first direct-detection channel uses ideal photon-counting receivers. The second direct-detection channel employs optical preamplification. The last channel uses coherent-detection receivers. Finally, we will experimentally verify, through hardware implementation, the benefit of receiver diversity using optically-preamplified, direct-detection receivers. Chapter 2 explains these channels in more detail.

1.2 Channel Capacity

Roughly speaking, channel capacity is the maximum rate of reliable communication [14, 21]. Although atmospheric losses are random, they are approximately constant

on intervals less than one to ten milliseconds [50]. Because typical data rates can exceed a billion bits per second, a block of several million bits can experience, on average, similar fading conditions.

This block fading model lends itself to many different information-theoretic notions of capacity [5, 44, 56]. Without any delay constraints, we can code over many channel realizations and achieve reliable communication rates up to the Shannon capacity, defined as the average maximum mutual information per unit time, where the average is taken with respect to the random path gains [24]. Denoting the path gains of an N transmit, M receive aperture system as $\alpha = \{\ \alpha_{nm} \mid 1 \leq n \leq N, 1 \leq m \leq M\ \}$, the ergodic capacity is the expectation with respect to the path gains of the instantaneous capacity, i.e., $E[C(\alpha)]$, [44, 66]. The ergodic assumption requires that communication occurs over several atmospheric coherence times, which allows coding across both "good" and "bad" channel realizations.

In practice, however, delay constraints may prevent coding over many channel realizations. In this case, the strict Shannon capacity is zero because there is a chance that the fading might be so egregious that the instantaneous capacity is below any desired rate [24, 5]. In this case, a more appropriate measure of capacity is the probability that the channel can support a desired rate. The capacity C_p per outage probability p is given by [5]

$$p = \Pr\{C(\alpha) \leq C_p\}. \tag{1.1}$$

In other words, the capacity per outage probability p is the p-th percentile of the instantaneous capacity, $C(\alpha)$, distribution. The channel can support data rates up to the outage capacity C_p with probability $1 - p$.

We will examine the ergodic and outage capacities for our three atmospheric communication channel models. We are particularly interested in how the capacity scales with the number of transmit and receive apertures. These capacities will also depend on whether the transmitter and/or receiver know the path gains.

1.3 Space-Time Coding

Space-time codes (STC) refer to multiple transmitters sending codewords to multiple receivers over multiple time periods. For example, let $x_n(t)$ represent the symbol sent on the n-th transmit aperture during the t-th time period. A space-time codeword is then a matrix

	Tx 1	Tx 2	$\cdots$	Tx N
$t=1$	$x_1(1)$	$x_2(1)$	$\cdots$	$x_N(1)$
$t=2$	$x_1(2)$	$x_2(2)$	$\cdots$	$x_N(2)$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$t=T$	$x_1(T)$	$x_2(T)$	$\cdots$	$x_N(T)$

Tarokh in [61] established space-time code design criteria for Rayleigh and Ricean fading channels. These design criteria specify the pairwise properties of codewords from the space-time code. We will develop similar criteria for the coherent detection channel model, and demonstrate the reliability improvement gained through STCs.

For example, Figure 1-1 shows the pairwise error probability for a two transmit, one receive antenna, Alamouti STC in Rayleigh fading as a function of signal-to-noise ratio [1]. The Alamouti STC uses two transmit apertures and two time-slots to send two complex symbols s_1 and s_2 according to the schedule

	Tx 1	Tx 2
t $=1$	s_1	s_2
t $=2$	$-s_2^*$	s_1^*

In Chapter 3, we will examine the performance of the Alamouti STC under lognormal fading for coherent detection receivers.

Tarokh demonstrated that the Alamouti STC is an example of a complex orthogonal design STC [60]. We will also show in Chapter 3 that orthogonal designs minimize the pairwise error probability for heterodyne systems using many apertures. Figure 1-2 shows that an orthogonal design space-time code can greatly reduce the effects of atmospheric turbulence on the error probability.

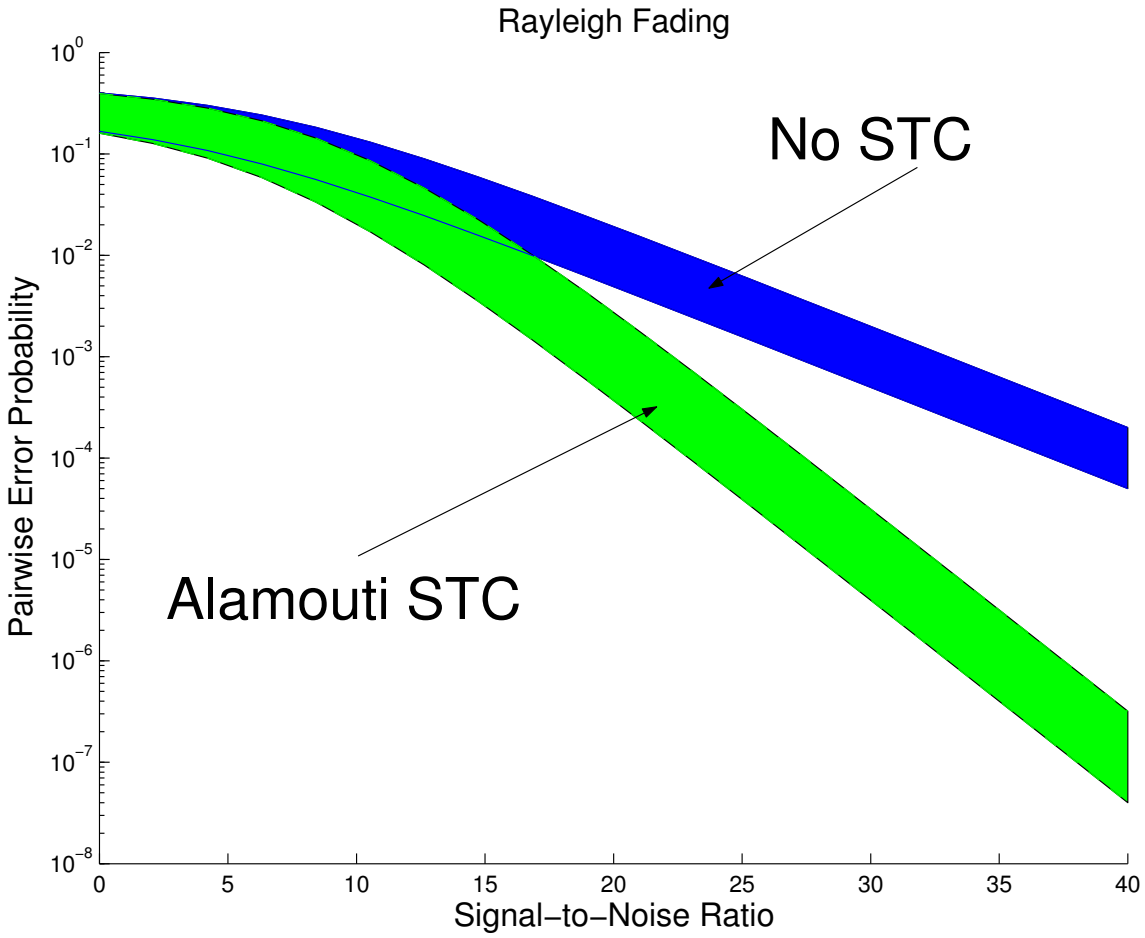

Figure 1-1: Space-time codes can improve the reliability of communication on fading channels. The shaded areas lie between upper and lower bounds on the pairwise error probability achieved in Rayleigh fading with and without an Alamouti space-time code. The bounds are plotted as a function of signal-to-noise ratio. In this case, the signal-to-noise ratio is the ratio of transmitted codeword energy difference to the receiver noise variance per real dimension.

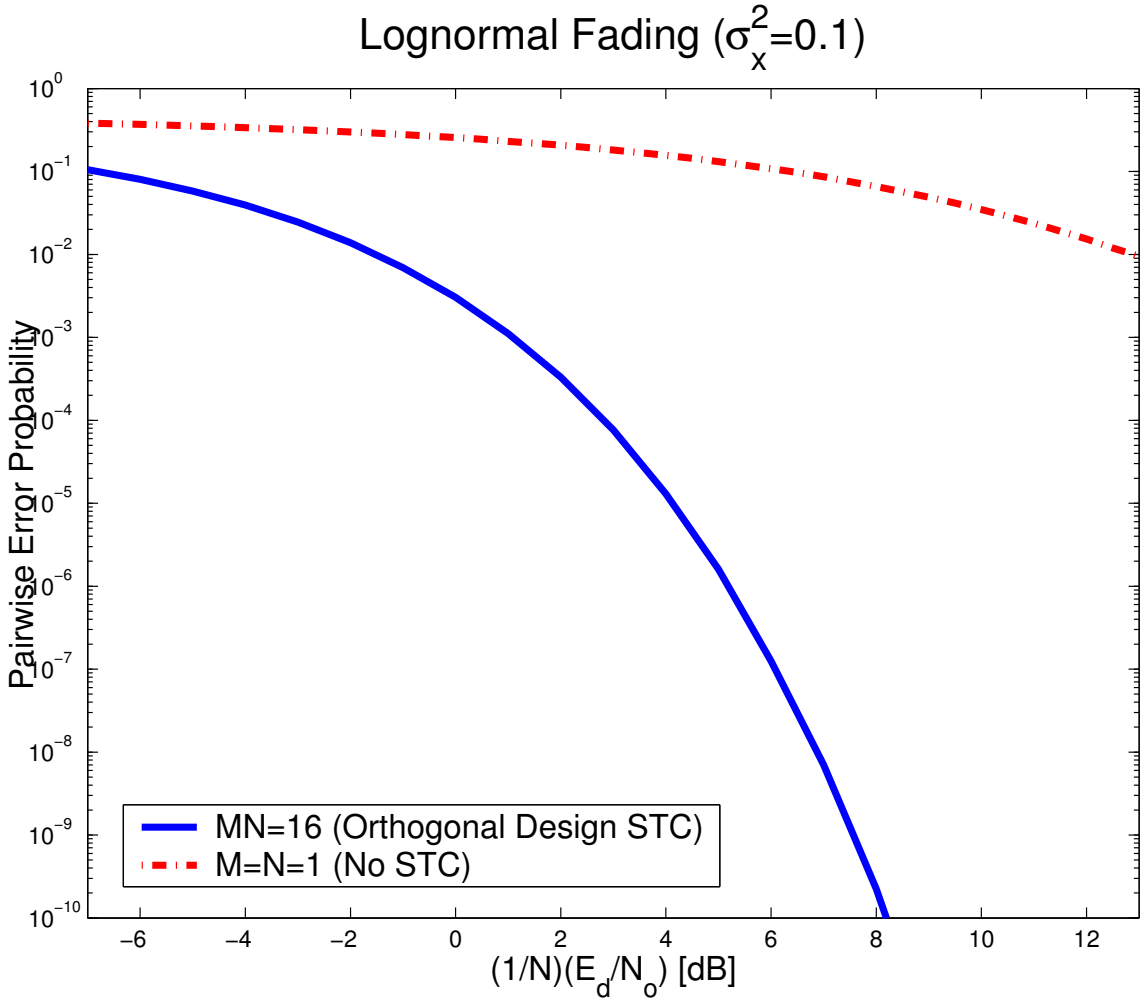

Figure 1-2: The pairwise error probability in moderate ($\sigma_\chi^2 = 0.1$) lognormal fading is shown for an orthogonal design STC using coherent detection receivers. The product of transmit (N) and receive (M) aperture numbers is $MN = 16$. The error probability is plotted against the ratio of energy difference between codewords at the transmitter (E_d) per transmit aperture (N) and receiver noise power spectral density (N_0). Also shown is the single transmit, single receive aperture ($M = N = 1$) error probability when no space-time code is used.

1.4 Summary of Main Results

The main theoretical results of this thesis appear in Chapters 3 through 5, categorized by the type of receiver structure. Chapter 3 explores quadrature amplitude modulation and coherent detection receivers. Chapters 4 and 5 examine intensity modulation and direct detection reception. Experimental results appear in Chapter 6. We present background material necessary for understanding the channel models in Chapter 2.

Figure 1-3 shows multiple lasers and detectors providing spatial diversity to combat the effects of atmospheric fading. All the results of this thesis are based on this multiple-input, multiple-output (MIMO) channel model.

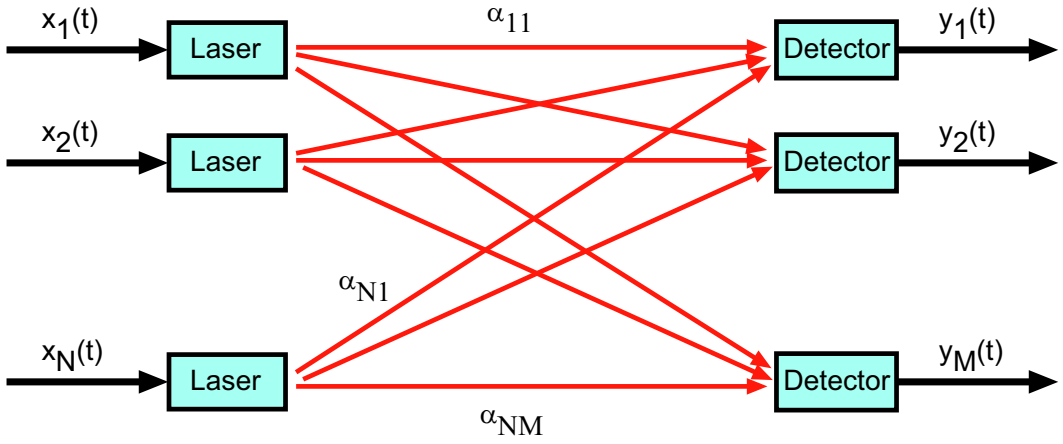

Figure 1-3: Providing the receiver with multiple copies of the transmitted message can improve the reliability of communication. This thesis explores the capacity of and coding for this multiple-input, multiple-output (MIMO) fading channel.

Roughly speaking, reliable communication occurs when not all of the paths in Figure 1-3 are deeply faded. This redundancy is an example of the old adage, "don't put all your eggs in one basket." We will demonstrate that "good" space-time codes for both coherent and direct detection cause the receiver to "see" the sum of path gain powers, i.e., the sum of squared magnitude complex field path gains.

Another theme of this thesis is to develop reliable communication systems that do not depend heavily on the tails of the fading distribution. Although we will argue in Chapter 2 that a lognormal distribution is an appropriate description of atmospheric

fading, experimental results indicate that atmospheric log-amplitude fluctuations are not Gaussian deep into the tails of its distribution [13]; for example, see Figures 6-4 and 6-5 in Chapter 6.

We will try to design communication systems that are insensitive to the tails of the fading distribution. In fact, many of our results only rely on the first and second moments of the fading distribution. Furthermore, several of our results are based on moment-matching approximations. By quantifying when these approximations are valid, we are essentially specifying the operating conditions in which the system design is insensitive to the distribution's tails.

1.4.1 Coherent Detection Receivers

In Section 3.1, we use Monte Carlo averaging to calculate the average capacity of the coherent detection channel, assuming the transmitter and receiver know the path gains. We show that the ergodic capacity is not very sensitive to the fading strength or distribution. We compare this to the average capacity when only the receiver knows the path gains. As with Rayleigh fading channels [22], the benefit of knowing the path gains at the transmitter is negligible for moderate numbers of apertures and transmit power.

In Section 3.2, we present a space-time channel coding technique for overcoming turbulence-induced fading in an atmospheric optical heterodyne communication system that uses multiple transmit and receive apertures. In particular, a design criterion for minimizing the pairwise probability of codeword error in a space-time code is developed from a central limit theorem approximation. This design criterion maximizes the mean-to-standard-deviation ratio of the received energy difference between codewords. It leads to STCs that are a subset of the previously reported STCs for Rayleigh channels, namely those created from orthogonal designs.

Our approach also extends to other fading channels with independent, zero-mean path gains. Consequently, for large numbers of transmit and receive antennas, STCs created from orthogonal designs minimize the pairwise codeword error probability for this larger class of fading channels. We published these space-time coding results in

[27].

1.4.2 Photon-Counting Receivers

In Section 4.1.1, we examine the Shannon capacity of the single-user, multiple-input, multiple-output Poisson channel with peak and average transmit power constraints. The MIMO Poisson channel is a good model for the physical layer of a multiple-aperture optical communication system that operates in the shot-noise-limited regime with known path gains. We derive upper and lower bounds on the capacity that coincide in a number of special cases. The capacity is bounded below by that of the MIMO channel with an additional on-off keying (OOK) transmitter constraint, and it is bounded above by that of parallel, independent, multiple-input, single-output (MISO) channels. We published these MIMO Poisson channel capacity results in [25].

We then consider the ergodic capacity and capacity-versus-outage probability of photon-counting, direct-detection optical communication through the turbulent atmosphere using multiple transmit and receive apertures. We assume shot-noise-limited operation in which detector outputs are doubly-stochastic Poisson processes whose rates are proportional to the sum of the transmitted powers, scaled by lognormal random fades, plus a background noise. With constraints on peak and average power per transmit aperture, we will show that at high signal-to-noise ratio, the ergodic capacity scales as the number of transmit apertures (N) times the number of receive apertures (M), and can be achieved with neither transmitter or receiver knowing the path gains. In the low signal-to-noise ratio regime, ergodic capacity scales as MN^2. In this regime, path-gain knowledge provides minimal capacity improvement when using a moderate number of transmit apertures. Furthermore, in the high and low signal-to-noise ratio regimes, we show that the ergodic capacity of this fading channel equals or exceeds that for a channel with deterministic path gains. In other words, we demonstrate that fading actually increases capacity.

We also develop expressions for the capacity-versus-outage probability in the high and low signal-to-noise ratio regimes by means of a moment-matching approximation

to the distribution of the sum of lognormal random variables. Monte Carlo simulations show that these capacity-versus-outage approximations are quite accurate for moderate numbers of apertures in moderate fading. These ergodic and outage capacity results are submitted for publication [26].

In Section 4.2, we examine space-time coding for photon-counting receivers. We show that a switching space-time code can perform as well as the capacity-achieving repetition spatial code.

1.4.3 Optically-Preamplified Receivers

Theory

In Chapter 5 we examine the use of optical amplifiers to improve communication reliability. We develop lower bounds to the capacity of this channel by constructing discrete-memoryless channel representations. These representations utilize repetition on-off keying (OOK) spatial coding transmitters and linear combining, threshold-decision receivers. We show that equally weighting the detector outputs minimizes the error probability when the average receive power is much greater than -56 dBm (using the nominal parameters of the 1.25 Gbps testbed in Chapter 6). For lower average receive powers, weighting the detector outputs in proportion to their signal-to-noise ratio, i.e., classical maximal-ratio combining, is the best linear combining strategy.

Experiment

We also build a 1.25 Gbps testbed using optical preamplification. We demonstrate the benefits of using two receivers and equal-gain combining with midpoint thresholding. In mild fading, this configuration requires about three decibels less power per receiver to maintain a 10^{-6} bit error rate as compared to a single aperture system. We also measure the distribution of the log-amplitude fluctuations on each receiver, and compare them to their theoretical Gaussian distributions.

1.4.4 Some Intuition: Paul Revere's Dilemma

In this thesis, we will show that equal-gain combining is a capacity-achieving receiver architecture for the photon-counting channel at high signal-to-noise ratio. Similarly, equal-gain combining minimizes the bit error rate for optically-preamplified receivers at high signal-to-noise ratio. The following anecdote captures the intuition behind these results.

Digital, wireless, optical communication is a very old form of communication. In fact, the Sexton Robert Newman used it to notify Paul Revere that the British were coming. By the presence or absence of lamps in the Old North Church, Newman signalled optically one of three messages, or $\log_2(3) \approx 1.6$ bits of information: the British are coming by land, they are coming by sea, or they are not coming at all.

Now suppose that Newman only cared about communicating whether or not the British were coming, but he was occasionally forgetful. If a lamp appears in the tower, then it is certain that the British are coming. But if no lamp appears, it means that either the British are not coming with probability $1 - p$, or that they are coming with probability p, and he simply forgot to light a lamp. Under these circumstances, how does Paul Revere know when to ride? What rule should he use to minimize the probability of making the wrong decision: either riding in vain, or failing to respond to the British invasion?

The decision rule that minimizes the probability of error, is to choose the most probable scenario (British coming or not coming) given the observation (lamp present or not present). If Paul sees a lamp, then he should definitely ride because a lamp indicates that the British are definitely coming. If he does not see a lamp, then by taking no action, there is a probability p of failing to respond to an invasion. On the other hand, if he does ride, there is a probability $1 - p$ that he does so without need. So, if Newman only occasionally forgets, i.e., $p < 1/2$, then he should not ride if he does not see a lamp.

To further complicate matters, suppose that the weather is bad that night, and that visibility is poor. As a result, Paul might not see the lamp when glancing up at

the tower, even if it is there. Now, what should he do? An obvious solution would be to do more than just glance, but to stare up at the tower. If rain obscures the tower for one moment, it might not do so the next. Averaging temporally over the weather conditions reduces the uncertainty that not seeing a lamp is due to poor visibility. In wireless communication systems, this form of redundancy is sometimes called temporal diversity.

But what if Paul needed to know right now, at this moment, whether or not he should ride? He could position other riders so that they each had a different view of the tower, and they could all glance up at the same time. If anyone sees the lamp, then Paul knows for sure that the British are coming, and that he should ride. If each vantage point has a different visibility, then the chances that no one will see the lamp, if it is indeed there, is small. This form of diversity is known as receiver spatial diversity in wireless communication systems.

If the British are coming, then Newman could also place another lamp in a different tower, separated sufficiently in distance, so that the visibility of each tower is most likely different. This redundancy in wireless communications is called transmitter spatial diversity. In fact, it is an on-off keying (OOK) repetition spatial code.

If Paul or any other rider see either lamp, then they should ride. Equivalently, they could add up the number of riders that saw a lamp, and ride if this sum is greater than or equal to one. This strategy is equivalent to an equal-gain combining, threshold-decision receiver in wireless communications. In this case, equal-gain combining with unity threshold, minimizes the probability of making a wrong decision.

How does this anecdote relate to the results of this thesis? For photon-counting receivers operating at high signal-to-noise ratio, an absence of light impinging on the photodetector results in no photon counts with certainty. In other words, if all transmit lasers turn on and off simultaneously, and if any detector sees a photon during a bit interval, then we are certain that all transmitters were on, i.e., the British are coming. Indeed, in Chapter 4 we will show that OOK repetition spatial coding and equal-gain combining with unity threshold detection is a capacity-achieving communication architecture for the atmospheric fading channel at high signal-to-noise ratio.

We observe a similar result for error probability in Chapter 5 for optically-preamplified receivers.

1.5 Notation and Abbreviations

Although we will try to remain consistent with notation throughout the thesis, we will redefine some notation between chapters for clarity. For example, we will always denote the path gain from transmitter n to receiver m as α_{nm}. For coherent detection channels, this path gain is the complex *field* path gain, while it is the real *power* path gain (magnitude squared of field gain) for direct detection channels. Thus, in the background and coherent detection chapters (Chapters 2 and 3), α_{nm} will denote the complex field path gain. For the direct detection chapters (Chapters 4 and 5), however, we will redefine α_{nm} to be the power path gain, instead of using the cumbersome notation $|\alpha_{nm}|^2$. We will similarly do so for the transmitted signal $x_n(t)$. Furthermore, the average number of photons per second is a more convenient measure of power for photon-counting receivers as it corresponds to the rate of the Poisson counting process. On the other hand, measuring power in Watts is more natural for optically-preamplified receivers because of the physical measurements recorded in Chapter 6. Tables 1.1 and 1.2 describe the major notation and abbreviations in this thesis.

Notation	Definition
N	Number of lasers
M	Number of detectors
$x_n(t)$	n-th transmitter symbol/waveform at time t
$y_m(t)$	m-th detector output at time t
α_{nm}	Path gain (field or power) from transmitter n to receiver m
α	Set of path gains $\{\alpha_{11}, \ldots, \alpha_{NM}\}$
χ_{nm}	Atmospheric log-amplitude fluctuation
ϕ_{nm}	Atmospheric phase fluctuation
σ_χ^2	Atmospheric log-amplitude variance (fading strength)
X, $\bar{X}$	Transmitted codewords; channel input
Y	Received signal; channel output
$Q(x)$	Area under upper tail of standard normal density function

Table 1.1: This table displays common symbols used throughout the thesis.

Abbreviation	Definition
ASE	Amplified spontaneous emission
AWGN	Additive, white, Gaussian noise
BER	Bit error rate (per channel use)
BPSK	Binary, phase-shift keying
CLT	Central limit theorem
DMC	Discrete, memoryless channel
EDFA	Erbium-doped, fiber amplifier
MIMO	Multiple-input, multiple-output
MISO	Multiple-input, single-output
OOK	On-off keying
OOK-LB	On-off keying, lower bound
PAM	Pulse amplitude modulation
PB-LB	Photon-bucket, lower bound
PC-UB	Parallel-channel, upper bound
QAM	Quadrature amplitude modulation
SIMO	Single-input, multiple-output
SISO	Single-input, single-output
SNR	Signal-to-noise ratio (quantified differently throughout)
STA	Short-time average
STC	Space-time code
TIA	Transimpedance amplifier

Table 1.2: This table displays common abbreviations used throughout the thesis.

Chapter 2

Background

In this chapter we will give a brief overview of optical communication through the atmosphere. For more complete references see [20] or [46].

2.1 Optical Communications

Conveying digital information via optical frequencies through the atmosphere is one of mankind's oldest forms of communication. For example, fire beacons lit on mountain peaks relayed news of Troy's fall in Aeschylus's play *Agamemnon*, written in 5th century B.C. Also, early naval communication relied heavily on signalling flags and shuttered lamps ([23], pg. 1).

With the advent of the laser, however, came a new era in optical communication systems. Figure 2-1 shows a block diagram of a modern optical communication system. An information source generates bits that the coder uses to modulate the optical field of a laser carrier. The resulting field propagates through a medium such as a fiber optic cable, free space, or the atmosphere. The detector converts the optical signal to an electrical signal, and the decoder tries to infer the transmitted codeword.

The term channel refers to the combined modulation, propagation, and demodulation processes that the transmitted codeword undergoes to reach the decoder. In this thesis, we will consider three atmospheric channels. All three channels will incorporate propagation through the turbulent atmosphere. This propagation causes

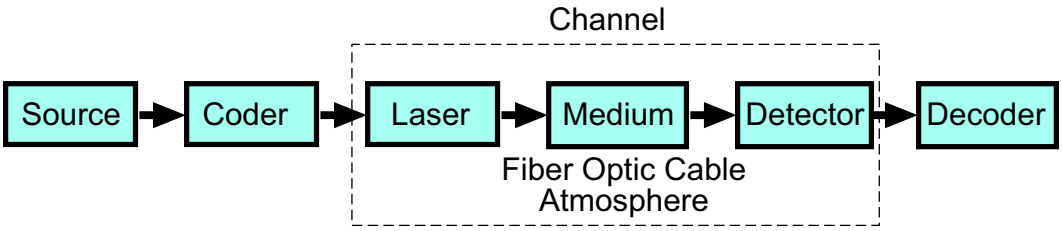

Figure 2-1: A modern optical communication system modulates a laser to convey information through a medium to a receiver. The receiver decodes the detected light and attempts to reconstruct the transmitted message.

fluctuations in the amplitude and phase of the received optical field. The extended Huygens-Fresnel principle models these fluctuations as a complex, lognormal random process [49].

The three channels we will study differ in their transmitter and receiver structures. Two channels use direct-detection (square-law or power) receivers, but vary in their models' idealizations. The first direct-detection channel uses amplitude (intensity) modulation at the transmitter and ideal photon-counting detectors at the receivers. In this case, each detector output is a doubly-stochastic Poisson counting process whose rate is proportional to the short-time average (STA) optical power impinging on the detector.

The second direct-detection channel uses intensity modulation and optically-preamplified, direct-detection receivers. We will model this more realistic channel output as a Gaussian process with a signal-dependent covariance. We will experimentally verify, through hardware implementation, the benefits of receiver diversity for this channel.

The final channel uses quadrature amplitude modulation (QAM) and heterodyne or coherent detection. Conditioned on the lognormal propagation fading, this channel behaves like an additive, white, Gaussian noise (AWGN) channel [20].

The next three subsections explain the modulation, demodulation, and propagation models in more detail.

2.2 Atmospheric Optical Propagation

2.2.1 General Propagation Effects

Light travelling through the atmosphere experiences a number of degradations. Aerosols, molecules[1], and thermal inhomogeneities in the atmosphere cause absorption and scattering of the transmitted optical field. Absorption and scattering also cause attenuation of the transmitted field, resulting in an irretrievable loss of signal energy. Scattering gives rise to beam, angular, multipath, and Doppler spread. Beam and angular spread are illustrated in Figure 2-2. Beam spread is the apparent increase of divergence angle as the beam propagates from transmitter to receiver. Angular spread is the apparent broadening of the angle subtended by the transmitter as seen at the receiver. Angular spread is sometimes called the "shower curtain" effect, referring to the broadening of a light source when viewed through a shower curtain. Multipath spread is the lengthening of the transmitted pulse shape, which could possibly lead to intersymbol interference in digital communications. For line-of-sight propagation in clear weather, multipath spread is at most a few picoseconds and we will neglect it in all that follows [49]. Doppler spread manifests as time-dependent fading. The atmospheric coherence time (reciprocal Doppler spread) is on the order of milliseconds, so that at gigabit per second data rates, the fading is flat over a great many symbols [49].

The refractive-index fluctuations induced by space- and time- varying thermal inhomogeneities are responsible for the twinkling of stars at night, and the shimmering above pavement on a hot summer day. As we shall see, these random amplitude fluctuations (called scintillation) can routinely be on the order of 10 dB, and last for several milliseconds. This receiver power outage leads to bursts of errors in wireless optical communication systems [49]. This thesis primarily focuses on mitigating atmospheric fading using multiple transmit and receive apertures.

[1]Examples of aerosols relevant to optical propagation are water droplets, ice, dust, and organic materials of size comparable to the optical wavelength. Depending on the communication wavelength, molecular constituents that influence optical propagation include H_2O, CO_2, O_3, O_2, and N_2 ([34], pg. 21).

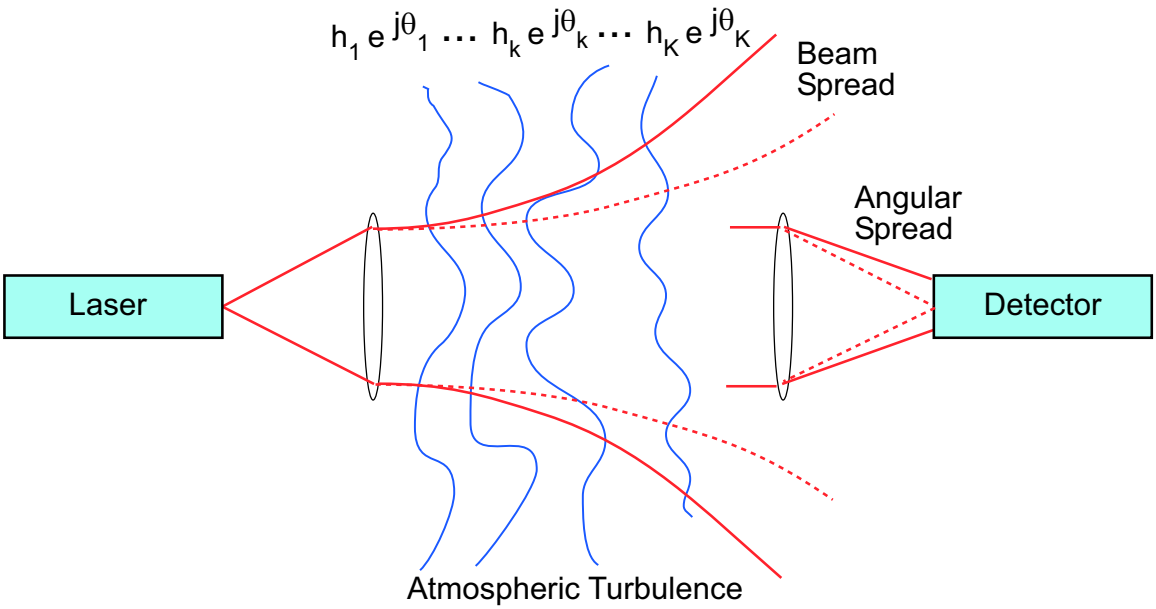

Figure 2-2: Light propagating through thin slabs of clear, turbulent atmosphere experience random amplitude and phase fluctuations. The cumulative effect of these variations is approximately lognormal in distribution due to the central limit theorem. Scattering also causes spreading of the transmitted beam, and an apparent increase in the angular extent of the source at the receiver.

2.2.2 Atmospheric Propagation Models

We will confine our attention to optical communication in clear weather conditions for which absorption is negligible. As noted earlier, an optical signal propagating through the clear atmosphere experiences random amplitude and phase fluctuations as it passes through thermal pockets that vary on the order of 1°K. The refractive index of clear air is temperature dependent. Consequently, as these thermal pockets mix and flow, they create eddies of refractive index turbulence. These eddies result in constructive and destructive interference of the propagating light.

The Thin-Screen Atmospheric Model

A simple, but useful, model of atmospheric propagation through turbulence is shown in Figure 2-2. This model divides the atmosphere into K thin slabs. Light propagating through each slab experiences a random amplitude and phase fluctuation, $h_k e^{j\theta_k}$.

Taken collectively, these variations result in the atmospheric path loss

$$\prod h_k e^{j\theta_k} = \exp\left(\sum \log h_k + j \sum \theta_k\right) \underset{K\to\infty}{\longrightarrow} \exp\left(\chi + j\phi\right). \tag{2.1}$$

The log-amplitude, $\chi = \sum \log h_k$, and phase, $\phi = \sum \theta_k$, are sums of random variables and hence tend to a jointly Gaussian distribution via the central limit theorem. As a result, the light's variation in amplitude and phase as it travels from transmitter to receiver is approximately lognormal in distribution.

We will generally assume that the atmospheric losses are random, but approximately constant during each codeword transmission. Or alternatively, that the codeword length is small compared to the coherence time of the channel, yet large enough that information-theoretic notions such as capacity are meaningful. We justify this assumption by noting that the coherence time for the turbulent atmosphere is on the order of 1 to 10 ms [50], and typical data rates for line-of-sight communication in clear weather are on the order of a gigabit per second. Hence, 1 to 10 million consecutive bits can experience on average similar fading conditions. This quasi-static fading model seems reasonable for our application.

We further assume that the turbulence-induced fading is frequency non-selective, i.e., there is a scalar multiplicative relationship between each transmitter and receiver path. The absence of multipath components at nanosecond durations in line-of-sight optical communication justifies this model [49].

Without loss of generality, we can separate the total atmospheric field attenuation, α, into two components, $\alpha = a_0 a$. The non-random component a_0 is due to the irretrievable power loss from absorption and scattering. The random component $a = \exp[\chi + j\phi]$ results from turbulence-induced fading.

The non-random component of atmospheric attenuation is described by

$$a_0 = e^{-\frac{1}{2}\sigma Z}, \tag{2.2}$$

where Z is the propagation distance in kilometers. The power-attenuation coefficient σ consists of scattering and absorption components. Usually, aerosol and Mie scatter-

ing factors dominate this coefficient [45]. Assuming that aerosol absorption is small compared to the Mie scattering [38], we have that

$$\sigma \approx \frac{3.91}{V}\left(\frac{\lambda}{550\text{ nm}}\right)^{-q(V)}, \tag{2.3}$$

where λ is the optical wavelength in nanometers, V is the visibility in kilometers, and $q(V)$ is the size distribution of the scattering particles given by

$$q(V) = \begin{cases} 1.6 & V > 50 \text{ km} & \text{(High Visibility)} \\ 1.3 & 50 \text{ km} \geq \text{V} > 6 \text{ km} & \text{(Average Visiblity)} \\ 0.585V^{1/3} & 6 \text{ km} > \text{V} & \text{(Low Visibility)} \end{cases}. \tag{2.4}$$

Figure 2-3 plots the atmospheric power-attenuation per kilometer in dB, i.e., $-20\log_{10}(a_0)/Z$, assuming that Mie scattering losses are the dominating factor at $\lambda = 1550$ nm. This figure illustrates that communicating in heavy fog can be extremely difficult due the hundreds of dB/km in attenuation. For more information on communicating through optical scattering channels see [35],([34], pg. 211), or ([20], pg. 291). In this thesis, we will ignore absorption and scattering, and set $a_0 = 1$.

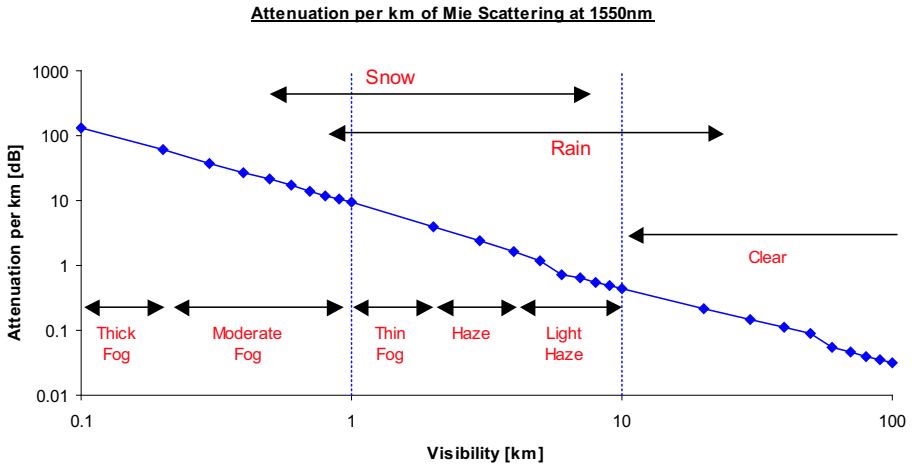

Figure 2-3: Mie scattering can cause an irretrievable loss in optical power. This figure shows the atmospheric power-attenuation per kilometer at the 1550 nm wavelength for a variety of visibility and weather conditions. For example, in clear weather, the visibility is greater than 10 km, and attenuation is less than one decibel per kilometer.

The random variable $a = \exp[\chi + j\phi]$ represents the fading component of the atmospheric attenuation α. We will choose the mean m_χ and variance σ_χ^2 of χ such that the fading does not, on average, attenuate or amplify the optical power, i.e., $E[|\alpha|^2] = a_0^2$. Doing so requires

$$E\left[|a|^2\right] = E\left[e^{2\chi}\right] = M_\chi(2) = 1, \tag{2.5}$$

where $M_\chi(s)$ is the moment-generating function of a Gaussian random variable given by

$$M_\chi(s) = \exp\left(m_\chi s + \frac{1}{2}\sigma_\chi^2 s^2\right). \tag{2.6}$$

Hence, choosing

$$m_\chi = -\sigma_\chi^2, \tag{2.7}$$

makes the average power loss due to atmospheric fading unity [50].

We can further simplify matters by assuming that the phase of the received optical field is uniformly distributed over [0,2π), and is independent of the amplitude fluctuations. This assumption is equivalent to making ϕ statistically independent of χ, with zero mean and a very large variance, i.e., var$[\phi] \gg 2\pi$.

The Extended Huygens-Fresnel Principle

The extended Huygens-Fresnel principle [49] models the diffractive nature of light, and provides the basis for a more thorough treatment of propagation through the turbulent atmosphere. Because the polarization-dependent effects of the atmospheric turbulence are negligible, we can assume that the electric field of the propagating optical signal is linearly polarized, i.e., it is a scalar function of space and time. We will represent this scalar field at a three-dimensional location $\mathbf{r} \in \mathcal{R}^3$, where $\mathcal{R}$ is the set of real numbers, and at a time t using the complex quasi-monochromatic notation

$$U(\mathbf{r}, t) = \Re\left\{u(\mathbf{r}, t)e^{-j2\pi f_c t}\right\}, \tag{2.8}$$

where $u(\mathbf{r},t)$ is a complex-valued scalar function whose temporal bandwidth is much less than the carrier frequency, f_c, and $\Re\{x\}$ denotes the real component of x. We also assume that the complex field $u(\mathbf{r},t)$ is normalized such that $\langle|u(\mathbf{r},t)|^2\rangle$ is the short-time average (STA) power of the optical field per unit area, i.e.,

$$\text{STA Power per Unit Area} \equiv \langle|u(\mathbf{r},t)|^2\rangle \equiv \frac{1}{T_{\text{STA}}}\int_{t-T_{\text{STA}}}^{t}|u(\mathbf{r},\tau)|^2 d\tau, \tag{2.9}$$

where $\langle|u(\mathbf{r},t)|^2\rangle$ has units Watts per meters squared. The integration period $T_{\text{STA}} > 0$ is much greater than the reciprocal of the optical carrier frequency and any radio frequency (RF) sub-carrier frequency differences, but much less than the reciprocal of the information-bearing bandwidth of the signal. This integration period will be made more precise later when we consider the bandwidth limitations of practical detectors.

Example:

Suppose $u(\mathbf{r},t)$ is the scalar optical field of a wavelength-division multiplexed signal with information bearing sub-carrier frequencies $f_1, f_2 \ll f_c$, i.e.,

$$u(\mathbf{r},t) = u_1(\mathbf{r},t)e^{-j2\pi f_1 t} + u_2(\mathbf{r},t)e^{-j2\pi f_2 t}, \tag{2.10}$$

where $u_1(\mathbf{r},t)$ and $u_2(\mathbf{r},t)$ have temporal bandwidths much less than their sub-carrier frequencies. The STA power of the optical field is

$$\begin{aligned}
\langle|u(\mathbf{r},t)|^2\rangle &= \frac{1}{T_{\text{STA}}}\int_{t-T_{\text{STA}}}^{t}\left|u_1(\mathbf{r},\tau)e^{-j2\pi f_1\tau} + u_2(\mathbf{r},t)e^{-j2\pi f_2\tau}\right|^2 d\tau \\
&= \frac{1}{T_{\text{STA}}}\int_{t-T_{\text{STA}}}^{t}\Big(|u_1(\mathbf{r},\tau)|^2 + |u_2(\mathbf{r},\tau)|^2 \\
&\quad +2\Re\left\{u_1(\mathbf{r},\tau)u_2^*(\mathbf{r},\tau)e^{-j2\pi(f_1-f_2)\tau}\right\}\Big)\, d\tau \\
&\approx |u_1(\mathbf{r},t)|^2 + |u_2(\mathbf{r},t)|^2,
\end{aligned}$$

where x^* denotes the complex conjugate of x, and $1/T_{\text{STA}}$ is much greater than the temporal bandwidths of $u_1(\mathbf{r},t)$ and $u_2(\mathbf{r},t)$, but much less than the frequency difference between sub-carriers, $f_1 - f_2$.

Figure 2-4 provides a reference frame for our multi-aperture communication system consisting of N transmit and M receive pupils. The n-th transmit pupil is located at $\sigma_n \in \mathcal{R}^2$ in the $z = 0$ plane. It produces a field that propagates in the $+z$ direction. Using the coordinates $\sigma \in \mathcal{R}^2$ in the $z = 0$ plane, denote this field as

$$s_n(\sigma, t) = \begin{cases} u([\sigma, 0], t) = p_n(\sigma) x_n(t), & \sigma \in \{n\text{-th Tx Pupil}\} \\ 0, & \text{otherwise} \end{cases}, \tag{2.11}$$

where we have assumed that the modulator at each transmitter changes the temporal characteristics of the optical field (amplitude and phase), but not its spatial characteristics. We can then separate the the spatial $p_n(\sigma)$ and temporal $x_n(t)$ field components over the n-th transmit aperture.

The total field in the $z = 0$ plane is the sum of the fields over each transmit aperture,

$$s(\sigma, t) = u([\sigma, 0], t) = \sum_{n=1}^{N} s_n(\sigma, t) = \sum_{n=1}^{N} p_n(\sigma) x_n(t). \tag{2.12}$$

The extended Huygens-Fresnel principle relates the complex field in the $z = 0$ plane to the complex field in the $z = Z$ plane. Denote the field in the receiver plane as

$$r(\rho, t) = u([\rho, Z], t), \tag{2.13}$$

where $\rho \in \mathcal{R}^2$ are the spatial coordinates in the $z = Z$ plane. Under the paraxial assumption that the propagation distance is much greater than the receiver and transmitter pupil diameters, then the field in the receiver ($z = Z$) plane is

$$r(\rho, t) = \int s(\sigma, t - Z/c)\, h_{\mathrm{FS}}(\rho - \sigma)\; e^{\chi(\sigma,\rho) + j\phi(\sigma,\rho)} d\sigma, \tag{2.14}$$

where c is the speed of light, $\{\, \chi(\sigma, \rho), \phi(\sigma, \rho) \,\}$ are jointly Gaussian random processes with known mean and covariance functions, and $h_{\mathrm{FS}}(\rho)$ is the paraxial free-space Green's function,

$$h_{\mathrm{FS}}(\rho) = \frac{1}{j\lambda Z} e^{jk\left(Z + \frac{\|\rho\|^2}{2Z}\right)}, \tag{2.15}$$

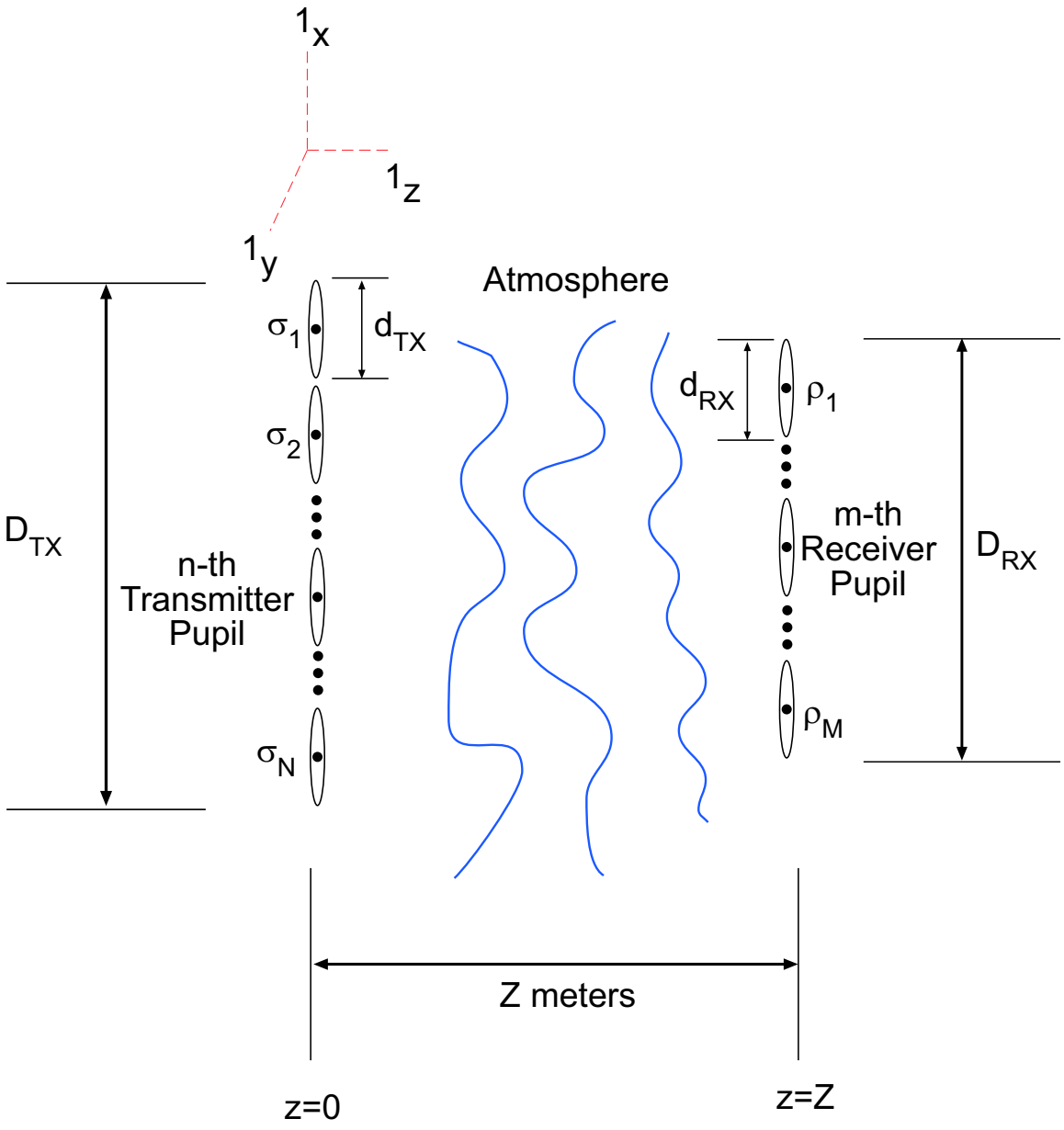

Figure 2-4: The geometry for the extended Huygens-Fresnel principle consists of the n-th transmitter pupil located at σ_n in the $z = 0$ plane and the m-th receiver pupil located at ρ_m in the $z = Z$ plane. In this diagram, 1_x, 1_y, and 1_z are unit vectors marking the origin.

where $\| \rho \|$ denotes the vector magnitude, and $k = 2\pi/\lambda$ is the wavenumber.

Over a horizontal path, the turbulence decorrelates spatially approximately every

$$\rho_0 = \left[1.09 k^2 C_{\mathrm{n}}^2 Z\right]^{-3/5} \qquad \text{[meters]}, \tag{2.16}$$

where C_{n}^2 is the refractive index structure constant that typically lies in the range $C_{\mathrm{n}}^2 \approx 5 \times 10^{-16}$ m$^{-2/3}$ for weak turbulence to $C_{\mathrm{n}}^2 \approx 5 \times 10^{-13}$ m$^{-2/3}$ for strong turbulence. Figure 2-5 plots the coherence length (2.16) as a function of path length for different turbulence strengths at the 1550 nm wavelength.

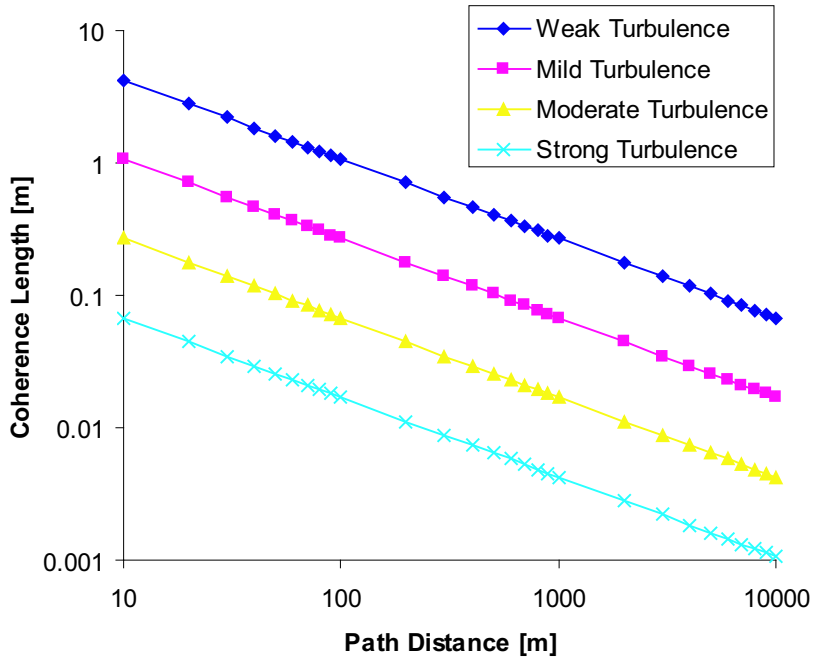

Figure 2-5: This figure plots the atmospheric coherence length (2.16) at the 1550 nm wavelength for different turbulence strengths: weak, $C_{\mathrm{n}}^2 = 5 \times 10^{-16}$ m$^{-2/3}$; mild, $C_{\mathrm{n}}^2 = 5 \times 10^{-15}$ m$^{-2/3}$; moderate, $C_{\mathrm{n}}^2 = 5 \times 10^{-14}$ m$^{-2/3}$; strong, $C_{\mathrm{n}}^2 = 5 \times 10^{-13}$ m$^{-2/3}$

We will assume that the transmit and receive apertures are small compared to this coherence length. As a result, we will approximate the random processes $\{\ \chi(\sigma, \rho),\ \phi(\sigma, \rho)\ \}$ as piecewise constant over each transmit and receive aperture. For σ in the n-th transmit aperture and ρ in the m-th receive aperture, this small

aperture approximation is

$$e^{\chi(\sigma,\rho)+j\phi(\sigma,\rho)} \approx e^{\chi(\sigma_n,\rho_m)+j\phi(\sigma_n,\rho_m)}, \tag{2.17}$$

where σ_n is the center of the n-th transmit pupil in the $z = 0$ plane, and ρ_m is the center of the m-th receive pupil in the $z = Z$ plane.

Define α_{nm} as the atmospheric path loss from the n-th transmit to m-th receive aperture under this small aperture approximation. Again, we can separate the non-random and random losses and write

$$\alpha_{nm} = a_0 a_{nm}, \tag{2.18}$$

where a_0 is the non-random loss in (2.2), and

$$a_{nm} = e^{\chi_{nm}+j\phi_{nm}} \tag{2.19}$$

is the normalized fading loss in the absence of beam spread.

If we further assume that the small apertures are separated by more than the atmospheric coherence length, ρ_0, we can model $\{\ \chi_{nm}, \phi_{nm} \mid 1 \le n \le N,\ 1 \le m \le M\ \}$ as independent Gaussian random variables. As in the thin-screen model, we set the means of the independent, identically-distributed, log-amplitudes $\{\ \chi_{nm} \mid 1 \le n \le N,\ 1 \le m \le M\ \}$ equal to minus their variance so that the atmosphere does not on average attenuate or amplify the transmitted power. For a horizontal path, the log-amplitude variance σ_χ^2 is related to the wavenumber, structure constant, and path length by

$$\sigma_\chi^2 = \min\left\{0.124\ k^{7/6} C_{\mathrm{n}}^2 Z^{11/6}, 0.5\right\}. \tag{2.20}$$

Notice that the log-amplitude variance saturates at 0.5. In this strong fading regime, the validity of our model becomes questionable [49]. Figure 2-6 plots the log-amplitude variance for different turbulence strengths. Again, we will assume uniform phase and make $E[\phi_{nm}] = 0$ with $\mathrm{var}[\phi_{nm}] \gg 2\pi$.

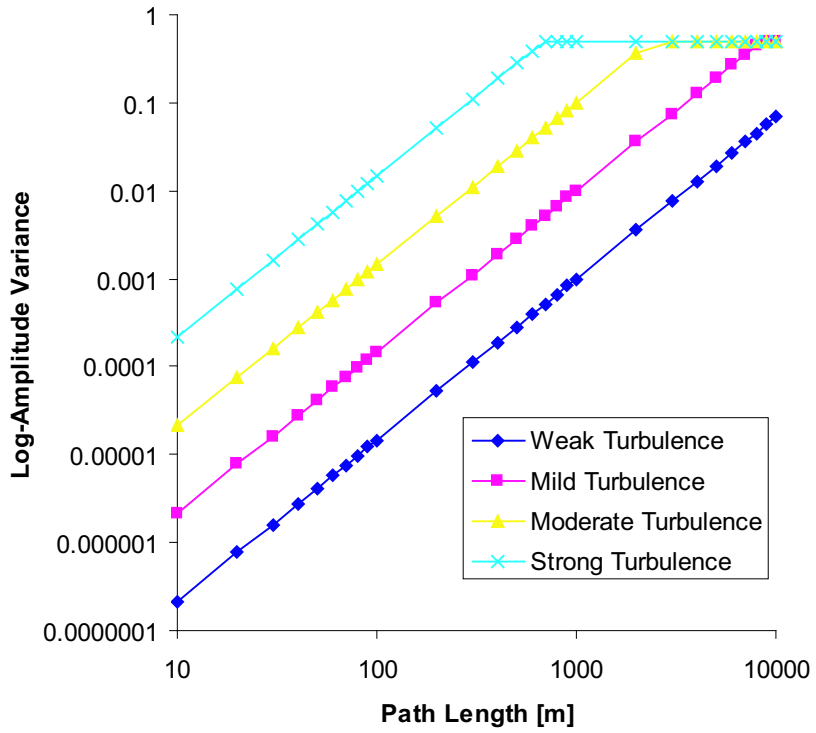

Figure 2-6: This figure plots the log-amplitude variance σ_χ^2 in (2.20) at the 1550 nm wavelength for different turbulence strengths: weak, $C_\mathrm{n}^2 = 5 \times 10^{-16}$ $\mathrm{m}^{-2/3}$; mild, $C_\mathrm{n}^2 = 5 \times 10^{-15}$ $\mathrm{m}^{-2/3}$; moderate, $C_\mathrm{n}^2 = 5 \times 10^{-14}$ $\mathrm{m}^{-2/3}$; strong, $C_\mathrm{n}^2 = 5 \times 10^{-13}$ $\mathrm{m}^{-2/3}$

Under this small aperture approximation, and ignoring the propagation delay Z/c, the received field in (2.14) over the m-th receive pupil becomes

$$r_m(\rho,t) = \begin{cases} r(\rho,t), & \rho \in \{m\text{-th Rx Pupil}\} \\ 0, & \text{otherwise} \end{cases} \tag{2.21}$$

$$= \sum_{n=1}^{N} \alpha_{nm}\beta_{nm}(\rho)\ x_n(t), \tag{2.22}$$

where

$$\beta_{nm}(\rho) = \begin{cases} \int p_n(\sigma)h_{\mathrm{FS}}(\rho-\sigma)\ d\sigma, & \rho \in \{m\text{-th Rx Pupil}\} \\ 0, & \text{otherwise} \end{cases}, \tag{2.23}$$

is the free-space diffraction pattern of the n-th transmitted field on the m-th receiver pupil.

Receiver Optics

Each channel model in this thesis will use slightly different receiver telescope optics as shown in Figures 2-7, 2-8, and 2-9. For all channels, we will assume that all transmitter and receiver pupils are circular and have diameters, d_{TX} and d_{RX}, respectively. Furthermore, we will make the paraxial assumption that the propagation distance Z is much greater than the transmit pupil diameter and receiver size, i.e., $d_{\mathrm{TX}}^2 \ll \lambda Z$ and $D_{\mathrm{RX}}^2 \ll \lambda Z$, where λ is the optical wavelength, and D_{RX} is the receiver size as shown in Figure 2-4.

Our ideal photon detector channel will use the simplest receiver optics shown in Figure 2-7. Here, the received field is focused onto the photo-sensitive portion of a photodetector. If the detector diameter, lens diameter, and lens focal length are chosen appropriately [7], then the short-time average (STA) power over the m-th

detector is approximately the STA power over the m-th receiver pupil, i.e.,

$$\begin{aligned}\int \langle |r_m(\rho,t)|^2 \rangle d\rho &= \sum_{n=1}^{N}\sum_{k=1}^{N} \alpha_{nm}\alpha_{km}^* \langle x_n(t)\ x_k^*(t)\rangle \int \beta_{nm}(\rho)\beta_{km}^*(\rho)d\rho \\ &\approx \sum_{n=1}^{N} |\alpha_{nm}|^2 |x_n(t)|^2 \int |\beta_{nm}(\rho)|^2 d\rho, \end{aligned} \tag{2.24}$$

where

$$\int |\beta_{nm}(\rho)|^2 d\rho \approx \frac{(\pi d_{\mathrm{TX}}^2/4)(\pi d_{\mathrm{RX}}^2/4)}{(\lambda Z)^2}, \tag{2.25}$$

is the diffraction-limited, free-space power loss expressed in terms of the pupil areas. In obtaining (2.24) we have assumed that each transmitter uses a collimated spatial field pattern, so that the free-space power loss is the same for all paths because all the receiver pupils lie within the main lobes of all the transmit field patterns.

In (2.24) we have also assumed that the optical powers from different transmitters add. This power addition assumption is valid if either of the following two conditions hold. First, if the separation between the n-th and k-th, $n \neq k$, transmit pupils satisfies $\parallel \sigma_n - \sigma_k \parallel d_{\mathrm{RX}} \gg \lambda Z$, then

$$\int \beta_{nm}(\rho)\beta_{km}^*(\rho)d\rho \approx 0, \quad n \neq k. \tag{2.26}$$

In other words, the free-space diffraction patterns from different transmit antenna are approximately orthogonal in the m-th receiver pupil because the angles of arrival are resolved by the diffraction limit of the receiver pupil.

The second case in which optical powers add is when each transmit laser operates at a different optical frequency. In this case,

$$\langle x_n(t)\ x_k^*(t)\rangle \approx |x_n(t)|^2\ \delta_{nk}, \tag{2.27}$$

where $\delta_{nk} = 1$ for $n = k$, and zero otherwise. For example, suppose each transmitter uses a different wavelength on the ITU-standard wavelength-division multiplexing grid separated by 50 GHz. If the STA interval T_{STA} is much greater than $1/(50$

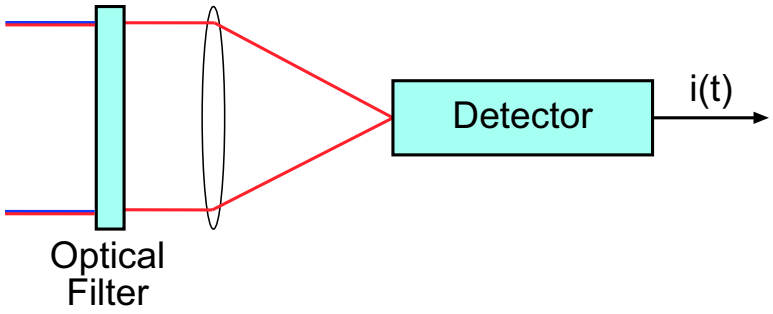

Figure 2-7: The ideal photon detector channel uses a lens to focus the received light onto the photosensitive detector. The optical filter passes the desired signal wavelengths while providing discrimination against extraneous light sources at other wavelengths.

GHz), then the received STA power at the m-th receive aperture is approximately the sum of the individual transmitted STA powers at the m-th receive aperture, as in the previous STA power example.

Under the assumption that optical powers add and that the free-space loss is the same for all transmit-receive aperture pairs, the STA power at the m-th receive aperture factors into temporal, $\sum_{n=1}^{N} |\alpha_{nm}|^2 |x_n(t)|^2$, and fixed spatial components, $\int |\beta_{nm}(\rho)|^2 d\rho$. Because the spatial component is independent of n and m by assumption, we can normalize the temporal component to include this fixed spatial component. In other words, we will assume that $x_n(t)$ is normalized to include the free-space losses. As a result, the STA power at the m-th receiver is

$$\text{STA Power at } m\text{-th Receiver} = \sum_{n=1}^{N} |\alpha_{nm}|^2 |x_n(t)|^2, \tag{2.28}$$

the scaled sum of transmitted powers measured at the receiver.

The optically-preamplified, direct-detection channel model uses the receiver optics shown in Figure 2-8. This receiver couples a single spatial mode of the received field into a single-mode fiber. In other words, if $\psi_m(\rho)$ is the propagating spatial mode of the fiber projected backwards to the receive pupil plane, then the coupled temporal

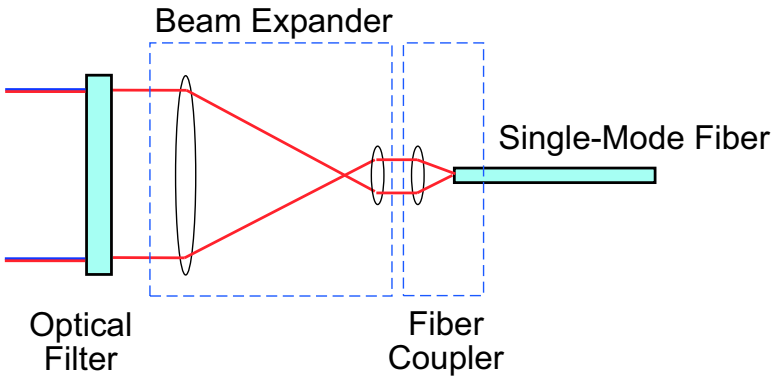

Figure 2-8: The optically-preamplified, direct-detection channel uses a telescope and objective lens to couple a single spatial mode into a single-mode fiber. Again, an optical filter passes the desired signal wavelengths while rejecting extraneous light sources at other wavelengths.

component of the optical field over the m-th aperture is the inner product

$$\begin{aligned} f_m(t) &= \int r_m(\rho,t)\psi_m^*(\rho)d\rho \\ &= \sum_{n=1}^{N} \alpha_{nm}x_n(t) \int \beta_{nm}(\rho)\psi_m^*(\rho)d\rho. \end{aligned} \tag{2.29}$$

For the optically-preamplified, direct-detection channel model we will also assume that the optical powers add. Because we can only couple a single spatial mode into the fiber, we cannot use transmitter separation to create this addition. Instead, we will assume that each transmit laser uses a different frequency, sufficiently separated, such that the optical powers add, i.e., (2.27) prevails. In this case, the received STA power coupled into the m-th receiver is

$$\left\langle \int |f_m(t)\psi_m(\rho)|^2 d\rho \right\rangle \approx \sum_{n=1}^{N} |\alpha_{nm}|^2 |x_n(t)|^2 \left| \int \beta_{nm}(\rho)\psi_m^*(\rho)d\rho \right|^2 \int |\psi_m(\rho)|^2 d\rho. \tag{2.30}$$

We will assume that the transmit pupils are spaced close enough, i.e., $D_{\mathrm{TX}} \ll \lambda Z$, that $\int \beta_{nm}(\rho)\psi_m^*(\rho)d\rho$ becomes independent of n and m. Furthermore, we will assume that $\int |\psi_m(\rho)|^2 d\rho$ is independent of m, e.g. $\psi_m(\rho) = \psi(\rho - \rho_m)$ for some pattern $\psi(\rho)$. As a result, we can again normalize the STA temporal component to include the free-space losses and make (2.28) hold.

In Chapter 3, we will consider the heterodyne receiver shown in Figure 2-9. This heterodyne structure is sometimes called coherent detection because the local oscillator is temporally and spatially in phase with the optical carrier. We will again make the assumption that $D_{\mathrm{TX}} \ll \lambda Z$ so that the free-space losses are independent of n and m.

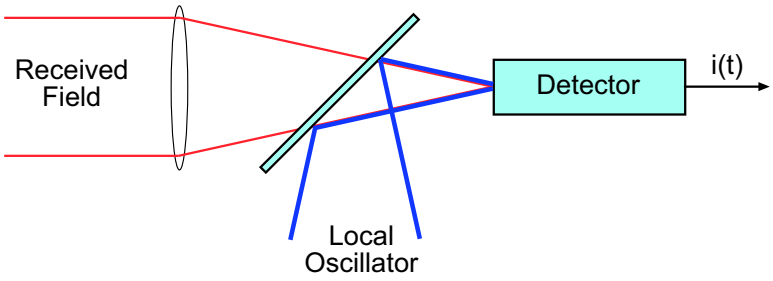

Figure 2-9: Heterodyne receivers mix the received optical field with a spatially and temporally coherent local oscillator, and extract the beat-frequency component in the resulting photocurrent.

2.3 Direct Detection

Direct detection is a term used to describe optical demodulation that responds to the short-time-average (STA) power of the optical field. In this section, we will develop the mathematical models for a single direct-detection receiver. We will later generalize our models to arrays of transmitters and receivers. We begin with a description of an ideal photon detector, then use this to build a practical optical receiver. Finally, we consider the effects of amplifying the optical signal with a low-noise, optical amplifier.

2.3.1 An Ideal Photon-Counting Detector

When a photon of energy exceeding the valence-to-conduction bandgap energy strikes a semiconductor diode junction, it can create an electron and hole pair through photoabsorption at an average rate of

$$\mu(t) = \frac{\eta p(t)}{h f_c} \quad [\text{photons/sec}], \tag{2.31}$$

where $0 \leq \eta \leq 1$ is the detector's quantum efficiency, $p(t)$ is the STA power of the impinging optical field, h is Planck's constant, and f_c is the optical carrier frequency.

When examining this electron (charge carrier) generation process over an extremely short time interval, $[t, t+\Delta)$, the probability of observing an electron is

$$\Pr\{1 \text{ electron in } [t, t+\Delta)\} \approx \mu(t)\Delta, \tag{2.32}$$

and the probability of not observing an electron is

$$\Pr\{0 \text{ electons in } [t, t+\Delta)\} \approx 1 - \mu(t)\Delta. \tag{2.33}$$

Consequently, the probability of observing two or more electrons in this small interval is approximately zero. In addition, the number of electrons observed in non-overlapping time intervals is statistically independent.

These observations imply that we can model the photon-generated electron occurrences as an inhomogeneous Poisson counting process with rate $\mu(t)$, conditioned on knowledge of the underlying STA optical power $p(\cdot) = \{\ p(t) \mid 0 \leq t \leq T\ \}$.

Let $y(t)$ be the number of counts on the interval $[0, t)$, where $y(0) = 0$. The photocurrent $i(t)$ created by the moving electrons is the electrical charge per unit time. Examining the photocurrent over the short interval $[t, t+\Delta)$, we find that the charge in this interval, $i(t)\Delta$, must equal the number of electrons in this interval $\delta y(t) = y(t+\Delta) - y(t)$ multiplied by the charge of each electron e. Hence, for small Δ

$$i(t) = e\frac{\delta y(t)}{\Delta} \xrightarrow[\Delta \to 0]{} e\frac{dy(t)}{dt}. \tag{2.34}$$

Because $y(t)$ is a discontinuous random process, the machinery required for a thorough analysis of these and subsequent limits is beyond the scope of this thesis. For a more precise treatment of point-process calculus see [6] or [58].

In practice, the photocurrent $i(t)$ will never be measured directly. Instead, we will always observe a filtered or integrated version of it due to electrical bandwidth limitations. However, a useful photocurrent representation of an *ideal* photon detector

on the interval [0,t) is

$$i(t) = \begin{cases} \sum_{k=1}^{y(t)} e\delta(t - t_k), & y(t) \geq 1 \\ 0, & y(t) = 0 \end{cases}, \tag{2.35}$$

or equivalently,

$$y(t) = \frac{1}{e} \int_0^t i(\tau) d\tau, \tag{2.36}$$

where $t_1 \leq \ldots \leq t_{y(t)}$ are the ordered arrival times of the electrons, and $\delta(t)$ is the impulse function. This detector model is ideal in the sense that it has infinite electrical bandwidth (we observe impulses), and an absence of optical power, $p(t) \equiv 0$, results in an absence of photocurrent, $i(t) \equiv 0$.

Statistically characterizing practical photodetectors requires modelling bandwidth limitations and mixtures of noise processes. In this thesis, we will only consider the first and second moments of practical photodetector outputs, and argue, through a central limit theorem approximation, that this filtered Poisson process output is a Gaussian process [46], [58].

As a starting point for modelling practical photodetectors, we want to find the mean and auto-covariance function of the photocurrent $i(t)$. We can then propagate these first and second moments through linear, time-invariant systems to model more realistic limitations.

This photocurrent, created by differentiating the counting process $y(t)$, is sometimes called a shot-noise process. Conditioned on the rate process $\mu(\cdot) \equiv \{ \mu(\tau) \mid 0 \leq \tau \leq T \}$, the mean of the photocurrent is ([20], pg. 102)

$$\begin{aligned} E[\, i(t) \mid \mu(\cdot) \,] &= e\frac{\partial}{\partial t} E[y(t) \mid \mu(\cdot)] \\ &= e\frac{\partial}{\partial t} \int_0^t \mu(\tau) d\tau = e\mu(t), \end{aligned} \tag{2.37}$$

where we have used the fact that $y(t)$ is a conditional Poisson random variable with conditional mean $\int_0^t \mu(\tau) d\tau$. Similarly the auto-covariance of $i(t)$ conditioned on the

rate is

$$\begin{aligned} K_{i|\mu}(t,s) &\equiv \mathrm{cov}\,[i(t), i(s) \mid \mu(\cdot)] \\ &= e^2 \frac{\partial^2}{\partial t \partial s}\ \mathrm{cov}\,[y(t), y(s) \mid \mu(\cdot)] \\ &= e^2 \frac{\partial^2}{\partial t \partial s}\ \mathrm{var}[y(\min\{t,s\}) \mid \mu(\cdot)] \\ &= e^2 \frac{\partial^2}{\partial t \partial s}\ \int_0^{\min\{t,s\}} \mu(\tau) d\tau \\ &= e^2 \frac{\partial}{\partial t}\ \mu(s) u(t-s) \\ &= e^2 \mu(t) \delta(t-s), \end{aligned} \tag{2.38}$$

where $u(t)$ is the unit step function, i.e., $u(t) = 1$ for $t \geq 0$, and zero elsewhere. In the above derivation, we have used the fact that conditioned on the rate, $\{\ y(\tau),\ 0 \leq \tau \leq t\ \}$ is an independent increments process, and hence its auto-covariance at two times is its variance evaluated at the minimum of these times.

2.3.2 A Practical Optical Receiver

The second direct-detection channel we consider incorporates more realistic assumptions about photodetectors. In practice, a photodiode, followed by a transimpedance amplifier (TIA), as shown in Figure 2-10, measures the STA optical power $p(t)$. The back-biased photodiode produces a flow of charge carriers, $i_{\mathrm{DET}}(t)$, through photoabsorption. This current flows through a transimpedance amplifier, and produces, assuming no electrical bandwidth limitations, a voltage, $v_{\mathrm{TIA}}(t) = \sqrt{G_{\mathrm{TIA}}} i_{\mathrm{DET}}(t)$, where $\sqrt{G_{\mathrm{TIA}}}$ is the resistor value of the operational amplifier feedback path.

Figure 2-11 shows a model of this practical photodetector. Real photodetectors differ from ideal photon-counting detectors in two significant ways. First, in the absence of an optical field, i.e., $p(t) \equiv 0$, real photodetectors will produce a small current, $i_{\mathrm{D}}(t)$, called the dark current. The dark current results from the random generation of charge carriers from the photodiode junction. We will model this current

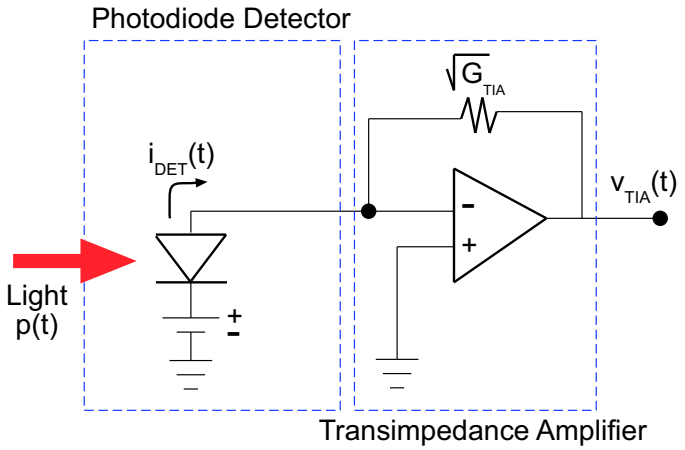

Figure 2-10: A common way to implement a direct-detection receiver is with a photodiode and a transimpedance amplifier. The photodiode produces a current, which the transimpedance amplifier (TIA) converts to a measurable voltage.

as a shot-noise process with constant underlying rate

$$\mu_{\mathrm{D}} = \frac{\eta P_{\mathrm{D}}}{h f_c}, \tag{2.39}$$

where P_{D} is a fictitious dark power. We assume that the dark current is independent of the ideal photon detector current, $i(t)$. In Chapter 4, we will examine the information-theoretic capacity of communication and coding using arrays of photon-counting detectors with dark current.

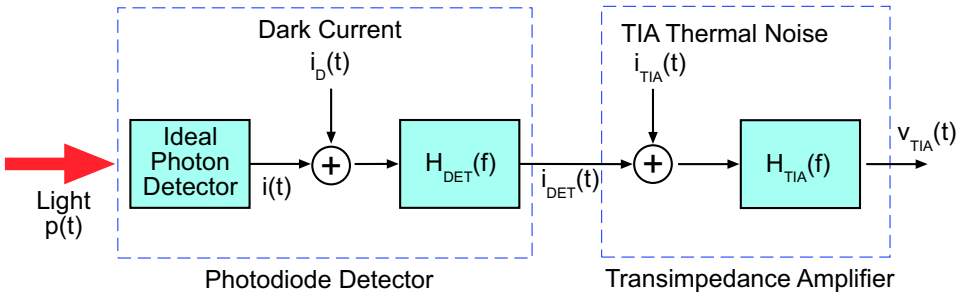

Figure 2-11: This diagram illustrates the noises and bandwidth limitations of a practical direct-detection receiver.

The second difference between real and ideal photodetectors is finite bandwidth. Junction capacitances and electron transit time in the photodiode limit its ability to resolve closely spaced electron emissions. We can model this limitation via a

linear, time-invariant filter $h_{\rm DET}(t) \leftrightarrow H_{\rm DET}(f) = \int h_{\rm DET}(t) \exp(-j2\pi ft)\ dt$, where $\leftrightarrow$ denotes a Fourier transform pair. Because the charge produced by one electron is the integral of the current created by that electron, (2.35) with $y(t) = 1$ implies

$$\int h_{\rm DET}(t)\ dt = H_{\rm DET}(0) = 1. \tag{2.40}$$

The detector photocurrent, $i_{\rm DET}(t)$, is then the filtered sum of the current from the ideal photon detector, $i(t)$, and the dark current, $i_{\rm D}(t)$, i.e.,

$$i_{\rm DET}(t) = \int [i_{\rm D}(\tau) + i(\tau)] h_{\rm DET}(t-\tau) d\tau \tag{2.41}$$

The transimpedance amplifier in Figure 2-11 adds a zero-mean, thermally-induced noise $i_{\rm TIA}(t)$, and filters the resulting current, creating the voltage $v_{\rm TIA}(t)$. The thermal noise is a wide-sense stationary (WSS), real Gaussian process that has a two-sided power spectral density $N_{\rm TIA}/2$ over the amplifier bandwidth. The filter has a response $h_{\rm TIA}(t) \leftrightarrow H_{\rm TIA}(f)$, with $H_{\rm TIA}(0) = \sqrt{G_{\rm TIA}}$, the transimpedance gain of the amplifier.

The voltage output of the transimpedance amplifier is then

$$v_{\rm TIA}(t) = \int [\ i_{\rm D}(\tau) + i(\tau)\ ] h_{\rm DET-TIA}(t-\tau)\ d\tau + \int i_{\rm TIA}(\tau) h_{\rm TIA}(t-\tau)\ d\tau, \tag{2.42}$$

where $h_{\rm DET-TIA}(t)$ is the composite impulse response of the detector and TIA filter,

$$h_{\rm DET-TIA}(t) = \int h_{\rm DET}(\tau) h_{\rm TIA}(t-\tau)\ d\tau. \tag{2.43}$$

Conditioned on knowledge of the ideal photocurrent $i(\cdot) \equiv \{\ i(t) \mid 0 \le t \le T\ \}$, the mean of the amplifier output is

$$E[\ v_{\rm TIA}(t) \mid i(\cdot)\ ] = \int [\ e\mu_{\rm D} + i(\tau)\ ] h_{\rm DET-TIA}(t-\tau)\ d\tau, \tag{2.44}$$

where we have assumed that $i_{\rm D}(t)$ is a shot-noise process with mean $e\mu_{\rm D}$ given by (2.37). Similarly, using the shot-noise auto-covariance in (2.38), the conditional auto-

covariance of the TIA output is

$$\begin{aligned} K_{\mathrm{TIA}|i}(t,s) &\equiv \mathrm{cov}\,[\, v_{\mathrm{TIA}}(t), v_{\mathrm{TIA}}(s) \mid i(\cdot)\,] \\ &= \frac{N_{\mathrm{TIA}}}{2}\int h_{\mathrm{TIA}}(t-\tau)h_{\mathrm{TIA}}(s-\tau)d\tau \\ &\quad + e^2\mu_{\mathrm{D}}\int h_{\mathrm{DET-TIA}}(t-\tau)h_{\mathrm{DET-TIA}}(s-\tau)d\tau, \end{aligned} \tag{2.45}$$

where we have assumed that the processes $\{\, i_{\mathrm{D}}(t),\ i_{\mathrm{TIA}}(t) \mid 0 \le t \le T \,\}$ are uncorrelated.

2.3.3 Optical Noise

In the preceding analysis, we have conditioned expectations on the rate process $\mu(\cdot)$ or the ideal photocurrent $i(\cdot)$. We will now examine the first- and second-order moments of the ideal photocurrent $i(t)$, when the rate process is created as in Figure 2-12. In this figure, $f(t)$ and $z(t)$ represent optical field temporal components, normalized such that the STA optical power is the square of the field magnitude. We interpret $f(t)$ as the signal-induced temporal component of the optical field mode that couples into the single-mode optical fiber after the telescope entrance optics in Figure 2-8, and (2.29).

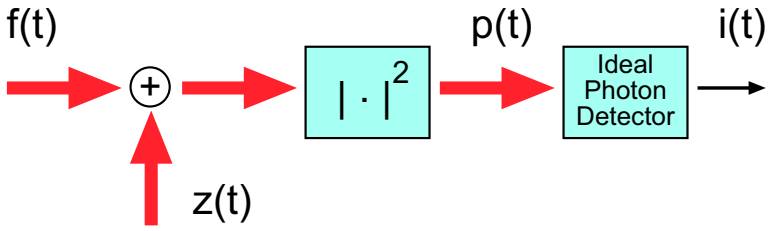

Figure 2-12: This block diagram shows an optical noise adding to the received optical field $f(t)$. For simplicity, we will assume the fields are normalized such that $p(t)$ is the STA optical power.

We interpret $z(t)$ as an optical noise produced by background radiation and/or amplified spontaneous emission (ASE) from an optical preamplifier. We will model this noise as a zero-mean, wide-sense stationary (WSS), complex-valued Gaussian process ([46], Appendix F), uncorrelated with the received field $f(t)$. We assume

that $E[z(t)z(s)] = 0$ and that $E[\Re\{z(t)\}\Im\{z(s)\}] = E[\Re\{z(s)\}\Im\{z(t)\}]$, where $\Im\{z\}$ denotes the imaginary part of z. As a result, the auto-covariance, $E[z(t)z^*(s)] = K_z(t-s)$, is a real function. We will now find the mean and auto-covariance of the photocurrent conditioned on $f(\cdot) \equiv \{\ f(t) \mid 0 \le t \le T\ \}$.

Because the STA optical power is the squared magnitude of the optical field, the rate of the photocurrent is from (2.31)

$$\mu(t) = \frac{\eta}{hf_c}\left(|f(t)|^2 + 2\Re\{f(t)z^*(t)\} + |z(t)|^2\right). \tag{2.46}$$

Using (2.37), iterated expectation, and the observation that conditioning on both $\mu(\cdot)$ and $f(\cdot)$ is equivalent to conditioning only on $\mu(\cdot)$, the conditional photocurrent mean is

$$\begin{aligned} E[\ i(t) \mid f(\cdot)\ ] &= E\{\ E[\ i(t) \mid \mu(\cdot)\ ] \mid f(\cdot)\ \} \\ &= \frac{e\eta}{hf_c}\left[|f(t)|^2 + K_z(0)\right]. \end{aligned} \tag{2.47}$$

The calculation of the auto-covariance is slightly more involved. Using iterated covariances, (2.37), (2.38), (2.46), the Gaussian moment factoring result $\mathrm{cov}[\ |z(t)|^2, |z(s)|^2\ ] = |K_z(t-s)|^2$, and the identity $2\Re\{a\}2\Re\{b\} = 2\Re\{ab + ab^*\}$ gives

$$\begin{aligned} K_{i|f}(t,s) &= \mathrm{cov}[\ i(t), i(s) \mid f(\cdot)\ ] \\ &= E\{\ \mathrm{cov}[\ i(t), i(s) \mid \mu(\cdot)\ ] \mid f(\cdot)\ \} \\ &\quad + \mathrm{cov}\{\ E[\ i(t) \mid \mu(\cdot)\ ], E[\ i(s) \mid \mu(\cdot)\ ] \mid f(\cdot)\ \} \\ &= \frac{e^2\eta}{hf_c}\left[|f(t)|^2 + K_z(0)\right]\delta(t-s) \\ &\quad + \left(\frac{e\eta}{hf_c}\right)^2 \left[2\Re\left\{f(t)f^*(s)\right\}K_z(t-s) + [K_z(t-s)]^2\right]. \end{aligned} \tag{2.48}$$

We can now combine Figures 2-11 and 2-12, and find the mean and auto-covariance of the TIA voltage output, conditioned on the signal-induced received field $f(t)$. Using

iterated expectation, (2.44), and (2.47), the conditional mean is

$$\begin{aligned} E[\, v_{\mathrm{TIA}}(t) \mid f(\cdot)\,] &= E\{\, E[\, v_{\mathrm{TIA}}(t) \mid i(\cdot)\,] \mid f(\cdot)\, \} \\ &= \frac{e\eta}{hf_c} \int \left[P_{\mathrm{D}} + K_z(0) + |f(\tau)|^2\right] h_{\mathrm{DET-TIA}}(t-\tau) d\tau. \end{aligned} \quad (2.49)$$

Using iterated covariances, (2.45), (2.48), (2.44), and assuming that the noise processes are mutually independent, the conditional auto-covariance is

$$\begin{aligned} K_{\mathrm{TIA}|f}(t,s) &= \mathrm{cov}\,[\, v_{\mathrm{TIA}}(t), v_{\mathrm{TIA}}(s) \mid f(\cdot)\,] \\ &= E\{\, \mathrm{cov}\,[\, v_{\mathrm{TIA}}(t), v_{\mathrm{TIA}}(s) \mid i(\cdot)\,] \mid f(\cdot)\, \} \\ &\quad + \mathrm{cov}\,\{\, E[\, v_{\mathrm{TIA}}(t) \mid i(\cdot)\,], E[\, v_{\mathrm{TIA}}(s) \mid i(\cdot)\,] \mid f(\cdot)\, \} \\ &= \frac{N_{\mathrm{TIA}}}{2} \int h_{\mathrm{TIA}}(t-\tau) h_{\mathrm{TIA}}(s-\tau) d\tau \\ &\quad + \frac{e^2\eta}{hf_c} \int \left[P_{\mathrm{D}} + K_z(0) + |f(\tau)|^2\right] h_{\mathrm{DET-TIA}}(t-\tau) h_{\mathrm{DET-TIA}}(s-\tau) d\tau \\ &\quad + \left(\frac{e\eta}{hf_c}\right)^2 \int\int 2\Re\left\{f(\tau) f^*(\sigma)\right\} K_z(\tau-\sigma) h_{\mathrm{DET-TIA}}(t-\tau) h_{\mathrm{DET-TIA}}(s-\sigma) d\tau d\sigma \\ &\quad + \left(\frac{e\eta}{hf_c}\right)^2 \int\int [K_z(\tau-\sigma)]^2 h_{\mathrm{DET-TIA}}(t-\tau) h_{\mathrm{DET-TIA}}(s-\sigma) d\tau d\sigma. \end{aligned} \quad (2.50)$$

The first term in the covariance is from the TIA thermal noise. The second term is sometimes called the "shot noise," and results from the inherent randomness of the underlying conditional Poisson process. The third and fourth terms represent the randomness of the Poisson process rate.

2.3.4 An Optically-Preamplified, Direct-Detection Receiver

In this thesis, the second direct-detection channel that we will study is the combination of intensity modulation[2], atmospheric propagation, and optically-preamplified demodulation, as shown in Figure 2-13.

We will now develop the statistical model for a single transmitter employing intensity modulation and a single optically-preamplified, direct-detection receiver. A

[2]Although we only consider intensity modulation for this channel, the following analysis extends without modification for amplitude and phase modulation.

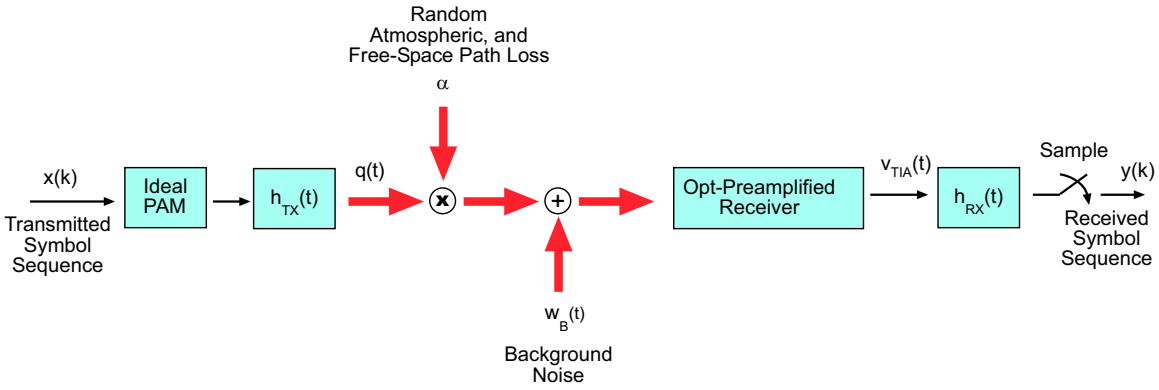

Figure 2-13: Our second direct-detection channel employs intensity, or pulse amplitude modulation (PAM), modulation, atmospheric propagation, and optically-preamplified demodulation.

codeword, consisting of the K real-valued symbols $x(1), \ldots, x(K) \in \mathcal{R}$, modulates a train of impulses separated by T_{SYM} seconds. These impulses pass through a real-valued transmit filter $h_{\mathrm{TX}}(t)$, producing the signal

$$q(t) = \sum_{k=1}^{K} x(k) h_{\mathrm{TX}}(t - kT_{\mathrm{SYM}}). \tag{2.51}$$

The atmosphere multiplies this signal by a random loss α and adds a background noise $w_{\mathrm{B}}(t)$. We model this background noise as a zero-mean, WSS, complex-valued Gaussian process with two-sided power spectral density N_{B} over the optical bandwidth. The optical bandwidth, $-B_{\mathrm{O}}/2 \leq f \leq B_{\mathrm{O}}/2$, is determined by an optical filter, $h_{\mathrm{O}}(t) \leftrightarrow H_{\mathrm{O}}(f)$, inside the optically-preamplified receiver.

Figure 2-14 shows a model of the optically-preamplified receiver. The optical amplifier multiplies the optical power by G_O, and adds a noise due to the amplification of randomly emitted photons. We model this amplified-spontaneous emission (ASE) noise as a WSS, complex-valued Gaussian process with two-sided power spectral density,

$$N_{\mathrm{ASE}} = \mathrm{n_{sp}} h f_c (G_{\mathrm{O}} - 1), \tag{2.52}$$

over the optical filter bandwidth. The spontaneous emission factor $\mathrm{n_{sp}} \geq 1$ depends on the population inversion within the amplifier ([46], pg. 184). This factor is equal

to one for ideal amplifiers. For convenience, we will group the background and ASE optical noises into a single, optically-filtered noise,

$$w_{\mathrm{O}}(t) = \int \left[\sqrt{G_{\mathrm{O}}} w_{\mathrm{B}}(\tau) + w_{\mathrm{ASE}}(\tau)\right] h_{\mathrm{O}}(t-\tau) d\tau. \tag{2.53}$$

The auto-covariance of this zero-mean optical noise is

$$\begin{aligned} K_{\mathrm{O}}(t-s) &\equiv \mathrm{cov}[w_{\mathrm{O}}(t), w_{\mathrm{O}}(s)] \\ &= (G_{\mathrm{O}} N_{\mathrm{B}} + N_{\mathrm{ASE}}) \int |H_{\mathrm{O}}(f)|^2 \, e^{j2\pi f(t-s)} df. \\ &= N_{\mathrm{O}} \int |H_{\mathrm{O}}(f)|^2 \, e^{j2\pi f(t-s)} df, \end{aligned} \tag{2.54}$$

where $N_{\mathrm{O}} \equiv G_{\mathrm{O}} N_{\mathrm{B}} + N_{\mathrm{ASE}}$ is the spectral density of the combined background and ASE optical noise.

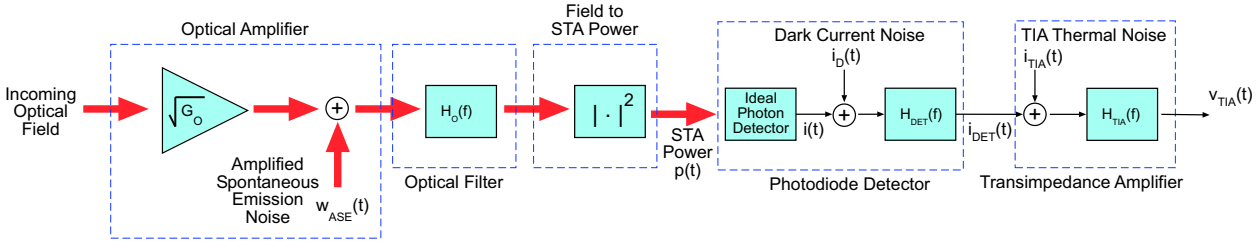

Figure 2-14: An optically preamplified receiver consists of an optical amplifier (e.g., an erbium-doped fiber amplifier (EDFA)), optical filter, photodiode detector, and transimpedance amplifier.

The TIA voltage passes through a real-valued filter $h_{\mathrm{RX}}(t)$ and is then sampled every T_{SYM} seconds. We will make the following approximations to simplify the analysis of this system:

- **Large Optical Bandwidth:** We assume that the optical bandwidth is much greater than the electrical bandwidth so that the optical filter does not appreciably distort the transmitted pulse, and that the optical noise appears white over the signal bandwidth.

- **No Transmitter Intersymbol Interference:** We assume that the transmitter uses rectangular pulses that do not overlap across symbol intervals.

- **No Receiver Intersymbol Interference:** We assume that the receiver output during the k-th symbol interval depends only on events that happened during that interval.

- **Ideal Lowpass Filters:** We assume that the optical filter, $H_{\rm O}(f) = 1$ for $-B_{\rm O}/2 \leq f \leq B_{\rm O}/2$, and zero otherwise. As a result $K_{\rm O}(0) = N_{\rm O}B_{\rm O}$. We also assume that the bandwidth of the detector and the bandwidth of the TIA are much larger than the bandwidth of the receive filter, and that their composite response is an ideal lowpass filter, $H_{\rm E}(f) = 1$ for $-B_{\rm E} \leq f \leq B_{\rm E}$, and zero otherwise.

In other words, we will ignore the complications that arise from intersymbol interference, and assume that the symbol samples are statistically independent. Furthermore, we approximate these sample statistics by those arising from the application of constant transmit power. These assumptions lead to the model in ([46],Appendix F).

Using (2.49) with $f(t) = \sqrt{G_{\rm O}}\ \alpha x(k)$ and $K_z(t) = K_{\rm O}(t)$, the mean of the k-th symbol sample, conditioned on the random path gain α and the transmitted symbol $x(k)$, is

$$E[\ y(k) \mid \alpha, x(k)\ ] = \frac{e\eta}{hf_c}\left[P_{\rm D} + N_{\rm O}B_{\rm O} + G_{\rm O}\ |\alpha|^2|x(k)|^2\right] \tag{2.55}$$

In practice, we can subtract out the non-signal dependent terms, and conclude that the average voltage is proportional to the $|\alpha|^2|x(k)|^2$.

From (2.50), the variance of the k-th symbol sample voltage consists of four components,

$$\text{var}[\ y(k) \mid \alpha, x(k)\ ] \approx \sigma^2_{\rm T} + \sigma^2_{\rm S}(k) + \sigma^2_{\rm SN}(k) + \sigma^2_{\rm NN}. \tag{2.56}$$

The first term is the variance of the TIA thermal noise given by

$$\sigma^2_{\rm T} = N_{\rm TIA}B_{\rm E}. \tag{2.57}$$

The second term is the signal-dependent shot-noise variance

$$\sigma^2_{\rm S}(k) = \frac{e^2\eta}{hf_c}\left[P_{\rm D} + N_{\rm O}B_{\rm O} + G_{\rm O}\ |\alpha|^2|x(k)|^2\right]2B_{\rm E}. \tag{2.58}$$

The third term is the cross term of the signal and optical noise

$$\sigma^2_{\mathrm{SN}}(k) = 2\left(\frac{e\eta}{hf_c}\right)^2 (G_{\mathrm{O}}\ |\alpha|^2|x(k)|^2)(2N_{\mathrm{O}}B_{\mathrm{E}}). \tag{2.59}$$

The last term is the from the optical noise beating against itself

$$\sigma^2_{\mathrm{NN}} = \left(\frac{e\eta}{hf_c}\right)^2 N^2_{\mathrm{O}}(2B_{\mathrm{O}} - B_{\mathrm{E}})B_{\mathrm{E}}. \tag{2.60}$$

This beat-noise term comes from the last term of (2.50), where we calculate the area under the optical noise power spectrum, $K_{\mathrm{O}}(\tau) \leftrightarrow S_{\mathrm{O}}(f)$, convolved with itself, over the electrical bandwidth.

2.4 Coherent Detection

Coherent-detection receivers add an optical local oscillator (LO) to the incoming field as shown in Figure 2-9. This local oscillator field is spatially and temporally coherent with the incoming field. We will only consider a single receiver in this section, and generalize our results to array reception in Chapter 3. Because the field's spatial and temporal components factor apart, we will, without loss of generality, examine only the temporal components as shown in Figure 2-15.

Let $u_{\mathrm{S}}(t)$ denote the complex-valued signal field with temporal bandwidth much less than the intermediate frequency (IF), i.e., $B_{\mathrm{S}} \ll f_{\mathrm{IF}}$. We represent the local oscillator's complex field as a complex-valued sinusoid at frequency f_{IF}, i.e., $u_{\mathrm{LO}}(t) = \sqrt{P_{\mathrm{LO}}}\exp(j2\pi f_{\mathrm{IF}}t)$. The sum of these two signals passes into the practical photodetector of Figure 2-11. We assume that the bandwidth of the detector and transimpedance amplifier is much greater than the intermediate frequency plus signal bandwidth, i.e., $B_{\mathrm{DET}}, B_{\mathrm{TIA}} \gg f_{\mathrm{IF}} + B_{\mathrm{S}}$. In other words, the detector and TIA pass the IF signal, and the IF filter, $h_{\mathrm{IF}}(t) \leftrightarrow H_{\mathrm{IF}}(f)$ determines the receiver spectral characteristics. We model the IF filter as an ideal bandpass filter with bandwidth $2B_{\mathrm{S}}$ centered at f_{IF}, i.e., $H_{\mathrm{IF}}(f) = 1$ for $|f - f_{\mathrm{IF}}| \leq B_{\mathrm{S}}$, and zero otherwise.

We will now find the mean and auto-covariance of the output $v(t)$. The STA

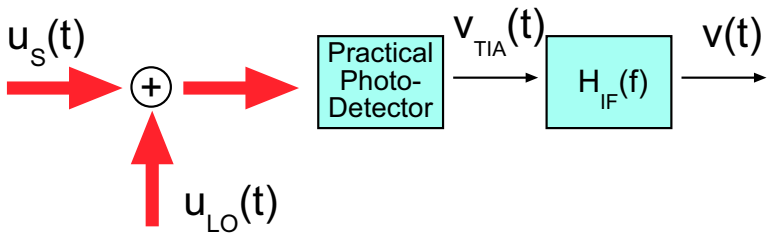

Figure 2-15: A coherent-detection receiver mixes a spatially- and temporally-coherent local oscillator with the incoming field. The STA power cross-term, which is proportional to the received field, then propagates through an ideal bandpass filter for subsequent processing.

power impinging on the detector is

$$|u_{\rm S}(t) + u_{\rm LO}(t)|^2 = |u_{\rm S}(t)|^2 + P_{\rm LO} + 2\Re\left\{\sqrt{P_{\rm LO}}u_{\rm S}(t)e^{-j2\pi f_{\rm IF}t}\right\}. \tag{2.61}$$

Using (2.49) with $f(t) = u_{\rm S}(t) + u_{\rm LO}(t)$ and $z(t) \equiv 0$, we have that the mean conditioned on $u_{\rm S}(\cdot) = \{\ u_{\rm S}(t) \mid 0 \le t \le T\ \}$,

$$\begin{aligned} & E[\ v(t) \mid u_{\rm S}(\cdot)\ ] \\ &= \frac{e\eta}{hf_c}\int\left[P_{\rm D} + |u_{\rm S}(\tau)|^2 + P_{\rm LO} + 2\Re\left\{\sqrt{P_{\rm LO}}u_{\rm S}(\tau)e^{-j2\pi f_{\rm IF}\tau}\right\}\right]h_{\rm IF}(t-\tau)d\tau \\ &= \left(\frac{2e\eta}{hf_c}\sqrt{P_{\rm LO}}\right)\Re\left\{u_{\rm S}(t)e^{-j2\pi f_{\rm IF}t}\right\}. \end{aligned} \tag{2.62}$$

The conditional auto-covariance comes from the first two terms in (2.50),

$$\begin{aligned} & \mathrm{cov}\left[\ v(t), v(s) \mid u_{\rm S}(\cdot)\ \right] \\ &= \frac{e^2\eta}{hf_c}\int\left[\left(\frac{N_{\rm TIA}}{2}\right)\left(\frac{hf_c}{e^2\eta}\right) + P_{\rm D} + |u_{\rm S}(\tau)|^2 + P_{\rm LO}\right. \\ &\qquad \left. +2\Re\left\{\sqrt{P_{\rm LO}}u_{\rm S}(\tau)e^{-j2\pi f_{\rm IF}\tau}\right\}\right]h_{\rm IF}(t-\tau)h_{\rm IF}(s-\tau)d\tau \\ &\approx \left(\frac{2e\eta}{hf_c}\sqrt{P_{\rm LO}}\right)^2\left(\frac{hf_c}{4\eta}\right)\int h_{\rm IF}(t-\tau)h_{\rm IF}(s-\tau)d\tau, \end{aligned} \tag{2.63}$$

where the last line comes from increasing the local oscillator power until the approx-

imation (2.63) holds, i.e., until

$$P_{\mathrm{LO}} \gg \left(\frac{N_{\mathrm{TIA}}}{2}\right)\left(\frac{hf_c}{e^2\eta}\right),\ P_{\mathrm{D}},\ |u_{\mathrm{S}}(t)|^2. \tag{2.64}$$

We can factor out and neglect the first parenthetical term in both (2.62) and (2.63), creating the equivalent model

$$\tilde{y}(t) = \Re\left\{u_{\mathrm{S}}(t)e^{-j2\pi f_{\mathrm{IF}}t}\right\} + \tilde{w}(t), \tag{2.65}$$

where

$$\tilde{y}(t) = v(t)\left(\frac{hf_c}{2e\eta}\right)\left(\frac{1}{\sqrt{P_{\mathrm{LO}}}}\right), \tag{2.66}$$

and $\tilde{w}(t)$ is a zero-mean, white noise with two-sided power spectral density $hf_c/4\eta$ over the IF filter passband. Because the local oscillator shot noise is created from the sum of many filtered photon arrivals, we can approximate its distribution as Gaussian via the central limit theorem [58]. An equivalent baseband model is then

$$y(t) = u_{\mathrm{S}}(t) + w(t), \tag{2.67}$$

where $\tilde{w}(t) = \Re\{w(t)\exp(-j2\pi f_{\mathrm{IF}}t)\}$, and $w(t)$ is a WSS, zero-mean, complex-valued, Gaussian process with auto-covariance $E[w(t)w^*(s)] = (hf_c/\eta)\delta(t-s)$, and $E[w(t)w(s)] = 0$. From (2.67), we conclude that coherent detection produces the familiar additive white Gaussian noise channel.

Chapter 3

Coherent Detection Receivers

This chapter explores the capacity of and coding for the coherent detection channel. We begin with this channel because of its similarity to the traditional microwave Rayleigh fading channel in additive, white Gaussian noise. In microwave communications, reception of multiple, randomly delayed and attenuated versions of the transmitted signal results in fading that is approximately Gaussian due to the central limit theorem. The fade magnitude is Rayleigh, and its phase is uniformly distributed [32]. As described in Section 2.2.2, propagation through atmospheric turbulence can be modelled as the product of random amplitude and phase fluctuations, resulting in a lognormal fade distribution, also from the central limit theorem. In both cases, the fades are normalized such that propagation does not on average attenuate or amplify the transmitted power. Unlike Rayleigh fading, however, atmospheric fading has an additional parameter, the log-amplitude variance σ_χ^2, that determines the severity of fading. We will see that the average capacity is not very sensitive to changes in the log-amplitude variance, or the fading distribution (Rayleigh versus lognormal). We will show that the probability of error, however, is quite sensitive to the log-amplitude variance. We first describe the channel model in more detail.

The detector output in optical heterodyne reception consists of a frequency downshifted version of the incident optical field plus an additive white Gaussian noise [20]. We will assume a quadrature-amplitude modulation (QAM) architecture where a space-time encoder maps a segment of bits from the information source to a code-

word. The N transmit apertures send the codeword

$$X = [x_1(1), x_1(2), \ldots, x_1(T), x_2(1), \ldots, x_2(T), \ldots, x_N(1), \ldots, x_N(T)], \tag{3.1}$$

over T non-overlapping adjacent discrete-time slots. During time-slot t, transmit aperture n sends $x_n(t)$, a symbol from the QAM signal constellation. In this chapter, for convenience, we will use t to denote discrete time, $t \in \{1, \ldots, T\}$.

We model the field path gain from transmit aperture n to receive aperture m as $\alpha_{nm} = \exp(\chi_{nm} + j\phi_{nm})$. Here: χ_{nm}, ϕ_{nm} are independent Gaussian random variables with moments $\mathrm{var}(\chi_{nm}) = \sigma_\chi^2$, $E(\chi_{nm}) = -\sigma_\chi^2$, $\mathrm{var}(\phi_{nm}) \gg 2\pi$, and $E(\phi_{nm}) = 0$, chosen so that $E[|\alpha_{nm}|^2] = 1$. The log-amplitude variance, σ_χ^2, is given in (2.20). We also assume that the spacing between elements of the receiver aperture array is large enough to ensure that the path gains for different (n, m) values are approximately independent.

We will use $w_m(t)$ to denote the additive Gaussian noise for receive aperture m during time slot t; it is a complex-valued, zero-mean, white Gaussian random process with $E[w_m(t_1)w_m^*(t_2)] = N_0\delta_{t_1 t_2}$ and $E[w_m(t_1)w_m(t_2)] = 0$. The noise variance is $N_0 = 2B_{\mathrm{S}}hf_c/\eta$, where B_{S} is the bandwidth of the information-bearing signal, h is Planck's constant, f_c is the optical carrier frequency, and η is the detector quantum efficiency.

Combining the fading and additive noise fluctuations, the signal at receive aperture $m \in \{1, \ldots, M\}$ during time slot $t \in \{1, \ldots, T\}$ is

$$y_m(t) = \sum_{n=1}^{N} \alpha_{nm} x_n(t) + w_m(t). \tag{3.2}$$

Equation (3.2) is our discrete-time, MIMO coherent-detection channel model.

3.1 Capacity

Telatar [63] derived formulas for the average capacity of the single-user MIMO Rayleigh fading Gaussian channel. He used a quasi-static, flat-fading channel model as in (3.2). The only difference was the distribution of the fading statistics. We will calculate the average capacity with and without the transmitter having path gain knowledge. We assume in both cases, that the receiver knows, and optimally uses the path gains to minimize the error probability. For comparison purposes, we will show the relationship between the average Rayleigh and lognormal fading capacities.

3.1.1 Path Gains Known at the Transmitter

We will first consider the case in which both the transmitter and receiver know the path gains. Consider the channel in (3.2), using the matrix notation for a single time-slot,

$$y = \mathcal{A}x + w, \tag{3.3}$$

where $\mathcal{A} = \{ \alpha_{nm}^* \}^\dagger \in \mathcal{C}^{M \times N}$ is the path gain matrix, $y = [y_1(1)^* \cdots y_M(1)^*]^\dagger \in \mathcal{C}^M$ is the output, $x = [x_1(1)^* \cdots x_N(1)^*]^\dagger \in \mathcal{C}^N$ is the input, and $w = [w_1(1)^* \cdots w_M(1)^*]^\dagger \in \mathcal{C}^M$ is the zero-mean complex-valued Gaussian noise with $E[ww^\dagger] = N_0 I$. In the above expressions, I is an identity matrix, and $\dagger$ denotes the conjugate-transpose operator, and $\mathcal{C}$ denotes the set of complex numbers. Following [63], an equivalent channel model using the singular value decomposition $\mathcal{A} = UDV^\dagger$ is

$$\tilde{y} = D\tilde{x} + \tilde{w}, \tag{3.4}$$

where $\tilde{y} = U^\dagger y$, $\tilde{x} = V^\dagger x$, and $\tilde{w} = U^\dagger w$. Here: $U \in \mathcal{C}^{M \times M}$ and $V \in \mathcal{C}^{N \times N}$ are unitary matrices, and $D \in \mathcal{C}^{M \times N}$ is a diagonal matrix containing the non-negative square roots of the eigenvalues of $\mathcal{A}\mathcal{A}^\dagger$. Because the rank L of $\mathcal{A}$ is at most $\min\{N, M\}$, at most $\min\{N, M\}$ coordinates of $\tilde{y}$ are signal dependent. Denoting the non-zero eigenvalues of $\mathcal{A}\mathcal{A}^\dagger$ by $\lambda_1, \ldots \lambda_L$, we can write the signal-dependent coordinates of

(3.4) as

$$\tilde{y}_l = \lambda_l^{1/2}\tilde{x}_l + \tilde{w}_l, \quad 1 \le l \le L \le \min\{N, M\}, \tag{3.5}$$

and ignore the remaining coordinates. Because U is unitary, $\tilde{w}$ has the same distribution as w. The conditional mutual information[1] of this $L \times L$ parallel Gaussian channel is bounded by ([14], pg. 250)

$$\mathcal{I}(\tilde{x}_1, \ldots, \tilde{x}_L; \tilde{y}_1, \ldots, \tilde{y}_L \mid \alpha) \le \sum_{l=1}^{L} \log\left(1 + \frac{\lambda_l P_l}{N_o}\right), \tag{3.6}$$

where $P_l = \mathrm{var}[\tilde{x}_l]$, and $\alpha = \{\alpha_{nm} \mid 1 \le n \le N, 1 \le m \le M\}$. Equality results when $\{\ \tilde{x}_l \mid 1 \le l \le L\ \}$ are independent Gaussian random variables with variances $\{\ P_l \mid 1 \le l \le L\ \}$, respectively. Hence maximizing the mutual information is a power allocation problem.

We can maximize the mutual information in (3.6) subject to the total average power constraint [14]

$$E[x^\dagger x] = E[\tilde{x}^\dagger \tilde{x}] = \sum_{l=1}^{L} P_l \le P, \tag{3.7}$$

using "water-filling." The capacity for a fixed set of path gains $\alpha \equiv \{\alpha_{nm} \mid 1 \le n \le N, 1 \le m \le M\}$ known to both transmitter and receiver is then

$$C(\alpha) = \sum_{l=1}^{L} \left(\log\left[\frac{\lambda_l \nu}{N_0}\right]\right)^+, \tag{3.8}$$

where $(x)^+ \equiv \max\{0, x\}$, and ν is chosen to satisfy the average power constraint

$$P = \sum_{l=1}^{L} \left(\nu - \frac{N_0}{\lambda_l}\right)^+. \tag{3.9}$$

Figure 3-1 and Table 3.1 show the "water-filling" transmitter capacity (3.8) averaged over 5,000 channel realizations for equal numbers of transmit and receive apertures ($N = M$). For comparison purposes, we also show the average channel ca-

[1] Unless otherwise noted, all logarithms are natural logarithms, and information is measured in nats.

$N = M$	Log-Amp. Var. (σ_χ^2) 0.01	0.1	0.35	Rayleigh
1	2.4	2.2	1.9	2.0
2	4.2	4.1	3.8	4.0
3	6.2	6.1	5.7	6.0
4	8.2	8.1	7.7	8.0
5	10.2	10.1	9.7	10.0
6	12.2	12.1	11.7	12.0
7	14.2	14.1	13.7	14.0
8	16.2	16.1	15.7	16.0
9	18.2	18.1	17.7	18.0
10	20.2	20.1	19.7	20.0

Table 3.1: Average capacity [nats/use] with path gain knowledge at the transmitter and receiver is shown as a function of aperture number ($M = N$), fading strength (σ_χ^2), and distribution (lognormal versus Rayleigh). The total transmit average power is constrained to be no greater than $P = 10$ dB.

pacity under Rayleigh fading. From this figure and table, we see that the the average capacity is not that sensitive to changes in the fading strength (σ_χ^2) or distribution (lognormal versus Gaussian).

If communication occurs over many coherence times of an ergodic channel, this average capacity is the maximum rate of reliable communication [66]. If time delays prevent coding over multiple channel realizations, the average capacity is still a figure of merit, but no longer a limit on reliable communication rate. In this case, the probability that the channel can support a given rate, or the complementary cumulative distribution function of the instantaneous capacity, is a better measure. We will explore this capacity versus outage probability further in Chapter 4 for photon-counting receivers.

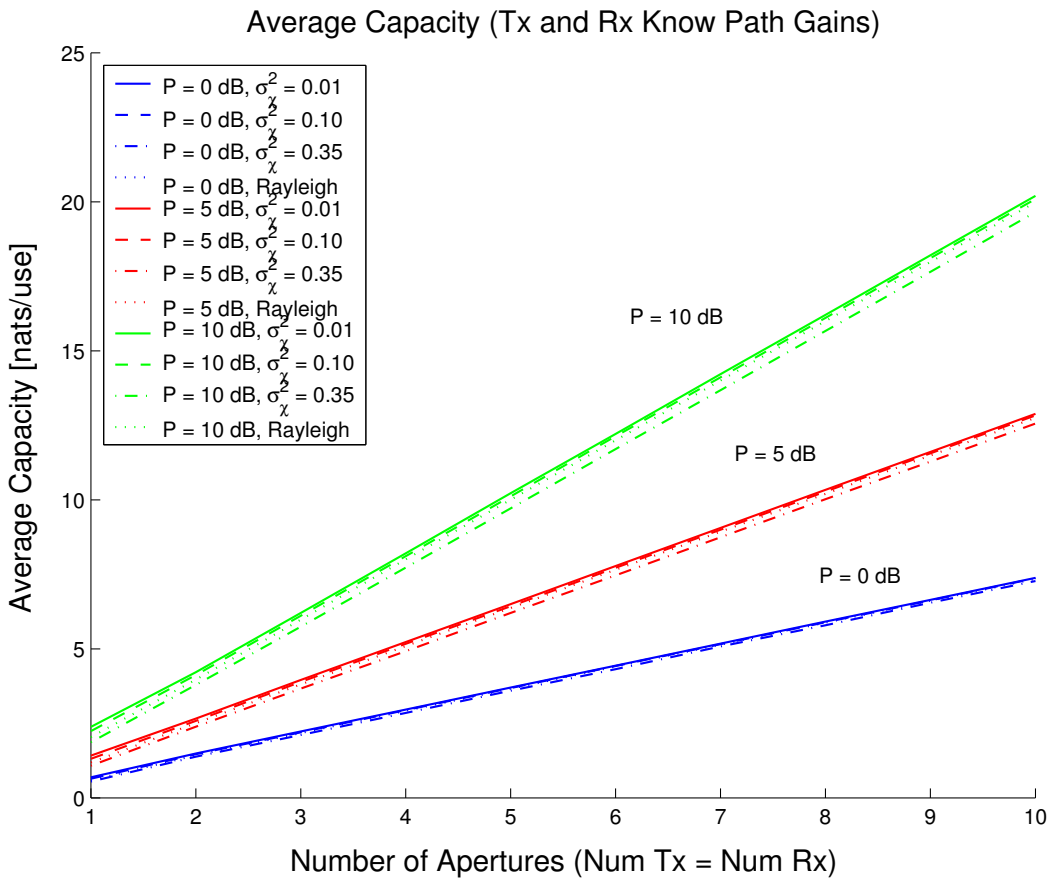

Figure 3-1: The average capacity when both transmitter and receiver know the path gains is plotted versus the number of transmit and receive apertures ($N = M$). We assume a unity receive noise power spectral density, i.e., $N_0 = 1$, and that the Rayleigh fading does not on average attenuate or amplify the transmitted power, i.e., $\text{var}[\Re\{\alpha_{nm}\}] = \text{var}[\Im\{\alpha_{nm}\}] = 1/2$. We constrain the total transmit average power, $\text{E}[x^\dagger x]$, to be no greater than P.

3.1.2 Path Gains Not Known at the Transmitter

The results of the preceding section assumed that the transmitter and receiver had path gain knowledge. If the only the receiver knows the path gains, then a sensible transmitter strategy is to use equal power, P/N, on each aperture. Furthermore, the maximum mutual information (3.6) occurs when the $\tilde{x}_l$ are independent, or equivalently, when $E[\tilde{x}\tilde{x}^\dagger] = V^\dagger E[xx^\dagger]V = (P/N)\, I$. Because the eigenvalues of $I + \gamma \mathcal{A}\mathcal{A}^\dagger$ are $1 + \gamma\lambda_l$, where λ_l are the eigenvalues of $\mathcal{A}\mathcal{A}^\dagger$, and γ is any constant, the resulting instantaneous capacity is [63]

$$C(\alpha) = \log \det \left(I + \frac{1}{N} \frac{\mathrm{P}}{N_0} \mathcal{A}\mathcal{A}^\dagger \right). \tag{3.10}$$

Figure 3-2 and Table 3.2 show the capacity (3.10) averaged over 5,000 channel realizations for equal numbers of transmit and receive apertures ($N = M$). Again, for comparison purposes, we show the Rayleigh fading average capacity. As with the "water-filling" average capacity, we see that the the average capacity is not that sensitive to changes in the fading strength (σ_χ^2) or distribution (lognormal versus Gaussian). Comparing Tables 3.1 and 3.2, we see that knowing the path gains at the transmitter does not appreciably increase the average capacity for moderate numbers of apertures and transmit power. For example, at a total transmit average power of $P = 10$ dB, the capacity increase from knowing the path gains at the transmitter is 6.5% for ten transmit and ten receive apertures in severe fading ($\sigma_\chi^2 = 0.35$). Similar results have been shown for the Rayleigh fading channel [36, 22, 5].

3.2 Coding

Tarokh, *et al.*, in [61] established space-time code (STC) design criteria for Rayleigh and Ricean fading channels. These design criteria specify the pairwise properties of codewords from the STC. In this section, we derive a similar design criterion for the lognormal fading channel based on a central limit theorem approximation [27]. Our criterion leads to STCs created from orthogonal designs, a subset of the previously

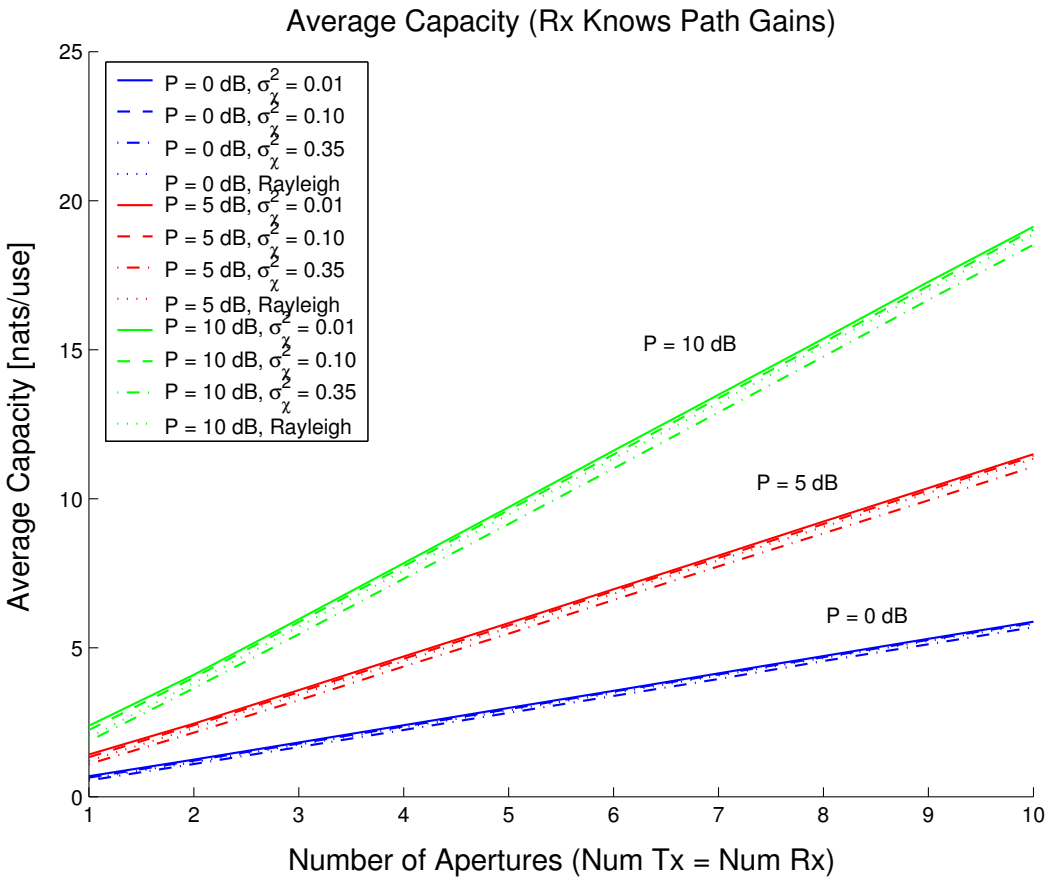

Figure 3-2: The average capacity when only the receiver knows the path gains is plotted versus the number of transmit and receive apertures ($N = M$). We assume a unity receive noise power spectral density, i.e., $N_0 = 1$, and that the Rayleigh fading does not on average attenuate or amplify the transmitted power, i.e., $\mathrm{var}[\Re\{\alpha_{nm}\}] = \mathrm{var}[\Im\{\alpha_{nm}\}] = 1/2$. We constrain the total transmit average power, $\mathrm{E}[x^\dagger x]$, to be no greater than P.

	Log-Amp. Var. (σ_χ^2)			
$N = M$	0.01	0.1	0.35	Rayleigh
1	2.4	2.2	1.9	2.0
2	4.1	4.0	3.6	3.8
3	6.0	5.9	5.5	5.7
4	7.8	7.7	7.3	7.6
5	9.7	9.6	9.2	9.5
6	11.6	11.5	11.0	11.4
7	13.5	13.4	12.9	13.2
8	15.4	15.3	14.8	15.1
9	17.3	17.1	16.7	17.0
10	19.1	19.0	18.5	18.9

Table 3.2: Average capacity [nats/use] with path gain knowledge at the receiver is shown as a function of aperture number ($M = N$), fading strength (σ_χ^2), and distribution (lognormal versus Rayleigh). The total transmit average power is constrained to be no greater than $P = 10$ dB.

reported STCs for Rayleigh channels. Tarokh, *et al.*, in [60] showed that such codes have a decoding algorithm requiring only linear processing at the receiver. We show that these STCs also maximize the mean-to-standard-deviation ratio of the received energy difference between codewords, a result analogous to maximal ratio combining.

Our derivation extends to other fading channels with independent, zero-mean path gains. In other words, we show that for large numbers of transmit and receive antennas, STCs created from orthogonal designs minimize the pairwise codeword error probability regardless of the individual path-gain fading distributions.

3.2.1 Problem Formulation

Given the received samples $\{\, y_m(t) : 1 \le m \le M, 1 \le t \le T \,\}$ and knowledge of the path gains $\alpha = \{\, \alpha_{nm} : 1 \le n \le N, 1 \le m \le M \,\}$, the minimum probability of error receiver chooses the codeword X from the STC $\mathbf{X}$ that minimizes the distance between the received samples and the transmitted codeword seen at the receiver,

$$\sum_{m=1}^{M} \sum_{t=1}^{T} \left| y_m(t) - \sum_{n=1}^{N} \alpha_{nm} x_n(t) \right|^2 . \tag{3.11}$$

The exact probability of error is difficult to calculate for a STC with more than two codewords. An upper bound on this probability of error comes from the union bound

$$P_e \le \sum_{X \in \mathbf{X}} \sum_{\substack{\bar{X} \in \mathbf{X} \\ X \neq \bar{X}}} \Pr(\, X \to \bar{X} \,) \Pr(X), \tag{3.12}$$

where $\Pr(\, X \to \bar{X} \,)$ is the probability of decoding codeword X as codeword $\bar{X}$ in the absence of all other codewords. This sum is usually dominated by the terms of the closest, or minimum distance, codeword pairs. The union bound estimate [18] of the codeword error probability is the sum of pairwise error probabilities of the minimum distance codeword pairs

$$P_e \approx K_{\min} \Pr(\, X \to \bar{X} \,)_{\min}, \tag{3.13}$$

where $K_{\min}$ is the average number of minimum-distance codeword neighbors and

$\Pr(\, X \rightarrow \bar{X}\,)_{\min}$ is the pairwise probability of erroneously decoding a pair of minimum distance codewords.

Given knowledge of the path gains and assuming equally-likely codewords, the pairwise probability of incorrectly decoding transmitted codeword X as codeword $\bar{X}$ is

$$\Pr(\, X \rightarrow \bar{X} \mid \alpha\,) = Q\left(\sqrt{\frac{d^2(X,\bar{X})}{2N_0}}\right), \tag{3.14}$$

where

$$d^2(X,\bar{X}) = \sum_{m=1}^{M}\sum_{t=1}^{T}\left|\sum_{n=1}^{N}\alpha_{nm}[\bar{x}_n(t) - x_n(t)]\right|^2 \tag{3.15}$$

is the squared distance between codewords at the receiver, and $Q(x)$ is the area under the upper tail of the standard normal density function. Averaging over α, the unconditional probability of incorrectly decoding X as $\bar{X}$ is therefore

$$\Pr(\, X \rightarrow \bar{X}\,) = \int \Pr(\, X \rightarrow \bar{X} \mid \alpha\,) p_\alpha(\alpha) d\alpha, \tag{3.16}$$

where $p_\alpha(\alpha)$ is the joint probability density function of the lognormal path gains.

An ultimate objective is to construct a space-time code that minimizes the exact code error probability, P_e. In this section, however, we will focus on minimizing $\Pr(X \rightarrow \bar{X})_{\min}$ in the union bound estimate of this probability. We will demonstrate in Section 3.2.3 that under certain operating conditions, an approximation to this pairwise error probability in the union bound estimate is a good proxy for the code error probability.

3.2.2 Design Criteria

The integral in the unconditional pairwise error probability (3.16) is very difficult to evaluate analytically because of the lognormal density function. We will attempt to simplify its evaluation using a central limit theorem (CLT) approximation.

Rewriting the squared distance between codewords at the receiver (3.15) as

$$d^2(X, \bar{X}) = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{k=1}^{N} \alpha_{nm} \alpha_{km}^* A_{nk}, \tag{3.17}$$

where

$$A_{nk} = \sum_{t=1}^{T} [\bar{x}_n(t) - x_n(t)][\bar{x}_k(t) - x_k(t)]^* \tag{3.18}$$

shows that $d^2(X, \bar{X})$ is the sum of MN^2 complex lognormal random variables.[2] We assume that the transmitter does not have path gain knowledge; therefore, A_{nk} does not depend on the path gains.

Because the coefficients $\{ A_{nk} : 1 \leq n, k \leq N \}$ and the central moments are bounded, we will assume that no single term dominates the sum. Thus, we will use the central limit theorem to approximate its distribution as a Gaussian[3] with mean μ and variance σ^2, truncated on the interval $d^2(X, \bar{X}) \geq 0$. Using this approximation, we can rewrite the unconditional pairwise error probability (3.16) as

$$\Pr(X \to \bar{X}) \approx \int_0^\infty Q\left(\sqrt{\frac{\bar{z}}{2N_0}} \right) p_{\bar{Z}|\bar{Z}\geq 0}(\bar{z} \mid \bar{Z} \geq 0) d\bar{z}, \tag{3.19}$$

where

$$p_{\bar{Z}}(\bar{z}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\bar{z}-\mu)^2}, \tag{3.20}$$

and

$$p_{\bar{Z}|\bar{Z}\geq 0}(\bar{z} \mid \bar{Z} \geq 0) = \frac{p_{\bar{Z}}(\bar{z})}{\Pr(\bar{Z} \geq 0)} \tag{3.21}$$

$$= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(\bar{z}-\mu)^2}}{1 - Q(\mu/\sigma)}, \text{ for } \bar{z} \geq 0. \tag{3.22}$$

Define A as the matrix with A_{nk} as its nk-th element. This matrix characterizes the relationship between a codeword pair of the space-time code. Our goal is to

[2]The scaled multiplication of lognormal random variables is also a lognormal random variable.

[3]When we discuss direct detection receivers in Chapter 4, we will argue that a Gaussian distribution is a poor approximation to the small sums of *real* lognormal random variables. A Gaussian distribution, however, is a good approximation to the sum of *complex* lognormal random variables because of their uniform phases.

derive properties of A that minimize the CLT approximation to the pairwise error probability $\Pr(\, X \rightarrow \bar{X}\,)$. We will do so by expressing the approximation (3.19) as a function of two normalized parameters that measure the fading strength and the signal-to-noise ratio. We then find bounds on the normalized fading strength based on the design matrix A. We demonstrate numerically that the CLT pairwise error probability (3.19) is unimodal as a function of this normalized fading strength. We then show that for large numbers of transmit and receive apertures, minimizing the normalized fading strength, or equivalently choosing A to be a scaled identity matrix, minimizes the CLT approximation to the pairwise probability of error.

Normalized Parameters

Our first step in minimizing the error probability approximation (3.19) is to rewrite $\Pr(\, X \rightarrow \bar{X}\,)$ in terms of normalized parameters. The first normalized parameter measures the strength of the fading. Define the normalized fading strength, η^2, to be the variance-to-mean-squared ratio of the energy difference between the codewords at the receiver, i.e.,

$$\eta^2 \equiv \frac{\mathrm{var}[d^2(X,\bar{X})]}{E[d^2(X,\bar{X})]^2} = \frac{\sigma^2}{\mu^2}. \tag{3.23}$$

This normalized fading strength gauges the STC's ability to mitigate fading. We will show that STCs with small normalized fading strength have good pairwise error performance in the CLT regime. The normalized fading strength η^2, therefore, acts as a "figure of merit" for space-time codes. The second normalized parameter measures the total received signal-to-noise ratio, and is defined as $\rho \equiv \mu/N_0$.

With the change of variables $z = \bar{z}/\mu$, the CLT pairwise error probability approximation (3.19) becomes

$$\Pr(\, X \rightarrow \bar{X}; \rho, \eta^2\,) \approx \int_0^\infty Q\left(\sqrt{\frac{1}{2}\,\rho z}\right) p_{Z|Z\geq 0}(z)dz, \tag{3.24}$$

where

$$p_{Z|Z\geq 0}(z) = \frac{\frac{1}{\sqrt{2\pi\eta^2}}\, e^{-\frac{1}{2\eta^2}(z-1)^2}}{1 - Q(1/\eta)}, \text{ for } z \geq 0, \tag{3.25}$$

and Z is a Gaussian random variable with unit mean and variance η^2.

Mean and Variance Calculations

To approximate $d^2(X, \bar{X})$ as Gaussian, we first must determine its mean μ and variance σ^2. Notice that because $\sigma_\phi^2 \gg 1$, we have that $E[\alpha_{nm}] \approx E[\alpha_{nm}^2] \approx 0$. Also, because $E[\chi_{nm}] = -\sigma_\chi^2$, we find that $E[|\alpha_{nm}|^2] = 1$ and $E[|\alpha_{nm}|^4] = e^{4\sigma_\chi^2}$. The mean of $d^2(X, \bar{X})$ is then

$$\mu \equiv E\left[\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{N}\alpha_{nm}\alpha_{km}^* A_{nk}\right] = M\text{tr}\{A\}, \tag{3.26}$$

where $\text{tr}\{A\} \equiv \sum_{n=1}^{N} A_{nn}$. We define the energy difference between transmitted codewords as

$$E_d \equiv \text{tr}\{A\} = \sum_{n=1}^{N}\sum_{t=1}^{T}|x_n(t) - \bar{x}_n(t)|^2. \tag{3.27}$$

We can then express the total signal-to-noise ratio, ρ, as the sum of signal-to-noise ratios at each receive aperture, i.e., $\rho = ME_d/N_0 = M\text{SNR}$, where $\text{SNR} \equiv E_d/N_0$ is the signal-to-noise ratio at each receive aperture.

The second moment of $d^2(X, \bar{X})$ is

$$E\left[d^4(X, \bar{X})\right] = \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{N}\sum_{\bar{m}=1}^{M}\sum_{\bar{n}=1}^{N}\sum_{\bar{k}=1}^{N} A_{nk}A_{\bar{n}\bar{k}}E\left[\alpha_{nm}\alpha_{km}^*\alpha_{\bar{n}\bar{m}}\alpha_{\bar{k}\bar{m}}^*\right]. \tag{3.28}$$

To evaluate this summation, we split it into two cases. For $m \neq \bar{m}$, we have that

$$\begin{aligned} E\left[\alpha_{nm}\alpha_{km}^*\alpha_{\bar{n}\bar{m}}\alpha_{\bar{k}\bar{m}}^*\right] &= E\left[\alpha_{nm}\alpha_{km}^*\right]E\left[\alpha_{\bar{n}\bar{m}}\alpha_{\bar{k}\bar{m}}^*\right] \\ &= \begin{cases} 1 & \text{if } n = k \text{ and } \bar{n} = \bar{k} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.29}$$

When $m = \bar{m}$, we find that

$$\begin{aligned} E\left[\alpha_{nm}\alpha_{km}^{*}\alpha_{\bar{n}\bar{m}}\alpha_{\bar{k}\bar{m}}^{*}\right] &= E\left[\alpha_{nm}\alpha_{km}^{*}\alpha_{\bar{n}m}\alpha_{\bar{k}m}^{*}\right] \\ &= \begin{cases} e^{4\sigma_{\chi}^{2}} & \text{if } n = k = \bar{n} = \bar{k} \\ 1 & \text{if } n = k \neq \bar{n} = \bar{k} \\ 1 & \text{if } \bar{n} = k \neq n = \bar{k} \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{3.30}$$

From these results it follows that the second moment of $d^2(X, \bar{X})$ is

$$\begin{aligned} & E\left[d^4(X,\bar{X})\right] \\ &= \sum_{m=1}^{M}\left[e^{4\sigma_{\chi}^{2}}\sum_{n=1}^{N}A_{nn}^{2} + \sum_{n=1}^{N}\sum_{\substack{k=1\\k\neq n}}^{N}A_{nn}A_{kk} + \sum_{n=1}^{N}\sum_{\substack{k=1\\k\neq n}}^{N}A_{kn}A_{nk}\right. \\ &\quad \left. + \sum_{\substack{\bar{m}=1\\\bar{m}\neq m}}^{M}\sum_{n=1}^{N}\sum_{k=1}^{N}A_{nn}A_{kk}\right] \\ &= M\left[\left(e^{4\sigma_{\chi}^{2}} - 2\right)\sum_{n=1}^{N}A_{nn}^{2} + \sum_{n=1}^{N}\sum_{k=1}^{N}|A_{nk}|^{2} + M\left(\mathrm{tr}\{A\}\right)^{2}\right]. \end{aligned} \tag{3.31}$$

The variance of the squared codeword difference at the receiver is, therefore,

$$\sigma^2 = \mathrm{var}\left[d^2(X,\bar{X})\right] = M\left[\left(e^{4\sigma_{\chi}^{2}} - 1\right)\sum_{n=1}^{N}A_{nn}^{2} + 2\sum_{n=1}^{N}\sum_{k=1}^{n-1}|A_{nk}|^{2}\right]. \tag{3.32}$$

Notice that although we have assumed that the path gains are lognormally distributed, we have only used the fact that they are independent and identically distributed with zero mean, unit variance, and finite fourth moment. Therefore, our method and results extend to all fading distributions that satisfy these weaker conditions.

Bounds on the Normalized Fading Strength

The mean μ and standard deviation σ of the squared codeword distance are tied to the design matrix A by (3.26) and (3.32), respectively. We will now derive bounds on the normalized fading strength, $\eta^2 = \sigma^2/\mu^2$, expressed in terms of the design matrix A.

A lower bound, obtained via the Cauchy-Schwarz inequality, is

$$\begin{aligned} \eta^2 &= \frac{M\left[\left(e^{4\sigma_\chi^2}-1\right)\sum_{n=1}^{N} A_{nn}^2 + 2\sum_{n=1}^{N}\sum_{k=1}^{n-1}|A_{nk}|^2\right]}{M^2\left(\sum_{n=1}^{N} A_{nn}\right)^2} \\ &\geq \frac{\left(e^{4\sigma_\chi^2}-1\right)\sum_{n=1}^{N} A_{nn}^2 + 2\sum_{n=1}^{N}\sum_{k=1}^{n-1}|A_{nk}|^2}{MN\sum_{n=1}^{N} A_{nn}^2}. \end{aligned} \tag{3.33}$$

Equality holds in (3.33) when $A_{nn} = \beta$, $n = 1, \ldots, N$, for some positive real number β. Furthermore, setting $A_{nk} = 0$ for $n \neq k$ minimizes the numerator in (3.33). Thus we get the bound

$$\eta^2 \geq \frac{e^{4\sigma_\chi^2}-1}{MN}, \tag{3.34}$$

with equality when $A = \beta I$, where I is the $N \times N$ identity matrix. Also, $A_{nn} = \sum_{t=1}^{T} |x_n(t) - \bar{x}_n(t)|^2 = \beta$, $n = 1, \ldots, N$, implies that $\beta = E_d/N$.

Orthogonal designs [60] provide a method to construct STCs that satisfy the design criterion $A = \frac{E_d}{N} I$ and provide easy decoding at the receiver. Therefore, STCs created from orthogonal designs maximize the mean-to-standard-deviation ratio of the received energy difference between codewords. In Section 3.2.3, we will demonstrate the error performance of an orthogonal design space-time code.

We start the upper bound derivation by noticing that A is positive semi-definite [29] because it has an $N \times T$ square-root matrix B with nt-th element $x_n(t) - \bar{x}_n(t)$ such that $A = BB^\dagger$ [61]. Thus, $\lambda_1, \ldots, \lambda_N$, the eigenvalues of A, are non-negative.

For $e^{4\sigma_\chi^2} - 2 \geq 0$, an upper bound on η^2 is found as follows:

$$\begin{aligned} \eta^2 &= \frac{\left(e^{4\sigma_\chi^2} - 2\right) \sum_{n=1}^{N} A_{nn}^2 + \sum_{n=1}^{N} \sum_{k=1}^{N} |A_{nk}|^2}{M \left(\sum_{n=1}^{N} A_{nn}\right)^2} \\ &\leq \frac{\left(e^{4\sigma_\chi^2} - 2\right) \sum_{n=1}^{N} \sum_{k=1}^{N} |A_{nk}|^2 + \sum_{n=1}^{N} \sum_{k=1}^{N} |A_{nk}|^2}{M \left(\sum_{n=1}^{N} A_{nn}\right)^2}, \end{aligned} \tag{3.35}$$

with equality when A is a diagonal matrix. Using $\mathrm{tr}\{A^2\} = \sum_{n=1}^{N} \sum_{k=1}^{N} |A_{nk}|^2 = \sum_{n=1}^{N} \lambda_n^2$, this upper bound becomes

$$\eta^2 \leq \frac{e^{4\sigma_\chi^2} - 1}{M} \sum_{n=1}^{N} \left(\frac{\lambda_n}{\sum_{k=1}^{N} \lambda_k} \right)^2 \leq \frac{e^{4\sigma_\chi^2} - 1}{M}, \tag{3.36}$$

with equality when A is a diagonal matrix of rank one. The last inequality follows from

$$\sum_{n=1}^{N} \left(\frac{\lambda_n}{\sum_{k=1}^{N} \lambda_k} \right)^2 \leq \sum_{n=1}^{N} \frac{\lambda_n}{\sum_{k=1}^{N} \lambda_k} = 1, \tag{3.37}$$

which is met with equality when exactly one of the eigenvalues is non-zero.

For $e^{4\sigma_\chi^2} - 2 < 0$, an upper bound on η^2 is found by suppressing the first term in σ^2:

$$\eta^2 \leq \frac{\sum_{n=1}^{N} \sum_{k=1}^{N} |A_{nk}|^2}{M \left(\sum_{n=1}^{N} A_{nn}\right)^2} = \frac{1}{M} \sum_{n=1}^{N} \left(\frac{\lambda_n}{\sum_{k=1}^{N} \lambda_k} \right)^2 \leq \frac{1}{M}. \tag{3.38}$$

The first inequality in (3.38) is tight when all the diagonal elements of A are zero. The second inequality in (3.38) is tight when A has rank one. There is no non-zero, positive semi-definite matrix that satisfies both of these conditions.

The bounds on the normalized fading strength η^2 are then

$$\frac{e^{4\sigma_\chi^2} - 1}{MN} \leq \eta^2 \leq \frac{\max\{1, e^{4\sigma_\chi^2} - 1\}}{M}. \tag{3.39}$$

The lower bound is achieved when $A = \frac{E_d}{N} I$. If $e^{4\sigma_\chi^2} - 1 \geq 1$, the upper bound is achieved when A has only one non-zero diagonal element. Although the upper bound

is unachievable when $e^{4\sigma_\chi^2} - 1 < 1$, we will soon show that it is asymptotically tight for large numbers of transmit apertures N.

Minimizing the Probability of Codeword Error

To our knowledge, the pairwise codeword error probability in (3.24) does not have a closed-form solution. In this section, we will analyze its asymptotic behavior, and demonstrate numerically that it is unimodal as a function of the fading strength η^2, i.e., that it has only one extremum, a maximum, for a fixed signal-to-noise ratio ρ.

First, we will fix the fading strength η^2 and examine the behavior of $\Pr(\, X \rightarrow \bar{X}; \rho, \eta^2\,)$ as we vary the signal-to-noise ratio ρ. We saw in the previous section that η^2 is closely tied to the STC design matrix; therefore, fixing a value of η^2 is in essence fixing a design matrix.

For small values of total receiver signal-to-noise ratio ρ, the probability of codeword error approaches one-half, i.e.,

$$\lim_{\rho \to 0} \Pr(\, X \rightarrow \bar{X}; \rho, \eta^2\,) = Q(0) = \frac{1}{2}. \tag{3.40}$$

As ρ increases without bound, the Q function becomes sharply peaked at zero, causing $\Pr(\, X \rightarrow \bar{X}; \rho, \eta^2\,)$ to decay as $1/\rho$, viz.,

$$\begin{aligned}
\Pr(\, X \rightarrow \bar{X};\ \text{large } \rho, \eta^2\,) &\approx \int_0^\infty Q\left(\sqrt{\frac{1}{2}\rho z}\right) p_{Z|Z\geq 0}(0) dz \\
&= \frac{e^{-\frac{1}{2\eta^2}}}{\sqrt{2\pi\eta^2}\,[1 - Q(1/\eta)]} \int_0^\infty Q\left(\sqrt{\frac{1}{2}\rho z}\right) dz \\
&= \left(\frac{e^{-\frac{1}{2\eta^2}}}{\sqrt{2\pi\eta^2}\,[1 - Q(1/\eta)]}\right) \frac{1}{\rho},
\end{aligned} \tag{3.41}$$

where $\int_0^\infty Q\left(\sqrt{\frac{1}{2}\rho z}\right) dz = \frac{1}{\rho}$ using integration by parts.

We will now fix the total receiver signal-to-noise ratio, ρ, and determine the probability of codeword error for different values of normalized fading strength, η^2, or equivalently, for different design matrices. As η^2 approaches zero, the Gaussian prob-

ability density function in (3.24) becomes sharply peaked around the mean value at $z = 1$. This sampling-like behavior results in

$$\lim_{\eta^2 \to 0} \Pr(\, X \to \bar{X}; \rho, \eta^2 \,) = Q\left(\sqrt{\frac{\rho}{2}}\right). \tag{3.42}$$

Furthermore, for any fixed value of ρ,

$$\lim_{\eta^2 \to \infty} \Pr(\, X \to \bar{X}; \rho, \eta^2 \,) = 0, \tag{3.43}$$

because the Gaussian density approaches zero for large values of η^2.

The behavior of (3.24) for intermediate values of η^2 is more difficult to evaluate analytically. We will, therefore, make the following conjecture as supported by numerical evaluations of $\Pr(\, X \to \bar{X}; \rho, \eta^2 \,)$:

Conjecture: For $0 < \eta^2 < \infty$, $\Pr(\, X \to \bar{X}; \rho, \eta^2 \,)$ has only one extremum, a maximum, for a given value of ρ. Plots of the pairwise error (3.24) for different values of ρ are shown in Figure 3-3 to support this conjecture. In these plots, we used trapezoidal integration to evaluate the integral in the pairwise error (3.24).

Assuming that $\Pr(\, X \to \bar{X}; \rho, \eta^2 \,)$ is unimodal in η^2, its minimum must occur on the boundary of the allowable range for η^2 in (3.39). In other words, if

$$\Pr\left(X \to \bar{X}; M\,\mathrm{SNR}, \frac{e^{4\sigma_\chi^2} - 1}{MN}\right) < \Pr\left(X \to \bar{X}; M\,\mathrm{SNR}, \frac{\max\{1, e^{4\sigma_\chi^2} - 1\}}{M}\right), \tag{3.44}$$

then the optimal design criterion, in terms of minimizing the CLT pairwise probability of codeword error, is $A = \frac{E_d}{N} I$, because this design matrix meets the lower bound on η^2 with equality. When (3.44) does not hold, and $e^{4\sigma_\chi^2} - 1 \geq 1$, then the optimal design criterion is to choose A to be all zero except for a single non-zero diagonal element. This design matrix, however, violates the CLT assumption that no single term dominates the summation in (3.17). Figure 3-4 shows the bounds on η^2 and the probability of codeword error curve.

Because of its relationship to the pairwise error probability, the normalized fading

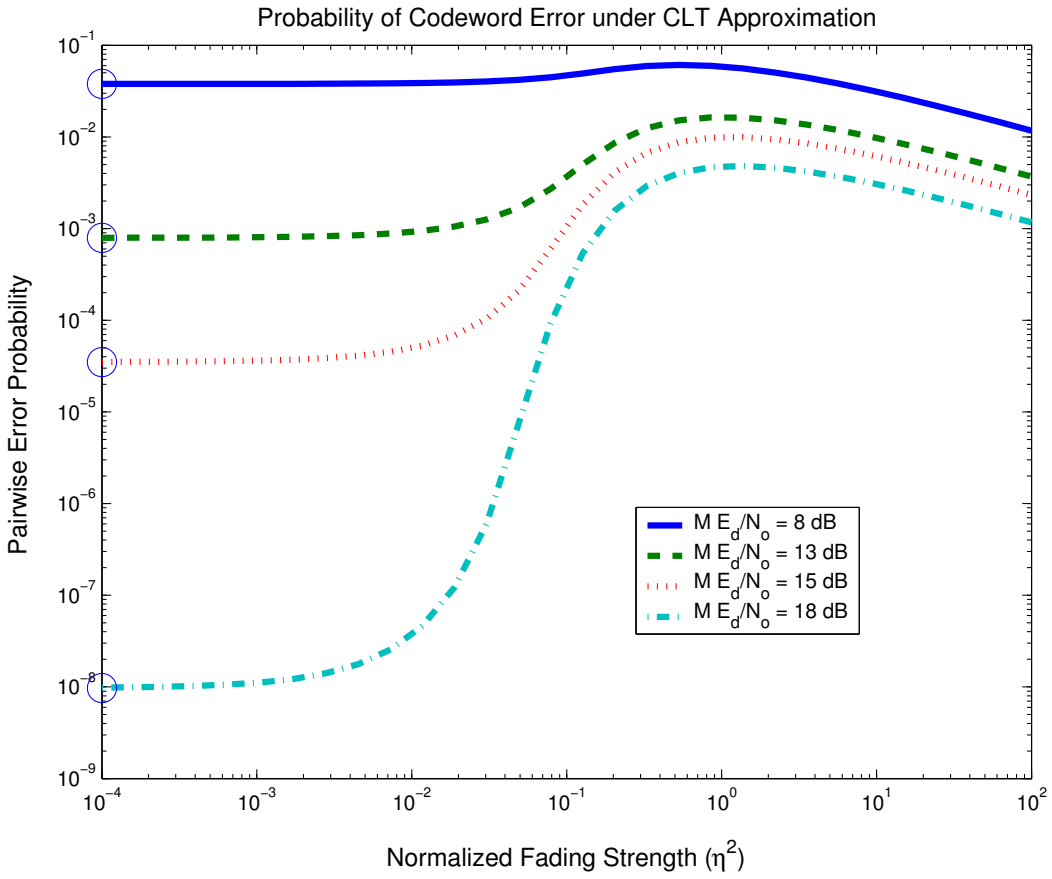

Figure 3-3: The probability of pairwise codeword error, $\Pr(X \to \bar{X}; \rho, \eta^2)$, as a function of the normalized fading strength η^2 for total signal-to-noise ratio $\rho = ME_d/N_0 =$ 8, 13, 15, and 18 dB. The limits as η^2 approaches zero (3.42) are shown as circles.

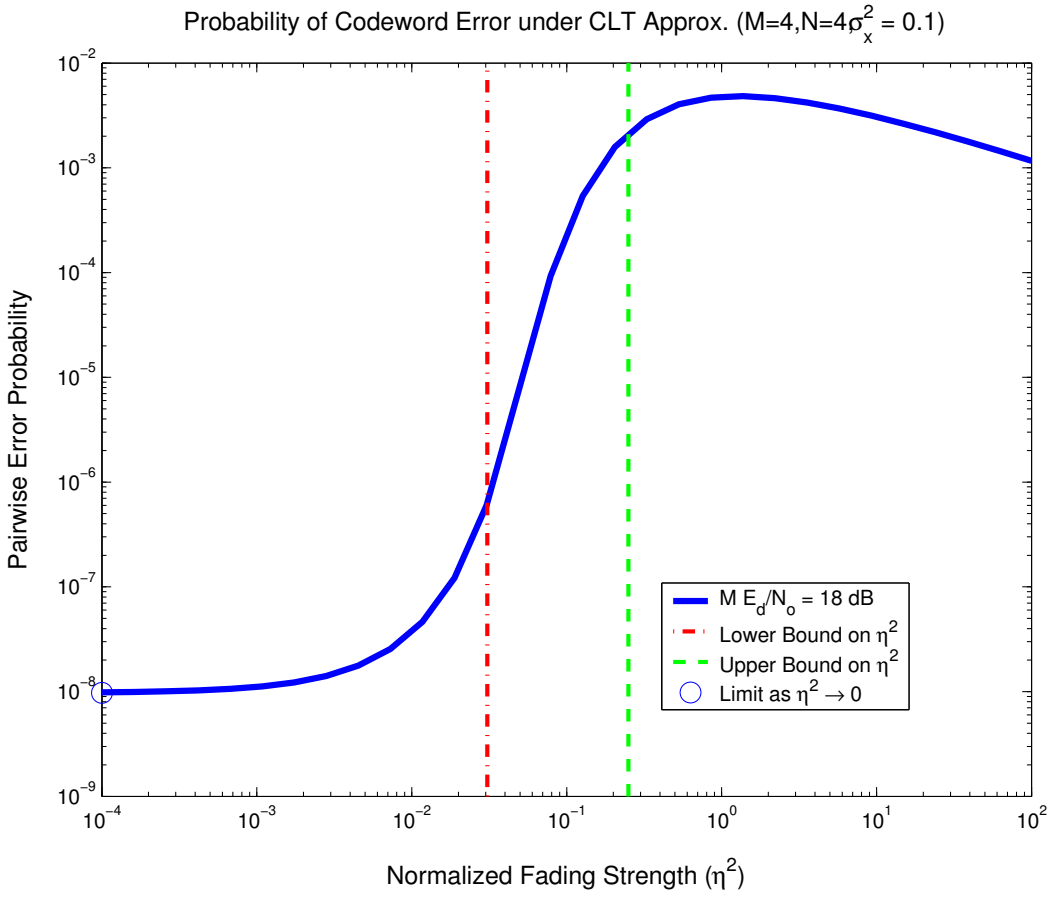

Figure 3-4: The CLT probability of pairwise codeword error, $\Pr(\, X \rightarrow \bar{X}; \rho, \eta^2\,)$, is plotted as a function of the normalized fading strength η^2 for $\rho = ME_d/N_0 = 18$ dB. The smallest achievable error probability occurs when $\eta^2 = \frac{e^{4\sigma_X^2}-1}{MN} \approx 3 \times 10^{-2}$, or equivalently, when $A = \frac{E_d}{N} I$.

strength η^2 is a figure of merit for space-time codes. We have demonstrated that space-time codes created from orthogonal designs minimize η^2, with

$$\eta^2 = \frac{e^{4\sigma_\chi^2} - 1}{MN}, \tag{3.45}$$

and, therefore, also minimize the pairwise error probability under the CLT approximation when (3.44) holds. When deriving the upper bound on the normalized fading strength η^2, however, we noted that when the design matrix A has only a few non-zero entries, the CLT approximation might not be valid for small numbers of receive apertures M. In this case, η^2 might not a good indicator of error probability.

Although the upper bound, $\eta^2 < 1/M$ for $e^{4\sigma_\chi^2} - 1 < 1$, cannot be met with equality, we can create a space-time code that matches it asymptotically for large numbers of transmit apertures N. Consider a repetition spatial code, where in each time-slot every transmitter sends the same symbol, i.e., $x_n(t) = x(t)$ for $1 \leq n \leq N$. In this case, elements of the design matrix A are $A_{nk} = E_d/N$ for all $1 \leq n \leq N$ and $1 \leq k \leq N$. The rank of the design matrix A is one, and according to the Tarokh's rank and determinant criterion [61], this space-time code will have a small diversity and coding advantage. We show that this small advantage is also reflected in the normalized fading strength η^2. For this repetition spatial code, the normalized fading strength is

$$\eta^2 = \frac{1}{M}\left(1 - \frac{1}{N}\right) + \frac{e^{4\sigma_\chi^2} - 1}{MN}. \tag{3.46}$$

We see that as we increase the number of transmit apertures N, the normalized fading strength approaches the upper bound $1/M$ for $e^{4\sigma_\chi^2} - 1 < 1$. Because A is a full matrix, the CLT approximation is valid for large numbers of transmit apertures, and this repetition spatial code is among the worst of space-time codes in terms of pairwise error probability.

For example, assuming $\rho = 18dB$, Figure 3-4 shows that the pairwise error probability for a repetition spatial STC ($\eta^2 \approx 2 \times 10^{-1}$) is approximately 2×10^{-3}, while that of an orthogonal design STC ($\eta^2 \approx 3 \times 10^{-2}$) is roughly 6×10^{-6}. The normalized fading strength of the repetition space-time code is very close to the upper bound

$1/M = 0.25$, and hence its poor performance. For a fixed total receiver signal-to-noise ratio $\rho = ME_d/N_0$, however, increasing the number of receive apertures M decreases the upper bound, until eventually the repetition spatial STC performs as well as the orthogonal design STC. In this case, transmitter diversity is useless. Alternatively, had we increased the number of transmitters, the smallest possible pairwise error probability for an orthogonal design is approximately 10^{-8} for $\rho = ME_d/N_0 = 18$ dB. We will examine this infinite diversity limit further in Section 3.2.3.

The number of transmit (N) and receive (M) apertures must be large in order for the energy difference between codewords, $d^2(X, \bar{X})$, to be approximately Gaussian in the central limit theorem regime. For a fixed total signal-to-noise ratio $\rho = ME_d/N_0$, increasing M and N will cause the bounds on the normalized fading strength η^2 given in (3.39) to decrease, until (3.44) eventually holds. We have also observed through numerical evaluation that the value of η^2 that maximizes $\Pr(\, X \to \bar{X}; \rho, \eta^2\,)$, i.e., the mode, increases with increasing ρ, see Figure 3-3. For a given value of signal-to-noise ratio E_d/N_0, increasing the number of receive apertures M, increases ρ, which in turn increases the mode of the pairwise error probability as a function of the normalized fading strength η^2. Increasing the number of receive apertures, therefore, has greater influence on making (3.44) hold, because it not only decreases the upper bound, but also increases the mode.

The values of M and N (rounded to the next greatest integers) that make (3.44) hold with equality are plotted in Figures 3-5 through 3-7 for different values of SNR and fading environments. For a given SNR, these plots show the smallest number of transmit and receive apertures required for $A = \frac{E_d}{N} I$ to be the optimal design matrix. From these plots, we conclude that in the central limit theorem regime (large values of M and N), $A = \frac{E_d}{N} I$ is the optimal design matrix in terms of pairwise error probability. As previously noted, we observe that increasing the number of receive apertures makes (3.44) hold more quickly than increasing the number of transmit apertures.

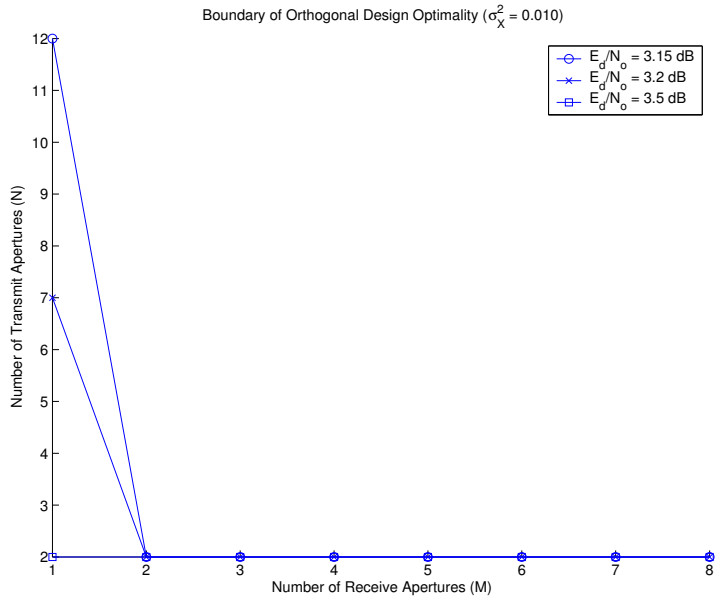

Figure 3-5: The smallest values of M and N such that (3.44) holds in mild fading ($\sigma_\chi^2 = 0.01$). In other words, orthogonal designs are optimal in the CLT regime for aperture numbers greater than these threshold values.

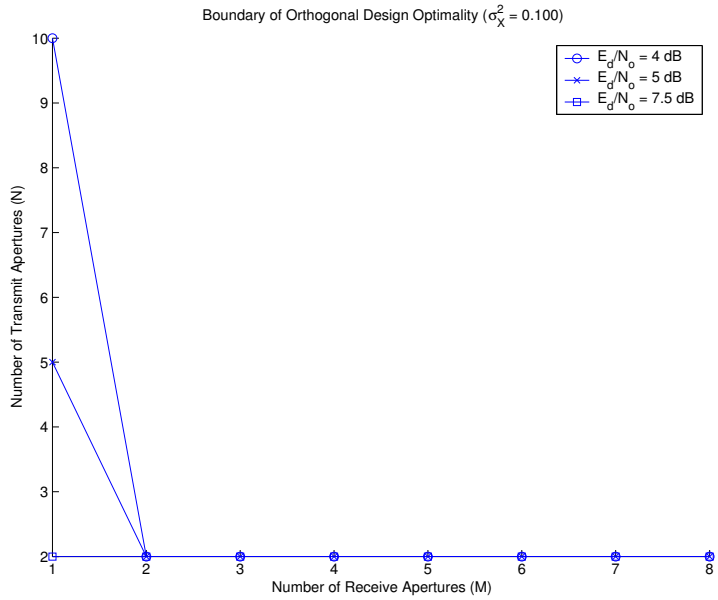

Figure 3-6: The smallest values of M and N such that (3.44) holds in moderate fading ($\sigma_\chi^2 = 0.1$).

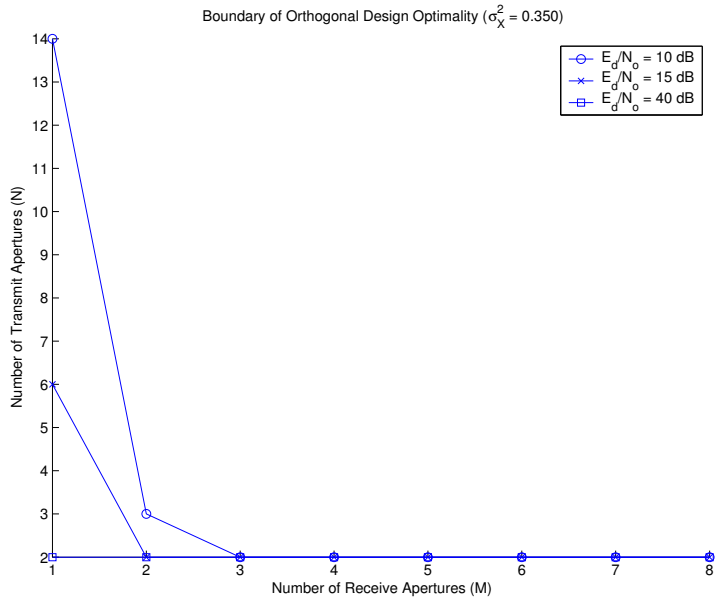

Figure 3-7: The smallest values of M and N such that (3.44) holds in severe fading ($\sigma_\chi^2 = 0.35$).

3.2.3 Performance

In this section we address the validity of the central limit theorem approximation and the performance of STCs on lognormal channels.

Performance Bounds for Orthogonal Design STCs

We will now derive the pairwise probability of decoding codeword X as codeword $\bar{X}$ assuming that the space-time code satisfies the design criterion $A = \frac{E_d}{N} I$, but without using the central limit theorem approximation. Under this design criterion, $d^2(X, \bar{X})$ becomes

$$d^2(X, \bar{X}) = \sum_{m=1}^{M} \sum_{n=1}^{N} \frac{E_d}{N} |\alpha_{nm}|^2 = \frac{E_d}{N} \sum_{k=1}^{MN} e^{2\chi_k}, \tag{3.47}$$

where χ_k, $k = 1, \ldots, MN$, are independent, identically distributed Gaussian random variables with $\text{var}(\chi_k) = \sigma_\chi^2$ and $E(\chi_k) = -\sigma_\chi^2$. Define $\chi = (\chi_1, \ldots, \chi_{MN})$. The probability of decoding X as $\bar{X}$ is then

$$\Pr(\, X \to \bar{X}\,) = \int_{-\infty}^{\infty} \Pr(\, X \to \bar{X} \mid \chi\,) p_\chi(\chi) d\chi, \tag{3.48}$$

where $p_\chi(\chi)$ is the multivariate Gaussian probability density function for χ, and

$$\Pr(\, X \to \bar{X} \mid \chi \,) = Q\left(\sqrt{\left[\frac{1}{2N}\frac{E_d}{N_0}\right][MN]\left[\frac{1}{MN}\sum_{k=1}^{MN} e^{2\chi_k}\right]}\right), \tag{3.49}$$

is the conditional pairwise error probability. We write the argument as three terms to emphasize that the error probability depends on the signal-to-noise ratio (E_d/N_0) per transmit aperture ($1/N$), aperture number product (MN), and the average fade power, $\sum_{k=1}^{MN}\exp(2\chi_k)/MN$. Using the bound $Q(x) \le \exp(-x^2/2)/2$ gives

$$\begin{aligned}
\Pr(\, X \to \bar{X} \,) &\le \int_{-\infty}^{\infty} \frac{1}{2}\frac{1}{(2\pi\sigma_\chi^2)^{MN/2}} \exp\left(-\frac{1}{N}\frac{E_d}{4N_0}\sum_{k=1}^{MN} e^{2\chi_k} - \frac{1}{2\sigma_\chi^2}\sum_{k=1}^{MN}(\chi_k + \sigma_\chi^2)^2\right) d\chi \\
&= \frac{1}{2}\left[\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_\chi^2}} \exp\left(-\frac{1}{N}\frac{E_d}{4N_0}e^{2x}\right)\exp\left(-\frac{1}{2\sigma_\chi^2}(x+\sigma_\chi^2)^2\right) dx\right]^{MN} \\
&= \frac{1}{2}\left[\mathrm{Fr}\left(\frac{1}{N}\frac{E_d}{4N_0}; -\sigma_\chi^2, \sigma_\chi^2\right)\right]^{MN} = \frac{1}{2}\left[\mathrm{Fr}\left(\frac{\mathrm{SNR}}{4N}; -\sigma_\chi^2, \sigma_\chi^2\right)\right]^{MN},
\end{aligned} \tag{3.50}$$

where $\mathrm{Fr}(a; m, s^2)$ is the lognormal density frustration function given by

$$\mathrm{Fr}(a; m, s^2) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s^2}} \exp\left(-ae^{2x}\right)\exp\left[-\frac{1}{2s^2}(x-m)^2\right] dx. \tag{3.51}$$

Using the bound $Q(x) \ge \exp(-x^2)/4$, gives a similar lower bound

$$\Pr(\, X \to \bar{X} \,) \ge \frac{1}{4}\left[\mathrm{Fr}\left(\frac{\mathrm{SNR}}{2N}; -\sigma_\chi^2, \sigma_\chi^2\right)\right]^{MN}. \tag{3.52}$$

A closed form evaluation of the frustration function does not exist; therefore, we use a saddle-point integration method developed by Halme in [28] to numerically evaluate it. For the design criterion $A = \frac{E_d}{N} I$, Figure 3-8 compares the probability of codeword error in (3.48), the central limit theorem approximation probability of codeword error in (3.24), its asymptotic behavior in (3.41), and the frustration function bounds in (3.50) and (3.52). This figure shows that for small values of SNR, or typical error probabilities of interest, the CLT approximation seems valid for $MN = 16$ in moderate fading. Asymptotically, however, the CLT probability of

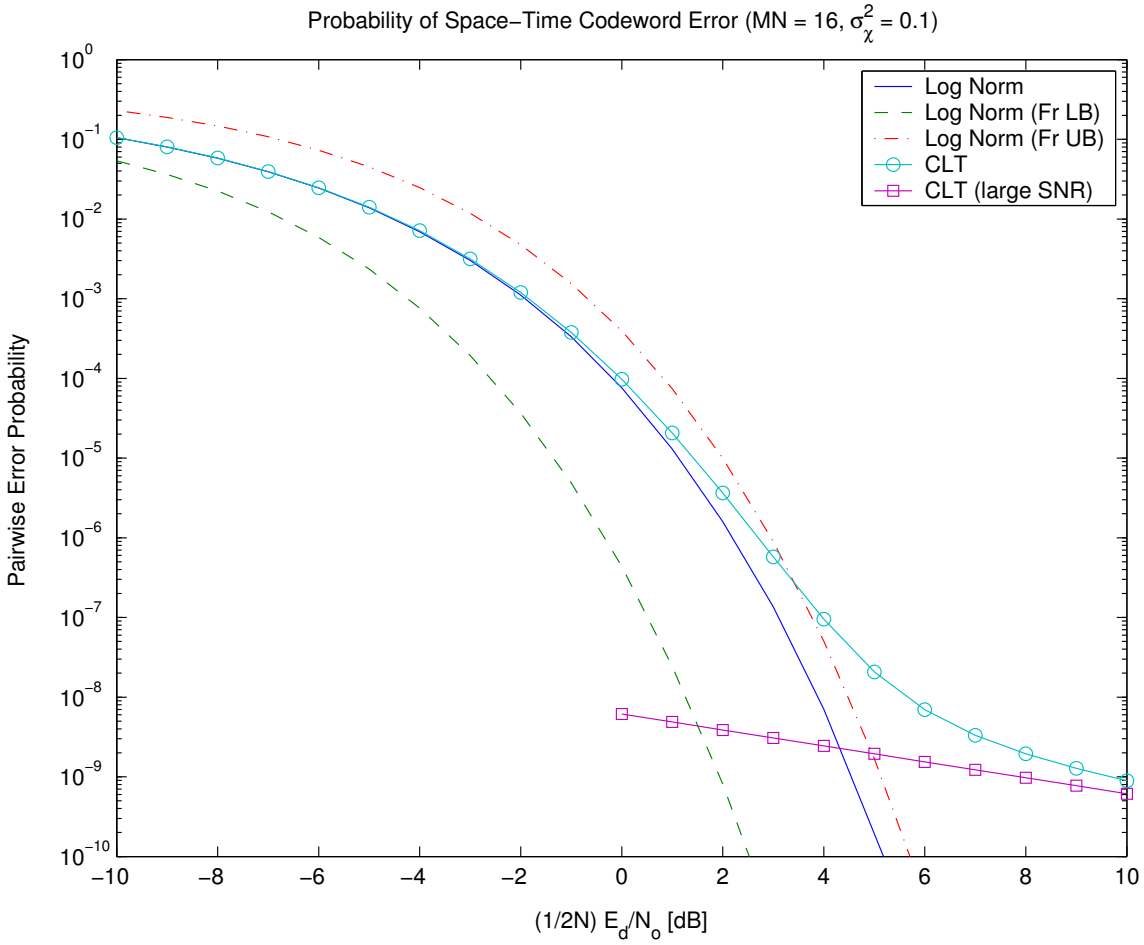

Figure 3-8: A comparison of the pairwise error probability for $A = \frac{E_d}{N} I$ STCs using the exact error probability in (3.48) computed via Monte Carlo averaging, the central limit theorem approximation (3.24) calculated via trapezoidal integration, its asymptotic behavior in (3.41), and the frustration function bounds in (3.50) and (3.52) computed via saddle-point integration.

codeword error decays slower than the actual error probability. From (3.41), we know that the CLT probability of codeword error decays as 1/SNR, whereas the frustration function bounds suggest the actual curve decays faster. This discrepancy arises from dissimilarities in the tails of the Gaussian distribution and the actual distribution as emphasized by large values of SNR. As this figure suggests, the CLT approximation tends to be a conservative estimate, or upper bound, to the pairwise error probability.

To measure the validity of the central limit theorem approximation, we examined the difference in SNR between the error probability expression in (3.48) and its approximation in (3.24) at a given error probability. For example, in Figure 3-8 for an error probability of $\Pr(\, X \to \bar{X}\,) = 10^{-6}$, the CLT approximation requires 0.5 dB more SNR than the actual lognormal curve. Figure 3-9 shows this spurious SNR for different aperture products (MN) in different fading environments ($\sigma_\chi^2 = 0.01, 0.1, 0.35$). From this figure, we see that the CLT approximation is accurate to fractions of a dB at 10^{-6} pairwise error probabilities in mild fading environments ($\sigma_\chi^2 = 0.01$) for all values of $MN \geq 2$. A larger number of apertures is required for more severe fading (roughly, $MN > 16$ for $\sigma_\chi^2 = 0.1$ and $MN > 64$ for $\sigma_\chi^2 = 0.35$). Regardless, it appears that the CLT approximation tends to over estimate the pairwise error, and acts as an upper bound.

A Lower Bound on the Probability of Codeword Error

In the previous section, we derived lower and upper bounds on the probability of incorrectly decoding codeword X as codeword $\bar{X}$ under the design criterion $A = \frac{E_d}{N} I$ without using the central limit theorem approximation for $d^2(X, \bar{X})$. In this section, we derive a lower bound on this probability of error without using the central limit theorem approximation that is valid for an arbitrary design matrix A. Using the

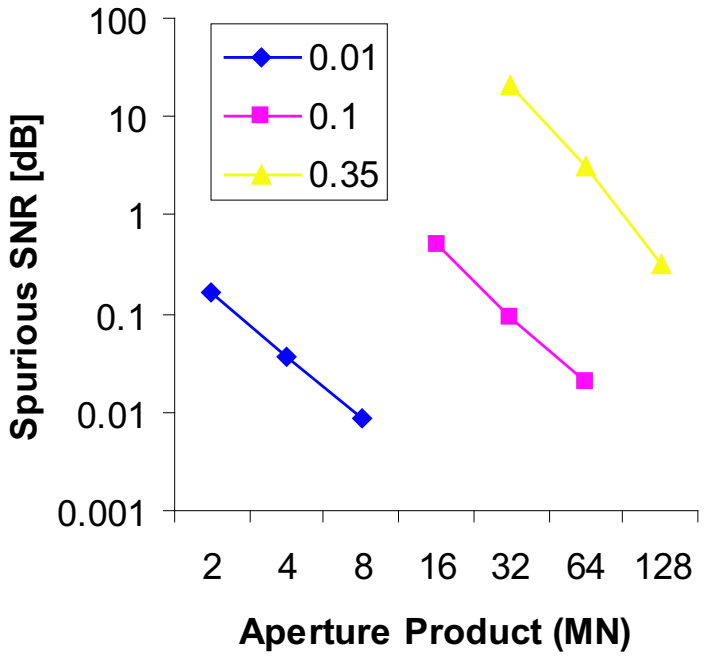

Figure 3-9: The difference in SNR required to achieve a 10^{-6} pairwise codeword error probability between the actual lognormal error expression in (3.48) and its central limit theorem approximation in (3.24) is shown for different fading strengths ($\sigma_\chi^2 = 0.01$, 0.1, and 0.35).

Cauchy-Schwarz inequality on (3.15) gives

$$\begin{aligned} d^2(X,\bar{X}) &\leq \sum_{m=1}^{M}\sum_{t=1}^{T}\sum_{n=1}^{N}|\alpha_{nm}|^2\sum_{k=1}^{N}|\bar{x}_k(t)-x_k(t)|^2 \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N}E_d\,|\alpha_{nm}|^2 = E_d\sum_{k=1}^{MN}e^{2\chi_k}, \end{aligned} \tag{3.53}$$

where we have renumbered the sum of the MN independent lognormal random variables as in the previous section. Following a similar derivation to that in the previous section, a lower bound on the probability of error for any design matrix is

$$\Pr(\, X \to \bar{X}\,) \geq \frac{1}{4}\left[\mathrm{Fr}\left(\frac{\mathrm{SNR}}{2};-\sigma_\chi^2,\sigma_\chi^2\right)\right]^{MN}. \tag{3.54}$$

For a large number of transmit apertures, N, this bound can be quite loose, cf. the orthogonal design bound in (3.52).

Infinite Diversity Performance Limit

If we fix the total receiver average signal-to-noise ratio, ρ, and have enough receive apertures, M, such that (3.44) holds, then $A = \frac{E_d}{N} I$ minimizes the pairwise error probability, and this design matrix gives $\eta^2 = [\exp(4\sigma_\chi^2) - 1]/MN$. As we increase the aperture number product, MN, we see that η^2 approaches zero, and hence (3.42) provides a performance limit for infinite diversity, i.e.,

$$\lim_{MN \to \infty} \Pr\left(X \to \bar{X}; M\mathrm{SNR}, \frac{e^{4\sigma_\chi^2} - 1}{MN} \right) = Q\left(\sqrt{\frac{M\mathrm{SNR}}{2}} \right). \tag{3.55}$$

We can also see this result from the conditional pairwise error (3.49). As the number of apertures MN increases, the sample average becomes the ensemble average by the law of large numbers, and we get the infinite diversity limit (3.55). These limits appear as circles in Figure 3-3 for $M\,\mathrm{SNR} = 8$, 13, 15, and 18 dB.

One can view this limit as the error probability of a one transmit, M receive aperture system with no fading. In other words, the large number of apertures mitigates the fading, and the only uncertainty in the decision process arises from the additive white Gaussian noise at each receiver.

We also note that for fixed $\rho = M\,\mathrm{SNR}$, increasing the number of receive apertures decreases the upper bound until all space-time codes have about the same pairwise error performance (3.55). For a fixed total receive signal-to-noise ratio ρ, a generally poor STC, such as the repetition spatial STC, will perform just as well as an "optimal" orthogonal design STC for large numbers of receive apertures.

An Orthogonal Design Example: The Alamouti Scheme

Alamouti in [1] proposed a simple transmit diversity technique using two transmit apertures ($N = 2$), two time-slots ($T = 2$), M receive apertures, and a complex QAM signal $\mathcal{S}$ constellation of size 2^b. During the first time-slot, $2b$ bits arrive, determining two signal constellation points, s_1 and s_2 that are transmitted simultaneously on the first and second apertures, respectively. During the second time-slot, the first

aperture transmits $-s_2^*$, while the second sends s_1^*. In other words, this STC consists of all the codewords of the form $x = [x_1(1), x_1(2), x_2(1), x_2(2)] = [s_1, -s_2^*, s_2, s_1^*]$ where s_1 and s_2 range over all possible signal constellation points. Tarokh in [60] showed that the Alamouti scheme is an example of a STC created from a complex orthogonal design.

The design matrix of this STC for two codewords $x = [x_1, -x_2^*, x_2, x_1^*]$ and $\bar{x} = [\bar{x}_1, -\bar{x}_2^*, \bar{x}_2, \bar{x}_1^*]$ satisfies our design criteria $A = \frac{E_d}{2} I$, where $E_d = 2|x_1 - \bar{x}_1|^2 + 2|x_2 - \bar{x}_2|^2$ is the energy difference between the codewords. The performance of this code for pairs of codewords is shown in Figure 3-8 for eight receive apertures ($M = 8$) in moderate fading ($\sigma_\chi^2 = 0.1$), and in Figures 3-12 and 3-11 for two receive apertures ($M = 2$) in mild and moderate fading, respectively.

Orthogonal designs [60] have the property that the symbol sequences on each aperture are orthogonal, i.e., $\sum_{t=1}^{T} x_n(t) x_k^*(t) = 0$ for $n \neq k$. As a result, space-time codes created from orthogonal designs, such as the Alamouti scheme, have a simple decoding algorithm. Rewriting the decision metric in (3.11) as

$$\begin{aligned} \hat{X} &= \underset{X \in \mathbf{X}}{\operatorname{argmin}} \sum_{m=1}^{M} \sum_{t=1}^{T} \left(|y_m(t)|^2 + \sum_{n=1}^{N} \sum_{k=1}^{N} \alpha_{nm} \alpha_{km}^* x_n(t) x_k^*(t) - 2\mathrm{Re} \left\{ y_m^*(t) \sum_{n=1}^{N} \alpha_{nm} x_n(t) \right\} \right) \\ &= \underset{X \in \mathbf{X}}{\operatorname{argmin}} \sum_{n=1}^{N} \sum_{t=1}^{T} \left[\left(\sum_{m=1}^{M} |\alpha_{nm}|^2 \right) |x_n(t)|^2 - 2\mathrm{Re} \left\{ \left(\sum_{m=1}^{M} y_m^*(t) \alpha_{nm} \right) x_n(t) \right\} \right] \end{aligned} \tag{3.56}$$

shows that joint detection of $[x_1(1), \ldots, x_N(T)]$ is equivalent to decoding each individual symbol, $x_n(t)$, separately. The structure of the Alamouti STC allows for further simplification, and the decision rules become [60]

$$\hat{s}_1 = \underset{s \in \mathcal{S}}{\operatorname{argmin}} \left[\left(-1 + \sum_{n=1}^{2} \sum_{m=1}^{M} |\alpha_{nm}|^2 \right) |s|^2 + \left| s - \sum_{m=1}^{M} [y_m(1) \alpha_{1m}^* + y_m^*(2) \alpha_{2m}] \right|^2 \right] \tag{3.57}$$

and

$$\hat{s}_2 = \underset{s\in\mathcal{S}}{\operatorname{argmin}} \left[\left(-1 + \sum_{n=1}^{2} \sum_{m=1}^{M} |\alpha_{nm}|^2 \right) |s|^2 + \left| s - \sum_{m=1}^{M} \left[y_m(1)\alpha_{2m}^* + y_m^*(2)\alpha_{1m} \right] \right|^2 \right]. \tag{3.58}$$

Figure 3-10 shows a simulation of the average code error probability, P_e, for the two transmit aperture Alamouti STC over 5×10^6 channel realizations. We use the binary, phase-shift, keying (BPSK) signal set, $\mathcal{S} = \{-\sqrt{E_b}, \sqrt{E_b}\}$. Because a codeword error occurs when either bit is in error, the codeword error rate is approximately twice the bit error rate.

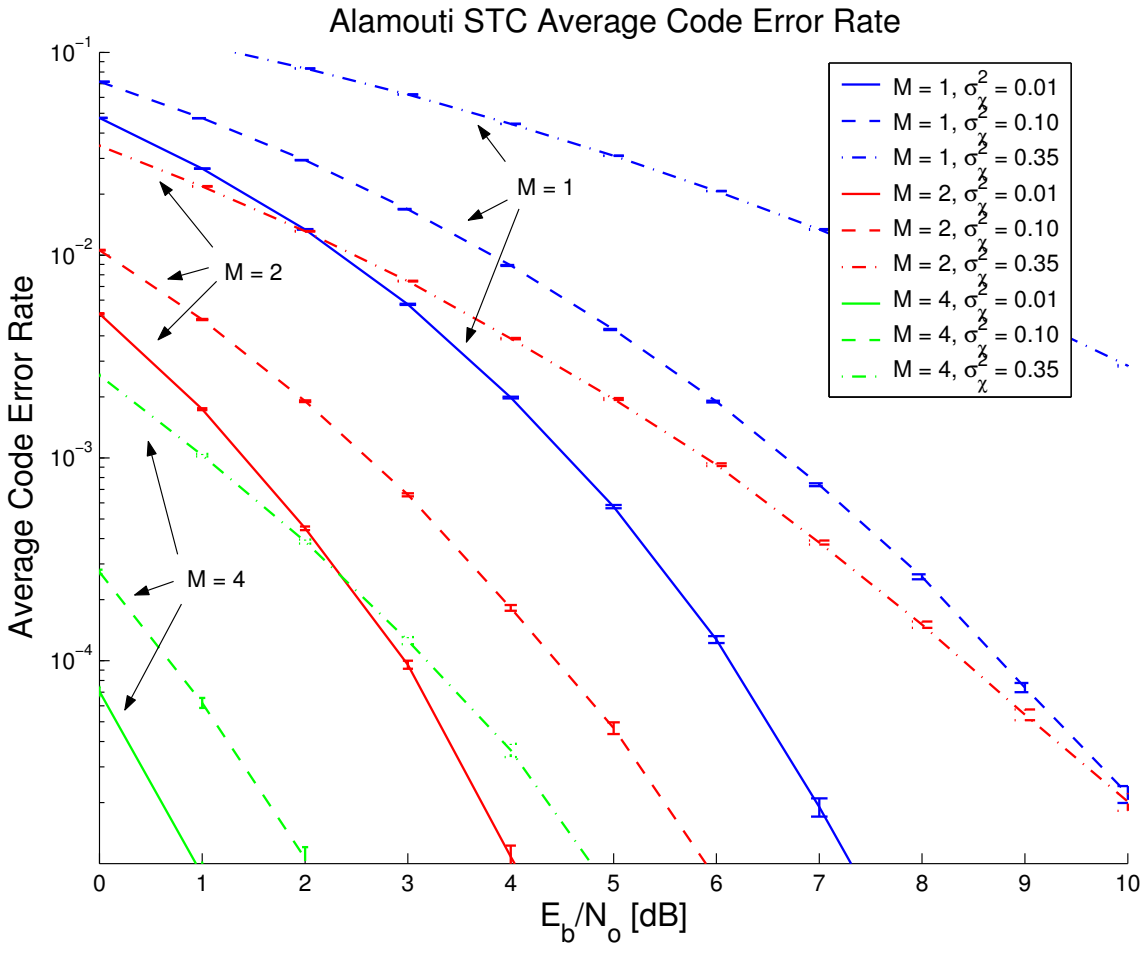

Figure 3-10: The two transmit aperture ($N = 2$), BPSK, Alamouti STC, average codeword error probability is plotted for different numbers of receive apertures ($M =$ 1, 2, and 4) and fading strengths ($\sigma_\chi^2 = 0.01$, 0.1, and 0.35). Error bars indicate the standard error of each estimate.

We now compare the union bound estimate (3.13) based on the pairwise error probability to the actual code error rate in Figure 3-10. After all, we used the union bound estimate as motivation for minimizing the pairwise error probability as a proxy to the actual code error rate. We would first like to verify that this union bound estimate based on the minimum distance pairwise error probability is a reasonable substitute for the code error rate. Second, we want to determine if the CLT pairwise error (3.24) is a good approximation to the actual pairwise error probability (3.48).

For example, we see in Figure 3-10 and Table 3.4 that the actual error rate for two receive apertures ($M = 2$) in moderate fading ($\sigma_\chi^2 = 0.1$) at $E_b/N_0 = 3$ dB is approximately 6.6×10^{-4}.

An Alamouti STC using a BPSK signal constellation has four codewords. Each codeword has two neighbors at a distance squared of $E_d^{\min} = 8E_b$, and one neighbor at $E_d^{\max} = 16E_b$. The average number of nearest neighbors per codeword is, therefore, $K_{\min} = 2$. The union bound estimate then for this Alamouti STC is twice the minimum distance pairwise error probability. From Figure 3-11 and Table 3.3, we see that the exact pairwise error probability in (3.48) for a minimum distance codeword pair at $E_b/N_0 = 3$ dB is 3.3×10^{-4}. The union bound estimate based on the exact pairwise error probability is 6.7×10^{-4} (Table 3.4), which is very close to the actual code error rate. We conclude that under these operating conditions, the union bound estimate is a good indication of code error rate.

The Alamouti STC is an example of an orthogonal design, which has a diagonal design matrix. The normalized fading strength in this case is

$$\eta^2 = \frac{e^{(4)(0.1)} - 1}{(2)(2)} \approx 0.12. \tag{3.59}$$

From Figure 3-3 we see that for $\eta^2 = 0.12$ and

$$M\frac{E_d^{\min}}{N_0} = 2\frac{8E_b}{N_0} \approx 15\text{dB}, \tag{3.60}$$

the pairwise error probability is approximately 1.6×10^{-3}. The CLT approximate union bound estimate of the codeword error probability is then 3.2×10^{-3}, which is

a factor of five from the actual probability of 6.6×10^{-4}.

As seen in Figure 3-11, the CLT approximation tends to over estimate the pairwise error probability. In fact, for a pairwise error probability of 3.3×10^{-4} (one-half the code error rate), the CLT approximation uses about 4 dB more SNR than the actual probability. In other words, the CLT minimum distance pairwise approximation at $E_b/N_0 = 7$ dB produces the same error rate as the true pairwise error at $E_b/N_0 = 3$ dB. This 4 dB spurious SNR indicates that for $M = N = 2$ in moderate fading $\sigma_\chi^2 = 0.1$, the CLT approximation to the pairwise error at 3.3×10^{-4} is not that good; hence, the factor of five in the CLT union bound estimate.

If we repeated the above analysis using mild fading ($\sigma_\chi^2 = 0.01$) instead of moderate fading ($\sigma_\chi^2 = 0.1$), then the CLT union bound estimate is very accurate. From Figure 3-10 and Table 3.4, the code error rate at $E_b/N_0 = 3$ dB in mild fading ($\sigma_\chi^2 = 0.01$) using two receive apertures ($M = 2$) is approximately 9.5×10^{-5}. The normalized fading strength of the Alamouti STC in mild fading is now $\eta^2 \approx 10^{-2}$. From Figure 3-3 and Table 3.3, the CLT pairwise error probability for $\rho = 15$ dB is 4.7×10^{-5}. The CLT union bound estimate is then 9.5×10^{-5}, which agrees with the actual code error rate.

As seen in Figure 3-11, for two transmitters and two receivers ($N = M = 2$) in mild fading ($\sigma_\chi^2 = 0.01$), the spurious SNR at 4.75×10^{-5} is roughly 0.01 dB. In this case, the central limit theorem approximation is very accurate. Furthermore, the union bound estimate tends to be more accurate at smaller error probabilities.

Also, note from Figure 3-3 that increasing the number of transmit apertures will not improve the error performance much for a fixed $\rho = ME_d/N_0 = 15$ dB. For this total signal-to-noise ratio, the infinite transmit diversity limit (3.55) is 3.5×10^{-5}, resulting in a union bound estimate of 7×10^{-5}. Furthermore, from Figure 3-5 we see that for $E_d^{\min}/N_0 = 8E_b/N_0 = 12$ dB, that orthogonal design STCs in mild fading are optimal in terms of the CLT pairwise error probability for any number of transmit and receive apertures. Because the CLT union bound estimate is very close to the true code error probability, and orthogonal designs minimize the pairwise error probability in the CLT regime, we conclude that the Alamouti STC under these

operating conditions[4] is the best STC in terms of code error rate, P_e. In other words, under these operating conditions, we have met our objective to develop a space-time code that minimizes the code error rate.

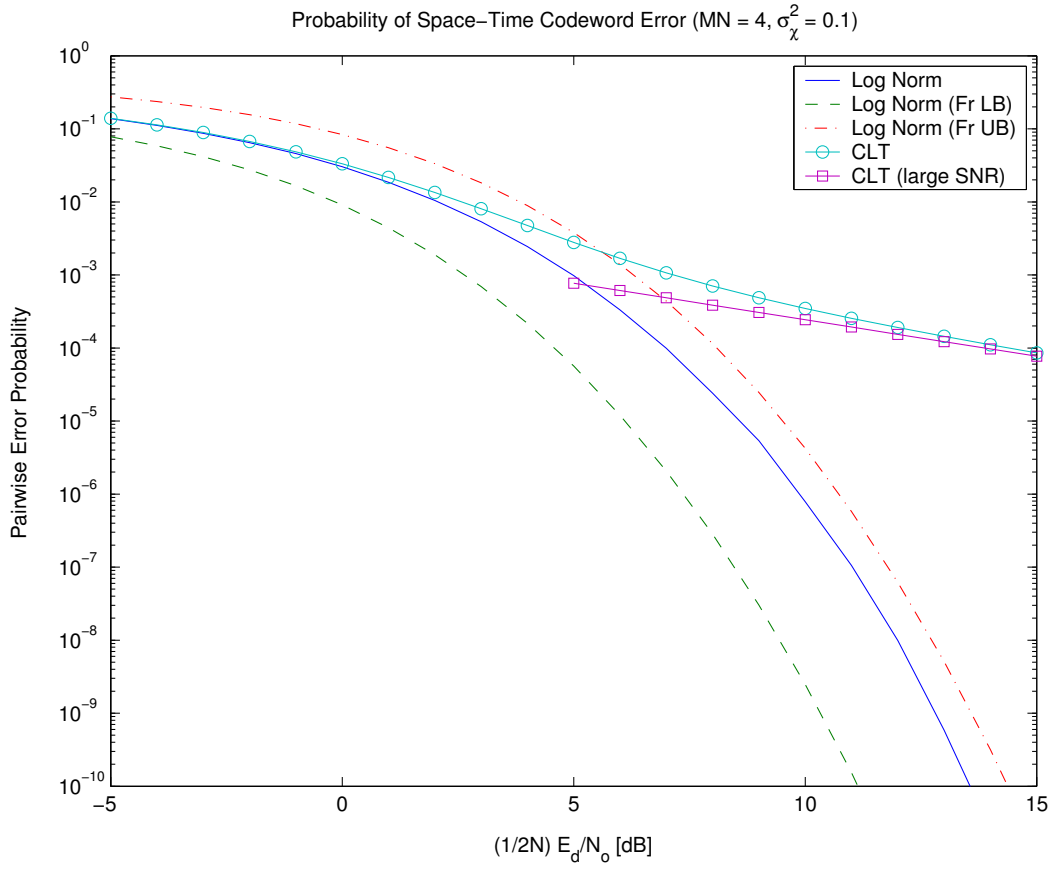

Figure 3-11: A comparison of the pairwise error probability for $A = \frac{E_d}{N} I$ STCs in moderate fading ($\sigma_\chi^2 = 0.1$) using the exact error probability in (3.48) computed via Monte Carlo averaging, the central limit theorem approximation (3.24) calculated via trapezoidal integration, its asymptotic behavior in (3.41), and the frustration function bounds in (3.50) and (3.52) computed via saddle-point integration.

[4]Under the latter operating conditions: BPSK signal set, $N = M = 2$, $\sigma_\chi^2 = 0.01$, $E_b/N_0 = 3$ dB

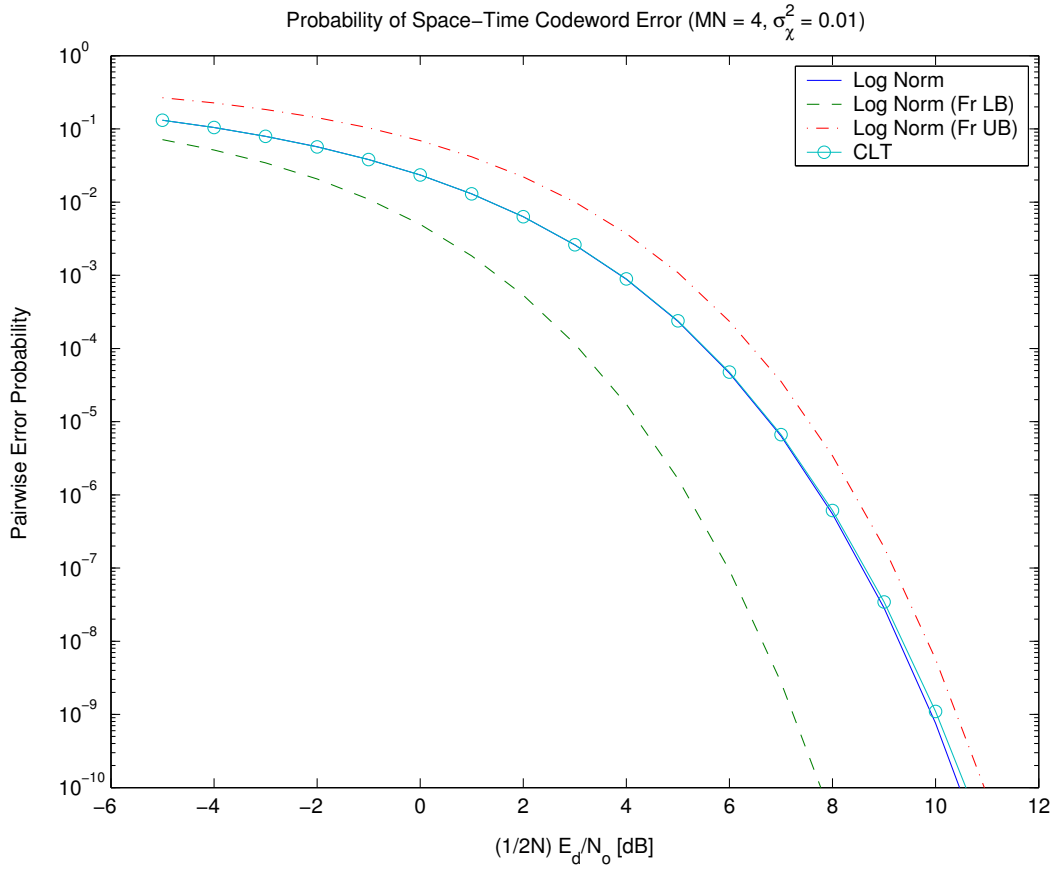

Figure 3-12: A comparison of the pairwise error probability for $A = \frac{E_d}{N} I$ STCs in weak fading ($\sigma_\chi^2 = 0.01$) using the exact error probability in (3.48) computed via Monte Carlo averaging, the central limit theorem approximation (3.24) calculated via trapezoidal integration, and the frustration function bounds in (3.50) and (3.52) computed via saddle-point integration.

σ_χ^2	Fr LB	LN	Fr UB	CLT
0.1	1.2E-05	3.3E-04	1.4E-03	1.6E-03
0.01	9.3E-08	4.6E-05	2.4E-04	4.7E-05

Table 3.3: Minimum distance pairwise error probability at $E_b/N_0 = 3$ dB is shown as a function of log-amplitude variance ($\sigma_\chi^2 = 0.1$ and 0.01) for the Alamouti STC using two transmit and two receive apertures ($N = M = 2$). The columns are as follows: the exact pairwise error probability (LN) in (3.48), the central limit theorem (CLT) approximation (3.24), and the frustration function bounds (Fr LB and Fr UB) in (3.50) and (3.52).

σ_χ^2	Fr LB UBE	LN UBE	Fr UB UBE	CLT UBE	Actual CER
0.1	2.4E-05	6.7E-04	2.8E-03	3.2E-03	6.6E-04
0.01	1.9E-07	9.2E-05	4.7E-04	9.5E-05	9.5E-05

Table 3.4: This table compares the Alamouti STC code error rate (Figure 3-10) at $E_b/N_0 = 3$ dB to the union bound estimates (twice the minimum distance pairwise error probability) using the pairwise error probabilities in Table 3.3.

Chapter 4

Photon-Counting Receivers

When the inherent randomness of photon arrivals, i.e., shot noise, has a larger variance than the electrical amplifier thermal noise, we say that the receiver is shot-noise limited. In this shot-noise limited regime, we can model the output of each detector as a doubly-stochastic, Poisson counting process [58]. For every receiver, we will assume as in [40] that the fields received from the multiple transmitters are sufficiently separated in frequency or angle of arrival to make the received power equal to the sum of the powers from the individual transmitters, i.e., (2.26) and/or (2.27) holds.

For simplicity, we assume that the receiver and transmitter geometry is such that the free-space losses are the same for all transmitter to receiver paths. Because the free-space losses are identical, we can normalize the optical field's temporal component such that the effective photon flux in (2.30) collected by the m-th receiver is,

$$\text{Effective STA photon flux at } m\text{-th Receiver} = \sum_{n=1}^{N} \alpha_{nm} x_n(t). \tag{4.1}$$

Because direct detection receivers respond to the impinging optical power, for convenience, we have redefined α_{nm} to be the *power* path gain from the n-th transmitter to the m-th detector.

Similarly, $x_n(t)$ is the effective photon flux waveform from the n-th transmitter[1].

[1]For simplicity of terminology, we shall use power instead of photon flux, in what follows. In essence, this amounts to saying that we are measuring power in units of photons per second instead of Watts.

The codeword waveform $x_n(t)$, is proportional to the optical power of the n-th transmitter during the interval $0 \leq t \leq T$. Physically, $(\eta/hf_c)^{-1}x_n(t)$ is the transmitted power in Watts from the n-th transmitter measured at each receiver in the absence of fading. Here: η is the detector quantum efficiency, h is Planck's constant, and f_c is the optical carrier frequency.

Under shot-noise-limited operation, the photon-count record of the m-th ($1 \leq m \leq M$) detector, $\{y_m(t),\, 0 \leq t \leq T\}$, conditioned on the knowledge of the transmitted codeword and path gains, can be taken to be a Poisson counting process with rate

$$\mu_m(t) = \sum_{n=1}^{N} \alpha_{nm} x_n(t) + \lambda_m, \tag{4.2}$$

where N is the number of transmit apertures, α_{nm} is the power path gain from the n-th transmit aperture to the m-th receive aperture, and $\lambda_m \geq 0$ is a dark noise rate for the m-th receiver. We can also incorporate an optical background noise in λ_m, if the background noise modes justify a Poisson approximation. This approximation occurs when the third and fourth terms in (2.50) are negligible. In this case, we can replace the stochastic background noise rate by its expected value as done in [12]. This situation arises in practice when the receiver collects many spatial and temporal optical noise modes. Regardless, will refer to λ_m as a background noise rate.

We normalize the power path gains

$$\alpha_{nm} = e^{2\chi_{nm}}, \tag{4.3}$$

such that the atmosphere does not, on average, attenuate or amplify the transmitted waveform, i.e., $E[\alpha_{nm}] = 1$, by setting $\text{var}[\chi_{nm}] = -E[\chi_{nm}] = \sigma_\chi^2$. Again, we note that α_{nm} is a real-valued power path gain. The log-amplitude variance σ_χ^2, given by (2.20), can be as small as zero when fading is negligible, and saturates at one-half in severe turbulence conditions. We also assume that the spacing between elements of the aperture arrays is large enough to ensure that the path gains for different (n, m) are approximately independent.

4.1 Capacity

In this section, we examine the capacity of this MIMO Poisson, fading channel. We start by assuming that the transmitter and receiver know and use the path gains optimally. We show that path gain knowledge at the transmitter does not increase capacity in the low and high signal-to-noise ratio regimes. Next, we calculate the instantaneous capacity of a system employing a receiver that simply adds the photon counts from each detector. This "photon-bucket" receiver does not require path gain knowledge and we will show that it does not reduce capacity in the high signal-to-noise ratio regime. Finally, we will examine the ergodic capacity and capacity-vs-outage probability, which take into account the channel's fading nature.

4.1.1 The MIMO Poisson Channel

We now examine the Shannon capacity of the MIMO Poisson channel with rate (4.2) subject to peak and average transmit power constraints [25]. We assume that the transmitter and receiver know the path gains, and use this information optimally to maximize the instantaneous capacity, i.e., the capacity for a given channel realization. In other words, in this section, we will treat the path gains as deterministic known constants. We derive upper and lower bounds on this capacity, and show that they coincide in a number of special cases. The capacity is bounded below by that of the MIMO channel with an additional on-off keying (OOK) transmitter constraint, and it is bounded above by that of parallel, independent multiple-input, single-output (MISO) channels.

The capacity of the single-input, single-output (SISO) Poisson channel is well understood. Kabanov [33] derived the information capacity of the SISO Poisson channel with a peak transmit power constraint using martingale techniques. Davis [17] considered the addition of an average transmit power constraint. Wyner [67] derived the capacity and error exponent from first principles, using a discrete memoryless channel approximation. Shamai (Shitz) [53, 52] derived the capacity with constraints on the transmitted pulse width. Frey [19] allowed for time-varying peak and average power

constraints, as well as for random noise intensities. Shamai (Shitz) and Lapidoth [55] considered general spectral constraints on the Poisson counting process rate process.

Recently, the multiple-user Poisson channel has received attention. Lapidoth and Shamai (Shitz) [40] computed the two-user capacity region of the multiple-access channel. Their concluding section included a comment on the capacity of the MISO channel in the absence of background noise and average power constraints. We will formalize this comment on the MISO Poisson channel. Bross [9] calculated the error exponent for the two-user Poisson multiple access channel. Lapidoth discusses the Poisson broadcast channel in [41].

In what follows, we shall derive upper and lower bounds on the capacity of the MIMO Poisson channel. We will show that our bounds coincide in a number of interesting special cases. These include the limits of low and high signal-to-noise ratio, and the MISO channel. We will also show that our lower bound gives the MIMO capacity for the single-input, multiple-output (SIMO) channel, and that, in general, our bounds are quite close.

Capacity and Mutual Information

We will force the (non-negative) codeword waveform $x_n(t)$, which is proportional to the power waveform sent from the n-th transmitter, to satisfy the peak power constraint,

$$0 \leq x_n(t) \leq A_n, \tag{4.4}$$

and the average power constraint,

$$\frac{1}{T}\int_0^T E[x_n(t)]dt \leq \sigma A_n, \quad \text{where } 0 \leq \sigma \leq 1. \tag{4.5}$$

Note that A_n has units of photons per second at the operating wavelength.

A realistic, practical constraint that we are not considering is that of bandwidth. We do so for two reasons. First, only bounds exist for the capacity of a SISO spectrally-constrained Poisson channel [55], but not for a multiple-aperture channel. Second, unlike microwave communication for which bandwidth is licensed and

expensive, communication bandwidth at optical wavelengths is unregulated and plentiful. Indeed, the bandwidth of an atmospheric optical link is limited by transmitter and receiver technology, rather than by atmospheric propagation. When discussing photon-bucket receivers, we will consider a lower bound to the channel capacity that imposes an ad hoc constraint on bandwidth.

The m-th detector photon-count record up to time t, $Y_m(t) \equiv \{y_m(\tau), 0 \leq \tau < t \leq T\}$, is completely described by the total number of arrivals occurring prior to time t, denoted by $y_m(t)$, and its ordered arrival times, $0 \leq t_{m1} \leq t_{m2} \leq \ldots \leq t_{my_m(t)} < t \leq T$. Note that $Y_m(t)$ is continuous from the left, viz., it includes all arrivals up to, but not including, time t. We can then equivalently define $Y_m(t) \equiv \{y_m(t), t_{m1}, \ldots, t_{my_m(t)}\}$. Note that when we refer to distributions or densities of the detector output processes, we mean distributions or densities on the ordered photon arrival times. Let $Y_m \equiv Y_m(T)$ represent the path of the m-th output process on the interval $[0, T)$. Denote the M detector output process paths up to time t as $Y(t) \equiv \{Y_1(t), \ldots, Y_M(t)\}$, and the entire path as $Y \equiv Y(T)$. Let $X \equiv \{x_1(t), \ldots, x_N(t);\, 0 \leq t < T\}$ represent the channel input.

Define $\mathcal{X}$ as all distributions on X that satisfy the transmitter peak and average power constraints in (4.4) and (4.5). The capacity of the MIMO Poisson channel in nats[2] per second is then

$$C(\alpha) = \sup_{p_X \in \mathcal{X}} \frac{1}{T} I(X;Y), \tag{4.6}$$

where the mutual information, $I(X;Y)$, is given by [14],

$$\begin{aligned} I(X;Y) &= E\left[\log\left(\frac{p_{Y|X}}{p_Y}\right)\right] = E\left[\log p_{Y|X}\right] - E\left[\log p_Y\right] \\ &= h(Y) - h(Y|X). \end{aligned} \tag{4.7}$$

Here: $h(Y) \equiv -E[\log p_Y]$ and $h(Y|X) \equiv -E[\log p_{Y|X}]$ are the unconditional and conditional entropies, respectively, of the output processes' ordered arrival times described by the densities p_Y and $p_{Y|X}$. The expectations in the unconditional and conditional entropies are taken with respect to the densities of the output processes

[2]Unless otherwise noted, all logarithms are assumed to be natural logarithms.

and the joint input and output processes, respectively. A more general derivation of the mutual information for point processes is given in [42] and [6] using martingale techniques. In what follows, we present a specialized derivation for doubly-stochastic Poisson processes [58], sufficient for our application, using ordered arrival time densities.

First consider the conditional entropy, $h(Y|X)$, of the output processes' ordered arrival times. Conditioned on the channel input X, the output processes $Y_1, \ldots, Y_M$ are independent, each with density ([58],Th. 2.3.2),

$$p_{Y_m|X} = \exp\left(\int_0^T \log[\mu_m(t)]dy_m(t) - \int_0^T \mu_m(t)dt\right), \tag{4.8}$$

where for a given function f,

$$\int_0^T f(t)dy_m(t) \equiv \begin{cases} 0, & y_m(T) = 0, \\ \sum_{k=1}^{y_m(T)} f(t_{mk}), & y_m(T) > 0, \end{cases} \tag{4.9}$$

and $0 \le t_{m1} \le t_{m2} \le \ldots \le t_{m\,y_m(T)} < T$ are the ordered arrival times of the m-th output process, Y_m.

We will use the following observation in both the conditional and unconditional entropy derivations. Let $\{P(t), 0 \le t < T\}$ be an inhomogeneous Poisson counting process with rate function $m(t)$. Then,

$$E\left[\int_0^T f(t)\, dP(t)\right] = \int_0^T f(t)\, m(t)\, dt, \tag{4.10}$$

where f is a given function. This observation essentially says that the expected number of arrivals in an interval of length dt is $E[dP(t)] = m(t)\, dt$. To prove this observation, we use the arrival time density in (4.8), symmetry about the $k!$ permutations of the k variables of integration, and the unity sum of the Poisson mass

function:

$$
\begin{aligned}
& E\left[\int_0^T f(t)\ dP(t)\right] \\
&= \sum_{k=1}^{\infty} \int_0^T \int_0^{t_k} \cdots \int_0^{t_2} \left(\prod_{i=1}^{k} m(t_i)\right) \exp\left(-\int_0^T m(t)dt\right) \left[\sum_{j=1}^{k} f(t_j)\right] dt_1 \cdots dt_k \\
&= \sum_{k=1}^{\infty} \frac{\exp\left(-\int_0^T m(t)dt\right)}{k!} \int_0^T \cdots \int_0^T \left(\prod_{i=1}^{k} m(t_i)\right) \left[\sum_{j=1}^{k} f(t_j)\right] dt_1 \cdots dt_k \\
&= \sum_{k=1}^{\infty} \frac{\exp\left(-\int_0^T m(t)dt\right)}{(k-1)!} \left(\int_0^T m(t)dt\right)^{k-1} \int_0^T f(t)\, m(t)\, dt \\
&= \int_0^T f(t)\, m(t)\, dt,
\end{aligned}
$$

where $0 \leq t_1 \leq \cdots \leq t_{P(t)} < T$ are the ordered arrival times up to time t.

Using (4.10) and (4.8), the conditional entropy of the ordered arrival times is

$$
\begin{aligned}
h(Y|X) &= -E\left[\log p_{Y|X}\right] \\
&= -E\left[\sum_{m=1}^{M} \log p_{Y_m|X}\right] \\
&= -\sum_{m=1}^{M} E\left[\int_0^T \log[\mu_m(t)]dy_m(t) - \int_0^T \mu_m(t)dt\right] \\
&= -\sum_{m=1}^{M} \left(E\left[E\left\{\int_0^T \log[\mu_m(t)]dy_m(t) \middle| \mu_m(t), 0 \leq t < T\right\}\right]\right. \\
&\qquad \left. - E\left[\int_0^T \mu_m(t)dt\right]\right) \\
&= -\sum_{m=1}^{M} \left(E\left[\int_0^T \mu_m(t) \log[\mu_m(t)]dt\right] - \int_0^T E[\mu_m(t)]dt\right), \qquad (4.11)
\end{aligned}
$$

where the expectations in (4.11) are with respect to the stochastic rate functions $\mu_m(t)$.

We now derive the unconditional entropy, $h(Y)$, of the ordered arrival times. The output Y is a multi-channel, doubly-stochastic, Poisson process with density given

by ([58],pg. 425):

$$p_Y = \exp\left[\sum_{m=1}^{M}\left(\int_0^T \log[\widehat{\mu}_m(t)]dy_m(t) - \int_0^T \widehat{\mu}_m(t)dt\right)\right], \tag{4.12}$$

where

$$\widehat{\mu}_m(t) \equiv E\left[\mu_m(t) \mid y_1(\tau), \ldots, y_M(\tau), 0 \le \tau < t\right] = E\left[\mu_m(t) \mid Y(t)\right], \tag{4.13}$$

is the causal least-squares estimator of $\mu_m(t)$ based on the M output processes.

Using (4.10), (4.12), and noting that $E[\widehat{\mu}(t)] = E[\mu(t)]$, the unconditional entropy is

$$\begin{aligned}
h(Y) &= -E\left[\log p_Y\right] \\
&= -\sum_{m=1}^{M} E\left[\int_0^T \log[\widehat{\mu}_m(t)]dy_m(t) - \int_0^T \widehat{\mu}_m(t)dt\right] \\
&= -\sum_{m=1}^{M}\left(E\left[E\left\{\int_0^T \log[\widehat{\mu}_m(t)]dy_m(t)\middle| \widehat{\mu}_m(t), 0 \le t < T\right\}\right]\right. \\
&\quad \left. - E\left[\int_0^T \widehat{\mu}_m(t)dt\right]\right) \\
&= -\sum_{m=1}^{M}\left(E\left[\int_0^T \widehat{\mu}_m(t)\log[\widehat{\mu}_m(t)]dt\right] - \int_0^T E[\mu_m(t)]dt\right),
\end{aligned} \tag{4.14}$$

where the expectations in (4.14) are with respect to the stochastic rate functions $\widehat{\mu}_m(t)$ and $\mu_m(t)$ as appropriate.

Substituting (4.11) and (4.14) into (4.7) gives the mutual information for the MIMO Poisson channel,

$$I(X;Y) = \sum_{m=1}^{M} E\left[\int_0^T \left\{\mu_m(t)\log[\mu_m(t)] - \widehat{\mu}_m(t)\log[\widehat{\mu}_m(t)]\right\}dt\right]. \tag{4.15}$$

Paralleling the SISO capacity derivation by Davis [17], we manipulate the mutual

information as follows. Define

$$\phi_m(z) \equiv (\lambda_m + z)\log(\lambda_m + z) - \lambda_m \log \lambda_m, \tag{4.16}$$

and note that

$$\widehat{\mu}_m(t) = E\left[\sum_{n=1}^{N} \alpha_{nm} x_n(t) + \lambda_m \middle| Y(t)\right] = \sum_{n=1}^{N} \alpha_{nm}\widehat{x}_n(t) + \lambda_m, \tag{4.17}$$

where $\widehat{x}_n(t) \equiv E\left[\, x_n(t) \mid Y(t)\right]$. The mutual information is then

$$\begin{aligned}
& I(X;Y) \\
&= \sum_{m=1}^{M} E\left[\int_0^T \left\{\phi_m\left(\sum_{n=1}^{N}\alpha_{nm}x_n(t)\right) - \phi_m\left(\sum_{n=1}^{N}\alpha_{nm}\widehat{x}_n(t)\right)\right\} dt\right] \\
&= \sum_{m=1}^{M} E\left[\int_0^T \left\{\frac{\sum_{n=1}^{N}\alpha_{nm}\widehat{x}_n(t)}{R_m}\phi_m(R_m) - \phi_m\left(\sum_{n=1}^{N}\alpha_{nm}\widehat{x}_n(t)\right)\right.\right. \\
&\qquad \left.\left. - \frac{\sum_{n=1}^{N}\alpha_{nm}x_n(t)}{R_m}\phi_m(R_m) + \phi_m\left(\sum_{n=1}^{N}\alpha_{nm}x_n(t)\right)\right\} dt\right] \\
&= \sum_{m=1}^{M} E\left[\int_0^T \left\{h_m\left(\sum_{n=1}^{N}\alpha_{nm}\widehat{x}_n(t)\right) - h_m\left(\sum_{n=1}^{N}\alpha_{nm}x_n(t)\right)\right\} dt\right],
\end{aligned} \tag{4.18}$$

with

$$h_m(z) \equiv \frac{z}{R_m}\phi_m(R_m) - \phi_m(z), \tag{4.19}$$

$$R_m \equiv \sum_{n=1}^{N}\alpha_{nm}A_n. \tag{4.20}$$

Equation (4.18) is the MIMO generalization of the SISO mutual information expression derived in [33] and [17]. The function $h_m(R_m p)$ is shown in Figure 4-1 as a function of p for illustrative high and low background noise rates.

We have not been able to find the supremum in (4.6), but instead have derived upper and lower bounds. The derivations of these bounds, given below, exploit the

following rewritten form of (4.18):

$$\begin{aligned} & I(X;Y) \\ &= \sum_{m=1}^{M} \int_0^T \left\{ h_m\left(\sum_{n=1}^{N} \alpha_{nm} E[x_n(t)]\right) - E\left[h_m\left(\sum_{n=1}^{N} \alpha_{nm} x_n(t)\right)\right]\right\} dt \\ &\quad - \sum_{m=1}^{M} \int_0^T \left\{ h_m\left(\sum_{n=1}^{N} \alpha_{nm} E[x_n(t)]\right) - E\left[h_m\left(\sum_{n=1}^{N} \alpha_{nm} \widehat{x}_n(t)\right)\right]\right\} dt. \end{aligned} \tag{4.21}$$

We note that $h_m(z)$ is concave, which makes both summations positive, and that mutual information is non-negative which makes the first sum greater than the second.

The Parallel-Channel Upper Bound

An upper bound on the capacity can be found by maximizing the first term in (4.21) and ignoring the second term. Because the supremum of a sum cannot exceed the sum of the supremums, we will upper bound the mutual information by maximizing each term of the summation, i.e.,

$$\begin{aligned} C(\alpha) \;\leq\; & \sum_{m=1}^{M} \sup_{p_X \in \mathcal{X}} \frac{1}{T} \int_0^T \left\{ h_m\left(\sum_{n=1}^{N} \alpha_{nm} E[x_n(t)]\right) \right. \\ & \left. - E\left[h_m\left(\sum_{n=1}^{N} \alpha_{nm} x_n(t)\right)\right]\right\} dt. \end{aligned} \tag{4.22}$$

To maximize each term, fix the argument $\sum_{n=1}^{N} \alpha_{nm} E[x_n(t)]$ and consider the distribution that minimizes $E\Big[h_m\Big(\sum_{n=1}^{N} \alpha_{nm} x_n(t)\Big)\Big]$. Because $h_m(0) = h_m(R_m) = 0$, this distribution is concentrated at $x_1(t) = \cdots = x_N(t) = 0$ and $x_1(t) = A_1, \cdots, x_N(t) = A_N$ at each time t. Notice that this distribution by construction satisfies the peak transmit power constraint, and makes this expectation term zero.

Let $p_m(t) = \Pr(x_1(t) = A_1, \cdots, x_N(t) = A_N)$ for the m-th term. Noting that $E[x_n(t)] = A_n p_m(t)$, we now find the argument,

$$\sum_{n=1}^{N} \alpha_{nm} E[x_n(t)] = R_m p_m(t), \tag{4.23}$$

or, equivalently, the function $p_m(t)$, that maximizes the m-th term

$$\frac{1}{T}\int_0^T h_m[R_m p_m(t)]\,dt. \tag{4.24}$$

The average power constraint requires that $p_m(t)$ obey

$$\frac{1}{T}\int_0^T p_m(t)dt = \frac{1}{T}\int_0^T \frac{E[x_n(t)]}{A_n}dt \le \sigma. \tag{4.25}$$

Using the concavity of $h_m[R_m\,p\,(t)]$, the Taylor series expansion with respect to $R_m p_m\,(t)$ about a point $R_m p_m^*$, $0 \le p_m^* \le 1$, yields

$$h_m[R_m p_m(t)] - h_m(R_m p_m^*) \le h_m'(R_m p_m^*)\,[R_m p_m(t) - R_m p_m^*]\,, \tag{4.26}$$

where

$$\begin{aligned}
h_m'(z) &\equiv \frac{d}{dz}h_m(z) \\
&= \frac{\phi_m(R_m)}{R_m} - \log\left[e(z+\lambda_m)\right] \\
&= \frac{(\lambda_m + R_m)\log(\lambda_m + R_m) - \lambda_m \log\lambda_m}{R_m} - \log\left[e(z+\lambda_m)\right] \\
&= (1+s_m)\log(1+s_m) - s_m\log s_m - \log\left[e\left(\frac{z}{R_m}+s_m\right)\right] \\
&= \log\left[\frac{(1+s_m)^{(1+s_m)}}{e s_m^{s_m}\left(\frac{z}{R_m}+s_m\right)}\right],
\end{aligned} \tag{4.27}$$

and $s_m \equiv \frac{\lambda_m}{R_m}$ is the ratio of background noise to peak received power at the m-th receiver, i.e., it is a reciprocal signal-to-noise ratio. Calling $1/s_m$ a signal-to-noise ratio is somewhat of a misnomer. Due to the inherent randomness of the photon arrivals, the signal itself has a noise-like component that is not included in this signal-to-noise ratio. The ratio simply relates the average peak receive power to the background noise power.

Averaging over the interval $[0, T)$ yields

$$\begin{aligned} \frac{1}{T}\int_0^T h_m\left[R_m p_m(t)\right] dt &- h_m(R_m p_m^*) \\ &\leq R_m h_m'(R_m p_m^*)\left(\frac{1}{T}\int_0^T p_m(t)dt - p_m^*\right). \end{aligned} \tag{4.28}$$

Let $p_m^* = p_m^{\text{opt}} \equiv \min(p_m^{\max}, \sigma)$ where

$$p_m^{\max} = \frac{(1+s_m)^{(1+s_m)}}{e s_m^{s_m}} - s_m, \tag{4.29}$$

is the value of p that maximizes $h_m(R_m p)$ found by solving $h_m'(R_m p_m^{\max}) = 0$ in (4.27). If $p_m^{\max} \leq \sigma$, then $p_m^{\text{opt}} = p_m^{\max}$. Consequently, $h_m'(R_m p_m^{\max}) = 0$ by definition, and the right hand side of (4.28) is zero, making

$$\frac{1}{T}\int_0^T h_m\left[R_m p_m(t)\right] dt \leq h_m(R_m p_m^{\text{opt}}), \tag{4.30}$$

with equality when $p_m(t) \equiv p_m^{\max} = p_m^{\text{opt}}$.

Similarly, when $\sigma < p_m^{\max}$, then $h_m'(R_m\,\sigma) > 0$ due to strict concavity. Furthermore, the average power constraint (4.25) makes the parenthetical term on the right hand side of (4.28) non-positive when $p_m^* = \sigma$. Consequently, (4.30) is also true, holding with equality when $p_m(t) \equiv \sigma = p_m^{\text{opt}}$.

We can interpret the probability p_m^{opt} as the optimal duty cycle of the m-th MISO channel. Without the average power constraint, $p_m^{\max}$ would be the optimal duty cycle. With a bound on average power, however, we must reduce the duty cycle to the minimum of σ and $p_m^{\max}$ to satisfy this constraint.

We now have our parallel-channel upper bound on the instantaneous capacity $C(\alpha)$, viz.,

$$C(\alpha) \leq C_{\text{PC-UB}}(\alpha) = \sum_{m=1}^{M} h_m\left(\sum_{n=1}^{N} \alpha_{nm} A_n p_m^{\text{opt}}\right), \tag{4.31}$$

where our parallel-channel terminology is due to this expression's being the sum of the capacities of M independent N-by-1 MISO channels, to be shown shortly. This terminology becomes clearer by expressing the mutual information via entropy chain

rule and conditioning inequalities [14] as,

$$\begin{aligned} I(Y_1,\ldots,Y_M;X) &= h(Y_1,\ldots,Y_M) - h(Y_1,\ldots,Y_M \mid X) \\ &= h(Y_1,\ldots,Y_M) - \sum_{m=1}^{M} h(Y_m \mid X) \\ &= \sum_{m=1}^{M} [h(Y_m \mid Y_1,\ldots,Y_{m-1}) - h(Y_m \mid X)] \\ &\leq \sum_{m=1}^{M} [h(Y_m) - h(Y_m \mid X)] \\ &= \sum_{m=1}^{M} I(Y_m;X). \end{aligned} \tag{4.32}$$

We could then maximize each individual MISO mutual information term.

Equation (4.32) brings up an interesting question regarding a possibly simpler derivation of the parallel-channel upper bound using entropy properties. This approach, however, requires calculating the capacity of a MISO Poisson channel with peak and average power constraints. Lapidoth in [40] made a brief comment on the capacity of a MISO Poisson channel in his conclusions. He compared the multiple-access Poisson channel maximum throughput to the single-user MISO channel capacity without regard to an average power constraint or background noise. Furthermore, the peak power constraint was an aggregate constraint at the receiver. We will formally calculate the MISO Poisson channel capacity by showing that our OOK lower bound coincides with the parallel-channel upper bound for a single receiver ($M = 1$).

We will simplify the function $h_m(R_m p_m^{\text{opt}})$ to emphasize the capacity bound depends on three quantities:

- Received peak power, R_m
- Background noise-to-signal ratio, s_m
- Optimal OOK duty cycle, p_m^{opt}

To see this, substitute (4.16) into (4.19) and simplify

$$\begin{aligned} h_m(R_m p) &= p\,[(\lambda_m + R_m)\log(\lambda_m + R_m) - \lambda_m \log \lambda_m] \\ &\quad - [(\lambda_m + R_m p\,)\log(\lambda_m + R_m p\,) - \lambda_m \log \lambda_m] \\ &= p\,R_m(1+s_m)\log[R_m(1+s_m)] - p\,R_m s_m \log(R_m s_m) \\ &\quad -R_m(p+s_m)\log[R_m(p+s_m)] + R_m s_m \log(R_m s_m) \\ &= R_m[p\,(1+s_m)\log(1+s_m) + (1-p\,)s_m \log s_m \\ &\quad -(p+s_m)\log(p+s_m)] \qquad (4.33) \\ &= R_m \mathcal{I}(p, s_m), \qquad (4.34) \end{aligned}$$

where

$$\mathcal{I}\,(p,r) \equiv p\,(1+r)\log(1+r) + (1-p\,)\,r\log r - (\,p+r\,)\log(\,p+r\,), \qquad (4.35)$$

is an information function with units of nats per photon for a given duty cycle p and noise-to-signal ratio r. We can now rewrite the parallel-channel upper bound as

$$C_{\mathrm{PC-UB}}(\alpha) = \sum_{m=1}^{M} R_m\,\mathcal{I}\left(p_m^{\mathrm{opt}}, s_m\right). \qquad (4.36)$$

The On-Off Keying Lower Bound

A lower bound on the channel capacity can be found by restricting the supremum in (4.6) to a subset of $\mathcal{X}$, namely those input distributions that, in addition to the peak and average power constraints, also satisfy $x_1(t) = A_1, \cdots, x_N(t) = A_N$ or $x_1(t) = \cdots = x_N(t) = 0$ for each time t. As in [33] and [17], switching between these "on" and "off" states arbitrarily fast reduces the causal least-squares estimator, $\widehat{x}_n(t)$, to the unconditional mean, $E[x_n(t)]$, which makes the second term in (4.21) vanish. For details, see [8].

Because $h_m(0) = h_m(R_m) = 0$, the mutual information for rapidly switching OOK

inputs reduces (4.21) to

$$I(X;Y) = \int_0^T \left(\sum_{m=1}^{M} h_m \left[R_m p(t) \right] \right) dt, \tag{4.37}$$

where $p(t) = \Pr(x_1(t) = A_1, \cdots, x_N(t) = A_N)$ is the probability that the inputs are "on" at time t. We now find the duty cycle $p(t)$ that maximizes the mutual information (4.37).

For convenience, we write the mutual information explicitly as a function of the duty cycle, defining

$$g[p\,(t)] \equiv \sum_{m=1}^{M} h_m \left[R_m p\,(t) \right]. \tag{4.38}$$

Expanding $g[p\,(t)]$ with respect to $p\,(t)$ about a point p^*, $0 \leq p^* \leq 1$, and using the concavity of $g[p\,(t)]$ gives

$$g[p\,(t)] - g(p^*) \leq g'(p^*)[p\,(t) - p^*], \tag{4.39}$$

where

$$\begin{aligned} g'(p) &= \frac{d}{dp} g(p) \\ &= \sum_{m=1}^{M} R_m h_m'\left(R_m p\right) \\ &= \sum_{m=1}^{M} R_m \log \left[\frac{(1+s_m)^{(1+s_m)}}{e s_m^{s_m} (p + s_m)} \right] \\ &= \log \left\{ \prod_{m=1}^{M} \left[\frac{(1+s_m)^{(1+s_m)}}{e s_m^{s_m} (p + s_m)} \right]^{R_m} \right\}. \end{aligned} \tag{4.40}$$

Let $p^* = p^{\text{opt}} \equiv \min(p^{\max}, \sigma)$ where $p^{\max}$ is the value of p that maximizes $g(p\,)$, found by solving

$$\prod_{m=1}^{M} \left[\frac{(1+s_m)^{(1+s_m)}}{e s_m^{s_m} (p^{\max} + s_m)} \right]^{R_m} = 1. \tag{4.41}$$

A closed form solution for this maximizing probability only exists in a few special cases (e.g. $M = 1$ or $s_m \equiv s$, $\forall m$). In general, numerical methods are needed to

calculate $p^{\max}$. We will see shortly, however, that in the low and high noise regimes, this maximizing probability converges to $1/e$ and $1/2$, respectively.

Similar to the derivation of the parallel-channel upper bound, we obtain

$$\frac{1}{T}\int_0^T \left(\sum_{m=1}^{M} h_m[R_m p\,(t)]\right) dt \leq \sum_{m=1}^{M} h_m(R_m p^{\text{opt}}), \tag{4.42}$$

with equality when $p\,(t) \equiv p^{\text{opt}}$.

We now have a lower bound on the MIMO capacity $C(\alpha)$ for a given channel realization,

$$C(\alpha) \geq C_{\text{OOK-LB}}(\alpha) = \sum_{m=1}^{M} h_m\left(\sum_{n=1}^{N} \alpha_{nm} A_n p^{\text{opt}}\right), \tag{4.43}$$

where $C_{\text{OOK-LB}}$ is the capacity of using OOK signaling with arbitrarily fast toggling. We can rewrite this lower bound using the normalized information function (4.35) as

$$C_{\text{OOK-LB}}(\alpha) = \sum_{m=1}^{M} R_m \mathcal{I}\left(p^{\text{opt}}, s_m\right). \tag{4.44}$$

Comparison of PC-UB and OOK-LB Bounds

Figure 4-1 shows the relationship between the upper and lower bounds on the MIMO channel capacity. Assuming no average energy constraint ($\sigma = 1$), the PC-UB is the sum of the concave functions $h_m(R_m p)$ evaluated at their respective maxima, whereas the OOK-LB is the maximum of their sum.

Davis [17] and Wyner [67] have shown (see Appendix A.1) that the maximum of $h_m(R_m p)$ occurs between $1/e$ for high signal-to-noise ratio ($s_m \to 0$), and $1/2$ for low signal-to-noise ratio ($s_m \to \infty$). By the concavity in p of the functions $h_m(R_m p)$, the maximizer of their scaled sum in the lower bound, $p^{\max}$, lies between the largest and smallest maximizers of the individual terms in the upper bound, i.e.,

$$\frac{1}{e} \leq \min_{1\leq m\leq M} p_m^{\max} \leq p^{\max} \leq \max_{1\leq m\leq M} p_m^{\max} \leq \frac{1}{2}, \tag{4.45}$$

which, in turn, implies that the PC-UB and OOK-LB are equal in both the high

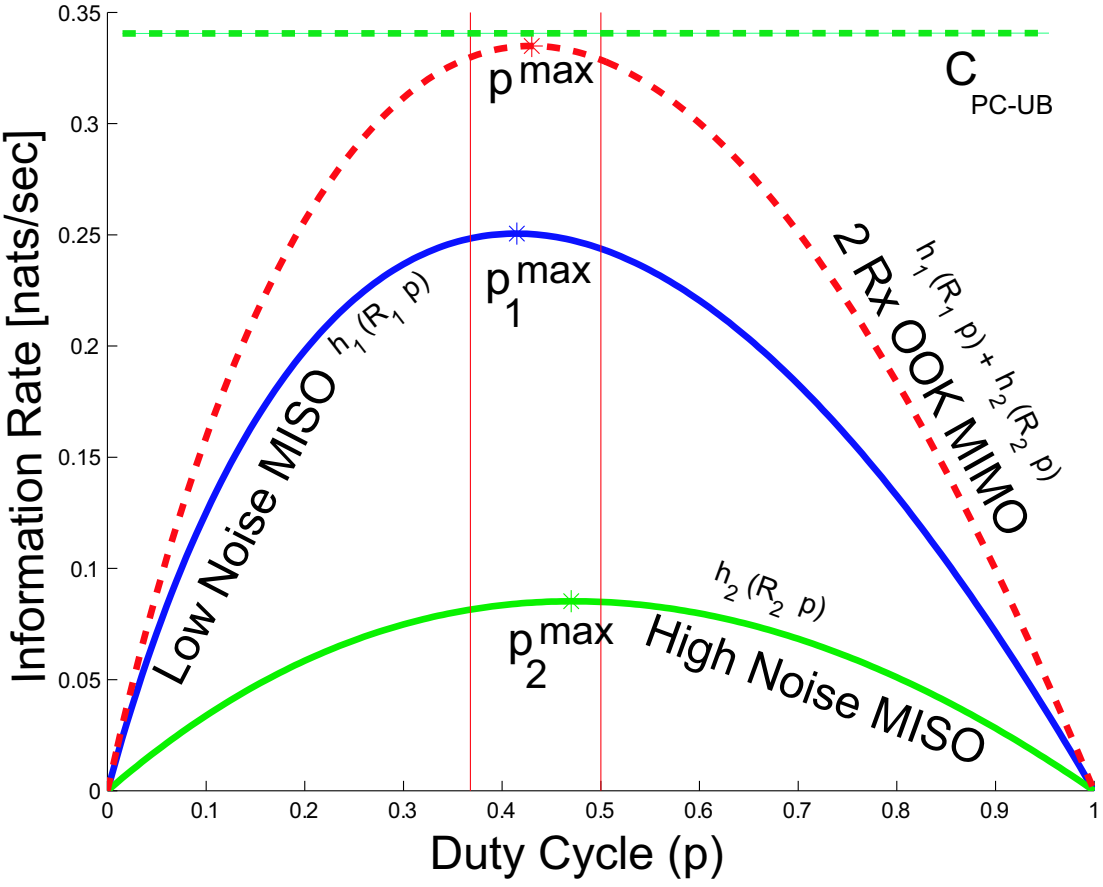

Figure 4-1: With no average power constraint, the the parallel-channel upper bound (PC-UB) is the sum of concave functions evaluated at their respective maxima, whereas the the OOK lower bound (OOK-LB) is the maximum of the their sum. In this two receive aperture ($M = 2$) example, $C_{\mathrm{PC-UB}} = h_1(R_1 p_1^{\max}) + h_2(R_2 p_2^{\max})$, and $C_{\mathrm{OOK-LB}} = h_1(R_1 p^{\max}) + h_2(R_2 p^{\max})$.

and low signal-to-noise ratio limits. Hence, in the low and high noise regimes, the upper and lower bounds converge, i.e., $p_m^{\max} = p^{\max}$ for $1 \leq m \leq M$, and we have a closed-form expression for the capacity of a given channel realization. See Figure 4-2 for a plot of the maximizing duty cycle $p_m^{\max}$ as a function of the noise-to-signal ratio s_m.

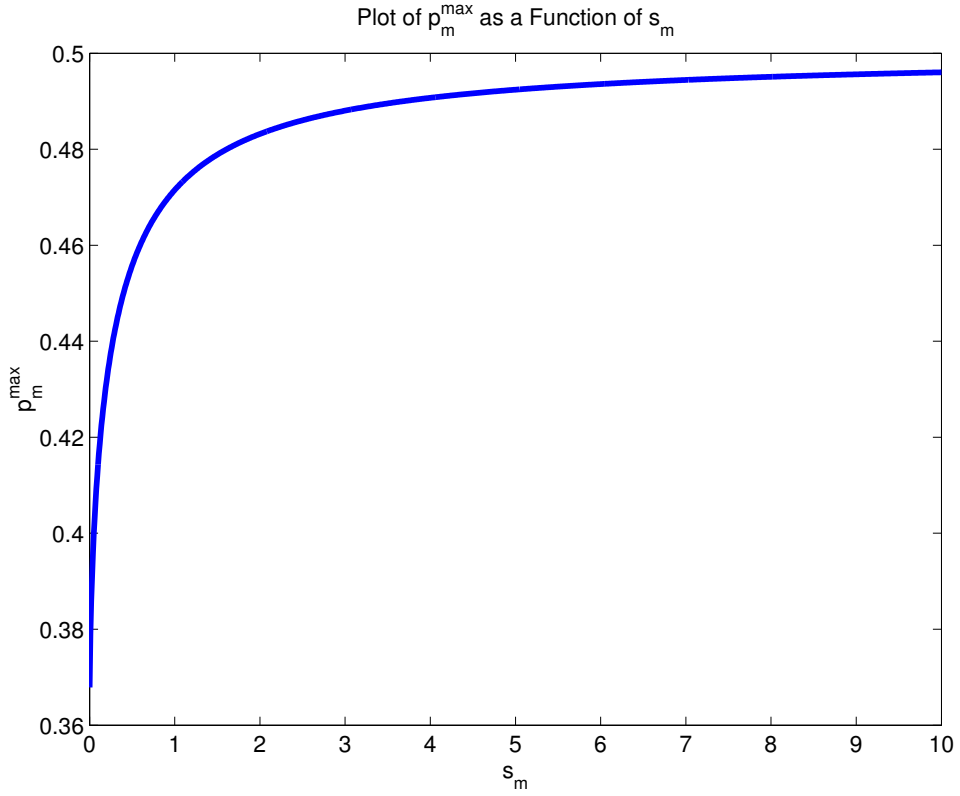

Figure 4-2: The maximum of $h_m(R_m p)$ as a function of s_m lies between $1/e \approx 0.3679$ for low noise ($s_m \to 0$) and $1/2$ for high noise ($s_m \to \infty$).

At high signal-to-noise ratio, i.e., the low noise regime in which $s_m \to 0$, $1 \leq m \leq M$, the information function in (4.35) becomes (see Appendix A.2)

$$\lim_{r \to 0} \mathcal{I}\,(p, r) = p \, \log 1/p, \tag{4.46}$$

and $p^{\text{opt}} \to \min(1/e, \sigma)$ becomes the optimal OOK duty cycle of the capacity-achieving distribution. The instantaneous capacity for a given channel realization known to

both the transmitter and receiver is therefore,

$$C(\alpha) = \left(p^{\text{opt}} \log 1/p^{\text{opt}}\right) \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha_{nm} A_n. \tag{4.47}$$

At low signal-to-noise ratio, i.e., the high noise regime in which $s_m \to \infty$, $1 \leq m \leq M$, the optimal transmit OOK duty cycle $p^{\text{opt}} \to \min(1/2, \sigma)$. In this regime, the information function becomes [17, 67] (see Appendix A.2)

$$\mathcal{I}\left(p, r\right) \approx p\left(1-p\right)/2r. \tag{4.48}$$

As a result, the instantaneous capacity in the high noise regime is

$$C(\alpha) = \frac{p^{\text{opt}}(1-p^{\text{opt}})}{2} \sum_{m=1}^{M} \frac{1}{\lambda_m} \left(\sum_{n=1}^{N} \alpha_{nm} A_n \right)^2. \tag{4.49}$$

In addition to the preceding high and low signal-to-noise regimes, there are several other special cases for which $C_{\text{OOK}-\text{LB}}(\alpha) = C(\alpha) = C_{\text{PC}-\text{UB}}(\alpha)$ prevails:

- the MISO channel ($M = 1$ case), in which case $p_1^{\text{opt}} = p^{\text{opt}}$;
- the flat signal-to-noise ratio case ($s_1 = \cdots = s_M$), in which case $p_1^{\text{opt}} = \cdots = p_M^{\text{opt}} = p^{\text{opt}}$;
- and the low average-power case ($\sigma \leq \min_{1 \leq m \leq M} p_m^{\max}$), in which case $p_1^{\text{opt}} = \cdots = p_M^{\text{opt}} = p^{\text{opt}} = \sigma$. Because imposing an average power constraint can only make the bounds closer, in many of the examples, we will not impose this constraint, and set $\sigma = 1$.

The parallel-channel upper bound also coincides with the capacity when the channel decouples, i.e., $\alpha_{nm} = g_n \delta_{nm}$, where g_n is the path gain for the n-th sub-channel, and δ_{nm} is the Kronecker delta.

For the SIMO Poisson channel (the $N = 1$ case), we have that $C(\alpha) = C_{\text{OOK}-\text{LB}}(\alpha)$, because the optimal distribution at any given time satisfies the OOK constraint. To see this, we can ignore the second term in (4.21) and write an upper bound on the

α_{nm}	$m=1$	$m=2$	$m=3$
$n=1$	0.2977	0.5760	0.1279
$n=2$	0.0692	0.6979	2.0322

Table 4.1: Path gains for the upper and lower bound comparison in Figure 4-3

mutual information as

$$I(X;Y) \leq \sum_{m=1}^{M} \int_0^T \{h_m(\alpha_{1m} E[x_1(t)]) - E[h_m(\alpha_{1m} x_1(t))]\}\, dt \tag{4.50}$$

To minimize this upper bound, we can fix $E[x_1(t)]$ and minimize the second term by choosing an OOK distribution on $x_1(t)$ with duty cycle $p(t) = \Pr(x_1(t) = 1)$. Following a similar development as the lower bound, we find that the optimal duty cycle is $p(t) \equiv p^{\text{opt}} = \min(p^{\max}, \sigma)$ where $p^{\max}$ is given by (4.41). This upper bound coincides the OOK lower bound for a single transmitter ($N = 1$).

In general, when the bounds are not equal, they are usually quite close, as can be seen from Figure 4-3, which was computed for the $N = 2$, $M = 3$ special case using the arbitrarily-chosen path gains given in Table 4.1.

Photon-Bucket Receivers

So far we have assumed that both the transmitter and receiver had knowledge of the path gains, and that they made optimal use of this information. We saw that at both high and low signal-to-noise ratios, however, the transmitter did not need to know the path gains to achieve capacity. We will now demonstrate that a "photon-bucket" or aperture-averaging receiver [49], which does not require path gain knowledge, can achieve capacity in the low noise regime and in the high noise regime for a moderate number of transmit apertures.

Instantaneous Capacity We will consider a photon-bucket (PB) receiver, shown in Figure 4-4, that adds the photon counts from the M detectors to form the doubly-

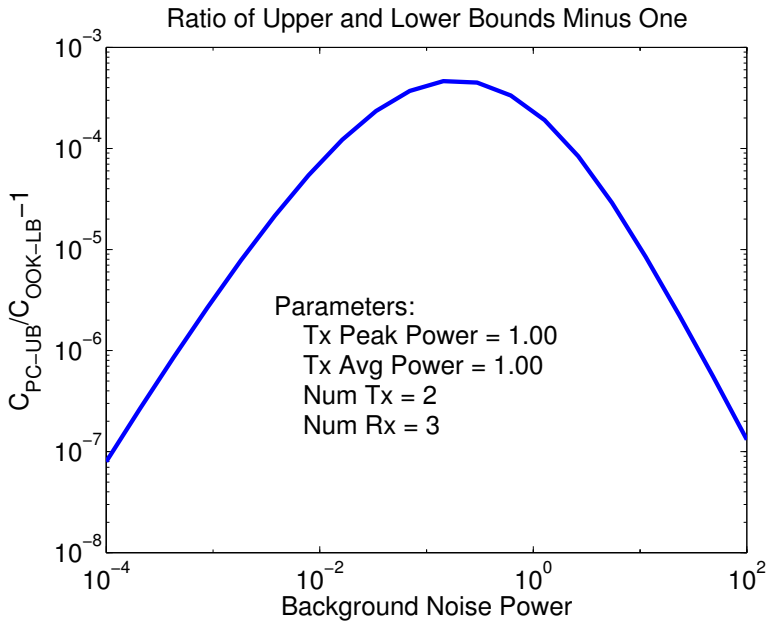

Figure 4-3: In general, the upper and lower bounds on channel capacity are quite close. This figure shows the fractional difference between the PC-UB and OOK-LB for the $N = 2$, $M = 3$ special case whose path gains are given in Table 1.

stochastic Poisson process $y(t) = \sum_{m=1}^{M} y_m(t)$ with rate

$$\mu(t) = \sum_{m=1}^{M} \mu_m(t) = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha_{nm} x_n(t) + \sum_{m=1}^{M} \lambda_m. \tag{4.51}$$

This receiver structure is the optical analog of the equal-gain combiner in microwave communications [32].

We will now determine a lower bound to capacity utilizing a photon-bucket receiver for a given channel realization known only to the transmitter. Following Wyner's approach in [67], we partition the interval $[0, T]$ into smaller, non-overlapping sub-intervals of length Δ, in which all transmit lasers are either on, $x_n(t) = A_n$, $1 \leq n \leq N$, or off $x_n(t) = 0$, $1 \leq n \leq N$. In each sub-interval, the receiver declares a "1" (all transmitters on) if exactly one photon count occurred in that interval, and a "0" (all transmitters off) otherwise. For a given channel realization α, this formulation reduces the channel to a binary-input, binary-output, discrete memoryless channel (DMC) [21] depicted in Figure 4-5.

Denote the binary input and output of this DMC during the k-th sub-interval,

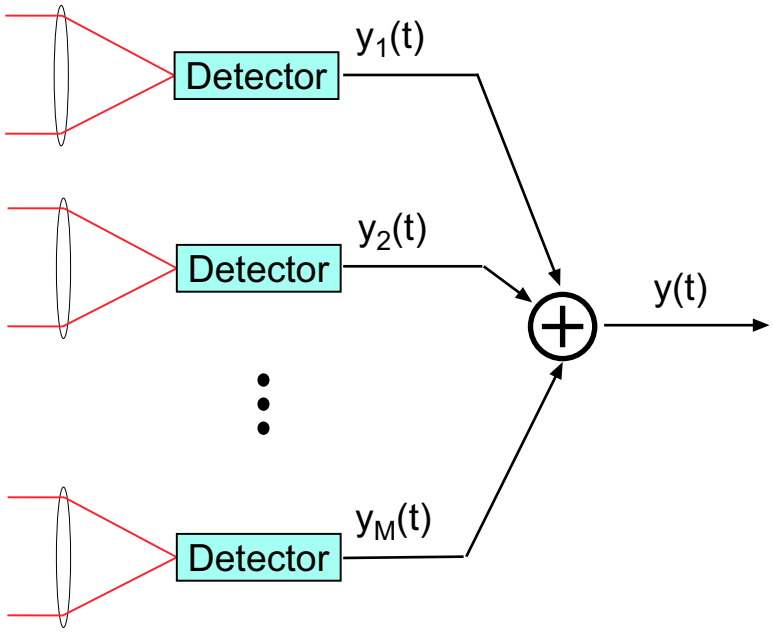

Figure 4-4: A photon-bucket receiver adds the photon counts from the M detectors to form a doubly-stochastic Poisson process.

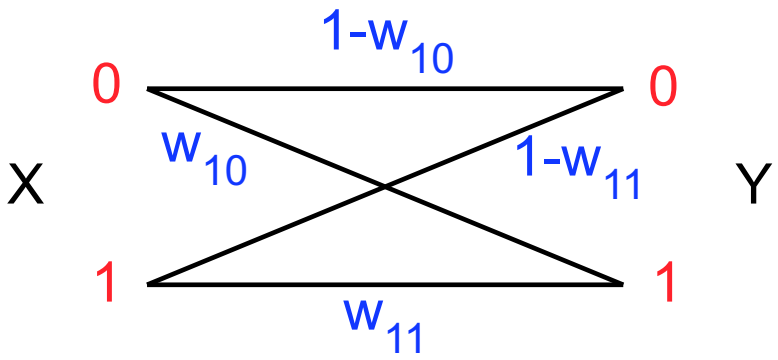

Figure 4-5: Using an OOK transmitter and photon-bucket receiver with a threshold decision rule creates a binary-input, binary-output discrete memoryless channel.

$(k-1)\Delta \leq t \leq k\Delta$, as $X_k \in \{0,1\}$ and $Y_k \in \{0,1\}$, respectively. For sufficiently small Δ, the resulting channel is described by the conditional probabilities

$$w_{10} \equiv \Pr(Y_k = 1 \mid X_k = 0) \approx \Delta L_1 = \Delta R s \tag{4.52}$$

$$w_{11} \equiv \Pr(Y_k = 1 \mid X_k = 1) \approx \Delta(R + L_1) = \Delta R(1+s), \tag{4.53}$$

where $R = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha_{nm} A_n$ is the aggregate received peak power, $s \equiv L_1/R$ is an aggregate noise-to-signal ratio, and $L_1 \equiv \sum_{m=1}^{M} \lambda_m$ is the aggregate background noise. Let q be the probability that all transmitters are on, i.e., $q \equiv \Pr(X_k = 1)$. We want choose the duty cycle q to maximize the mutual information of the binary input, binary output DMC, namely

$$I(X;Y) = \mathcal{H}(w_{10}(1-q) + w_{11}q) - \mathcal{H}(w_{10})(1-q) - \mathcal{H}(w_{11})q, \tag{4.54}$$

subject to the average power constraint

$$\frac{1}{T}\int_0^T E[x_n(t)/A_n]dt = \frac{1}{K}\sum_{k=1}^{K} E[X_k] = q \leq \sigma, \tag{4.55}$$

where $K \equiv T/\Delta$ is the number of sub-intervals and $\mathcal{H}(u) \equiv -u\log u - (1-u)\log(1-u)$ is the binary entropy function. By construction, the OOK transmitted waveform satisfies the peak power constraint, $0 \leq x_n(t) \leq A_n$, for $1 \leq n \leq N$.

As we reduce the sub-interval length Δ to zero, the channel probabilities $w_{10} \approx \Delta R s$ and $w_{11} \approx \Delta R(1+s)$ also approach zero. We note that for small values of u that the approximation $\mathcal{H}(u) \approx -u\log u + u$ holds. Using this approximation and

simplifying, we can write the mutual information for small Δ as

$$\begin{aligned} I(X;Y) &\approx -[\Delta Rs(1-q)+\Delta R(1+s)q]\log[\Delta Rs(1-q)+\Delta R(1+s)q] \\ &\quad +\Delta Rs(1-q)+\Delta R(1+s)q \\ &\quad +\{\Delta Rs\log[\Delta Rs]-\Delta Rs\}(1-q) \\ &\quad +\{\Delta R(1+s)\log[\Delta R(1+s)]-\Delta R(1+s)\}q \\ &= \Delta R\,[\,(s+q)\{-\log[\Delta R(s+q)]+1\} \\ &\quad +s\{\log[\Delta Rs]-1\}(1-q)+(1+s)\{\log[\Delta R(1+s)]-1\}]q \\ &= \Delta R\,[\,q\,(1+s)\log(1+s)+(1-q\,)\,s\log s-(s+q\,)\log(s+q\,)\,] \\ &= \Delta R\,\mathcal{I}(q,s), \end{aligned} \tag{4.56}$$

where

$$\mathcal{I}\,(p,r)\equiv p\,(1+r)\log(1+r)+(1-p\,)r\log r-(r+p\,)\log(r+p\,), \tag{4.57}$$

as previously defined. Because ΔR has the units of photons, from this expression we get our interpretation that $\mathcal{I}(p,r)$ is an information function with units of nats per photon for a given duty cycle p and noise-to-signal ratio r.

The instantaneous capacity of this OOK transmitter and photon-bucket receiver configuration is the maximum mutual information per unit time for a given channel realization subject to the average power constraint $q \leq \sigma$. Because the OOK transmitter constraint can only decrease capacity, we have the lower bound

$$C_{\mathrm{PB-LB}}(\alpha) \geq \max_{0\leq q\leq\sigma}\lim_{\Delta\to 0} I(X;Y)/\Delta = \max_{0\leq q\leq\sigma} R\,\mathcal{I}(q,s). \tag{4.58}$$

We emphasize that the right-hand side is a lower bound to the actual photon-bucket receiver capacity, $C_{\mathrm{PB-LB}}(\alpha)$, because we imposed an additional transmitter OOK constraint. Furthermore, on the left-hand side, we use the subscript notation LB to emphasize that this photon-bucket receiver capacity is a lower bound to the instantaneous capacity $C(\alpha)$ with no imposed transmitter or receiver structure, i.e.,

$C(\alpha) \geq C_{\text{PB-LB}}(\alpha)$.

The unconstrained maximum of $\mathcal{I}(q,s)$ with respect to q, found by differentiation with respect to q, is

$$q^{\max} = \frac{(1+s)^{(1+s)}}{es^s} - s. \tag{4.59}$$

Furthermore, $\mathcal{I}(q,s)$ is a concave function of q; therefore its constrained maximum occurs at the minimum of σ and $q^{\max}$, i.e., $q^{\text{opt}} = \min(q^{\max}, \sigma)$. As a result, the instantaneous photon-bucket capacity is bounded by

$$C_{\text{PB-LB}}(\alpha) \geq R\mathcal{I}(q^{\text{opt}}, s). \tag{4.60}$$

We can show that the additional OOK transmitter constraint did not actually decrease capacity, i.e., $C_{\text{PB-LB}}(\alpha) \leq R\mathcal{I}(q^{\text{opt}}, s)$, by using the capacity result for a SISO Poisson channel [33, 17, 67]. Because all transmitters are either on or off at the same time, and the photon-bucket receiver adds the counts from each detector, we essentially have created a SISO Poisson channel with rate of $\bar{x}(t) = \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}x_n(t)$ plus a background noise $L_1 = \sum_{m=1}^{M}\lambda_m$. The capacity of a SISO Poisson channel with peak and average power constraints $0 \leq \bar{x}(t) \leq R = \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A_n$ and $\frac{1}{T}\int_0^T E[\bar{x}(t)]dt \leq \sigma R$, where $0 \leq \sigma \leq 1$, is $R\mathcal{I}(q^{\text{opt}}, s)$, assuming a given channel realization α.

The "aggregate" power constraints considered in this SISO case, however, are different than our original "per transmitter" peak and average transmit constraints. We notice, though, that waveforms satisfying our peak power constraint $0 \leq x_n(t) \leq A_n$ for $1 \leq n \leq N$, automatically satisfy the SISO peak power constraint $0 \leq \bar{x}(t) \leq R$. Similarly, waveforms that satisfy our average power constraint $\frac{1}{T}\int_0^T E[x_n(t)]dt \leq \sigma A_n$ for $1 \leq n \leq N$ also satisfy the SISO average power constraint $\frac{1}{T}\int_0^T E[\bar{x}(t)]dt \leq \sigma R$. We see that our original transmitter power constraints are stricter than those required for the SISO Poisson channel capacity result. Hence, this SISO Poisson

channel capacity is an upper bound to the photon-bucket receiver capacity, and

$$C_{\mathrm{PB-LB}}(\alpha) = R\mathcal{I}(q^{\mathrm{opt}}, s) = \left(\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A_n\right)\mathcal{I}(q^{\mathrm{opt}}, s). \tag{4.61}$$

At high signal-to-noise ratio, i.e., when $s \to 0$, we use the limit (4.46) in the photon-bucket instantaneous capacity (4.61) to show that it converges to the optimal receiver capacity (4.47). In this low noise regime, therefore, using photon-bucket reception does not reduce capacity.

At low signal-to-noise ratio, i.e., when $s \to \infty$, we use the approximation (4.48) to write the photon-bucket instantaneous capacity as

$$C_{\mathrm{PB-LB}}(\alpha) = \frac{q^{\mathrm{opt}}(1-q^{\mathrm{opt}})/2}{\sum_{m=1}^{M}\lambda_m}\left(\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A_n\right)^2. \tag{4.62}$$

In this regime, photon-bucket receivers are suboptimal. The photon-bucket capacity in high noise is related to the actual capacity (4.49) through Jensen's inequality,

$$\begin{aligned}
C_{\mathrm{PB-LB}}(\alpha) &= \frac{q^{\mathrm{opt}}(1-q^{\mathrm{opt}})/2}{\sum_{m=1}^{M}\lambda_m}\left(\sum_{m=1}^{M}\left[\frac{\lambda_m}{\sum_{l=1}^{M}\lambda_l}\right]\left[\frac{\sum_{l=1}^{M}\lambda_l}{\lambda_m}\right]\sum_{n=1}^{N}\alpha_{nm}A_n\right)^2 \\
&\leq \frac{q^{\mathrm{opt}}(1-q^{\mathrm{opt}})/2}{\sum_{m=1}^{M}\lambda_m}\sum_{m=1}^{M}\left[\frac{\lambda_m}{\sum_{l=1}^{M}\lambda_l}\right]\left(\left[\frac{\sum_{l=1}^{M}\lambda_l}{\lambda_m}\right]\sum_{n=1}^{N}\alpha_{nm}A_n\right)^2 \\
&= \frac{q^{\mathrm{opt}}(1-q^{\mathrm{opt}})}{2}\sum_{m=1}^{M}\frac{1}{\lambda_m}\left(\sum_{n=1}^{N}\alpha_{nm}A_n\right)^2 \\
&= C(\alpha).
\end{aligned} \tag{4.63}$$

Because squaring is a strictly convex function, equality occurs when ([14], pg. 25) for $1 \leq m \leq M$

$$\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A_n = \left[\frac{\sum_{l=1}^{M}\lambda_l}{\lambda_m}\right]\sum_{n=1}^{N}\alpha_{nm}A_n. \tag{4.64}$$

Even though photon-buckets are suboptimal in the high noise regime, we will show in Section 4.1.2 that the average capacity gained through optimal path gain use at the receiver vanishes when using moderate numbers of transmit apertures. This result

Bound	Formula	High SNR Approx.	Low SNR Approx.
PC-UB	$\sum_{m=1}^{M} R_m \mathcal{I}\left(p_m^{\text{opt}}, s_m\right)$	$\left(p^{\text{opt}} \log 1/p^{\text{opt}}\right) \sum_{m=1}^{M} R_m$	$\frac{p^{\text{opt}}(1-p^{\text{opt}})}{2} \sum_{m=1}^{M} \frac{R_m^2}{\lambda_m}$
OOK-LB	$\sum_{m=1}^{M} R_m \mathcal{I}\left(p^{\text{opt}}, s_m\right)$	Same as PC-UB	Same as PC-UB
PB-LB	$\left(\sum_{m=1}^{M} R_m\right) \mathcal{I}(q^{\text{opt}}, s)$	Same as PC-UB	$\frac{q^{\text{opt}}(1-q^{\text{opt}})}{2} \frac{\left(\sum_{m=1}^{M} R_m\right)^2}{\sum_{m=1}^{M} \lambda_m}$

Table 4.2: This table shows bounds on the MIMO Poisson channel capacity for a given channel realization. The photon-bucket lower bound assumes a receiver structure that does not use path gain knowledge. The other two bounds assume the receiver knows and uses the path gains optimally. All three bounds assume that the transmitter optimally uses path gain knowledge. In the low and high signal-to-noise ratio (SNR) regimes, however, the transmit optimal duty cycles do not require path gain knowledge. In the high and low signal-to-noise ratio regimes, the optimal duty cycles converge, i.e., $p_m^{\text{opt}} = p^{\text{opt}} = q^{\text{opt}}$, and equal $\min(1/e, \sigma)$ and $\min(1/2, \sigma)$, respectively.

becomes intuitive by examining the equality condition (4.64). For identical transmitters, $A_n \equiv A$, and receivers, $\lambda_m \equiv \lambda$, increasing the number of transmit apertures (N) causes the sum $\sum_{n=1}^{N} \alpha_{nm} A/N$ to converge to its expected value, $E[\alpha_{nm}]A = A$. The left- and right-hand sides of (4.64) are then roughly equal for large numbers of transmit apertures. For large numbers of receive apertures and small numbers of transmit apertures, photon-buckets are suboptimal, however.

Table 4.2 highlights the similarities and differences between the parallel-channel upper bound (PC-UB), OOK lower bound (OOK-LB), and photon-bucket lower bound (PB-LB) on the MIMO Poisson channel capacity for a given channel realization.

Examining the Fast Toggling Assumption The preceding capacity developments did not impose a bandwidth constraint on the transmit waveform. As a result, we saw that the capacity achieving waveforms were OOK with arbitrarily fast toggling. Shamai in [53] considered the capacity of a SISO Poisson channel with limits on how fast the input waveform can toggle. Also, Shamai and Lapidoth in [55] developed capacity bounds for general spectral constraints.

The goal of this section is not to find the capacity of the spectrally-constrained MIMO Poisson channel. Instead, we seek to examine the behavior of the photon-

bucket capacity result (4.61) as the interval size Δ approaches zero in the absence of fading. This capacity is essentially the capacity of the SISO Poisson channel created from OOK transmission and unit-threshold decisions.

We start by reexamining the mutual information of the binary-input, binary-output, discrete-memoryless channel (4.54). The input distribution $q = \Pr(X_k = 1)$ that maximizes the mutual information is found by differentiating it, setting the result equal to zero, and solving for q. We note that the derivative of the binary entropy function is

$$\frac{d\mathcal{H}(u)}{du} \equiv \mathcal{H}'(u) = \log\left(\frac{1-u}{u}\right). \tag{4.65}$$

Consequently, the derivative of the mutual information with respect to the duty cycle is

$$\begin{aligned}\frac{d\mathcal{I}(X;Y)}{dq} &= \mathcal{H}'(w_{10}(1-q)+w_{11}q)(w_{11}-w_{10}) + \mathcal{H}(w_{10}) - \mathcal{H}(w_{11}) \\ &= \log\left(\frac{1-w_{10}-q(w_{11}-w_{10})}{w_{10}+q(w_{11}-w_{10})}\right)(w_{11}-w_{10}) - [\mathcal{H}(w_{11}) - \mathcal{H}(w_{10})].\end{aligned}$$

Equating the above derivative to zero and solving for the duty cycle gives

$$q^{\max} = \frac{[1+\exp(\xi)]^{-1} - w_{10}}{w_{11}-w_{10}}, \tag{4.66}$$

where

$$\xi \equiv \frac{\mathcal{H}(w_{11}) - \mathcal{H}(w_{10})}{w_{11}-w_{10}}. \tag{4.67}$$

To check this result, notice that for small w_{11} and w_{10} we can use the approxima-

tion $\mathcal{H}(u) \approx -u \log u + u$ to show that

$$\begin{aligned}
\xi &\approx \frac{-w_{11} \log w_{11} + w_{11} + w_{10} \log w_{10} - w_{10}}{w_{11} - w_{10}} \\
&= 1 + \log \left\{ \left[\frac{w_{10}^{w_{10}}}{w_{11}^{w_{11}}} \right]^{1/(w_{11}-w_{10})} \right\} \\
&= \log \left\{ e(w_{11} - w_{10})^{-1} \frac{\left(\frac{w_{10}}{w_{11}-w_{10}} \right)^{\left(\frac{w_{10}}{w_{11}-w_{10}} \right)}}{\left(\frac{w_{11}}{w_{11}-w_{10}} \right)^{\left(\frac{w_{11}}{w_{11}-w_{10}} \right)}} \right\},
\end{aligned}$$

and, therefore,

$$q^{\max} \approx \left(w_{11} - w_{10} + \frac{e \left(\frac{w_{10}}{w_{11}-w_{10}} \right)^{\left(\frac{w_{10}}{w_{11}-w_{10}} \right)}}{\left(\frac{w_{11}}{w_{11}-w_{10}} \right)^{\left(\frac{w_{11}}{w_{11}-w_{10}} \right)}} \right)^{-1} - \frac{w_{10}}{w_{11} - w_{10}}. \tag{4.68}$$

For the decision rule, decide one, if exactly one count occurs, and zero otherwise, we saw that $w_{10} \approx \Delta R s$ and $w_{11} \approx \Delta R(1+s)$ for small Δ. These probabilities imply that $w_{11} - w_{10} \approx \Delta R$, $w_{10}/(w_{11} - w_{10}) \approx s$, $w_{11}/(w_{11} - w_{10}) \approx 1 + s$, resulting in

$$q^{\max} \approx \left(\Delta R + \frac{e s^s}{(1+s)^{(1+s)}} \right)^{-1} - s \approx \frac{(1+s)^{(1+s)}}{e s^s} - s, \tag{4.69}$$

for small Δ, which agrees with (4.59).

We will now consider a slightly different decision rule, that declares a one if one or more counts occur. In other words, for an arbitrary interval length Δ, we will decide zero if no counts occur during this interval, and one otherwise. The channel probabilities are then characterized as

$$w_{00} \equiv \Pr(Y_k = 0 \mid X_k = 0) = 1 - w_{10} = e^{-\Delta R s} \tag{4.70}$$

$$w_{01} \equiv \Pr(Y_k = 0 \mid X_k = 1) = 1 - w_{11} = e^{-\Delta R(1+s)}. \tag{4.71}$$

Using these probabilities, the mutual information becomes

$$I(X;Y) = \mathcal{H}(w_{00}(1-q) + w_{01}q) - \mathcal{H}(w_{00})(1-q) - \mathcal{H}(w_{01})q, \tag{4.72}$$

The duty cycle that maximizes the mutual information is

$$q^{\max} = \frac{[1+\exp(\xi)]^{-1} - w_{00}}{w_{01} - w_{00}}, \tag{4.73}$$

where

$$\xi \equiv \frac{\mathcal{H}(w_{01}) - \mathcal{H}(w_{00})}{w_{01} - w_{00}}. \tag{4.74}$$

Notice that for small Δ this decision rule is identical to our previous rule, because the probability of two or more counts in an interval is negligible.

We have already examined the behavior of the mutual information and optimal duty cycle for small interval lengths Δ. We now examine them for small noise-to-signal ratios s. For small s, the channel probabilities become $w_{00} \to 1$ and $w_{01} \to e^{-\Delta R}$. These limits describe a "Z-channel" for this very simple transmitter and receiver architecture. As the noise-to-signal ratio s approaches zero we have for a given duty cycle q

$$\lim_{s\to 0} I(X;Y) = \mathcal{H}\left(\{1-e^{-\Delta R}\}q\right) - \mathcal{H}(e^{-\Delta R})q, \tag{4.75}$$

and the duty cycle that maximizes the mutual information as $s \to 0$ is

$$\lim_{s\to 0} q^{\max} = \frac{\left[1+\exp\left(\frac{\mathcal{H}(e^{-\Delta R})}{e^{-\Delta R}-1}\right)\right]^{-1} - 1}{e^{-\Delta R} - 1}. \tag{4.76}$$

We plot the capacity-achieving (without average power constraint) duty cycle (4.73) in Figure 4-6 as a function of noise-to-signal ratio s and interval length Δ for $R = 1$. In Figure 4-7 we plot the capacity $\max_q I(X;Y)/\Delta$ for this OOK transmitter and unit-threshold receiver for different interval lengths Δ. In both figures, choosing $R = 1$ photons/second makes the interval length Δ indicative of the number of photons per bit interval. Also, by setting $R = 1$ we are essentially ignoring fading, and reducing the MIMO channel to a SISO channel.

From these figures, we see that for low signal-to-noise ratios (high s) and large numbers of photons per bit interval (large Δ), our simple transmitter and receiver architecture performs poorly. This behavior is not surprising because of our simplistic decision rule. We could have done better had we chosen a decision threshold greater than one photon.

For small numbers of photons per bit and high signal-to-noise ratios, however, this simple transmitter and receiver structure is not that bad. For example, when $R = 1$, $s = 0.001$ and $\Delta = 1$, the capacity of this OOK transmitter and unit-threshold receiver is 0.2987 nat/sec. As $\Delta \to 0$, the capacity of using optimal transmitters and receivers is $1/e \approx 0.3679$ nats/sec. In this case, the optimal receiver (OOK is already optimal as $\Delta \to 0$) capacity is 23% greater than the unit-threshold receiver.

We now explain the intuition regarding why this simple receiver performs just as well as an optimal receiver at high signal-to-noise ratio and small numbers of signal photons per bit. When the background noise rate and the interval length are small enough, the probability of a photon count on any detector is small. In the case of no background noise, if *any* detector records a photon, then the receiver knows for certain that all transmitters were "on." After all, there is a zero probability that a photon will arrive when the Poisson rate function is zero, i.e., all transmitters are "off."

In fact, the receiver that decides one if one or more counts occurred at any detector, is the minimum probability of error receiver. When background noise is negligible, a photon-bucket is optimal because adding photon counts and using a unit-threshold decision rule is equivalent to asking whether a photon arrived at any detector.

At low signal-to-noise ratio, however, optimally combining photon counts from the different detectors is better than a photon-bucket receiver, because less emphasis should be placed on the counts from detectors that have a low signal-to-noise ratio. Less emphasis should be placed on these photon counts, because it is more likely that the background noise generated them instead of the signal power.

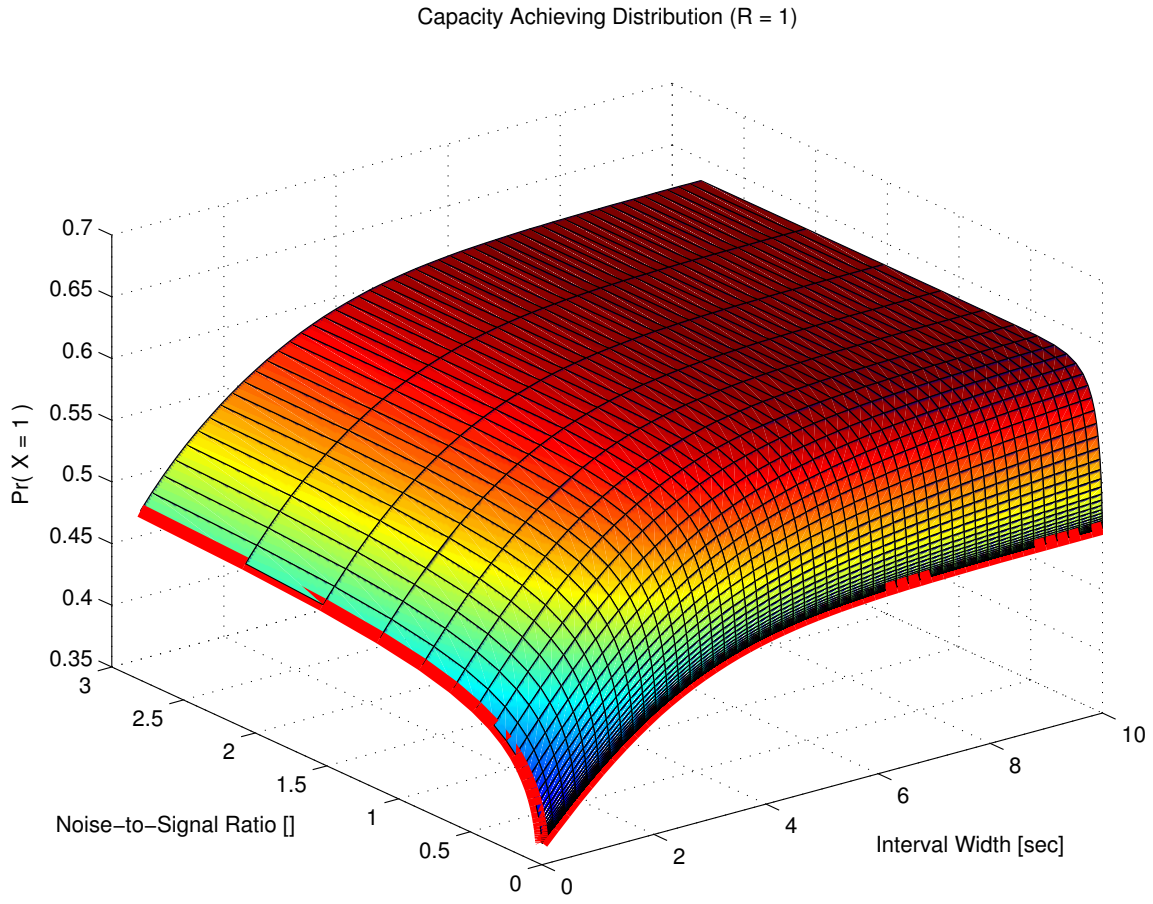

Figure 4-6: This figure shows the maximizing duty cycle (4.73) as a function of noise-to-signal ratio s and interval length Δ. The solid red lines are the asymptotes (4.69) and (4.76). These two asymptotes coincide at $1/e$ as $s \to 0$ and $\Delta \to 0$.

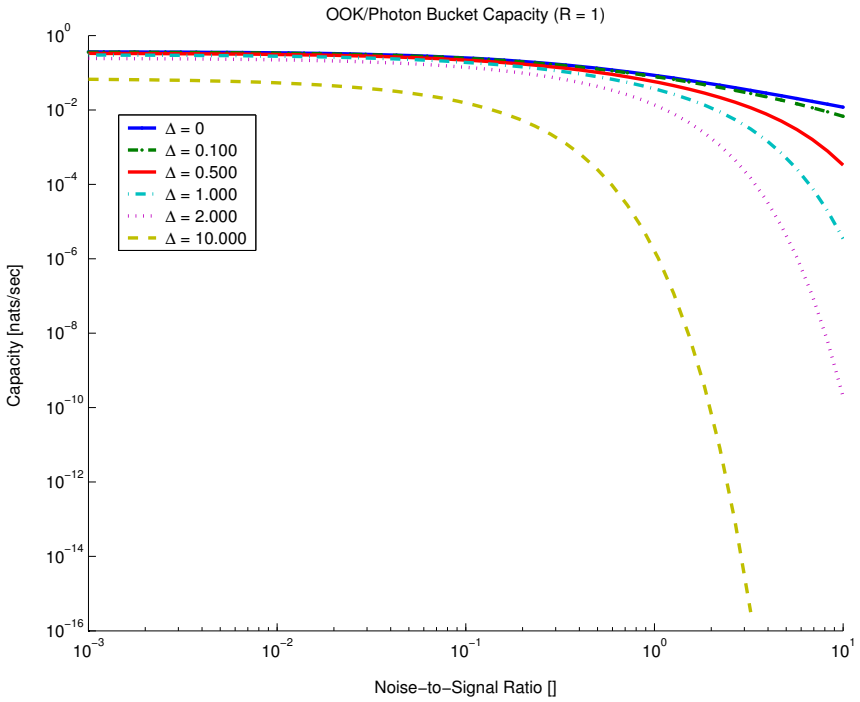

Figure 4-7: This figure shows the capacity of the OOK transmitter and unit-threshold receiver for different interval lengths Δ and noise-to-signal ratios s. From top to bottom, the curves correspond to $\Delta = 0$, 0.1, 0.5, 1, 2, and 10 seconds.

4.1.2 Ergodic Capacity

We now consider the ergodic (average) capacity of photon-counting, direct-detection optical communication through the turbulent atmosphere using multiple transmit and receive apertures [26]. In the high and low signal-to-noise ratio regimes, we show that the ergodic capacity of this fading channel equals or exceeds that for a channel with deterministic path gains. We will show that at high signal-to-noise ratio, the ergodic capacity scales as the number of transmit apertures (N) times the number of receive apertures (M), and can be achieved with neither transmitter or receiver knowing the path gains. In the low signal-to-noise ratio regime, ergodic capacity scales as MN^2. Moreover, in this regime, knowing the path gains at the transmitter does not increase capacity, and knowing them at the receiver does not appreciably increase capacity when employing moderate numbers of transmit apertures.

Optimal Receivers

We first consider systems in which both transmitter and receiver optimally use path gain knowledge. Although the transmitter knows the path gains, in the low and high signal-to-noise ratio regimes, the transmitter does not need to use the path gains to set the optimal duty cycle. Furthermore, in these regimes, the parallel-channel upper bound and OOK lower bound coincide, and the instantaneous capacity is given by (4.47) and (4.49).

Recalling that we normalized the path gains such that $E[\alpha_{nm}] = 1$, the average capacity at high signal-to-noise ratio is then

$$\begin{aligned} E[C(\alpha)] &= \left(p^{\text{opt}} \log 1/p^{\text{opt}}\right) \sum_{m=1}^{M} \sum_{n=1}^{N} E[\alpha_{nm}] A_n \\ &= \left(p^{\text{opt}} \log 1/p^{\text{opt}}\right) M \sum_{n=1}^{N} A_n \qquad (4.77) \\ &= M P_1 \; p^{\text{opt}} \log 1/p^{\text{opt}} \qquad (4.78) \\ &= MNA \; p^{\text{opt}} \log 1/p^{\text{opt}}, \qquad (4.79) \end{aligned}$$

where $P_k \equiv \sum_{n=1}^{N} A_n^k$. The last equality (4.79) is for identical transmit peak power constraints $A = A_1 = \cdots = A_N$. In this case, the average capacity in the low noise regime scales as the product of the number of transmit and receive apertures. In general—without equal limits on peak power—the average capacity scales linearly with the number of receive antennas and the total transmit peak power P_1.

Notice that the expression for capacity in the low noise regime does not depend on the log-amplitude variance σ_χ^2. In fact, had we replaced the path gains with their expected values, i.e., set $\alpha_{nm} = 1$, then the capacity would not have changed. We conclude that in the low noise regime, fading does not increase or decrease capacity compared to a channel with unit path gains. We will show next that in the high noise regime, fading actually increases capacity.

In the low signal-to-noise ratio regime, averaging over the path gains and using

$E[\alpha_{nm}^2] = e^{4\sigma_\chi^2}$ gives

$$
\begin{aligned}
E[C(\alpha)] &= \frac{p^{\text{opt}}(1-p^{\text{opt}})}{2} \sum_{m=1}^{M} \frac{1}{\lambda_m} \sum_{n=1}^{N} \sum_{k=1}^{N} E[\alpha_{nm}\alpha_{km}] A_n A_k && (4.80)\\
&= \frac{p^{\text{opt}}(1-p^{\text{opt}})}{2} \left(\sum_{m=1}^{M} \frac{1}{\lambda_m} \right) \left([e^{4\sigma_\chi^2} - 1] \sum_{n=1}^{N} A_n^2 + \left[\sum_{n=1}^{N} A_n \right]^2 \right) && (4.81)\\
&= L_{-1}(P_2 S_4 + P_1^2) p^{\text{opt}}(1-p^{\text{opt}})/2, && (4.82)
\end{aligned}
$$

where $L_k \equiv \sum_{m=1}^{M} \lambda_m^k$ and $S_k \equiv \exp(k\sigma_\chi^2) - 1$. Assuming identical transmit peak power constraints and receiver background noise $\lambda = \lambda_1 = \cdots = \lambda_M$, the ergodic capacity in the high noise regime is

$$
\begin{aligned}
E[C(\alpha)] &= \frac{M(NA)^2}{2\lambda} \left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N} \right) p^{\text{opt}}(1-p^{\text{opt}}) && (4.83)\\
&\approx M(NA)^2 p^{\text{opt}}(1-p^{\text{opt}})/2\lambda, && (4.84)
\end{aligned}
$$

where the last approximation holds for a moderate number of transmit apertures, i.e., $N \gg e^{4\sigma_\chi^2} - 1 \approx 6$ for severe fading ($\sigma_\chi^2 = 0.5$). We also see from this last expression that the average capacity of this fading channel ($\sigma_\chi^2 > 0$) is greater than the capacity of the deterministic channel with unit path gains ($\sigma_\chi^2 = 0$) in the high noise regime. In fact, (4.84) is the capacity of a unit path gain channel in the high noise regime.

The average capacity of the fading channel in the high noise regime is greater than the unit gain channel because of convexity. In this regime, the instantaneous capacity (4.49) is the sum of squared sums of random variables. Because the expected value of a random variable squared is never less than the square of its expected value, i.e., variance is non-negative, the average capacity in the high noise regime is greater than the unit gain capacity. A similar argument holds for photon-bucket capacity, (4.62). In the low noise regime, we saw that there was no difference in average capacity between the fading and unit path gain channels.

In our capacity formulation, the transmitter may need path gain knowledge to set the optimal OOK duty cycle (e.g. (4.41) and (4.29)). But in the low and high noise

regimes, i.e., $s_m \to 0$ and $s_m \to \infty$, the optimal duty cycle without average power constraint converges to $1/e$ and $1/2$, respectively. As a result, in these regimes the transmitter does not need to know the path gains. Because the noise-to-signal ratio s_m depends on the random path gains, however, we must be careful to qualify what "low and high noise-to-signal ratio regime" means.

Taking the limit of the background noise rates λ_m to zero or infinity will result in the appropriate regime with probability one for a fixed log-amplitude variance σ_χ^2. Alternatively, we can vary the peak power constraints $\{A_n\}$ to ensure that we are in one of these regimes. In this way, the transmitter can determine the optimal duty cycle without knowing the actual path gains. For convenience, we will refer to the low (high) signal-to-noise ratio regime as the high (low) noise regime, and remind the reader that "low" and "high" noise power is with respect to the signal power.

Photon-Bucket Receivers

We now come back to the question of how much a photon-bucket receiver degrades the capacity performance compared to an optimal receiver. Surprisingly, we will show that we did not lose much by using this suboptimal receiver structure.

We already observed that the capacity of a photon-bucket receiver (4.61) in the high signal-to-noise ratio regime ($s \to 0$) is identical the the optimal receiver capacity (4.47). In fact, in low noise, photon bucket reception is the minimum error probability receiver structure for OOK repetition spatial coding. Thus, in the high signal-to-noise ratio regime, photon buckets are capacity-achieving receiver structures.

In the low signal-to-noise ratio regime, averaging over the path gains in (4.62) gives

$$E[C_{\mathrm{PB-LB}}(\alpha)] = \frac{q^{\mathrm{opt}}(1-q^{\mathrm{opt}})/2}{L_1/M}\left[S_4 P_2 + M P_1^2\right]. \tag{4.85}$$

Specializing this result for identical transmitters and detectors, the average capacity

in this high noise regime becomes

$$E[C_{\text{PB-LB}}(\alpha)] = \frac{M(NA)^2}{2\lambda}\left(1+\frac{e^{4\sigma_\chi^2}-1}{MN}\right)q^{\text{opt}}(1-q^{\text{opt}}) \quad (4.86)$$

$$\approx M(NA)^2 q^{\text{opt}}(1-q^{\text{opt}})/2\lambda, \quad (4.87)$$

where the last approximation is valid when $MN \gg e^{4\sigma_\chi^2}-1 \approx 6$ in severe fading ($\sigma_\chi^2 = 0.5$). Comparing the photon-bucket capacity (4.86) with the optimal receiver capacity (4.83), we see that the only difference is a factor of M that discounts the path gain variance $e^{4\sigma_\chi^2}-1$. For large numbers of transmitters, the expressions are identical.

Notice that this last approximation (4.87) in the high noise regime is independent of the log-amplitude variance σ_χ^2, and is the capacity of the unit path gain channel (4.84). We conclude that for large numbers of transmit apertures in the low and high noise regimes, both optimal and photon-bucket receivers mitigate the fading, and achieve the same capacity as the unit path gain channel. As seen with the optimal receiver capacity in the high noise regime (4.83), the average photon-bucket capacity is greater than the unit path gain capacity.

Furthermore, comparing the photon-bucket average capacity (4.86) with the optimal receiver average capacity (4.83), we see that the capacity improvement, in the high noise regime, realized by knowing the path gains at the receiver is at most

$$\frac{E[C(\alpha)]}{E[C_{\text{PB-LB}}(\alpha)]} = \frac{e^{4\sigma_\chi^2}-1+N}{(e^{4\sigma_\chi^2}-1)/M+N}, \quad (4.88)$$

which is close to unity for moderate numbers of transmit apertures. In the worst-case scenario, when $\sigma_\chi^2 = 0.5$, $N = 1$, and $M \to \infty$, the photon-bucket receiver is within a factor of $e^{4\sigma_\chi^2} \approx 7.4$ of the optimal receiver in terms of average capacity. On the other hand, for a two-transmit and two-receive aperture ($N = M = 2$) system in moderate fading $\sigma_\chi^2 = 0.1$, using an optimal receiver with perfect path gain knowledge increases the average capacity by only 11% over the photon-bucket receiver, which does not require path gain knowledge and is simpler to implement. In mild fading, $\sigma_\chi^2 = 0.01$,

the average capacity gained is only 1%. A photon bucket, therefore, is a capacity-achieving receiver structure in the low noise regime, and also in the high noise regime for moderate numbers of transmit apertures.

Figure 4-8 shows the average capacity for a two-transmit and three-receive aperture system in moderate fading. Notice that the parallel-channel upper bound (PC-UB) from (4.36) and the OOK lower bound (OOK-LB) from (4.44) are nearly indistinguishable, and in fact become identical in the low and high noise regimes.

We conclude this section by noting that our ergodic capacity development only used the first and second moments of the fading distribution, never utilizing the fact the fades were actually lognormal in distribution. This is important, as it is unwise to assume that the distribution for atmospheric fading is lognormal deep into its tails.

4.1.3 Capacity-Versus-Outage Probability

When delay constraints prevent averaging over good and bad channel realizations, a more appropriate measure of capacity is the probability that the channel can support a given rate [5], i.e., $\Pr\{C(\alpha) > R_0\}$. In general, characterizing the exact distribution of the instantaneous capacity is difficult because only bounds exist on $C(\alpha)$, and these bounds, (4.36) and (4.44), are complicated functions of the path gains. In the high and low noise regimes, however, we saw that these bounds converge. Furthermore, the capacity expressions become proportional to the scaled sums of lognormal random variables, see (4.47) and (4.49). Recall that these high and low noise capacity expressions assume that the receiver knows and uses the path gains optimally. In these regimes, the transmitter does not need to know the path gains to set the optimal duty cycle.

The distribution of the sum of independent, real, lognormal random variables, however, does not have a nice closed-form expression [2]. As a result, we will develop approximations for the outage capacity in the high and low noise regimes by taking advantage of the "permanence" of the lognormal distribution in the sum of lognormal random variables [43]. That is, the sum of real lognormal random variables converges very slowly in distribution to a Gaussian, maintaining its lognormal "character" along

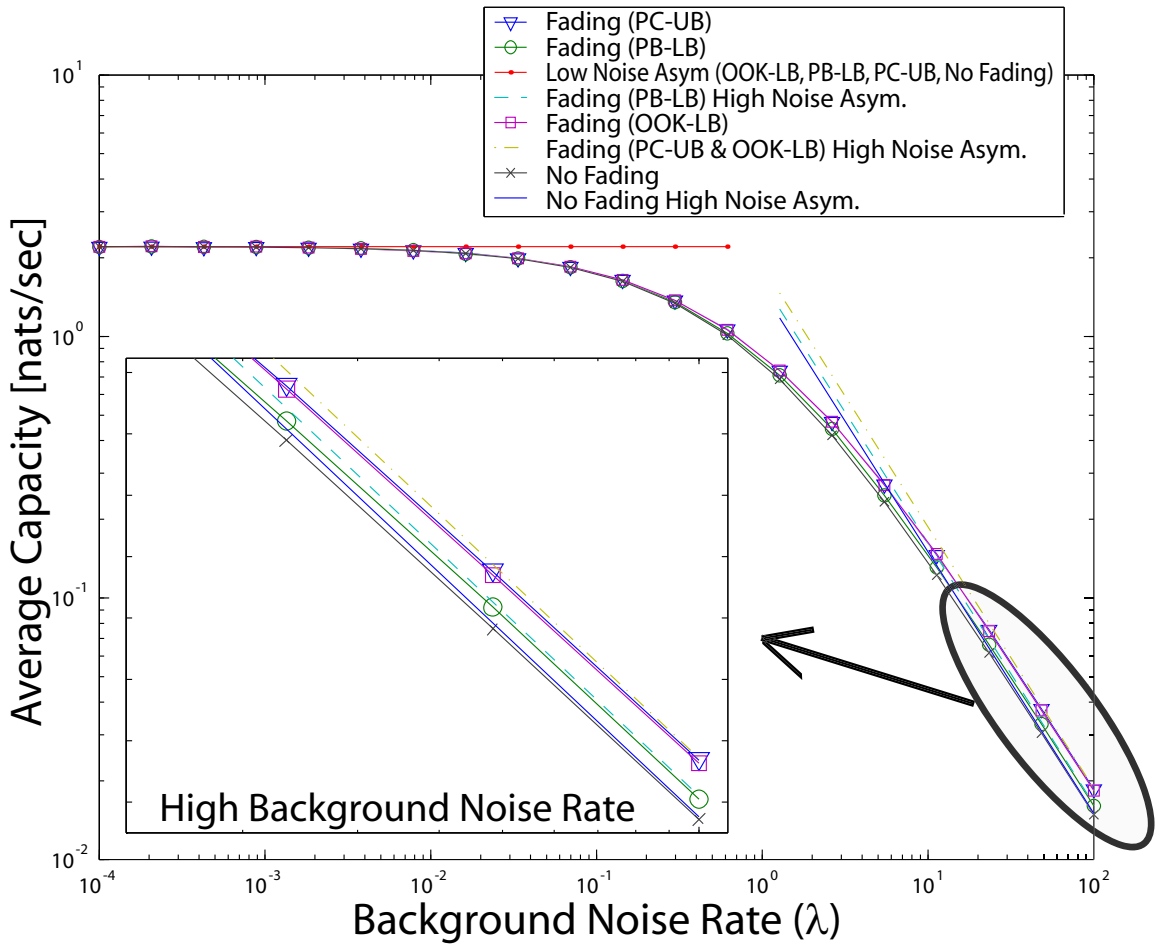

Figure 4-8: The average capacity without average power constraint ($\sigma = 1$) for two ($N = 2$) identical transmitters ($A_1 = A_2 = 1$) and three ($M = 3$) receivers ($\lambda = \lambda_1 = \lambda_2 = \lambda_3$) is shown as a function of background noise power λ. The parallel-channel upper bound (PC-UB) and OOK lower bound (OOK-LB) from [25] are shown along with the photon-bucket lower bound (PB-LB). All of these bounds have been averaged over 20,000 channel realizations of moderate fading intensity ($\sigma_\chi^2 = 0.1$). The average capacity results for the high and low noise regimes, viz., (4.79), (4.83), and (4.86), are shown as lines, along with the capacity of a unit path gain channel ($\sigma_\chi^2 = 0$) and its high noise asymptote (4.84).

the way[3]. A natural approach of handling sums of real lognormal random variables is then to approximate the sums themselves as lognormal random variables.

Several different methods exist to make this approximation. See [3] for a comparison of several popular methods. We will use a simple first- and second-order moment-matching approach, known as Wilkinson's method, which works well for the small log-amplitude variances in our application [48].

We wish to approximate the sum $S \equiv \sum_{k=1}^{K} \exp(u_k)$ by $Z \equiv \exp(u)$, where the u_k's are jointly Gaussian with $E[u_k] \equiv \mu_k$, $\mathrm{cov}[u_k, u_j] \equiv \nu_{kj}$ and u is Gaussian with $E[u] \equiv \mu$, $\mathrm{var}[u] \equiv \nu^2$. We make this approximation by matching the first and second moments, i.e., by choosing the mean μ and variance ν^2 such that $E[S] = E[Z] \equiv m$ and $\mathrm{var}[S] = \mathrm{var}[Z] \equiv v^2$. Equating the first and second moments, and solving for the μ and ν^2 gives the approximation (see Appendix A.3 for details)

$$\mu \equiv E[u] = \log\left(\frac{m}{\sqrt{1 + v^2/m^2}}\right) \tag{4.89}$$

$$\nu^2 \equiv \mathrm{var}[u] = \log\left(1 + \frac{v^2}{m^2}\right), \tag{4.90}$$

where

$$m \equiv E[S] = E[Z] = \sum_{k=1}^{K} e^{\mu_k + \nu_{kk}/2} \tag{4.91}$$

$$v^2 \equiv \mathrm{var}[S] = \mathrm{var}[Z] = \sum_{k=1}^{K}\sum_{l=1}^{K} e^{\mu_k + \mu_l + [\nu_{kk} + \nu_{ll}]/2} \left(e^{\nu_{kl}} - 1\right). \tag{4.92}$$

From (4.47) and (4.49), we can write the instantaneous capacity in the low and high signal-to-noise ratio regimes as being proportional to the scaled sum of lognormal random variables,

$$C(\alpha) = \gamma S(\alpha). \tag{4.93}$$

[3]Unlike sums of real lognormal random variables, sums of complex lognormal random variables tend to a Gaussian much faster because of the random uniform phase, see for example Section 3.2.3.

In the low noise regime,

$$\gamma = p^{\text{opt}} \log 1/p^{\text{opt}} \tag{4.94}$$

$$p^{\text{opt}} = \min(1/e, \sigma), \tag{4.95}$$

and

$$S(\alpha) = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha_{nm} A_n. \tag{4.96}$$

In the high noise regime,

$$\gamma = p^{\text{opt}}(1 - p^{\text{opt}})/2 \tag{4.97}$$

$$p^{\text{opt}} = \min(1/2, \sigma), \tag{4.98}$$

and

$$S(\alpha) = \sum_{m=1}^{M} \frac{1}{\lambda_m} \left(\sum_{n=1}^{N} \alpha_{nm} A_n \right)^2 = \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{k=1}^{N} \frac{A_n A_k}{\lambda_m} \alpha_{nm} \alpha_{km}. \tag{4.99}$$

Because the scaled product of lognormal random variables is still a lognormal random variable, $S(\alpha)$ in each regime is the sum of lognormal random variables. In the low noise regime it is the sum of MN independent lognormal random variables, and in the high noise regime it is the sum of MN^2 dependent lognormal random variables. In either case, we will approximate $S(\alpha) \approx Z \equiv e^u$, where u is Gaussian with mean $E[u] \equiv \mu$ and variance $\text{var}[u] \equiv \nu^2$.

These log-moments μ and ν^2 will depend on the mean and variance of $S(\alpha)$ through (4.89) and (4.90). In the low noise regime (see Appendix A.4),

$$m \equiv E[S(\alpha)] = MP_1 = MNA \tag{4.100}$$

$$v^2 \equiv \text{var}[S(\alpha)] = MS_4P_2 = MNA^2 \left(e^{4\sigma_\chi^2} - 1 \right), \tag{4.101}$$

where $P_k \equiv \sum_{n=1}^{N} A_n^k$ and $S_k \equiv \exp(k\sigma_\chi^2) - 1$ as previously defined. The last equalities are for identical transmitters ($A_1 = \cdots = A_N = A$). In the high noise regime (see

Appendix A.5.1),

$$m \equiv E[S(\alpha)] = L_{-1}[S_4P_2 + P_1^2] = \frac{M(NA)^2}{\lambda}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right) \tag{4.102}$$

$$\begin{aligned} v^2 &\equiv \mathrm{var}[S(\alpha)] = L_{-2}[(S_8+1)S_{16}P_4 + 4(S_4+1)S_8(P_1P_3 - P_4) \\ &\quad +2S_8(P_2^2 - P_4) + 4S_4(P_1^2P_2 - P_2^2 - 2P_1P_3 + 2P_4)] \qquad (4.103) \\ &= \frac{MNA^4}{\lambda^2}\left[e^{8\sigma_\chi^2}\left(e^{16\sigma_\chi^2} - 1\right) + 2\left(2e^{4\sigma_\chi^2} + 1\right)\left(e^{8\sigma_\chi^2} - 1\right)(N-1)\right. \\ &\quad \left. + 4\left(e^{4\sigma_\chi^2} - 1\right)(N-1)(N-2)\right], \qquad (4.104) \end{aligned}$$

where $L_k \equiv \sum_{m=1}^{M} \lambda_m^k$ as previously defined. The last equalities are again for identical transmitters and receivers. The lognormal sum variance expression simplifies for large numbers of transmitters to

$$v^2 \approx 4MN^3A^4\left(e^{4\sigma_\chi^2} - 1\right)/\lambda^2 \qquad \text{(large } N\text{)}. \tag{4.105}$$

A slightly less messy approximation to the lognormal sum $S(\alpha)$ in the high noise regime is to approximate R_m as lognormal, square the resulting lognormal random variable, and then approximate the sum of these squared lognormal random variables as being lognormal. The details of this method are in Appendix A.5.2. This approximate moment matching method approximates the first and second moments of the sum (4.99) as

$$m \equiv E[S(\alpha)] \approx L_{-1}\left[S_4P_2 + P_1^2\right] = \frac{M(NA)^2}{\lambda}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right) \tag{4.106}$$

$$v^2 \equiv \mathrm{var}[S(\alpha)] \approx L_{-2}\, P_1^4\left(1 + \frac{S_4P_2}{P_1^2}\right)^2\left[\left(1 + \frac{S_4P_2}{P_1^2}\right)^4 - 1\right] \tag{4.107}$$

$$= \frac{M(NA)^4}{\lambda^2}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^2\left(\left[1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right]^4 - 1\right), \tag{4.108}$$

where again the last equalities are for identical transmitters and receivers. Comparing (4.106) and (4.102) we see that this approximation preserves the mean of the lognormal sum. This observation is not surprising because the mean of the lognormal sum

$S(\alpha)$ in the high-noise regime depends only on the second moments of R_m, and this latter approximation matches the first and second moments of R_m. Furthermore, we see that the variance of the lognormal sum for large numbers of transmit apertures becomes

$$v^2 \approx 4MN^3A^4\left(e^{4\sigma_\chi^2}-1\right)/\lambda^2 \qquad \text{(large } N\text{)}, \tag{4.109}$$

using the approximation $(1+\epsilon)^4 \approx 1+4\epsilon$ for small $\epsilon = (e^{4\sigma_\chi^2}-1)/N$. Comparing (4.105) and (4.109), we see that both high noise approximations are the same for large numbers of transmit apertures.

Regardless of the method used to calculate the moments of the lognormal sum, we will approximate $S(\alpha) \approx e^u$, where u is Gaussian with mean $E[u] \equiv \mu$ and variance $\text{var}[u] \equiv \nu^2$ given by (4.89) and (4.90). Consequently, the approximation $C(\alpha) = \gamma S(\alpha) \approx \gamma e^u$ leads to the following expression in the low and high noise regimes for the probability that the channel can support a desired rate,

$$\Pr\{C(\alpha) > R_0\} \;\approx\; \Pr\{\gamma e^u > R_0\} = Q\left(\frac{\log R_0 - \log\gamma - \mu}{\nu}\right), \tag{4.110}$$

where $Q(x) = \int_x^\infty dz \exp(z^2/2)/\sqrt{2\pi}$ is the tail area under the standard normal distribution, ν and μ are given by (4.89) and (4.90), and γ will depend on the regime, see (4.94) and (4.97). The capacity-versus-outage probability is one minus this probability.

Notice that in the low and high noise regimes, our outage capacity approximation (4.110) becomes a step function at the ergodic capacity,

$$E[C(\alpha)] = \gamma E[S(\alpha)] = \gamma m, \tag{4.111}$$

when the instantaneous capacity variance-to-mean-squared ratio,

$$\frac{v^2}{m^2} = \frac{\text{var}[S(\alpha)]}{E[S(\alpha)]^2} = \frac{\text{var}[C(\alpha)]}{E[C(\alpha)]^2}, \tag{4.112}$$

approaches zero. In other words, $v^2/m^2 \to 0$ causes the log-variance $\nu^2 \to 0$ by (4.90),

and the outage capacity approximation becomes $\Pr\{C(\alpha) > R_0\} = 1$ if $R_0 < \gamma e^{\mu}$, and zero otherwise. But, as $v^2/m^2 \to 0$, the log-mean $\mu \to \log m$ by (4.89). Consequently, as $v^2/m^2 \to 0$, the outage capacity approximation $\Pr\{C(\alpha) > R_0\} = 1$ if $R_0 < \gamma m = E[C(\alpha)]$, and zero otherwise. This behavior indicates that as $v^2/m^2 \to 0$, the randomness of the instantaneous capacity is negligible, and reliable communication can occur at all rates below the ergodic capacity. We will now examine scenarios in which the variance-to-mean-squared ratio $v^2/m^2 \to 0$, providing a consistency check to our outage capacity approximation.

The first scenario is when fading is negligible. From (4.101), (4.104), and (4.108) we see that as $\sigma_\chi^2 \to 0$, the variance of the sum $S(\alpha)$ converges to zero, i.e., $v^2 \to 0$. The sum mean μ, on the other hand, remains finite, see (4.100), (4.102), and (4.106). Consequently, $v^2/m^2 \to 0$ as $\sigma_\chi^2 \to 0$ and the approximation to the outage capacity becomes a step function at the ergodic capacity, (4.79) and (4.84), which in this case is the unit path gain capacity. Of course, when $\sigma_\chi^2 = 0$, there is no fading and the capacity at all times equals the unit path gain capacity, so that the preceding step function behavior is a consistent check on our approximation.

The same step function behavior also occurs for large numbers of transmit and receive apertures. In the low and high noise regimes, the ratio v^2/m^2 is proportional to $1/N$ and $1/M$ for large numbers transmit and receive apertures, respectively. Consequently, as $N \to \infty$ and/or $M \to \infty$, the outage capacity approximation $\Pr\{C(\alpha) > R_0\} = 1$ if $R_0 < E[C(\alpha)]$, and zero otherwise. In the low and high noise regimes, the ergodic capacity converges to the unit path gain capacity for large numbers of transmit apertures, see (4.79) and (4.83). In this case, the incoherent averaging of the received power contributions from the many transmit apertures mitigates the fading, resulting in the unit path gain capacity. For large numbers of receive apertures, the ergodic capacity from optimal combining will in general be greater than the unit path gain capacity, see (4.83).

The intuition behind why increasing the number of transmitters and receivers decreases the instantaneous capacity variance-to-mean-squared ratio comes from examining the expressions for the instantaneous capacity and loosely applying the law

of large numbers. Without loss of generality, we will consider identical transmit power constraints and receiver background noise. In the low noise regime, the instantaneous capacity (4.47) shows that increasing either the number of transmit or receive apertures has the same effect. For example, as the number of receive apertures M becomes large, the law of large numbers suggests that the path gain sum $\frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A$ becomes close to its average NA. Writing,

$$C(\alpha) \propto M\left(\frac{1}{M}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A\right) \approx ME\left[\sum_{n=1}^{N}\alpha_{nm}A\right] = MNA, \tag{4.113}$$

we see that the capacity is proportional to MNA, in the sense that fractional deviations from this proportionality become negligible for large numbers of receive apertures. In other words, the variance-to-mean-squared ratio

$$\frac{\mathrm{var}[C(\alpha)]}{E[C(\alpha)]^2} = \frac{e^{4\sigma_\chi^2}-1}{MN}, \tag{4.114}$$

decays to zero as the number of receive and/or transmit apertures increases without bound.

In the high noise regime, however, the instantaneous capacity (4.49) behaves differently as the number of transmit or receive apertures increases. Increasing the number of transmit apertures causes the instantaneous capacity to scale as

$$C(\alpha) \propto \sum_{m=1}^{M} N^2\left(\frac{1}{N}\sum_{n=1}^{N}\alpha_{nm}A\right)^2 \approx MN^2(AE[\alpha_{nm}])^2 = M(NA)^2, \tag{4.115}$$

which agrees with the ergodic capacity for large N and the unit path gain capacity (4.84). For large N, we see that the capacity variance-to-mean-squared ratio in the high noise regime

$$\frac{\mathrm{var}[C(\alpha)]}{E[C(\alpha)]^2} = \frac{4\left(e^{4\sigma_\chi^2}-1\right)}{MN}, \tag{4.116}$$

decays four times slower than in the low noise regime, see (4.114).

On the other hand, increasing the number of receive apertures causes the instan-

taneous capacity to scale as

$$C(\alpha) \propto M\left[\frac{1}{M}\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}A\right)^2\right] \tag{4.117}$$

$$\approx ME\left[\left(\sum_{n=1}^{N}\alpha_{nm}A\right)^2\right] \tag{4.118}$$

$$= M(NA)^2\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right), \tag{4.119}$$

which agrees the ergodic capacity (4.83). This ergodic capacity for large numbers of receive apertures M is in general greater than the capacity of a unit path gain channel in the high noise regime. Using either high noise variance expression (4.104) or (4.108), the capacity variance-to-mean-squared ratio decays as $1/M$ for increasing numbers of receive apertures. We conclude that increasing either the number of transmit or receive apertures forces the instantaneous capacity variance-to-mean-squared ratio to zero, providing another consistency check to our approximation.

Figures 4-9 through 4-12 plot the outage capacity and the moment matching approximations for moderate fading. Figure 4-9 plots the probability that the channel can support a given rate in the low noise regime. The solid lines are the moment matching approximation (4.110) and the symbols are the empirical complementary cumulative distribution function from Monte Carlo realizations of the channel capacity. Figure 4-10 plots the rate the channel can support 99% of the time in the low noise regime versus the number of transmit and receive apertures. Figures 4-11 and 4-12 plot similar quantities, but for the high noise regime.

From these figures, we see that the outage capacity is symmetric with respect to the number of transmit and receive apertures in the low noise regime. Like the ergodic capacity, however, the outage capacity in the high noise regime improves more from increasing the number of transmit apertures than increasing the number of receive apertures, cf. the ($N = 3$, $N = 2$) and ($N = 2$, $M = 3$) curves.

The outage capacity approximations tend to be worse in the high noise regime because the lognormal sum consists of MN^2 components, compared to the MN com-

ponents in the low noise regime. As seen in Figures 4-11 through 4-14, in the range of outage probabilities of interest, the approximate moment matching method tends to be more accurate. Both approximation methods fail in the high noise regime for moderate numbers of apertures and strong fading as demonstrated in Figure 4-13. For large numbers of apertures, however, fractional deviations from the ergodic capacity become negligible, and both approximations become quite good, see Figure 4-14. For example, the variance-to-mean-squared ratio of the 300,000 channel capacity realizations in Figure 4-13 was approximately 0.43, while that of the 3,000 realizations in Figure 4-14 was 0.0014. Notice that for large numbers of transmit apertures, (4.116) predicts that this latter ratio should be 0.0012.

4.2 Coding

Space-time codes for heterodyne wireless communications such as the Alamouti scheme [1] have the property that the transmitted codewords add incoherently at the receiver. In other words, the received signal during each time slot is the sum of transmitted symbols multiplied by magnitude-squared path gains plus noise. Poor reception occurs when this sum of magnitude-squared path gains is small compared to the noise.

By their very nature, direct-detection systems lend themselves well to space-time coding. In a sense, we gain some space-time coding for "free" by construction. We have assumed that the transmitters are sufficiently separated in angle or frequency such that optical powers add. This assumption leads to channel models in which the received signal is the sum of transmitted powers scaled by the power path gains plus noise. Like the Alamouti scheme, poor performance occurs when all the paths are in deep fades. We will take advantage of this construction, and examine space-time codes that have good error probability characteristics and low complexity.

Spatial and temporal coding can improve the reliability of communication through the turbulent atmosphere. Chan in [12] considered the benefit of using coded pulse-position modulation to reduce the required average transmit power for a desired error probability. Davidson and Koh analyzed interleaved convolutional [15] and concate-

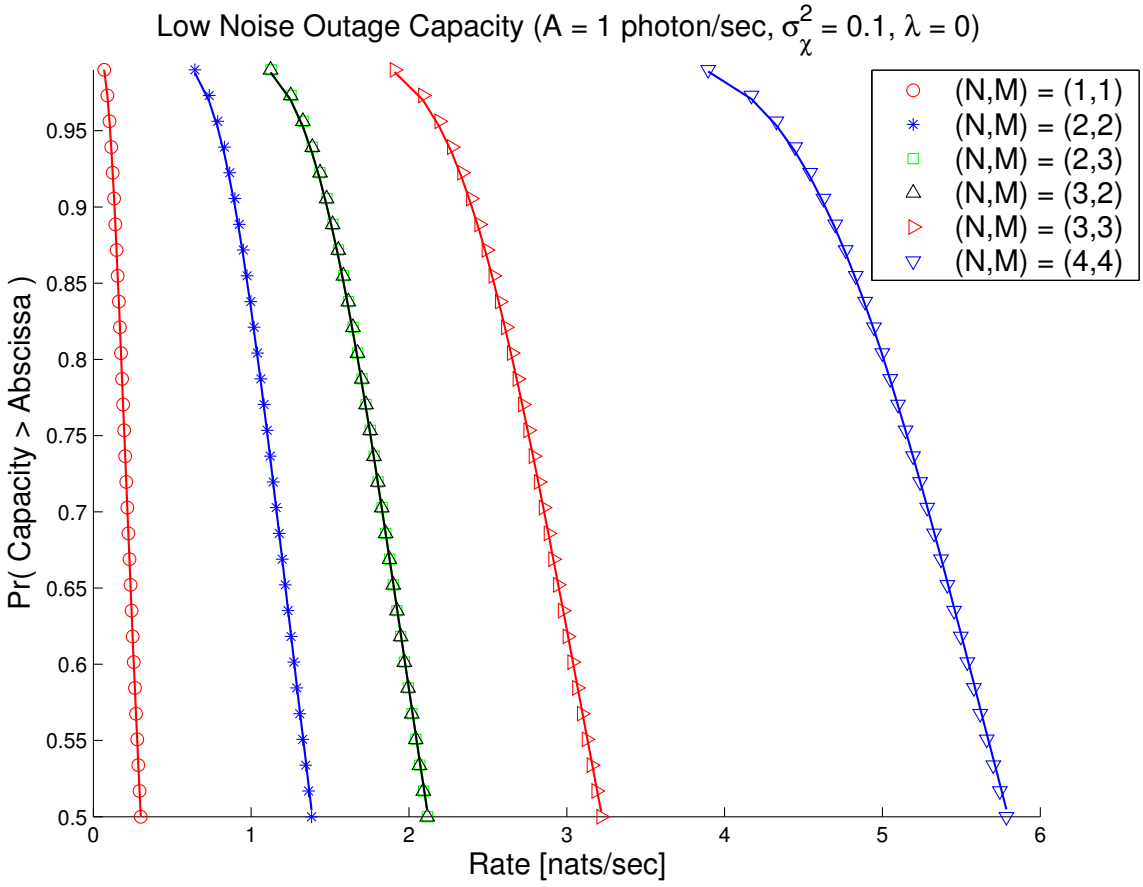

Figure 4-9: The probability that the channel can support a given rate is shown for the low noise regime in moderate fading ($\sigma_\chi^2 = 0.1$) with no average power constraint ($\sigma = 1$). The solid lines are the lognormal approximation of (4.110) and the symbols are the empirical complementary cumulative distribution of 300,000 channel capacity realizations. We assume that the identical transmitters ($A_1 = \cdots = A_N = 1$) and receivers ($\lambda = \lambda_1 = \cdots = \lambda_M$) know and use the path gains optimally.

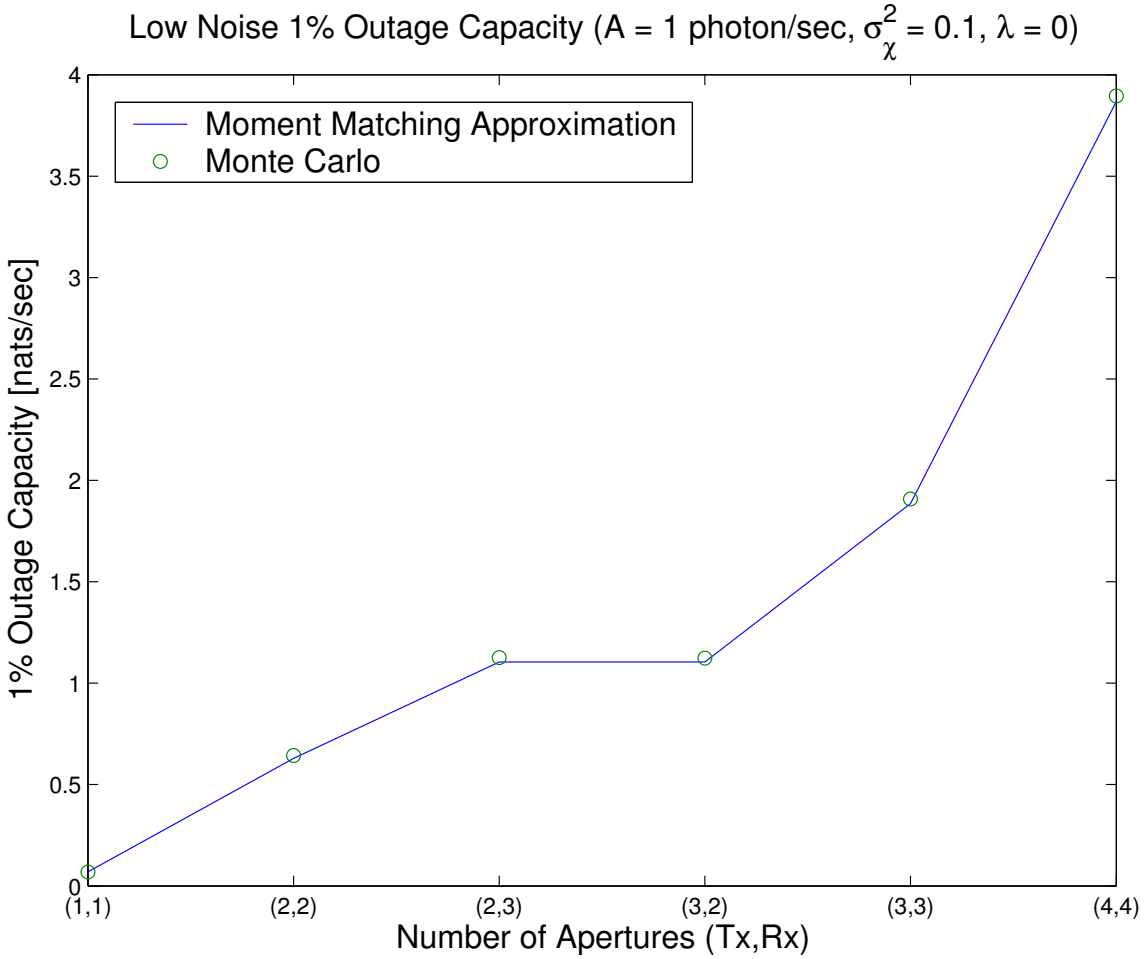

Figure 4-10: The rate achieved 99% of the time is plotted versus the number of transmit and receive apertures in the low noise regime in moderate fading ($\sigma_\chi^2 = 0.1$) with no average power constraint ($\sigma = 1$).

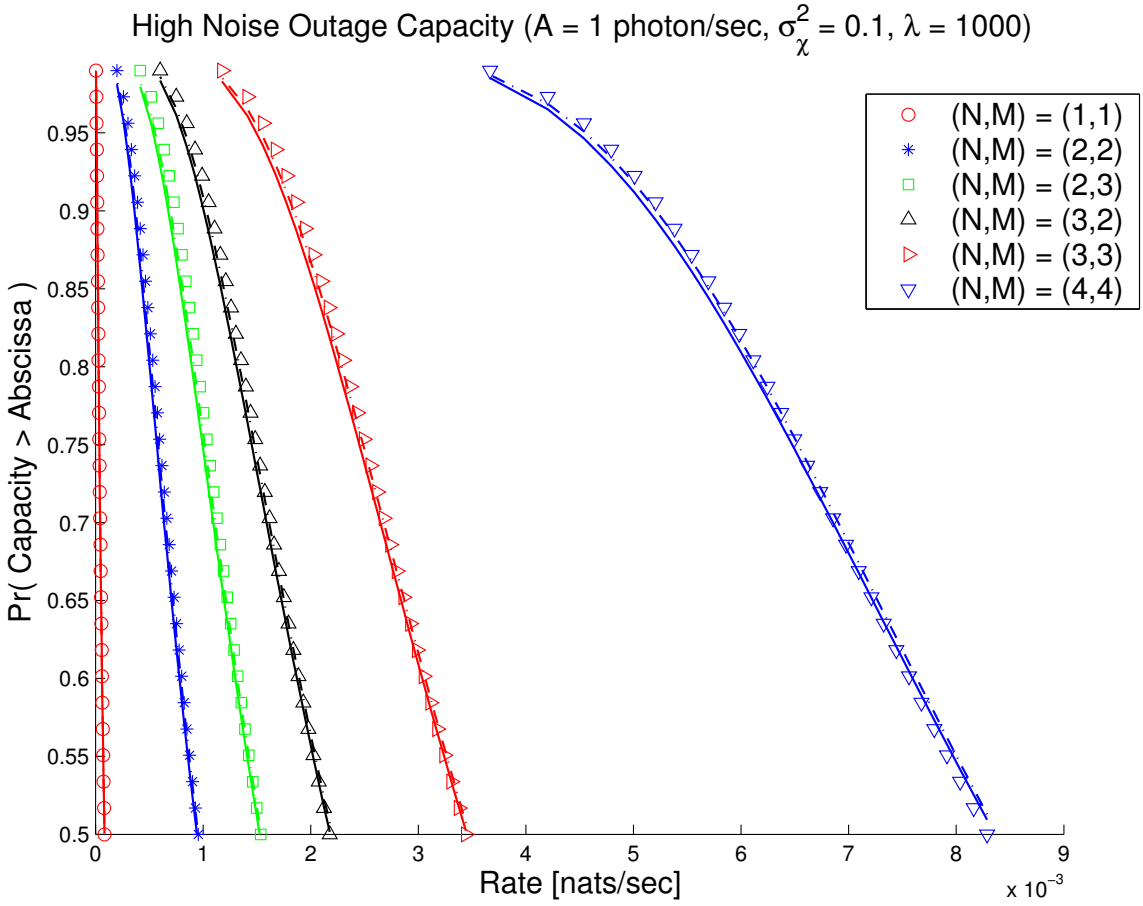

Figure 4-11: The probability that the channel can support a given rate is shown for the high noise regime in moderate fading ($\sigma_\chi^2 = 0.1$) with no average power constraint ($\sigma = 1$). The solid lines are the lognormal approximation of (4.110) using the exact moments of the lognormal sum, see (4.102) and (4.104). The dashed lines are also the lognormal approximation using approximate moments of the lognormal sum, see (4.106) and (4.108). The symbols are the empirical complementary cumulative distribution of 300,000 channel capacity realizations. We assume that the identical transmitters ($A_1 = \cdots = A_N = 1$) and receivers ($\lambda = \lambda_1 = \cdots = \lambda_M$) know and use the path gains optimally.

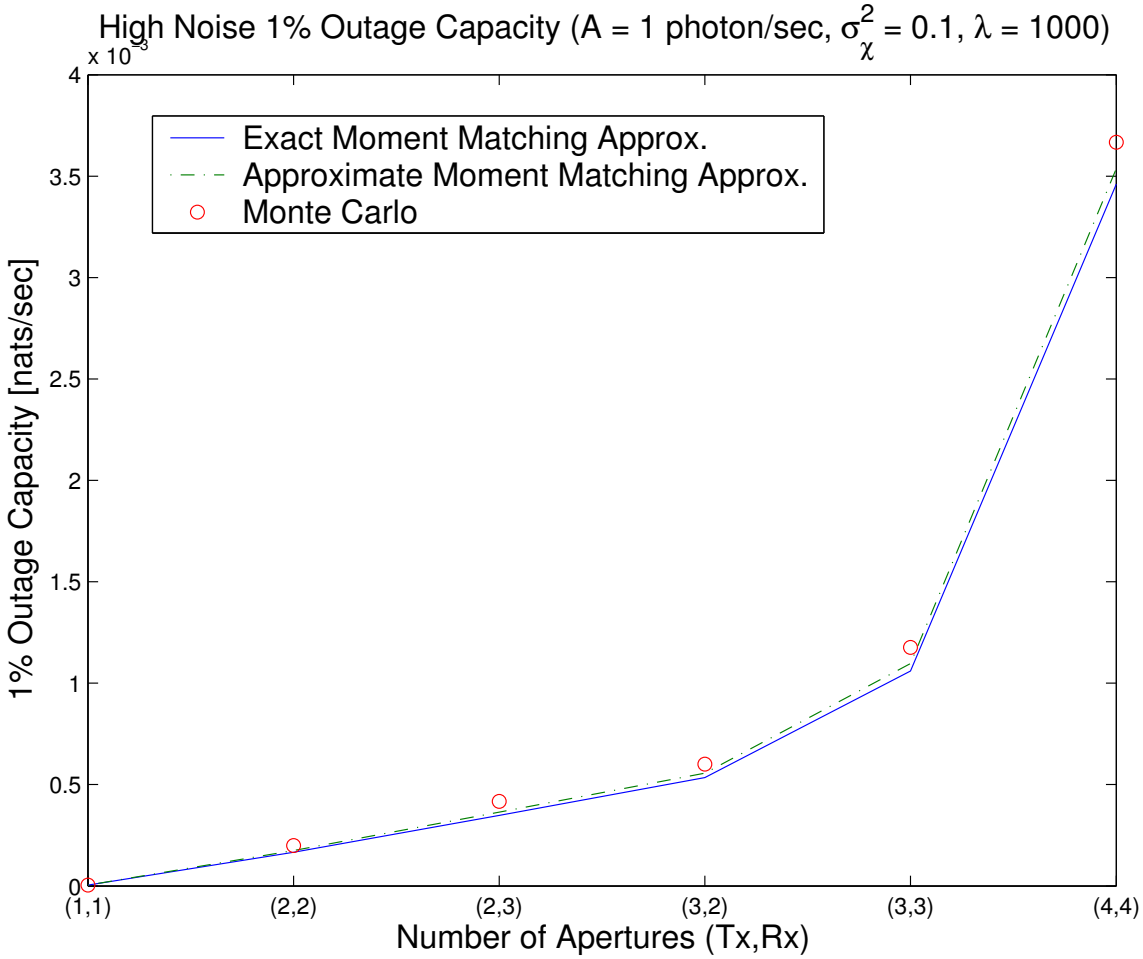

Figure 4-12: The rate achieved 99% of the time is plotted versus the number of transmit and receive apertures in the high noise regime in moderate fading ($\sigma_\chi^2 = 0.1$) with no average power constraint ($\sigma = 1$)

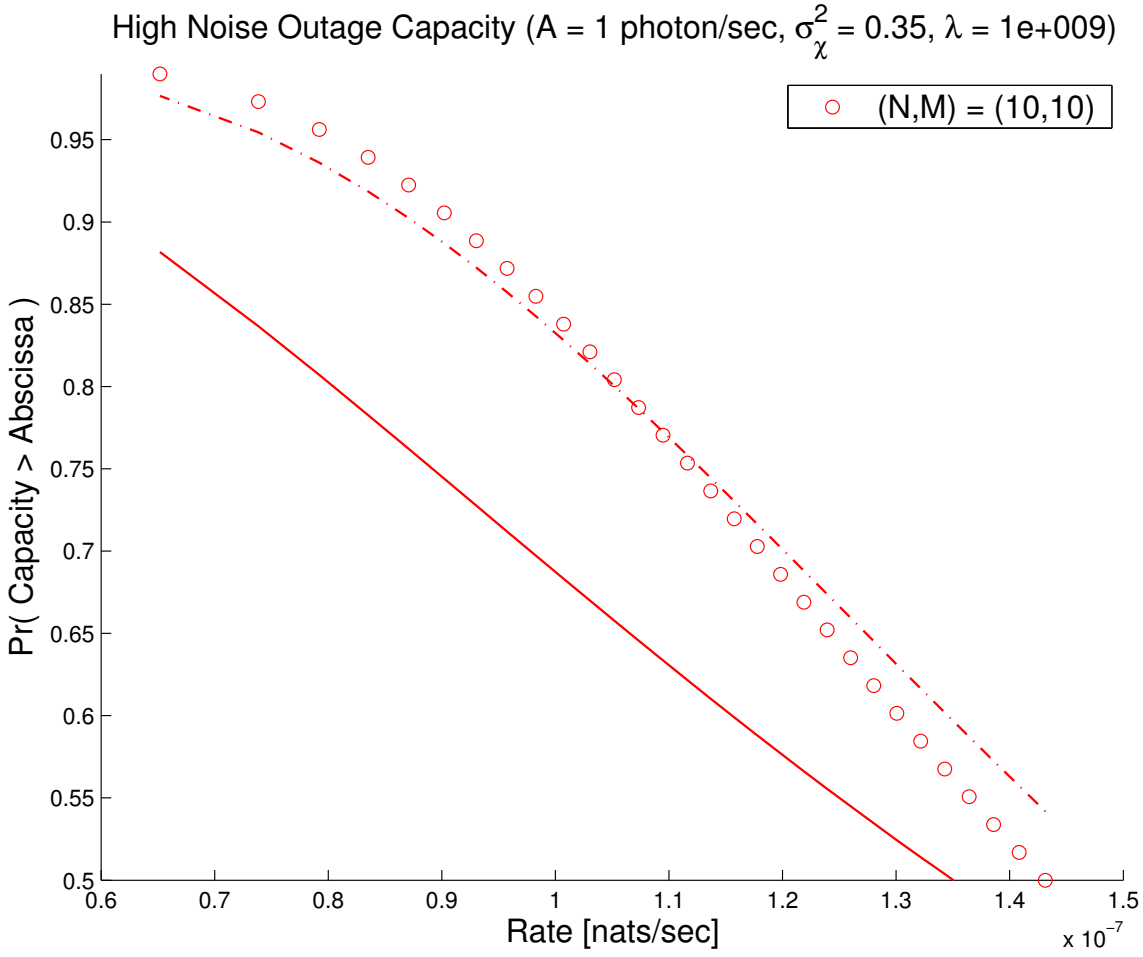

Figure 4-13: The probability that the channel can support a given rate for ten transmit and ten receive apertures ($N = M = 10$) in the high noise regime for strong fading ($\sigma_\chi^2 = 0.35$) with no average power constraint ($\sigma = 1$). The solid line is the lognormal approximation of (4.110) using the exact moments of the lognormal sum, see (4.102) and (4.104). The dashed line is also the lognormal approximation using approximate moments of the lognormal sum, see (4.106) and (4.108). The symbols are the empirical complementary cumulative distribution of 300,000 channel capacity realizations.

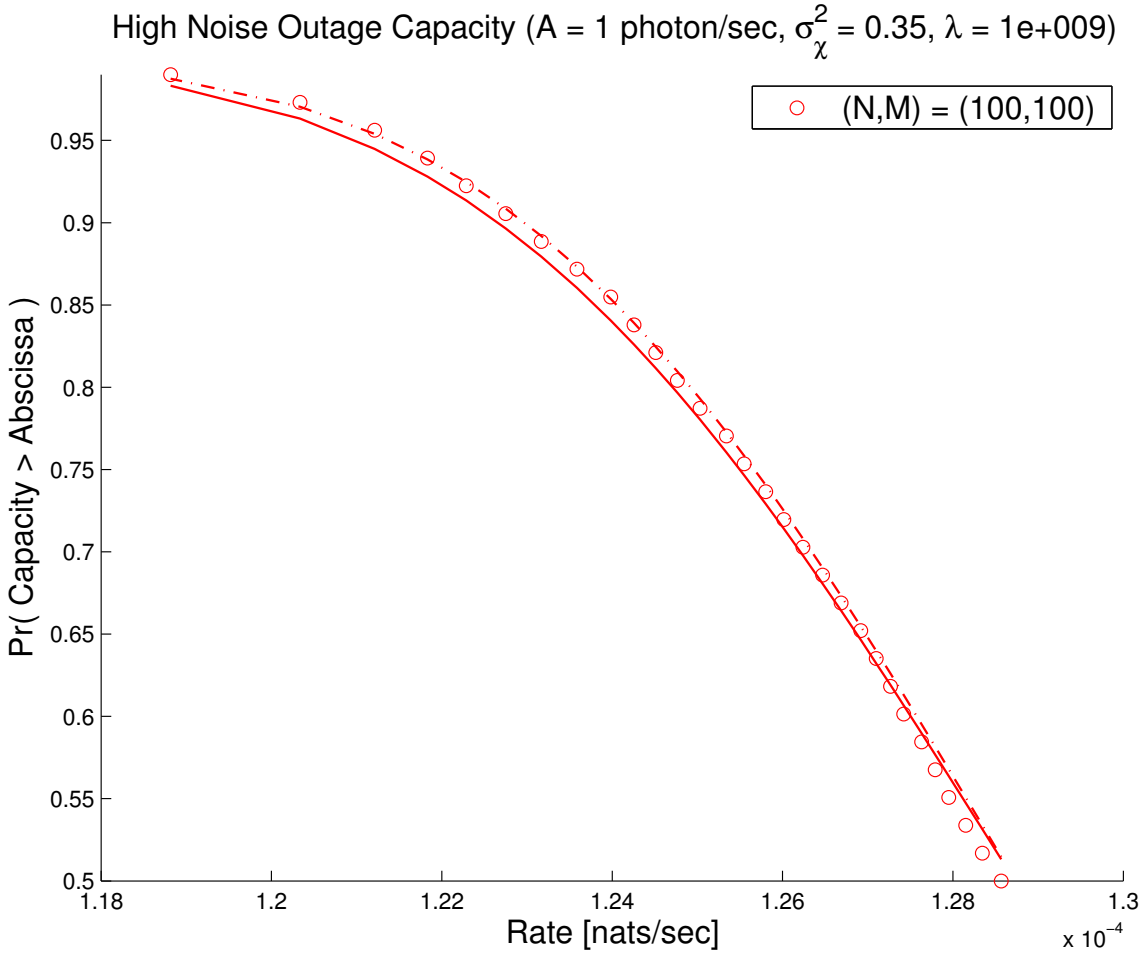

Figure 4-14: The probability that the channel can support a given rate for one hundred transmit and one hundred receive apertures ($N = M = 100$) in the high noise regime for strong fading ($\sigma_\chi^2 = 0.35$) with no average power constraint ($\sigma = 1$). The solid line is the lognormal approximation of (4.110) using the exact moments of the lognormal sum, see (4.102) and (4.104). The dashed line is also the lognormal approximation using approximate moments of the lognormal sum, see (4.106) and (4.108). The symbols are the empirical complementary cumulative distribution of 3,000 channel capacity realizations (only 3,000 realizations were used due to limited computational resources).

nated coding [39] through the turbulent atmosphere, and experimentally verified its performance. Shapiro and Harney in [50] calculated error probability bounds for block and convolutional coded transmission using multiple receive apertures. Spatial diversity at the receiver [30] and transmitter [10, 37] can also help mitigate fading due to atmospheric turbulence.

In this section, we will develop error probability bounds for space-time codes employing a minimum probability of error receiver. We will then examine the performance of simple space-time codes such as repetition spatial coding and switching diversity.

4.2.1 Minimum Probability of Error Decoding

Consider a space-time codeword waveform $X \equiv \{x_1(t), \ldots, x_N(t) \mid 0 \leq t \leq T\}$ from a space-time code $\mathbf{X}$. This codeword will induce a Poisson process at the m-th detector $y_m(t)$ with rate

$$\mu_m(t) \equiv \sum_{n=1}^{N} \alpha_{nm} x_n(t) + \lambda_m, \tag{4.120}$$

Conditioned on the codeword and path gains, each detector's Poisson process is independent of the others. Using the arrival time density (4.8), the log-likelihood of observing the detector waveforms $Y \equiv \{y_1(t), \ldots, y_M(t) \mid 0 \leq t \leq T\}$ given the path gains $\alpha \equiv \{\alpha_{nm} \mid 1 \leq n \leq N, 1 \leq m \leq M\}$ and that codeword X was sent is

$$\log p_{Y|X,\alpha} = \sum_{m=1}^{M} \log p_{Y_m|X,\alpha}, \tag{4.121}$$

where $Y_m \equiv \{y_m(t) \mid 0 \leq t \leq T\}$ is the m-th detector's count record and

$$\log p_{Y_m|X,\alpha} = \int_0^T \log[\mu_m(t)] dy_m(t) - \int_0^T \mu_m(t) dt, \tag{4.122}$$

is the ordered, arrival time density, with stochastic integration defined by (4.9). Assuming that the transmitter is equally likely to send each codeword, the minimum probability of error receiver will choose the codeword that was most likely to have

produced the observed photon record Y, i.e.,

$$\widehat{X} = \arg\max_{X \in \mathbf{X}} \ \log p_{Y|X,\alpha}. \tag{4.123}$$

In the above expression, we have assumed that the receiver knows and uses the path gains in making its decision.

The probability of error conditioned on a path gain realization α is

$$P_e(\alpha) \equiv \frac{1}{|\mathbf{X}|} \sum_{X \in \mathbf{X}} \sum_{\substack{\bar{X} \in \mathbf{X} \\ \bar{X} \neq X}} \Pr(\text{ Decide } \bar{X} \mid X \text{ sent}, \alpha\,), \tag{4.124}$$

where $|\mathbf{X}|$ denotes the number of codewords in the space-time code $\mathbf{X}$. The unconditional error probability is then

$$P_e \equiv E[P_e(\alpha)], \tag{4.125}$$

where the expectation is with respect to the path gains.

4.2.2 Bounds on Pairwise Error Probability

We can use the union bound to upper bound the error probability (4.124) in terms of pairwise error probabilities,

$$P_e(\alpha) \leq \frac{1}{|\mathbf{X}|} \sum_{X \in \mathbf{X}} \sum_{\substack{\bar{X} \in \mathbf{X} \\ \bar{X} \neq X}} \Pr(\, X \to \bar{X} \mid \alpha\,), \tag{4.126}$$

where $\Pr(\, X \to \bar{X} \mid \alpha\,)$ is the probability of decoding codeword X as $\bar{X}$ in the absence of other codewords. This sum is usually dominated by terms of codeword pairs with largest pairwise error. We will, therefore, focus on analyzing the pairwise error probability.

In terms of the decision rule, the pairwise error probability is

$$\Pr(\, X \to \bar{X} \mid \alpha\,) \equiv \Pr(\log p_{Y|\bar{X},\alpha} > \log p_{Y|X,\alpha} \mid X, \alpha\,). \tag{4.127}$$

We will develop an upper bound on this pairwise error probability conditioned on the path gains using the Chernoff bound [21]. This bound relates the probability of a random variable exceeding a particular value to its moment-generating function. Specifically, for any $\tau \geq 0$ and random variable Z,

$$\Pr(Z > z) \leq E\left[e^{\tau Z}\right] e^{-\tau z}. \tag{4.128}$$

In our context, the pairwise error probability is

$$\begin{aligned} \Pr(\, X \to \bar{X} \mid \alpha\,) &= \Pr\left(\sum_{m=1}^{M} \int_0^T \left[\log \bar{\mu}_m(t) - \log \mu_m(t)\right] dy_m(t) \right. \\ &\qquad \left. > \sum_{m=1}^{M} \int_0^T \left[\bar{\mu}_m(t) - \mu_m(t)\right] dt \;\middle|\; X, \alpha \right), \end{aligned} \tag{4.129}$$

where $\bar{\mu}_m(t)$ is the rate function induced by the codeword $\bar{X}$. Defining,

$$Z \equiv \sum_{m=1}^{M} \int_0^T \left[\log \bar{\mu}_m(t) - \log \mu_m(t)\right] dy_m(t), \tag{4.130}$$

and

$$z \equiv \sum_{m=1}^{M} \int_0^T \left[\bar{\mu}_m(t) - \mu_m(t)\right] dt, \tag{4.131}$$

we can use the Chernoff bound to find an upper bound on the pairwise error probability. Because the detector processes are conditionally independent given the path gains and transmitted codeword, the conditional moment generating function of Z is

$$E\left[e^{\tau Z} \mid X, \alpha\right] = \prod_{m=1}^{M} E\left[\exp\left\{ \int_0^T \log\left[\frac{\bar{\mu}_m(t)}{\mu_m(t)}\right]^{\tau} dy_m(t) \right\} \;\middle|\; X, \alpha \right]. \tag{4.132}$$

Letting $0 \leq t_{m1} \leq t_{m2} \leq \ldots \leq t_{m\, y_m(T)} < T$ denote the photon arrivals at the m-th

detector, the m-th term of the above product is

$$
\begin{aligned}
& E\left[\exp\left\{\int_0^T \log\left[\frac{\bar{\mu}_m(t)}{\mu_m(t)}\right]^\tau dy_m(t)\right\} \;\middle|\; X,\alpha\right] \\
&= E\left[\exp\left\{\int_0^T \log\left[\frac{\bar{\mu}_m(t)}{\mu_m(t)}\right]^\tau dy_m(t)\right\} \;\middle|\; X,\alpha,y_m(T)=0\right]\Pr[y_m(T)=0] \\
&\quad + E\left[\exp\left\{\int_0^T \log\left[\frac{\bar{\mu}_m(t)}{\mu_m(t)}\right]^\tau dy_m(t)\right\} \;\middle|\; X,\alpha,y_m(T)>0\right]\Pr[y_m(T)>0] \\
&= \exp\left(-\int_0^T \mu_m(t)dt\right) \\
&\quad + \sum_{k=1}^{\infty}\int_0^T\int_0^{t_k}\cdots\int_0^{t_2}\left(\prod_{j=1}^{k}\mu_m(t_j)\left[\frac{\bar{\mu}_m(t_j)}{\mu_m(t_j)}\right]^\tau\right)\exp\left(-\int_0^T\mu_m(t)dt\right)dt_1\cdots dt_k \\
&= \exp\left(-\int_0^T \mu_m(t)dt\right) \\
&\quad + \sum_{k=1}^{\infty}\frac{1}{k!}\int_0^T\int_0^{T}\cdots\int_0^{T}\left(\prod_{j=1}^{k}\mu_m(t_j)^{1-\tau}\bar{\mu}_m(t_j)^\tau\right)\exp\left(-\int_0^T\mu_m(t)dt\right)dt_1\cdots dt_k \\
&= \sum_{k=0}^{\infty}\frac{1}{k!}\left(\int_0^T\mu_m(t)^{1-\tau}\bar{\mu}_m(t)^\tau dt\right)^k\exp\left(-\int_0^T\mu_m(t)dt\right) \\
&= \exp\left(\int_0^T\mu_m(t)^{1-\tau}\bar{\mu}_m(t)^\tau dt-\int_0^T\mu_m(t)dt\right) \qquad (4.133)
\end{aligned}
$$

For any $\tau \geq 0$, we have via (4.128), (4.132), (4.133), and (4.129), a bound on the conditional pairwise error probability,

$$\Pr(\, X \to \bar{X} \mid \alpha\,) \leq \prod_{m=1}^{M}\exp\left(\int_0^T\{\mu_m(t)^{1-\tau}\bar{\mu}_m(t)^\tau-(1-\tau)\mu_m(t)-\tau\bar{\mu}_m(t)\}dt\right). \qquad (4.134)$$

The bound is valid for any choice of non-negative τ, and we could in theory choose the τ that minimizes this upper bound. Performing this optimization, however, is difficult, and we will, for convenience, choose $\tau = 1/2$. This choice of τ yields the bound

$$\Pr(\, X \to \bar{X} \mid \alpha\,) \leq \prod_{m=1}^{M}\exp\left(-\frac{1}{2}\int_0^T\left\{\sqrt{\mu_m(t)}-\sqrt{\bar{\mu}_m(t)}\right\}^2 dt\right), \qquad (4.135)$$

or, equivalently

$$\Pr(\, X \to \bar{X} \mid \alpha \,) \leq \exp\left(-\frac{1}{2} \sum_{m=1}^{M} \int_0^T \left\{ \sqrt{\sum_{n=1}^{N} \alpha_{nm} x_n(t) + \lambda_m} - \sqrt{\sum_{n=1}^{N} \alpha_{nm} \bar{x}_n(t) + \lambda_m} \right\}^2 dt \right). \quad (4.136)$$

Equation (4.136) is the MIMO generalization to the Chernoff bound commonly used in the photon-counting receiver literature, e.g., [12, 15, 11]. In the following sections, we will evaluate this bound for a few simple space-time coding techniques.

4.2.3 Repetition Spatial Coding

We saw in Section 4.1.1 that an OOK repetition spatial code, i.e., all transmit lasers turn on or off in unison, is a capacity-achieving signalling scheme in both the low and high signal-to-noise ratio regimes. We now will evaluate the Chernoff bound (4.136) for an OOK repetition spatial code with minimum probability of error detection (4.123).

We will divide the transmission block $[0, T]$ into K non-overlapping intervals of width $\Delta = T/K$. In each of these intervals, all transmit lasers are either on or off, i.e., $x_n(t) = A_n X_k$, $(k-1)\Delta \leq t < k\Delta$, and $X_k \in \{0, 1\}$. Let $X \equiv \{X_1, \ldots, X_K\}$ and $\bar{X} \equiv \{\bar{X}_1, \ldots, \bar{X}_K\}$ be two codewords from this OOK repetition spatial code. Using this transmission scheme, the Chernoff bound on pairwise error probability becomes

$$\begin{aligned} \Pr(\, X \to \bar{X} \mid \alpha \,) &\leq \exp\left(-\frac{\Delta}{2} \sum_{m=1}^{M} \sum_{k=1}^{K} R_m \left\{ \sqrt{X_k + s_m} - \sqrt{\bar{X}_k + s_m} \right\}^2 \right) && (4.137) \\ &= \exp\left(-\frac{\Delta \mathcal{D}[X, \bar{X}]}{2} \sum_{m=1}^{M} R_m \left\{ \sqrt{1 + s_m} - \sqrt{s_m} \right\}^2 \right) && (4.138) \end{aligned}$$

where $\mathcal{D}[X, \bar{X}] = \sum_{k=1}^{K} (X_k - \bar{X}_k)^2$ is the Hamming distance between the two codewords, $R_m \equiv \sum_{n=1}^{N} \alpha_{nm} A_n$ is the peak received power at the m-th aperture, and $s_m = \lambda_m / R_m$ is a noise-to-signal ratio at the m-th detector. We will now simplify

this bound on the pairwise error probability for the low and high signal-to-noise ratio regimes.

High Signal-to-Noise Ratio Regime

In the high signal-to-noise ratio regime, i.e., $s_m \to 0$, we have

$$\begin{aligned} \Pr(\, X \to \bar{X} \mid \alpha \,) &\leq \exp\left(-\frac{\Delta R}{2} \sum_{k=1}^{K} \left\{ X_k - \bar{X}_k \right\}^2 \right) \\ &= \exp\left(-\frac{\Delta R}{2} \mathcal{D}[X, \bar{X}] \right), \end{aligned} \tag{4.139}$$

where $R = \sum_{m=1}^{M} R_m = \sum_{m=1}^{M} \sum_{n=1}^{N} \exp(2\chi_{nm}) A_n$ is the aggregate peak received power. Notice that this pairwise error probability is very similar to the pairwise error of an orthogonal design STC for coherent detection receivers (3.49). The unconditional pairwise error probability is then bounded by

$$\begin{aligned} \Pr(\, X \to \bar{X} \,) &= E[\Pr(\, X \to \bar{X} \mid \alpha \,)] \\ &\leq \prod_{m=1}^{M} \prod_{n=1}^{N} E\left[\exp\left(-\frac{\Delta A_n \mathcal{D}[X, \bar{X}]}{2} \alpha_{nm} \right) \right] \\ &= \prod_{n=1}^{N} \left[\mathrm{Fr}\left(\frac{\Delta A_n \mathcal{D}[X, \bar{X}]}{2}; -\sigma_\chi^2, \sigma_\chi^2 \right) \right]^M \qquad (4.140) \\ &= \left[\mathrm{Fr}\left(\frac{\Delta A \mathcal{D}[X, \bar{X}]}{2}; -\sigma_\chi^2, \sigma_\chi^2 \right) \right]^{MN} \qquad (4.141) \end{aligned}$$

where the last equality holds for identical transmitters ($A_n = A$), and

$$\mathrm{Fr}(a; m, s^2) \equiv \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi s^2}} \exp\left(-a e^{2x} \right) \exp\left(-\frac{1}{2s^2} (x - m)^2 \right) dx, \tag{4.142}$$

is the lognormal density frustration function [28].

Low Signal-to-Noise Ratio Regime

We can also express the Chernoff bound approximately in terms of the frustration function in the low signal-to-noise ratio regime. Using the approximation $\sqrt{1+\epsilon} \approx$

$1+\epsilon/2$ for small ϵ, we have for $s_m \gg 1$

$$\begin{aligned}
\Pr(\, X \to \bar{X} \mid \alpha \,) &\leq \exp\left(-\frac{\Delta}{2}\sum_{m=1}^{M}\sum_{k=1}^{K} R_m s_m \left\{\sqrt{1+\frac{X_k}{s_m}} - \sqrt{1+\frac{\bar{X}_k}{s_m}}\ \right\}^2\right) \\
&\approx \exp\left(-\frac{\Delta}{2}\sum_{m=1}^{M}\sum_{k=1}^{K} R_m s_m \left\{1+\frac{X_k}{2s_m} - \left[1+\frac{\bar{X}_k}{2s_m}\right]\ \right\}^2\right) \\
&= \exp\left(-\frac{\Delta}{8}\sum_{m=1}^{M}\frac{R_m^2}{\lambda_m}\mathcal{D}[X,\bar{X}]\right). \qquad (4.143)
\end{aligned}$$

Averaging over the path gains, the unconditional pairwise error probability is approximately bounded by

$$\Pr(\, X \to \bar{X} \,) \lesssim \prod_{m=1}^{M} E\left[\exp\left(-\frac{\Delta\mathcal{D}[X,\bar{X}]}{8\lambda_m}\ R_m^2\right)\right]. \qquad (4.144)$$

We can approximate the sum of independent lognormal random variables $R_m = \sum_{n=1}\alpha_{nm}A_n$ as being lognormal using moment matching (see Appendix A.5.2). Specifically, we approximate $R_m \approx \exp(u_m)$, where u_m is Gaussian with mean and variance

$$E[u_m] = \log\left(\frac{P_1}{\sqrt{1+\frac{S_4P_2}{P_1^2}}}\right) = \log\left(\frac{NA}{\sqrt{1+\frac{e^{4\sigma_\chi^2}-1}{N}}}\right) \qquad (4.145)$$

$$\mathrm{var}[u_m] = \log\left(1+\frac{S_4P_2}{P_1^2}\right) = \log\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right), \qquad (4.146)$$

where the last equalities are for identical transmitters. Using these log-moments, results in an approximate upper bound to the unconditional pairwise error probability at low signal-to-noise

$$\begin{aligned}
\Pr(\, X \to \bar{X} \,) &\lesssim \prod_{m=1}^{M} E\left[\exp\left(-\frac{\Delta\mathcal{D}[X,\bar{X}]}{8\lambda_m}\ e^{2u_m}\right)\right] \\
&= \left[\mathrm{Fr}\left(\frac{\Delta\mathcal{D}[X,\bar{X}]}{8\lambda}; E[u_m], \mathrm{var}[u_m]\right)\right]^M, \qquad (4.147)
\end{aligned}$$

where the last expression holds for identical receivers ($\lambda_m = \lambda$). Notice that for large

numbers of transmit apertures $N \gg e^{4\sigma_\chi^2} - 1$, the log-variance $\mathrm{var}[u_m]$ approaches zero, the log-mean $E[u_m]$ approaches $\log NA$, and we have

$$\Pr(\, X \to \bar{X}\,) \lesssim \exp\left(-\frac{\Delta\mathcal{D}[X,\bar{X}]M(NA)^2}{8\lambda}\right). \tag{4.148}$$

For large numbers of identical detectors, the sum $\sum_{m=1}^{M} R_m^2/\lambda M$ approaches its mean (4.102). The error probability for identical transmitters, and large numbers of identical receivers, is approximately less than

$$\Pr(\, X \to \bar{X}\,) \lesssim \exp\left(-\frac{\Delta\mathcal{D}[X,\bar{X}]M(NA)^2}{8\lambda}\left[1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right]\right). \tag{4.149}$$

As with capacity, we see that increasing the number of transmit apertures in the low signal-to-noise ratio regime is more beneficial (up to the accuracy of this approximate bound) than increasing the number of receive apertures. In the high signal-to-noise ratio regime, the bound improves the same amount when changing either the number of transmit apertures or receive apertures.

4.2.4 Switching Diversity

Traditionally, switching diversity refers to using a branch (e.g., a laser and detector pair) until the branch signal-to-noise ratio drops below some specified threshold, then "switching" to the next branch [32]. We will examine a slightly different version of switching diversity, in which we will continually switch between transmit apertures, regardless of signal-to-noise ratio, and decode using a minimum probability of error receiver.

Motivation

We examine this form of switching diversity because of intuition that we gained from the development of space-time codes for coherent detection (see Chapter 3). For coherent detection, the space-time code that minimized the pairwise error probability satisfied a particular design criterion. We will see that the conditional pairwise er-

ror probability for photon-counting receivers in the low signal-to-noise regime has a similar structure to that of coherent detection receivers.

In the low signal-to-noise regime ($\lambda_m \to \infty$), we can use the approximation $\sqrt{1+\epsilon} \approx 1+\epsilon/2$ for small ϵ to write the conditional pairwise error probability (4.136) as

$$\begin{aligned}
&\Pr(\, X \to \bar{X} \mid \alpha \,) \\
&\quad \lesssim \quad \exp\left(-\frac{1}{8} \sum_{m=1}^{M} \frac{1}{\lambda_m} \int_0^T \left\{ \sum_{n=1}^{N} \alpha_{nm} [x_n(t) - \bar{x}_n(t)] \right\}^2 dt \right) \\
&\quad = \quad \exp\left(-\frac{1}{8} \sum_{m=1}^{M} \sum_{n=1}^{N} \sum_{k=1}^{N} \alpha_{nm} \alpha_{km} A_{nk} / \lambda_m \right), \qquad (4.150)
\end{aligned}$$

where

$$A_{nk} \equiv \int_0^T [x_n(t) - \bar{x}_n(t)][x_k(t) - \bar{x}_k(t)] dt. \qquad (4.151)$$

Notice the similarities between the coherent detection pairwise error probability (3.14), (3.17), and (3.18) with that of photon-counting in high noise, (4.150) and (4.151). In the coherent detection case, we saw that choosing $A_{nk} = \beta \delta_{nk}$, for some constant β, minimized the unconditional pairwise error probability. Although, we will not perform a similar optimization for photon-counting receivers, we will use the coherent detection result to motivate a similar choice for A_{nk} here.

Code Construction and Performance

We will now construct a signalling scheme that satisfies $A_{nk} = \beta_n \delta_{nk}$. In this scheme, the transmitter will continually switch between transmit apertures as follows. Divide the transmission block $[0, T]$ into N intervals of width Δ. During the n-th interval, transmit aperture n will send $x_n(t) = A_n X_n$, $(n-1)\Delta \le t < n\Delta$, $X_n \in \{0, 1\}$, while the other transmit apertures send nothing. For two codewords $X \equiv \{X_1, \ldots, X_N\}$ and $\bar{X} \equiv \{\bar{X}_1, \ldots, \bar{X}_N\}$, this "switching" space-time code satisfies

$$A_{nk} = A_n^2 [X_n - \bar{X}_n]^2 \Delta \delta_{nk}. \qquad (4.152)$$

The receiver performs minimum probability of error detection with knowledge of the path gains.

Low Signal-to-Noise Ratio Regime In the low signal-to-noise regime ($\lambda_m \to \infty$), the conditional pairwise error probability bound (4.150) becomes

$$\begin{aligned}\Pr(\, X \to \bar{X} \mid \alpha\,) &\lesssim \exp\left(-\frac{\Delta}{8}\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{N}\alpha_{nm}\alpha_{km}A_n^2[X_n-\bar{X}_n]^2\delta_{nk}/\lambda_m\right)\\ &= \exp\left(-\frac{\Delta}{8}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}^2A_n^2[X_n-\bar{X}_n]^2/\lambda_m\right), \qquad (4.153)\end{aligned}$$

for our switching diversity scheme. The unconditional error probability is then

$$\begin{aligned}\Pr(\, X \to \bar{X}\,) &\lesssim \prod_{m=1}^{M}\prod_{n=1}^{N}E\left[\exp\left(-\frac{\Delta}{8}\alpha_{nm}^2A_n^2[X_n-\bar{X}_n]^2/\lambda_m\right)\right]\\ &= \prod_{m=1}^{M}\prod_{\substack{n=1\\ n:X_n=\bar{X}_n}}^{N}\mathrm{Fr}\left(\frac{\Delta A_n^2}{8\lambda_m};-2\sigma_\chi^2,4\sigma_\chi^2\right) \qquad (4.154)\\ &= \left[\mathrm{Fr}\left(\frac{\Delta A^2}{8\lambda};-2\sigma_\chi^2,4\sigma_\chi^2\right)\right]^{M\mathcal{D}[X,\bar{X}]}, \qquad (4.155)\end{aligned}$$

where the last equality holds for identical transmitters and receivers.

The error probability for identical transmitters, and large numbers of identical receivers, is approximately less than

$$\Pr(\, X \to \bar{X}\,) \lesssim \exp\left(-\frac{\Delta\mathcal{D}[X,\bar{X}]MA^2}{8\lambda}\; e^{4\sigma_\chi^2}\right). \qquad (4.156)$$

High Signal-to-Noise Ratio Regime In the high signal-to-noise regime ($\lambda_m = 0$), the conditional pairwise error probability bound from (4.136) becomes

$$\begin{aligned}\Pr(\, X \to \bar{X} \mid \alpha\,) &\le \exp\left(-\frac{\Delta}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\left\{\sqrt{\alpha_{nm}A_nX_n}-\sqrt{\alpha_{nm}A_n\bar{X}_n}\;\right\}^2\right)\\ &= \exp\left(-\frac{\Delta}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A_n\left\{X_n-\bar{X}_n\;\right\}^2\right), \qquad (4.157)\end{aligned}$$

for our switching diversity scheme. The unconditional pairwise error is then bounded by

$$\begin{aligned}
\Pr(\, X \to \bar{X}\,) &\leq \prod_{\substack{n=1 \\ n:X_n \neq \bar{X}_n}}^{N} \left\{ E\left[\exp\left(-\frac{\Delta A_n}{2} \alpha_{nm} \right) \right] \right\}^M \\
&= \prod_{\substack{n=1 \\ n:X_n \neq \bar{X}_n}}^{N} \left[\mathrm{Fr}\left(\frac{\Delta A_n}{2}; -\sigma_\chi^2, \sigma_\chi^2 \right) \right]^M \qquad (4.158) \\
&= \left[\mathrm{Fr}\left(\frac{\Delta A}{2}; -\sigma_\chi^2, \sigma_\chi^2 \right) \right]^{M\mathcal{D}[X,\bar{X}]}, \qquad (4.159)
\end{aligned}$$

where the last equality is for identical transmitters and receivers.

4.2.5 Comparison of Repetition and Switching Diversity

Repetition and switching diversity are both simple space-time coding schemes. Repetition diversity requires multiple lasers operating at different wavelengths (or separated sufficiently in angle) driven by a common modulator. Switching diversity, if done optically, only requires one laser, however. The cost of a high-speed optical switch, though, could negate this advantage.

As proposed, the OOK repetition spatial space-time code uses N times the total transmit power of the OOK switching code. To compare the error performance of these two schemes, let P be the average power per receiver per bit, assuming equally likely ones and zeros. For repetition spatial coding, the average power per receiver is $P = NA/2$. Because only one transmitter is active during each bit interval, the average power per receiver for switching diversity is $P = A/2$. In the low noise regime, the Chernoff bound on switching diversity pairwise error probability is (4.159) with $A = 2P$,

$$\Pr(\, X \to \bar{X}\,) \leq \left[\mathrm{Fr}\left(\Delta P; -\sigma_\chi^2, \sigma_\chi^2 \right) \right]^{M\mathcal{D}[X,\bar{X}]}. \qquad (4.160)$$

The bound on repetition diversity (4.141) with $A = 2P/N$ in the low noise regime is

$$\Pr(\, X \to \bar{X}\,) \leq \left[\mathrm{Fr}\left(\Delta P \, \frac{\mathcal{D}[X, \bar{X}]}{N}; -\sigma_\chi^2, \sigma_\chi^2 \right) \right]^{MN} \tag{4.161}$$

To compare these two bounds, suppose that the number of intervals in the repetition code is equal to the number of transmit apertures, i.e., $K = N$. Furthermore, suppose that we are interested in codewords that are maximally separated, i.e., $\mathcal{D}[X, \bar{X}] = N$. In this case, the bounds on pairwise error probability become identical.

A relevant example, in which $\mathcal{D}[X, \bar{X}] = N = 2$, is binary pulse-position modulation with two transmit apertures. In this case, the codewords are $X = \{1, 0\}$ and $\bar{X} = \{0, 1\}$. To the accuracy of the Chernoff bounds, repetition and switching diversity have the same error performance in the high signal-to-noise ratio regime. Essentially, the switching space-time code averages over all the path gains in time, while the repetition space-time code spatially averages the path gains.

In the low signal-to-noise ratio regime, we can compare the error probability approximate upper bounds (4.149) and (4.156) for large numbers of receive apertures. Setting $A = 2P/N$ for the repetition spatial code we have,

$$\Pr(\, X \to \bar{X}\,) \lesssim \exp\left(-\frac{\Delta\mathcal{D}[X, \bar{X}] M P^2}{2\lambda} \left[1 + \frac{e^{4\sigma_\chi^2} - 1}{N} \right] \right). \tag{4.162}$$

The switching bound for large receive apertures using $A = 2P$ is

$$\Pr(\, X \to \bar{X}\,) \lesssim \exp\left(-\frac{\Delta\mathcal{D}[X, \bar{X}] M P^2}{2\lambda} \; e^{4\sigma_\chi^2} \right). \tag{4.163}$$

Because $1 + (\exp[4\sigma_\chi^2] - 1)/N \leq \exp[4\sigma_\chi^2]$, to within the accuracy of these approximate Chernoff bounds, switching diversity appears to be slightly better at low signal-to-noise ratio than repetition spatial coding, especially for large numbers of transmit and receive apertures.

Chapter 5

Optically-Preamplified Receivers

The second direct-detection channel consists of intensity modulation, atmospheric propagation, and optically-preamplified direct detection. Optical amplification pulls the received signal power above the transimpedance amplifier thermal noise floor, and creates a channel sharing characteristics of both a photon counting channel and an additive, white, Gaussian noise channel. With optical amplification, the required power to achieve a given bit error rate is much less than without amplification. For example, without optical amplification, the single-transmitter, single-receiver, 1.25 Gbps direct detection testbed in Chapter 6 requires -25 dBm average receive power per bit to achieve 10^{-9} error rates. With amplification, however, it only requires one-hundredth the power, -45 dBm.

We model our optically-preamplified direct detection channel as follows. The n-th transmitter sends a sequence of symbols (intensities) $\{ x_n(t) \mid 1 \leq t \leq T \}$ over T, non-overlapping time-slots. We assume that the overlap between symbols at the transmitter and receiver is negligible. In other words, we avoid the complications that arise from intersymbol interference, and assume that the receiver samples conditioned on the path gains and transmitted symbols are statistically independent. Furthermore, we approximate these sample statistics by those arising from the application of constant transmit power. In this chapter, we use the index t to denote discrete time.

Generalizing the derivation of Section 2.3.4 to an N transmit and M receive aperture system, we model the output of the m-th receiver during the t-th symbol

period as

$$y_m(t) = \sum_{n=1}^{N} \alpha_{nm} x_n(t) + w_m(t), \tag{5.1}$$

where $w_m(t)$ is a real-valued Gaussian noise with zero mean and signal-dependent variance

$$\sigma_m^2(t) = \sigma^2 + \gamma \sum_{n=1}^{N} \alpha_{nm} x_n(t). \tag{5.2}$$

The constants in this expression come from (2.57), (2.58), (2.59), and (2.60),

$$\begin{aligned} \sigma^2 &= \frac{1}{G_{\mathrm{O}}^2} \left\{ \overbrace{\left(\frac{e\eta}{hf_c}\right)^{-2} N_{\mathrm{TIA}} B_{\mathrm{E}}}^{\text{Thermal Noise}} + \overbrace{D_{\mathrm{pol}} N_{\mathrm{O}}^2 (2B_{\mathrm{O}} - B_{\mathrm{E}}) B_{\mathrm{E}}}^{\text{Beat Noise}} \right. \\ &\qquad \left. + \overbrace{e \left(\frac{e\eta}{hf_c}\right)^{-1} \left[P_{\mathrm{D}} + D_{\mathrm{pol}} N_{\mathrm{O}} B_{\mathrm{O}}\right] 2B_{\mathrm{E}}}^{\text{Shot Noise}} \right\}, \end{aligned} \tag{5.3}$$

and,

$$\gamma = \frac{4 N_{\mathrm{O}} B_{\mathrm{E}}}{G_{\mathrm{O}}} \left[\overbrace{\frac{hf_c}{2\eta N_{\mathrm{O}}}}^{\text{Shot Noise}} + \overbrace{1}^{\text{Beat Noise}} \right]. \tag{5.4}$$

We have also included the number of noise polarization modes D_{pol} in the variance.

Because we are coupling into a single mode fiber (see Figure 2-8), each transmitter must use a different carrier frequency, spaced sufficiently apart, such that their optical powers add, i.e., (2.28) holds. Consequently, the optical bandwidth B_{O} must scale with the number of transmit apertures in a system designed to minimize collection of extraneous light.

For the communication systems considered in this chapter, however, we will use a fixed optical bandwidth of 41 GHz. We do so for two reasons. First, this bandwidth corresponds to the nominal bandwidth of the optical filter in the optical carrier (OC)-24 testbed of Chapter 6. With data rates of 1.25 Gbps, this optical bandwidth can

easily accommodate several transmit apertures. Using a fixed optical bandwidth will slightly bias our results in favor of transmit diversity, because adding receive apertures also increases the noise, but adding transmit apertures does not. In practice, reducing the optical bandwidth will reduce the fixed component of the noise variance, σ^2, improving communication reliability.

Second, if in the future, multi-mode optical amplifiers become prevalent, we could use transmitter angle separation to create the incoherent addition of the optical fields. In this case, the optical bandwidth would not need to scale with the number of transmit apertures.

As in Chapter 4, we use the notation α_{nm} to represent the *power* path gain from transmitter n to receiver m. The path gains are normalized so that atmospheric propagation does not on average attenuate or amplify the transmitted power, i.e., $E[\alpha_{nm}] = E[\exp(2\chi_{nm})] = 1$, by setting $\text{var}[\chi_{nm}] = -E[\chi_{nm}] = \sigma_\chi^2$. Also, $x_n(t)$ is the n-th transmit *power* in Watts measured at each receive aperture in the absence of fading[1].

Under our no intersymbol interference approximation, conditioned on the transmitted symbols and path gains, we assume that the noise is independent from symbol to symbol and across receivers, i.e., $\{\ w_m(t) \mid 1 \le m \le M, 1 \le t \le T\ \}$ are conditionally independent.

5.1 Capacity

Although this channel is still an additive noise channel [21], the signal-dependent noise makes evaluation of its capacity difficult. Furthermore, a capacity comparison with the photon-counting receiver channel of (5.1) under peak- and average-power constraints is difficult. A closed form solution for the capacity of the peak- and average-power constrained Gaussian channel does not exist in general [57, 54], let alone for signal-dependent noise. Furthermore, without a peak-power constraint, the

[1] Although a quadrature amplitude modulator could produce the transmitted symbols, the detector only responds to the squared magnitude of the optical field. For this reason, we only consider intensity modulation.

capacity of the MIMO Poisson channel is infinite.

Smith [57] and Shamai [54] show that for peak- and average-power constrained Gaussian channels, the capacity achieving distribution is discrete. This observation will motivate us to examine discrete signalling schemes such as on-off keying.

Ben-Eli, *et. al.*, derived bounds on the cutoff rate for the optically pre-amplified, direct detection channel in the shot-noise limited regime for OOK signalling [4]. While they did not use to a conditional Gaussian approximation, they neglect background and thermal noise.

In what follows, we will develop lower bounds based on discrete memoryless channel (DMC) representations of the MIMO, atmospheric propagation, optically-preamplified, direct detection channel (5.1). These DMC representations assume an OOK spatial repetition transmitter and a linear combining, threshold-decision receiver. We will show that equal-gain combining minimizes the error probability for this architecture class when the fixed component of the noise variance is much smaller than the signal-dependent component. When the fixed component dominates, however, maximal ratio combining is the best weighting scheme. Under nominal fading conditions, the average received power threshold that delineates these two regimes is -56 dBm for the 1.25 Gbps testbed in Chapter 6. When using more much power than this threshold, equal-gain combining is the best linear combining strategy.

We then develop approximations to the probability that the channel can support a desired bit error rate. We also show how these approximations are useful in reducing the number of computations necessary to simulate the error performance for equal-gain combining receivers and selection diversity schemes.

5.1.1 Discrete Memoryless Channel Representations

We will develop a simple lower bound to the capacity of our optically-preamplified channel (5.1) under a peak power constraint, that reflects current, technological implementations. We will develop a lower bound by restricting ourselves to OOK signalling, satisfying a peak power constraint. We have already seen in Chapter 4 that OOK repetition spatial coding is capacity achieving for photon-counting receivers. While

repetition spatial coding might not be optimal for this channel, we will, nonetheless, restrict ourselves to this transmitter structure, i.e., $x_n(t) = A_n X_t$, where $X_t \in \{0, 1\}$.

Furthermore, we will use a linear-combining receiver structure with threshold detection. We further assume that the receiver knows and uses the path gains to set the threshold ξ and combining weights c_m. During each time-slot the receiver will make a hard decision as to whether a one or zero was sent based on the received sample,

$$y(t) = \sum_{m=1}^{M} c_m(\alpha) y_m(t), \tag{5.5}$$

and threshold, yielding

$$Y_t = \begin{cases} 1, & y(t) > \xi(\alpha) \\ 0, & \text{otherwise.} \end{cases} \tag{5.6}$$

The pairwise error probabilities for this OOK spatial repetition transmitter, and linear-combining, threshold-decision receiver are ([46], pg. 188)

$$\Pr(Y_t = 0 \mid X_t = 1, \alpha) = Q\left(\frac{m_1 - \xi}{s_1}\right), \tag{5.7}$$

and

$$\Pr(Y_t = 1 \mid X_t = 0, \alpha) = Q\left(\frac{\xi - m_0}{s_0}\right), \tag{5.8}$$

where, for $x \in \{0, 1\}$,

$$m_x = E[\, y(t) \mid X_t = x, \alpha \,] = x \sum_{m=1}^{M} c_m \sum_{n=1}^{N} \alpha_{nm} A_n, \tag{5.9}$$

and

$$s_x^2 = \text{var}[\, y(t) \mid X_t = x, \alpha \,] = \sigma^2 \sum_{m=1}^{M} c_m^2 + x\gamma \sum_{m=1}^{M} c_m^2 \sum_{n=1}^{N} \alpha_{nm} A_n \tag{5.10}$$

are the conditional means and variances, respectively. We assume that $s_1, s_0 > 0$. We will address the case when $s_0 = 0$ in Section 5.1.2.

The threshold that minimizes the probability of error, assuming equally likely ones

and zeros, is ([46], pg. 198)

$$\xi = \frac{m_0 s_1^2 - m_1 s_0^2 + s_0 s_1 \sqrt{(m_1 - m_0)^2 + 2(s_1^2 - s_0^2)\log(s_1/s_0)}}{s_1^2 - s_0^2}, \tag{5.11}$$

for $s_1 \neq s_0$. For $s_1 = s_0$, the optimal threshold is the average (or midpoint) of the conditional means,

$$\xi = \frac{m_0 + m_1}{2}. \tag{5.12}$$

The instantaneous bit error rate using midpoint thresholding is

$$\Pr(Y_t \neq X_t \mid \alpha) = \frac{1}{2} Q\left(\frac{m_1 - m_0}{2s_1}\right) + \frac{1}{2} Q\left(\frac{m_1 - m_0}{2s_0}\right), \tag{5.13}$$

assuming equally likely ones and zeros.

We will now examine two DMC representations of this channel. The first uses a threshold that makes the probability of decoding a zero as a one the same as the probability of decoding a one as a zero. This threshold is near optimal in the "high signal-to-noise ratio regime," to be quantified shortly. The second channel represents the case when the fixed component of the noise vanishes. In this case, the "Z-Channel" is a good model for communication.

Binary Symmetric Channel Representation

When $(m_1 - m_0)^2 \gg 2(s_1^2 - s_0^2)\log(s_1/s_0)$, as is the case for high signal-to-noise ratios, the optimum threshold becomes

$$\xi \approx \frac{m_0 s_1 + m_1 s_0}{s_1 + s_0}. \tag{5.14}$$

This threshold makes the pairwise error probabilities equal, and

$$\epsilon \equiv \Pr(Y_t = 1 \mid X_t = 0, \alpha) = \Pr(Y_t = 0 \mid X_t = 1, \alpha) = Q\left(\frac{m_1 - m_0}{s_1 + s_0}\right) \tag{5.15}$$

$$= Q\left(\frac{\sum_{m=1}^{M} c_m R_m}{\sqrt{\sigma^2 \sum_{m=1}^{M} c_m^2} + \sqrt{\sigma^2 \sum_{m=1}^{M} c_m^2 + \gamma \sum_{m=1}^{M} c_m^2 R_m}}\right), \tag{5.16}$$

is the cross-over probability for a binary symmetric channel (BSC) [14], where $R_m \equiv \sum_{n=1}^{N} \alpha_{nm} A_n$ is the received peak power on the m-th receive aperture. The instantaneous capacity of this OOK spatial repetition transmitter, and linear-combining, near-optimal threshold detection receiver is, therefore,

$$C_{\text{BSC}}(\alpha) = \log 2 - \mathcal{H}(\epsilon), \qquad \text{[nats/use]} \tag{5.17}$$

where $\mathcal{H}$ is the binary entropy function [14].

Z-Channel Representation

When the fixed component of the noise variance is small compared to the signal-dependent component, i.e.,

$$\sigma^2 \sum_{m=1}^{M} c_m^2 \ll \gamma \sum_{m=1}^{M} c_m^2 R_m, \tag{5.18}$$

we say that we are operating in the "low noise regime." We call this the low noise regime because the fixed component of the noise variance contains the thermal noise, background and amplified spontaneous emission (ASE) shot noise, ASE-plus-background beat noise, and dark current shot noise, see (5.3). The signal-dependent component of the noise contains the signal and ASE-plus-background beat noise and the signal shot noise, see (5.4).

Nominally, for $A_n \equiv A$ and $R_m \approx NA$, being in the low noise regime means,

$$\sigma^2 \ll \gamma NA = 2P\gamma, \tag{5.19}$$

where $P \equiv NA/2$ is the average power per bit at each receiver, assuming equally likely ones and zeros.

Although, technically, being in the "low noise" regime depends on the random path gains and combining weights, we can vary the transmit peak power A, until we are confident that the fixed component of the noise variance is much less than the signal-dependent component. Furthermore, the optimal duty cycle might not be one-half

for non-symmetric channels. Regardless, we will use (5.19) to nominally characterize the "low noise" regime. For example, using the parameters of the 1.25 Gbps testbed in Table 5.1, an average receive power much greater than $P \gg \sigma^2/2\gamma \approx -56$ dBm results in low noise operation most of the time.

As $\sigma \to 0$, the optimum threshold (5.11) approaches zero. Noting that $m_0 = 0$ and $s_0 \propto \sigma$, the probability of decoding a zero as a one vanishes,

$$\begin{aligned}\lim_{s_0\to 0} \Pr(Y_t = 1 \mid X_t = 0, \alpha) &= \lim_{s_0\to 0} Q\left(\frac{-m_1 s_0^2 + s_0 s_1 \sqrt{m_1^2 + 2(s_1^2 - s_0^2)\log(s_1/s_0)}}{(s_1^2 - s_0^2)s_0}\right) \\ &= Q\left(\frac{\lim_{s_0\to 0}\sqrt{m_1^2 + 2(s_1^2 - s_0^2)\log(s_1/s_0)}}{s_1}\right) \\ &= 0. \end{aligned} \tag{5.20}$$

On the other hand, the probability of decoding a one as a zero persists

$$\delta \equiv \lim_{s_0\to 0} \Pr(Y_t = 0 \mid X_t = 1, \alpha) = Q\left(\frac{m_1}{s_1}\right) = Q\left(\frac{\sum_{m=1}^{M} c_m R_m}{\sqrt{\gamma \sum_{m=1}^{M} c_m^2 R_m}}\right) \tag{5.21}$$

These conditional probabilities describe the "Z-channel", often associated with optical communications [59]. The capacity in nats per use of this channel is [59, 64],

$$C_{\mathrm{Z}}(\alpha) = \mathcal{H}\left(\frac{1}{e^{\kappa}+1}\right) - \frac{\kappa}{e^{\kappa}+1} = \log\left(1 + e^{-\kappa}\right) \approx \log 2 - \frac{1}{2}\mathcal{H}(\delta), \tag{5.22}$$

where $\kappa \equiv \mathcal{H}(\delta)/(1-\delta)$, and the last approximation holds for small δ. Comparing the Z-channel capacity with the binary symmetric channel capacity in the low noise regime,

$$\lim_{\sigma\to 0} C_{\mathrm{BSC}}(\alpha) = \log 2 - \mathcal{H}(\delta), \tag{5.23}$$

we see that, capacity is maximized in both cases by maximizing the ratio m_1/s_1.

Notice that the optimum receiver using a likelihood ratio test as $\sigma \to 0$ decides zero if the received sample $y(t) = 0$, and decides one if $y(t) \neq 0$. Because the probability of a continuous random variable taking on a given value is zero, the error probabilities are zero. By restricting ourselves to threshold receivers, we will sometimes decode

a one as a zero erroneously. Having this restriction, however, is more realistic, as practical receivers cannot sample with infinite precision.

5.1.2 Choosing the Combining Weights

In the BSC representation, for cross-over probabilities less than one-half, to maximize the capacity, we want to minimize the cross-over probability, or equivalently, maximize the Q-function argument of (5.16)

$$J \equiv \frac{m_1 - m_0}{s_1 + s_0} \tag{5.24}$$

$$= \frac{\sum_{m=1}^{M} c_m R_m}{\sqrt{\sigma^2 \sum_{m=1}^{M} c_m^2} + \sqrt{\sigma^2 \sum_{m=1}^{M} c_m^2 + \gamma \sum_{m=1}^{M} c_m^2 R_m}}. \tag{5.25}$$

Our goal is to find the combining weights to maximize this Q-function argument. We will show that when the signal-dependent component dominates the noise variance, i.e., $\sigma^2 \ll 2P\gamma$, equal-gain combining is the optimal linear combining strategy for optimal (5.11) and near-optimal (5.14) threshold receivers. When the fixed component dominates, i.e., $\sigma^2 \gg 2P\gamma$, maximal ratio combining is best.

Low Noise Regime

When the fixed component of the noise variance is small compared to the signal-dependent component, i.e., $\sigma^2 \ll 2P\gamma$, the Q-function argument in both the BSC (near-optimal threshold) and Z-channel (optimal threshold) representations is bounded via the Cauchy-Schwarz inequality[2],

$$J \approx \frac{\sum_{m=1}^{M} c_m R_m}{\sqrt{\gamma \sum_{m=1}^{M} c_m^2 R_m}} \leq \sqrt{\frac{\sum_{m=1}^{M} R_m}{\gamma}}, \tag{5.26}$$

with equality when $c_m \equiv \beta$ for some scalar β. Therefore, the optimal combining strategy in the low noise regime is to equally weight the detector outputs, e.g., choose

[2] An alternate proof is to differentiate with respect to the combining weight c_m, set it equal to zero, and solve for this weight. This optimal weight will be independent of the receive aperture index m.

$c_m \equiv 1$.

In the low noise regime, the probability of decoding a one as a zero for an equal-gain combining receiver is, therefore,

$$\delta \equiv \lim_{\sigma \to 0} \Pr(Y_t = 0 \mid X_t = 1, \alpha) = Q\left(\sqrt{\frac{R}{\gamma}}\right), \tag{5.27}$$

where $R = \sum_{m=1}^{M} R_m$ is the aggregate peak received power.

Notice that when the fixed component of the variance is much smaller than the signal-dependent component, the variance of the one bit is much greater than the variance of the zero bit, i.e., $s_1^2 \gg s_0^2$. As a result, the bit error rate for midpoint thresholding is dominated by the probability of decoding a one as a zero. Consequently, the best combining strategy when using midpoint thresholding is again equal-gain combining, which minimizes the probability of decoding a one as a zero, i.e., the first term of (5.13).

High Noise Regime

When the fixed component dominates the noise variance, i.e., $\sigma^2 \gg 2P\gamma$, the conditional variances of the one and zero bit are roughly equal, and the argument of the Q-function using the optimal midpoint threshold (5.12) is bounded by

$$J \approx \frac{\sum_{m=1}^{M} c_m R_m}{2\sigma\sqrt{\sum_{m=1}^{M} c_m^2}} \leq \frac{1}{2\sigma}\sqrt{\sum_{m=1}^{M} R_m^2}, \tag{5.28}$$

with equality when $c_m = \beta R_m$ for some scalar β. In the high noise regime, we see that classical maximal ratio combining [32] is the best linear combining strategy. The maximal ratio combining, optimal threshold (5.12) receiver, in the high noise regime, yields a BSC with cross-over probability

$$\epsilon \approx Q\left(\sqrt{\frac{\sum_{m=1}^{M} R_m^2}{4\sigma^2}}\right). \tag{5.29}$$

Parameter	Symbol	Value	Units
Opt. BW	B_{O}	41	GHz
Elect. BW	B_{E}	750	MHz
ASE Factor	n_{sp}	1	
Opt. Gain	G_{O}	40	dB
Opt. Wavelength	λ	1550	nm
Num. Pol. Modes	D_{pol}	2	
Det. Quant. Eff.	η	1	
Dark Power	P_{D}	0	W
TIA PSD	N_{TIA}	0	A^2/Hz
Backgrd. Power	N_{B}	0	W/Hz
Fixed Noise Comp.	σ^2	2.00×10^{-18}	W^2
Sig. Dep. Noise Factor	γ	3.85×10^{-10}	W

Table 5.1: This table shows the nominal parameters of the 1.25 Gbps testbed used in this chapter. These parameters represent a best case scenario with negligible background and thermal noise, ideal quantum efficiency, and minimum ASE noise.

Maximal ratio combining is also optimal in the high noise regime for midpoint thresholding, because midpoint thresholding is the optimal threshold in this regime.

5.1.3 Ergodic Capacity

We now examine the ergodic or average capacity of the binary symmetric channel lower bound on the optically preamplified, direct detection channel. Because the optimal duty cycle of the BSC is one-half, the transmitter does not need to know the path gains. We assume that the receiver can use the path gains to set the optimal threshold and choose the linear combining weights. We will examine this lower bound using the nominal parameters (Table 5.1) of the 1.25 Gbps experimental testbed in Chapter 6.

Figures 5-1 and 5-2 and Tables 5.2 and 5.3 show the ergodic capacity (averaged

over 500,000 channel realizations) for the OOK spatial repetition transmitter and linear combining, threshold-decision receiver, BSC representation. The receiver in Figure 5-1 uses equal-gain combining, while the receiver in Figure 5-2 uses maximal ratio combining. The average capacity is shown as a function of transmit and receive aperture numbers ($N = M$) and average optical power per bit at each receiver, $P = NA/2$, for identical transmitters, $A_n \equiv A$.

Both receivers use the threshold (5.14), which makes the probability of decoding a one as a zero the same as the probability of decoding a zero as a one. This threshold is near optimal when $(m_1 - m_0)^2 \gg 2(s_1^2 - s_0^2)\log(s_1/s_0)$. For example, using $R_m \approx NA$, $P = -65$ dBm average receive power, equal-gain combining, and $M = 1$, we see that this condition requires

$$2MP \gg \gamma \log\left(1 + \frac{2P\gamma}{\sigma^2}\right), \tag{5.30}$$

and evaluates to

$$6.3 \times 10^{-10} > 4.4 \times 10^{-11}, \tag{5.31}$$

with the left-hand side an order of magnitude greater than the right-hand side. For small $2P\gamma/\sigma^2 \approx 0.12$ ($P = -65$ dBm), the optimality condition becomes $M \gg \gamma^2/\sigma^2 \approx 0.07$, which holds for $M \geq 1$ under these operating conditions.

Notice from Figures 5-1 and 5-2 that the capacity grows linearly as a function of aperture number for small average receive power ($P = -65$ dBm). From Tables 5.2 and 5.3 we see that for $P = -62$ dBm, maximal ratio combining is slightly better than equal-gain combining (at most an 18% improvement for $M = N = 3$ in severe fading $\sigma_\chi^2 = 0.35$). For large enough power ($P = -55$ dBm), the error probability becomes small enough that even in severe fading ($\sigma_\chi^2 = 0.35$), five or six apertures is enough to attain the largest capacity for this channel ($\log 2 \approx 0.693$ nats/use). Also notice that an increase in log-amplitude variance increases the capacity for small power, but decreases it for large power.

In practice, for an optical amplifier gain greater than one, the noise variance fixed component σ^2 will always contain the ASE shot noise and the ASE-ASE beat noise, even in the absence of background, thermal, and dark current noises. The parameters

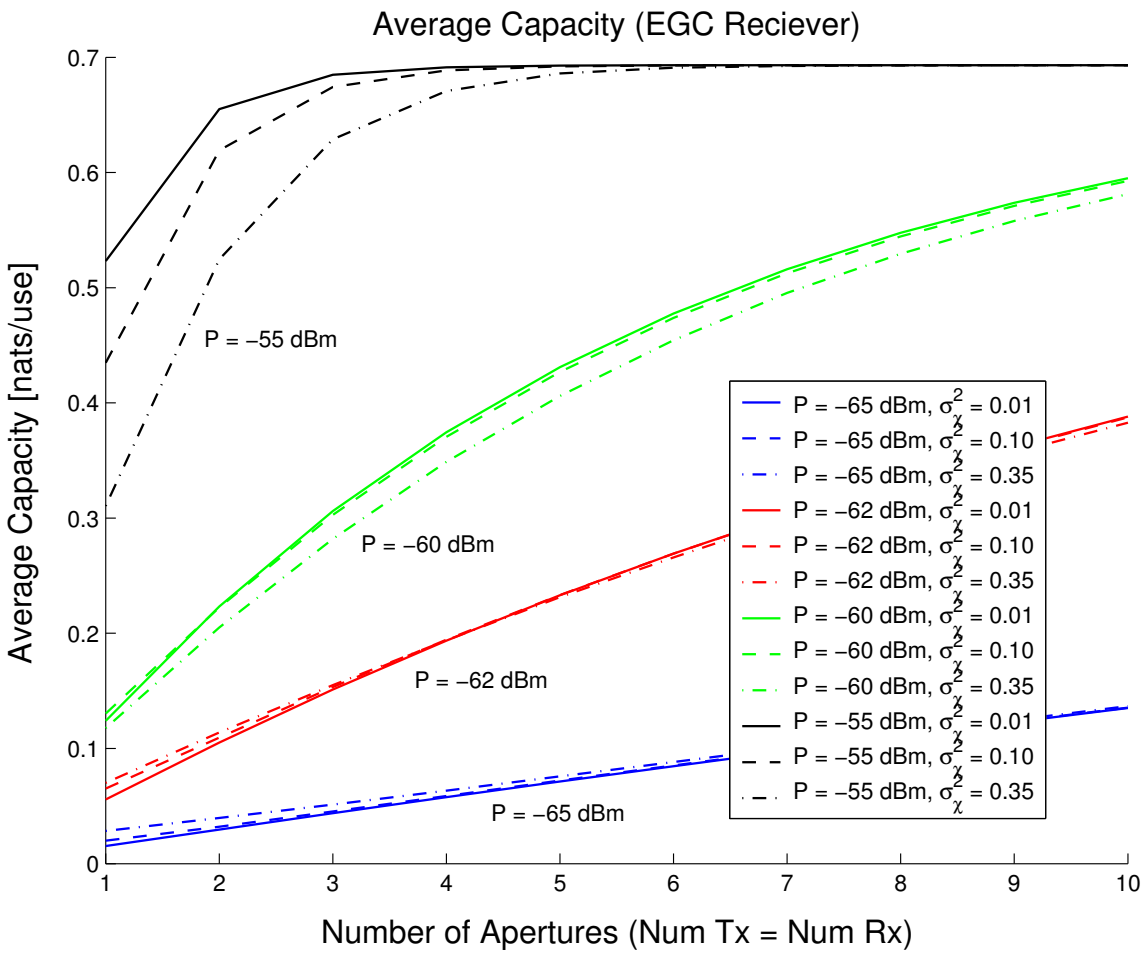

Figure 5-1: The average capacity for OOK spatial repetition transmitters and equal-gain combining receivers using near-optimal thresholds is shown as a function of log-amplitude variance and average optical power per bit at each receiver, i.e., $P = NA/2$. The standard error (sample standard deviation divided by the square root of the number of samples) on each estimate is less than 10^{-3}.

	Log-Amp. Var. (σ_χ^2)		
$N = M$	0.01	0.1	0.35
1	0.056	0.065	0.070
2	0.105	0.110	0.114
3	0.151	0.153	0.155
4	0.194	0.195	0.194
5	0.233	0.233	0.231
6	0.269	0.269	0.266
7	0.303	0.302	0.299
8	0.333	0.333	0.329
9	0.362	0.361	0.357
10	0.388	0.387	0.383

Table 5.2: Equal-gain combining average capacity [nats/use] from Figure 5-1 for an average optical power per bit at each receiver of $P = -62$ dBm. The standard error (sample standard deviation divided by the square root of the number of samples) on each estimate is less than 10^{-3}.

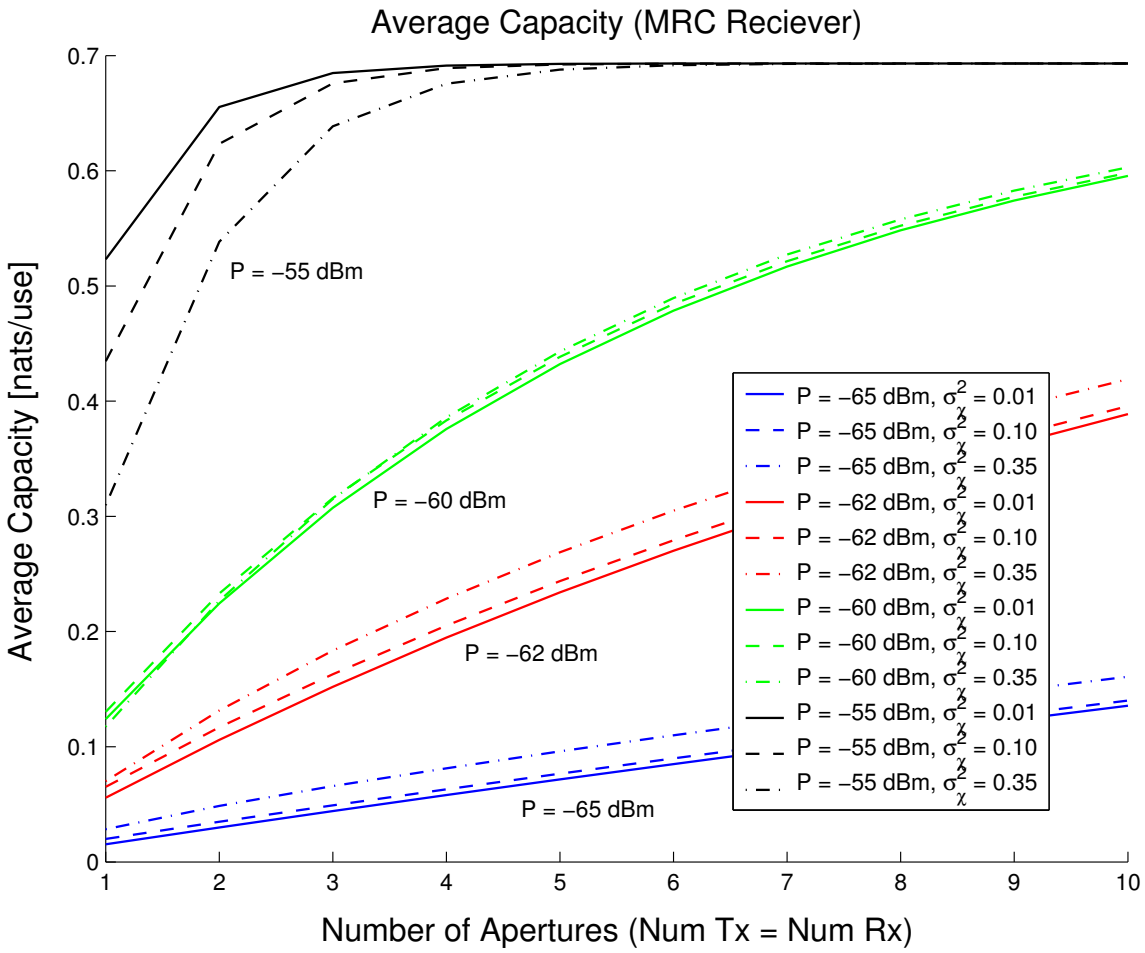

Figure 5-2: The average capacity for OOK spatial repetition transmitters and maximal ratio combining receivers using near-optimal thresholds is shown as a function of log-amplitude variance and average optical power per bit at each receiver, i.e., $P = NA/2$. The standard error on each estimate is less than 10^{-3}.

	Log-Amp. Var. (σ_χ^2)		
$N = M$	0.01	0.1	0.35
1	0.056	0.065	0.070
2	0.106	0.117	0.131
3	0.152	0.163	0.183
4	0.195	0.205	0.229
5	0.234	0.244	0.269
6	0.270	0.279	0.305
7	0.304	0.312	0.337
8	0.334	0.342	0.368
9	0.363	0.370	0.394
10	0.389	0.396	0.419

Table 5.3: Maximal ratio combining average capacity [nats/use] from Figure 5-1 for an average optical power per bit at each receiver of $P = -62$ dBm. The standard error on each estimate is less than 10^{-3}.

in Table 5.1 reflect these ideal operation conditions.

We now examine the low and high noise regimes as the optical gain varies. Had we set the optical gain to one, $G_{\mathrm{O}} = 1$, we would have eliminated the ASE noise, and made the fixed component zero, in the absence of thermal, dark, and background noises. The only remaining noise would have been the signal shot noise, and we would be operating in the "low-noise" regime. In practice, however, without optical amplification, thermal, dark, and background noises can be much greater than the signal shot noise.

On the other hand, increasing the optical gain, increases the ASE noise, until the ratio of ASE-ASE and ASE-signal beat noises (5.19) becomes

$$\lim_{G_{\mathrm{O}} \to \infty} \frac{\sigma^2}{2P\gamma} = \frac{D_{\mathrm{pol}} n_{\mathrm{sp}} h f_c (2B_{\mathrm{O}} - B_{\mathrm{E}})}{8P} \approx \frac{D_{\mathrm{pol}} n_{\mathrm{sp}} h f_c B_{\mathrm{O}}}{4P}, \tag{5.32}$$

where the last approximation is for optical bandwidths much greater than the electrical bandwidth, $B_{\mathrm{O}} \gg B_{\mathrm{E}}$. Alternatively, the average receive power must be large enough such that

$$P \gg \frac{D_{\mathrm{pol}} n_{\mathrm{sp}} h f_c B_{\mathrm{O}}}{4} \approx -55.80 \text{ dBm}, \tag{5.33}$$

to be in the low noise regime for high optical gain, e.g., $G_{\mathrm{O}} = 40$ dB. The last evaluation is for the nominal parameters in Table 5.1. Notice that this high optical gain threshold is within one tenth of a decibel of the nominal regime threshold $\sigma^2/2\gamma \approx -55.85$ dBm.

In theory, because the receiver knows the path gains, it can switch between equal gain and maximal ratio combining (or any weighting in between) when operating close to this regime threshold. For simplicity, however, in our simulations, we assume that the receive always uses either equal gain or maximal ratio combining.

Because the received powers in the ergodic capacity results were at or below this threshold, we conclude that we were barely operating in the high noise regime. At these power levels, the ergodic capacity results suggest that MRC is slightly better than EGC.

5.2 Coding

In this section, we will further examine the error probabilities of the OOK spatial repetition transmitter and linear combining, threshold-decision receivers developed in the previous section. As with outage capacity in Chapter 4, we will characterize the distribution of the bit error rate in the low and high noise regimes. We will also examine the average bit error rate for equal gain combining, maximal ratio combining, and selection diversity receivers using near-optimal and midpoint thresholding. These average bit error rate curves are indicative of the ideal performance of the testbed in Chapter 6 using perfect optical amplifiers ($n_{\text{sp}} = 1$).

5.2.1 Bit Error Rate Versus Outage Probability

When delay requirements prevent averaging over many channel realizations, the distribution of the instantaneous error probability, i.e., the error probability conditioned on the path gains, is a useful measure of communication reliability. That is, we want to characterize the probability that fading conditions yield a bit error rate worse than a desired level,

$$p_{\text{out}} = \Pr\left\{\Pr(Y_t \neq X_t \mid \alpha) > \text{BER}\right\}, \tag{5.34}$$

where, assuming equally likely ones and zeros, and near-optimal threshold (5.14), the conditional error probability (5.16) is

$$\Pr(Y_t \neq X_t \mid \alpha) = Q\left(\frac{\sum_{m=1}^{M} c_m R_m}{\sqrt{\sigma^2 \sum_{m=1}^{M} c_m^2} + \sqrt{\sigma^2 \sum_{m=1}^{M} c_m^2 + \gamma \sum_{m=1}^{M} c_m^2 R_m}}\right) \tag{5.35}$$

We now develop approximations to this outage probability in the low and high noise regimes.

Low Noise Regime

We saw in Section 5.1.2 that in low noise, i.e., $\sigma^2 \ll 2P\gamma$, equal-gain combining, $c_m \equiv 1$, minimizes the instantaneous error probability (5.35), and

$$\Pr(Y_t \neq X_t \mid \alpha) \approx Q\left(\sqrt{\frac{\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A_n}{\gamma}}\right). \tag{5.36}$$

As with the outage capacity development in Chapter 4, we will approximate the sum (4.96) as being lognormal with log-moments (4.89) and (4.90).

We will shortly see that in the high noise regime with maximal ratio combining, we can also express the outage probability in a similar form as equal gain combining in the low noise regime,

$$p_{\text{out}} \approx \Pr\left\{Q\left(\sqrt{\frac{S(\alpha)}{\tau}}\right) > \text{BER}\right\}, \tag{5.37}$$

where in the low-noise regime, $\tau \equiv \gamma$ and $S(\alpha) \equiv \sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}A_n \equiv R$. Using the approximation $S(\alpha) \approx e^u$, where $E[u] \equiv \mu$ and $\text{var}[u] \equiv \nu^2$ gives

$$\begin{aligned} p_{\text{out}} &\approx \Pr\left\{S(\alpha) \leq \tau[Q^{-1}(\text{BER})]^2\right\} && (5.38)\\ &\approx \Pr\left\{u \leq \log\tau + 2\log Q^{-1}(\text{BER})\right\} && (5.39)\\ &= 1 - Q\left(\frac{\log\tau + 2\log Q^{-1}(\text{BER}) - \mu}{\nu}\right). && (5.40)\end{aligned}$$

For identical transmitters, $A_n \equiv A$, these log-moments are (see (4.89), (4.90), (4.100) and (4.101)),

$$\begin{aligned} \mu &\equiv E[u] = \log\left(\frac{MNA}{\sqrt{1+\frac{e^{4\sigma_\chi^2}-1}{MN}}}\right) && (5.41)\\ \nu^2 &\equiv \text{var}[u] = \log\left(1+\frac{e^{4\sigma_\chi^2}-1}{MN}\right). && (5.42)\end{aligned}$$

Figure 5-3 shows the low noise outage probability approximation (5.37) using the

parameters in Table 5.1. The moment matching approximation is quite accurate for small numbers of apertures and small log-amplitude variances.

Although not shown, at -50 dBm average receive power, the low noise approximation (5.36) to the bit error rate is not that accurate. In other words, at this power level, because our optical bandwidth is so large, the fixed component of the noise variance is not negligible, and these figures represent an optimistic estimate of the bit error rate versus outage probability. For the parameters in Table 5.1, operating deep in the low noise regime, e.g., $P = -40$ dBm, yields extremely small error probabilities. For the sake of displaying more interesting error rates, we chose to use less power.

High Noise Regime

In the high noise regime, i.e., $\sigma^2 \gg 2P\gamma$, the maximal ratio combiner minimized the error probability (5.29), resulting in

$$\Pr(Y_t \neq X_t \mid \alpha) \approx Q\left(\sqrt{\frac{\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}A_n\right)^2}{4\sigma^2}}\right). \tag{5.43}$$

Again, we see that the outage probability has the same form as that of equal-gain combining in low noise (5.37). Here, the sum is

$$S(\alpha) = \sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}A_n\right)^2, \tag{5.44}$$

and the constant is $\tau \equiv 4\sigma^2$. Notice this sum is the same sum that we encountered with the outage capacity expressions in the high noise regime for photon-counting receivers (4.99) with the background noise rate set to one, $\lambda_m \equiv 1$. As in Section 4.1.3, we can use exact or approximate moment matching (see Appendices A.5.1 and A.5.2) to approximate this sum as lognormal with exact moments (4.102) and (4.104) or approximate moments (4.106) and (4.108).

Figure 5-4 shows the outage probability in the high noise regime using maximal-

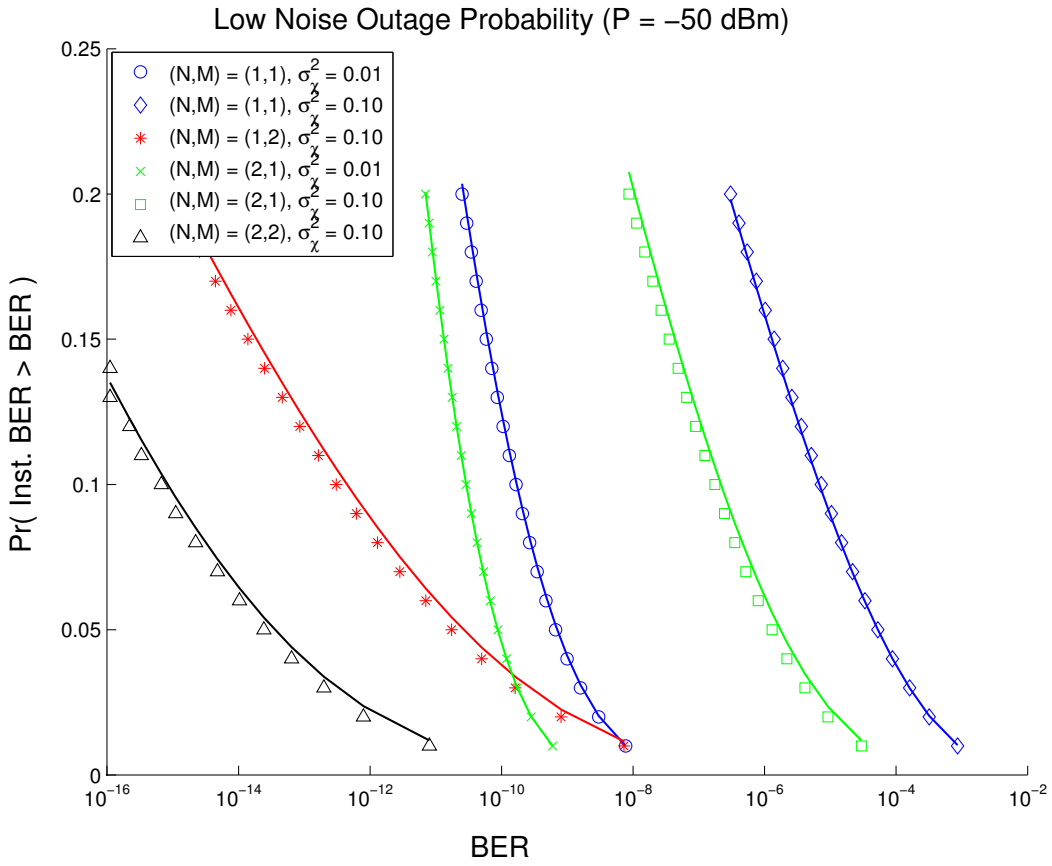

Figure 5-3: The bit error rate versus outage probability in the low noise regime using equal-gain combining and near-optimal thresholding is shown as a function of log-amplitude variance and aperture number. The solid lines depict the approximation (5.37), while the symbols are the empirical complementary cumulative distribution function of 100,000 bit error rate realizations. Note that the two transmitter, single receiver curve is not equal to the one transmitter, two receiver curve because we have defined P to be the average power *per receiver*. Consequently, the two transmitter system transmits half the power of the two receiver system.

ratio combining. At this power level, however, the signal-dependent component of the noise variance is not negligible, and (5.43) is not a good approximation to the instantaneous bit error rate (5.35). For the sake of displaying meaningful error rates, however, we chose to use this higher power.

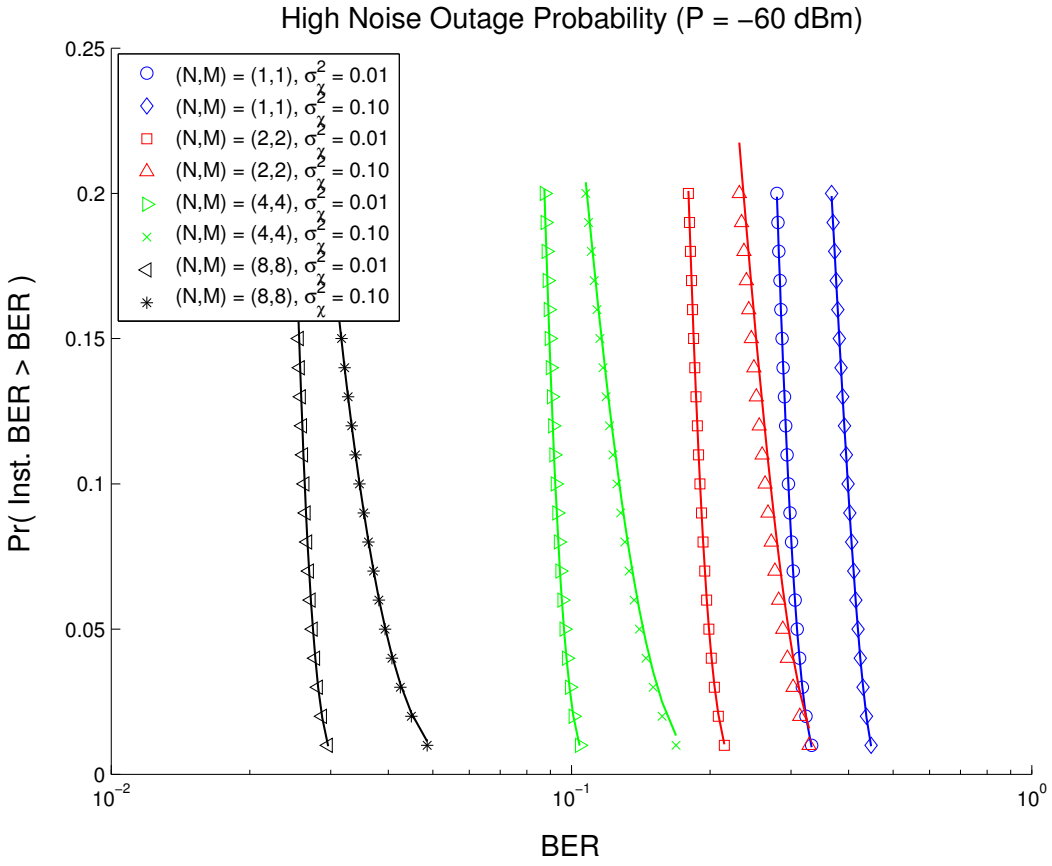

Figure 5-4: The bit error rate versus outage probability in the high noise regime using maximal ratio combining and near-optimal thresholding is shown as a function of log-amplitude variance and aperture number. The solid lines depict the approximation (5.37) using approximate moment matching, while the symbols are the empirical complementary cumulative distribution function of 100,000 bit error rate realizations.

5.2.2 Average Bit Error Rate

In this section we will examine the average bit error rate of equal gain combining, maximal ratio combining, and selection diversity techniques. We will quantify the performance gained through the use of the near-optimal threshold (5.14) versus the

easier to implement, midpoint threshold (5.12). In this section, we will assume that each transmitter uses an identical transmit power, $A_n \equiv A$.

Equal Gain Combining

An equal gain combining receiver simply adds, with equal weighting, the detector outputs. The instantaneous bit error rate for an OOK spatial repetition transmitter and an equal gain combining, near-optimal thresholding (5.14) receiver is (5.35)

$$\Pr(Y_t \neq X_t \mid \alpha) = Q\left(\frac{2(P/N)\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}}{\sqrt{M\sigma^2} + \sqrt{M\sigma^2 + 2\gamma(P/N)\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}}}\right), \quad (5.45)$$

where $P \equiv NA/2$ is the average power per receive aperture, assuming equally likely ones and zeros. We are also assuming that all transmitters use the same peak power $A_n \equiv A$.

A midpoint threshold (5.12) in practice is easier to implement than an optimal threshold, as it corresponds to AC-coupling the combiner output to a simple positive/negative comparator. The instantaneous bit error rate using midpoint thresholding is

$$\begin{aligned}\Pr(Y_t \neq X_t \mid \alpha) &= \frac{1}{2}\,Q\left(\frac{(P/N)\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}}{\sqrt{M\sigma^2}}\right) \\ &\quad + \frac{1}{2}\,Q\left(\frac{(P/N)\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}}{\sqrt{M\sigma^2 + 2\gamma(P/N)\sum_{m=1}^{M}\sum_{n=1}^{N}\alpha_{nm}}}\right). \quad (5.46)\end{aligned}$$

Figures 5-5 through 5-10 show the instantaneous bit error rates averaged over a million channel realizations for various diversity schemes under different fading conditions.

Maximal Ratio Combining

A maximal ratio combining receiver weights each detector output in proportion to the signal power on that branch. The instantaneous bit error rate for an OOK repetition spatial transmitter and maximal ratio combining, near-optimal thresholding receiver

is

$$\Pr(Y_t \neq X_t \mid \alpha) = Q\left(\frac{2(P/N)\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^2}{\sqrt{\sigma^2\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^2}+\sqrt{\sigma^2\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^2+2\gamma(P/N)\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^3}}\right). \tag{5.47}$$

Using midpoint thresholding the instantaneous bit error rate is

$$\begin{aligned}\Pr(Y_t \neq X_t \mid \alpha) &= \frac{1}{2}\,Q\left(\frac{P}{\sigma N}\sqrt{\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^2}\right)\\ &\quad+\frac{1}{2}\,Q\left(\frac{(P/N)\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^2}{\sqrt{\sigma^2\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^2+2\gamma(P/N)\sum_{m=1}^{M}\left(\sum_{n=1}^{N}\alpha_{nm}\right)^3}}\right).\end{aligned} \tag{5.48}$$

Figures 5-5 through 5-10 compare maximal ratio combining against other diversity schemes.

Selection Diversity

Transmitter and Receiver Selection Diversity A simple alternative to equal-gain combining and maximal ratio combining is to use the best transmitter and receiver pair. This architecture assumes that both the transmitter and receiver know the path gains, and can use the best path for communication. For simulation performance, we can use a single uniform random number to generate this path gain. The maximum of the MN independent, identically, distributed random path gains α_{nm} is distributed as

$$F(a) = p \equiv \Pr(\max\{\alpha_{nm}\} \leq a) = \left[\Phi\left(\frac{1}{2\sigma_\chi}\log a + \sigma_\chi\right)\right]^{NM}, \tag{5.49}$$

where $\Phi(x) \equiv \int_{-\infty}^{x} \exp(z^2/2)/\sqrt{2\pi}\; dz$ is the standard normal cumulative distribution function. Evaluating the inverse of this cumulative distribution function,

$$\max\{\alpha_{nm}\} = a = F^{-1}(p) = \exp\left[2\{\sigma_\chi\,\Phi^{-1}(p^{1/MN}) - \sigma_\chi^2\}\right], \tag{5.50}$$

at a uniform random number p on the interval $[0, 1]$ provides a realization of the best path gain. Notice that for $M = N = 1$ this equation produces a lognormal fade. As MN increase, the uniform random variable p is raised to a smaller fractional power, resulting in larger values for the greatest path gain.

The instantaneous bit error rate for selection diversity at both the transmitter and receiver using near-optimal thresholding is

$$\Pr(Y_t \neq X_t \mid \alpha) = Q\left(\frac{2P\max\{\alpha_{nm}\}}{\sqrt{\sigma^2} + \sqrt{\sigma^2 + 2\gamma P\max\{\alpha_{nm}\}}}\right), \tag{5.51}$$

Using midpoint thresholding gives an instantaneous BER of

$$\begin{aligned}\Pr(Y_t \neq X_t \mid \alpha) &= \frac{1}{2}\,Q\left(\frac{P\max\{\alpha_{nm}\}}{\sigma}\right) \\ &\quad + \frac{1}{2}\,Q\left(\frac{P\max\{\alpha_{nm}\}}{\sqrt{\sigma^2 + 2\gamma P\max\{\alpha_{nm}\}}}\right).\end{aligned} \tag{5.52}$$

In the above equations, P is still the average power per receiver. Because only one transmitter is in use at a time, however, the average power per receiver is $P = A/2$.

Receiver Selection Diversity If path gain knowledge at the transmitter is not known, the receiver can choose the detector with the largest signal power. In this case, the instantaneous bit error rate for near-optimal thresholding is

$$\Pr(Y_t \neq X_t \mid \alpha) = Q\left(\frac{2(P/N)\max\left\{\sum_{n=1}^{N}\alpha_{nm}\right\}}{\sqrt{\sigma^2} + \sqrt{\sigma^2 + 2\gamma(P/N)\max\left\{\sum_{n=1}^{N}\alpha_{nm}\right\}}}\right) \tag{5.53}$$

Using midpoint thresholding gives an instantaneous BER of

$$\Pr(Y_t \neq X_t \mid \alpha) = \frac{1}{2}\, Q\left(\frac{(P/N)\max\left\{\sum_{n=1}^{N}\alpha_{nm}\right\}}{\sigma}\right) + \frac{1}{2}\, Q\left(\frac{(P/N)\max\left\{\sum_{n=1}^{N}\alpha_{nm}\right\}}{\sqrt{\sigma^2 + 2\gamma(P/N)\max\left\{\sum_{n=1}^{N}\alpha_{nm}\right\}}}\right). \quad (5.54)$$

Figures 5-5 through 5-10 compare selection diversity against other diversity schemes.

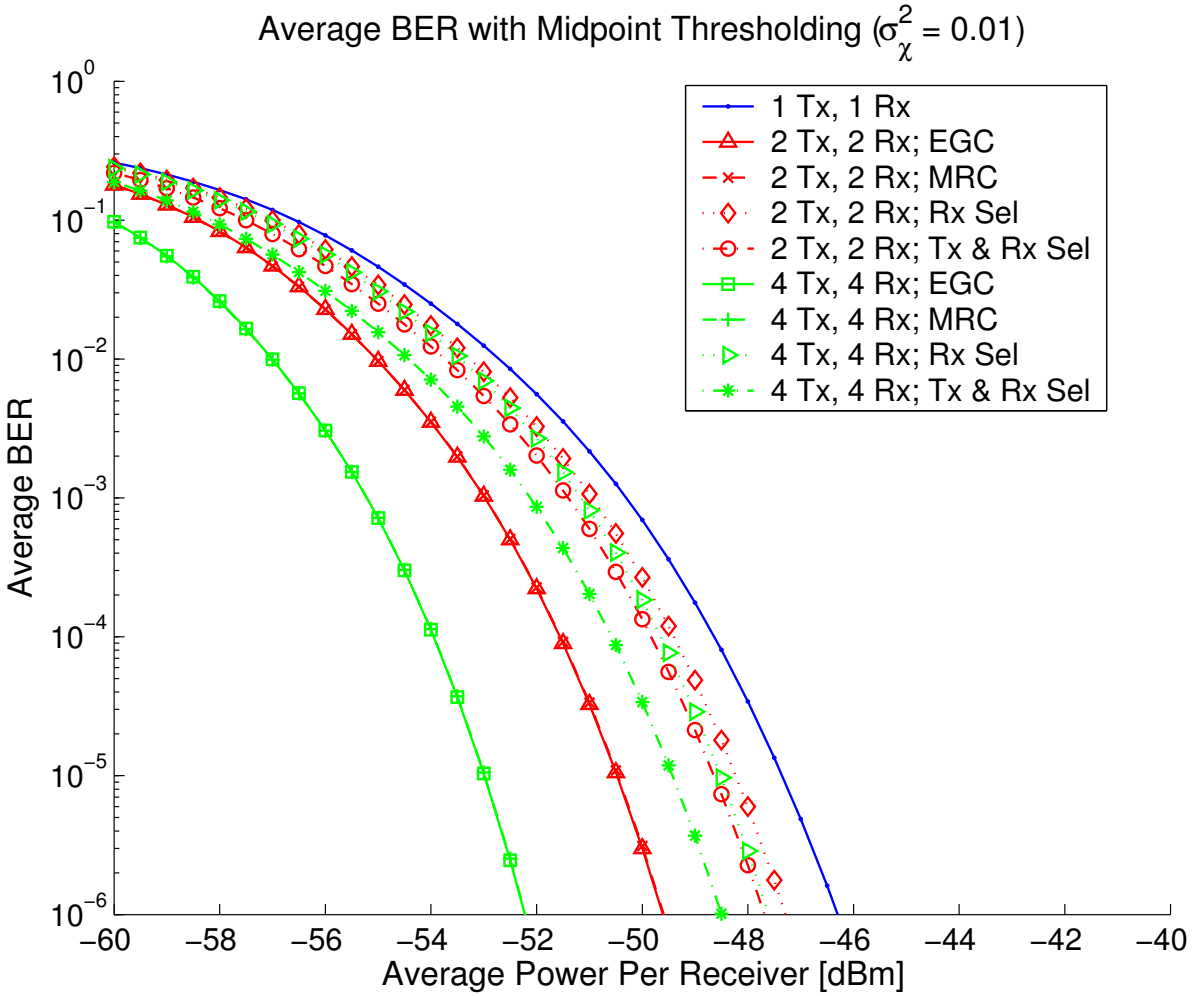

Figure 5-5: The average bit error rate for different transmitter and receiver diversity schemes with midpoint thresholding (5.12) is shown as a function of average power per receiver and number of apertures in mild fading ($\sigma_\chi^2 = 0.01$).

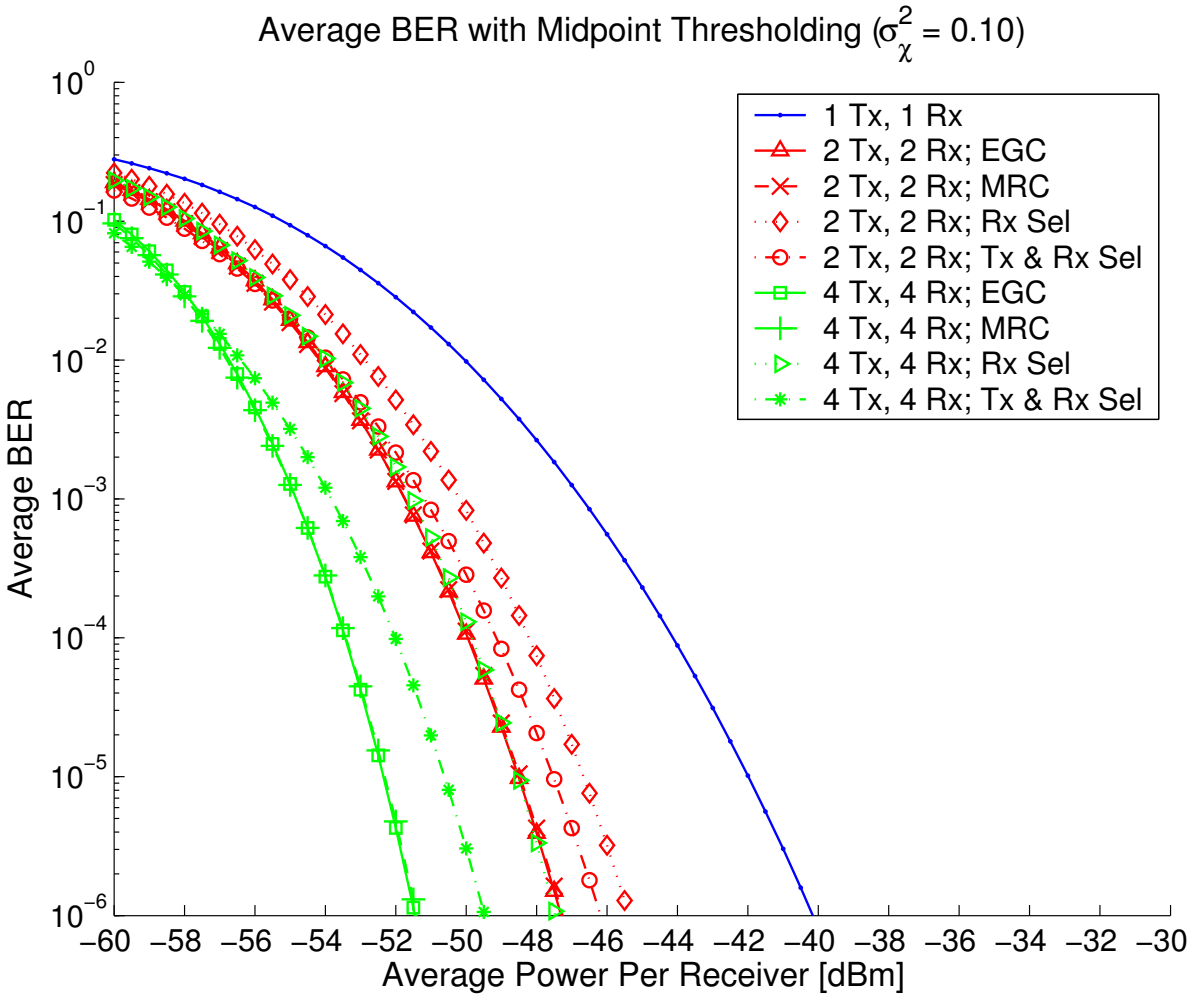

Figure 5-6: The average bit error rate for different transmitter and receiver diversity schemes with midpoint thresholding (5.12) is shown as a function of average power per receiver and number of apertures in moderate fading ($\sigma_\chi^2 = 0.1$).

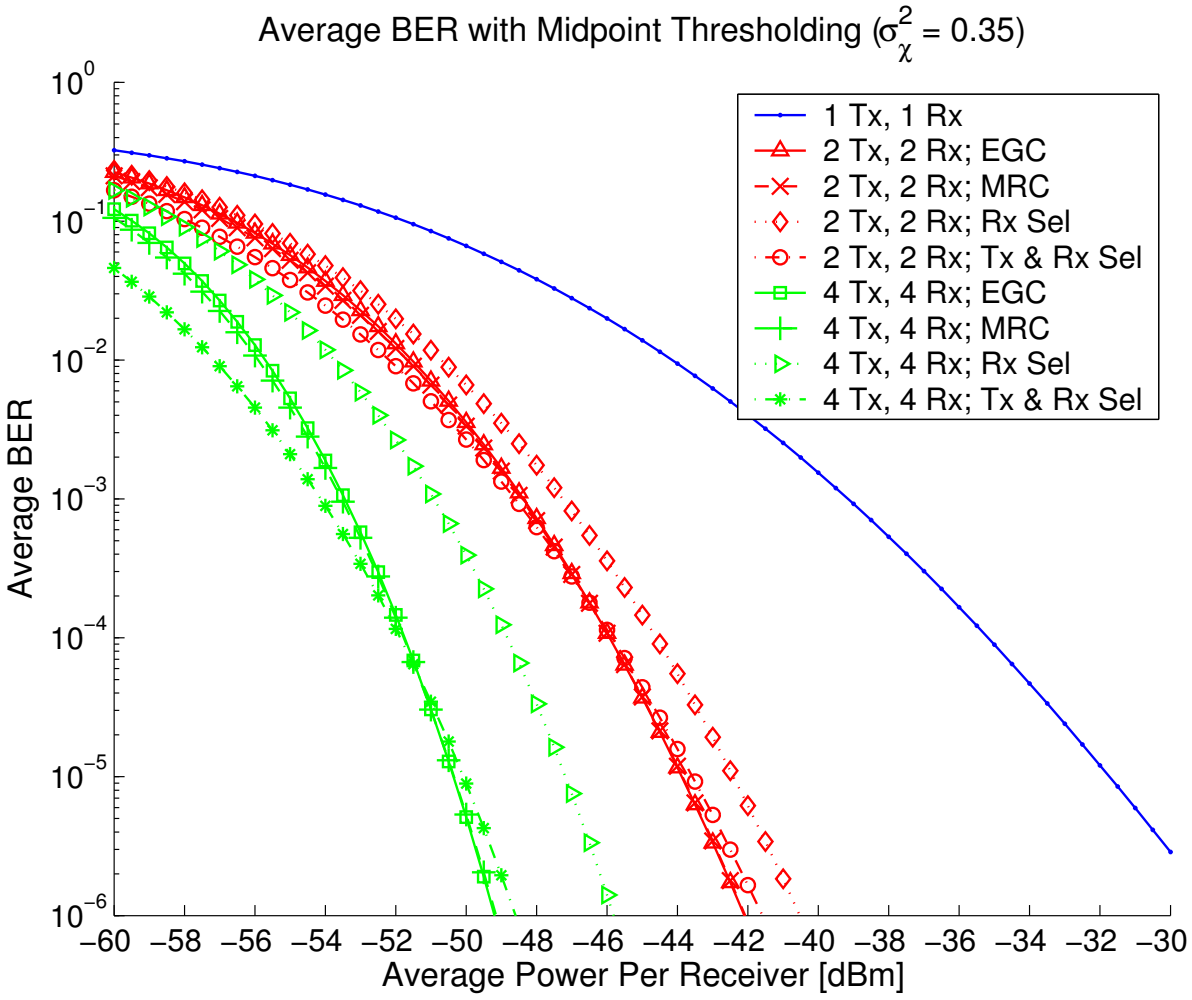

Figure 5-7: The average bit error rate for different transmitter and receiver diversity schemes with midpoint thresholding (5.12) is shown as a function of average power per receiver and number of apertures in strong fading ($\sigma_\chi^2 = 0.35$).

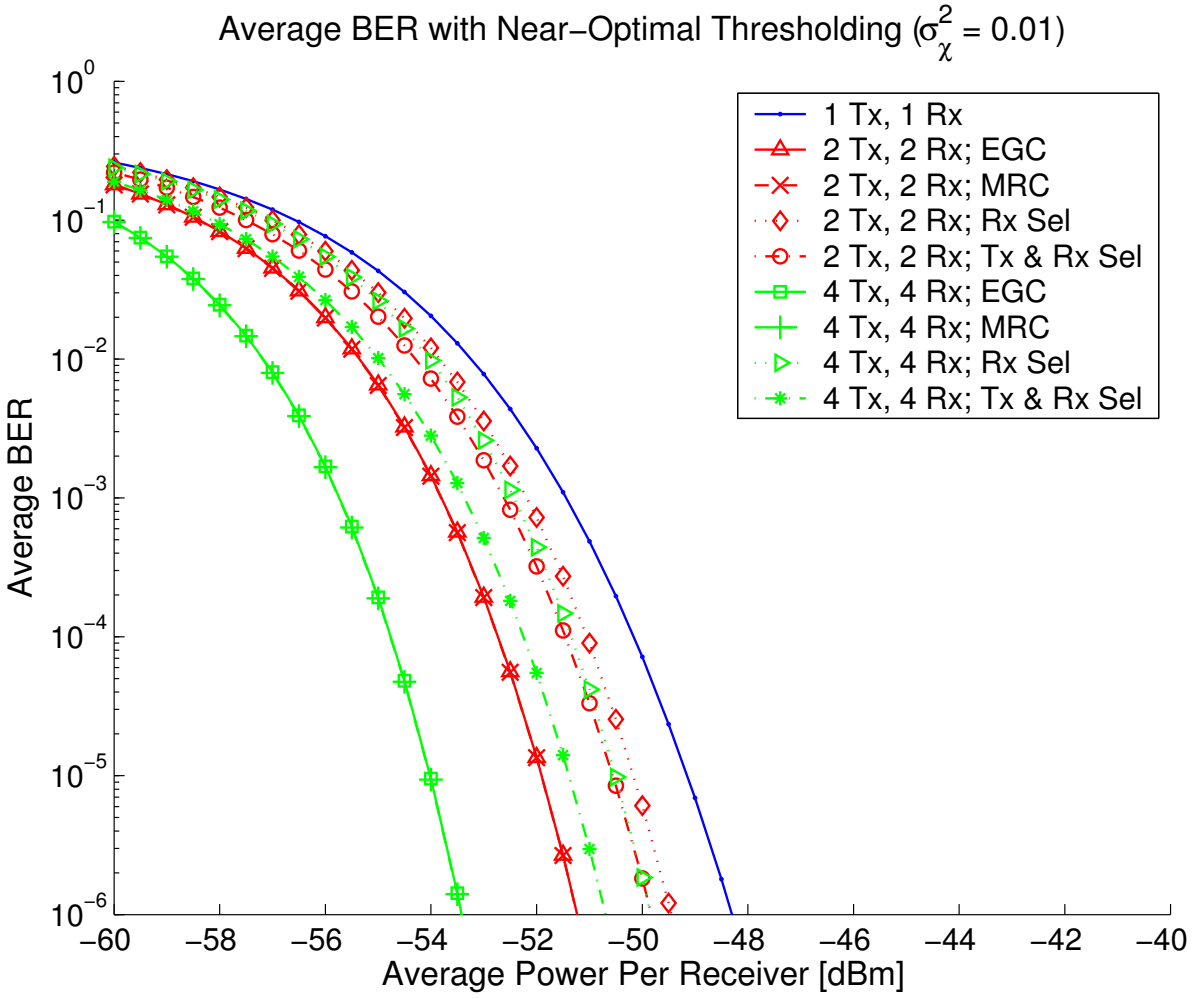

Figure 5-8: The average bit error rate for different transmitter and receiver diversity schemes with near-optimal thresholding (5.14) is shown as a function of average power per receiver and number of apertures in mild fading ($\sigma_\chi^2 = 0.01$).

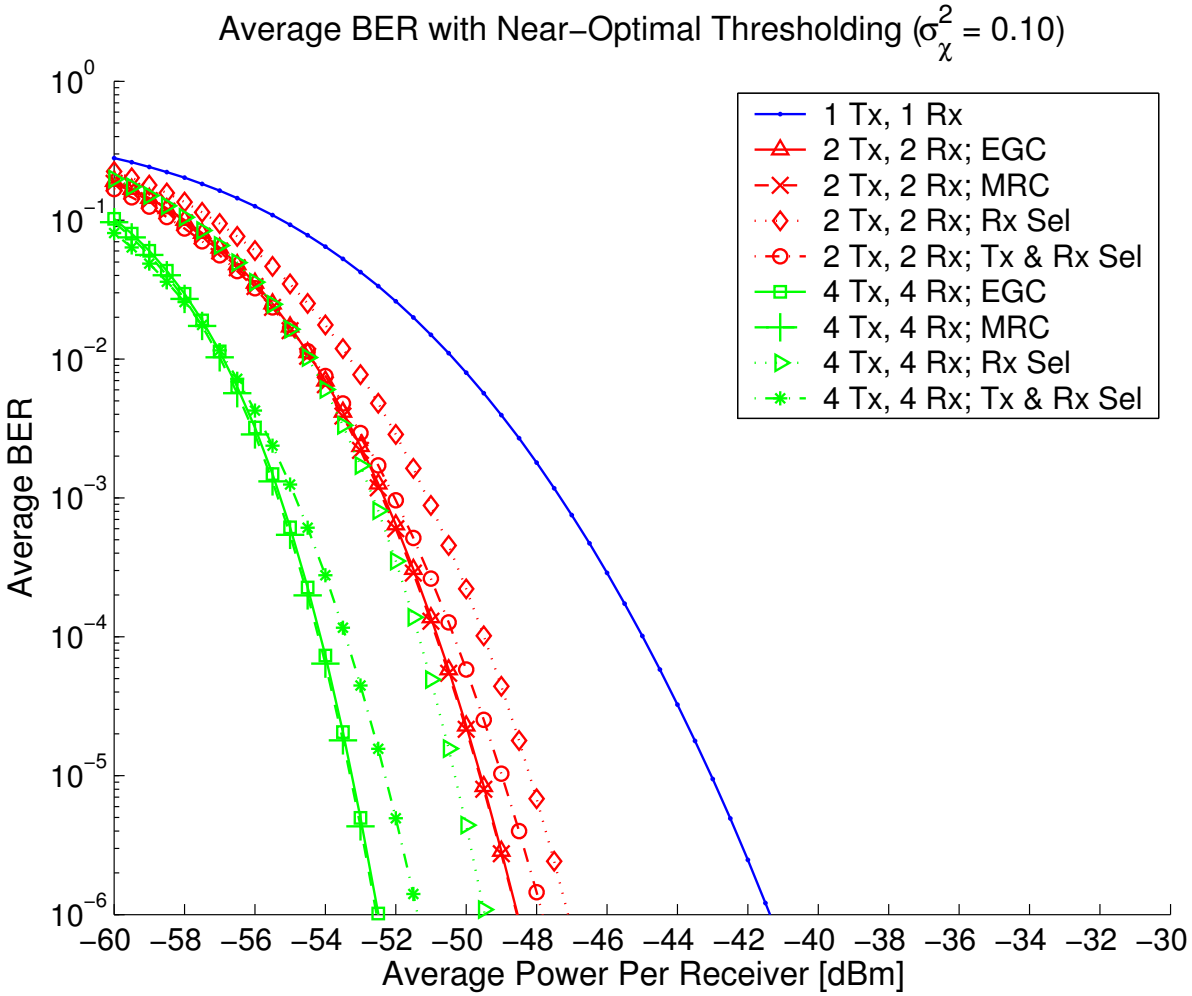

Figure 5-9: The average bit error rate for different transmitter and receiver diversity schemes with near-optimal thresholding (5.14) is shown as a function of average power per receiver and number of apertures in moderate fading ($\sigma_\chi^2 = 0.1$).

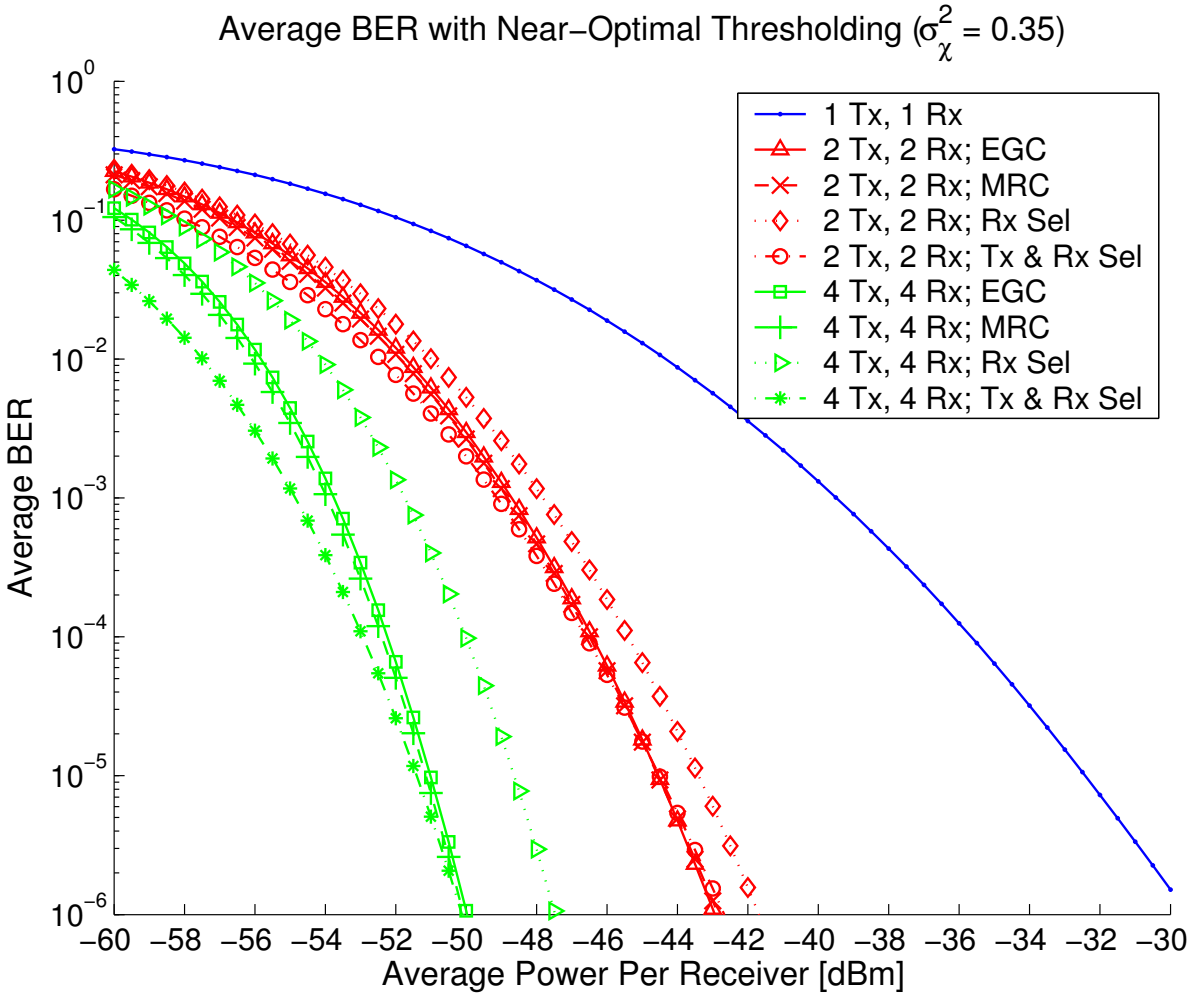

Figure 5-10: The average bit error rate for different transmitter and receiver diversity schemes with near-optimal thresholding (5.14) is shown as a function of average power per receiver and number of apertures in strong fading ($\sigma_\chi^2 = 0.35$).

An Approximate Method to Calculate EGC BER

We have seen that for OOK spatial repetition transmitters and linear combining, threshold-decision receivers, if the signal-dependent component of the variance dominates the fixed component, then equal-gain combining using optimal thresholds, minimizes the error probability. We now examine some simple Monte Carlo averaging techniques to calculate the average error probability for equal-gain combining receivers using near-optimal thresholds (5.14) and midpoint thresholds (5.12) for the nominal parameters in Table 5.1. Our error rate approximations can be implemented using spreadsheets such as Excel.

For equal-gain combining receivers, the error probability, (5.46) and (5.45), is a function of the sum of MN lognormal random fades, $R = \sum_{m=1}^{M} \sum_{n=1}^{N} \alpha_{nm} A$. We will approximate this sum as being lognormal. Instead of generating a channel of NM lognormal random variables, we only generate one lognormal random variable with appropriately chosen moments. This lognormal random variable can be generated from a single uniform random variable on the interval $[0, 1]$, using the transformation:

$$L(p) = \exp\left[\nu\, \Phi^{-1}(p) + \mu\right], \tag{5.55}$$

where the log-moments μ and ν^2 are given by (5.41) and (5.42), respectively.

A comparison of this approximate Monte Carlo averaging method (solid lines) and the exact bit error rate (symbols) averaged over one million channel realizations is shown in Figures 5-11 through 5-13 under various fading conditions. From these figures we see that this approximation is quite accurate for mild and moderate fading. For severe fading, this approximate method tends to give conservative estimates of the average bit error rate.

Increase the Power or the Number of Apertures?

We have seen in Figures 5-5 through 5-10 that atmospheric turbulence greatly impacts the the BER performance of a single-transmit, single-receive aperture system. For example, from Figures 5-8 and 5-10 we see that we lose the equivalent of about 17

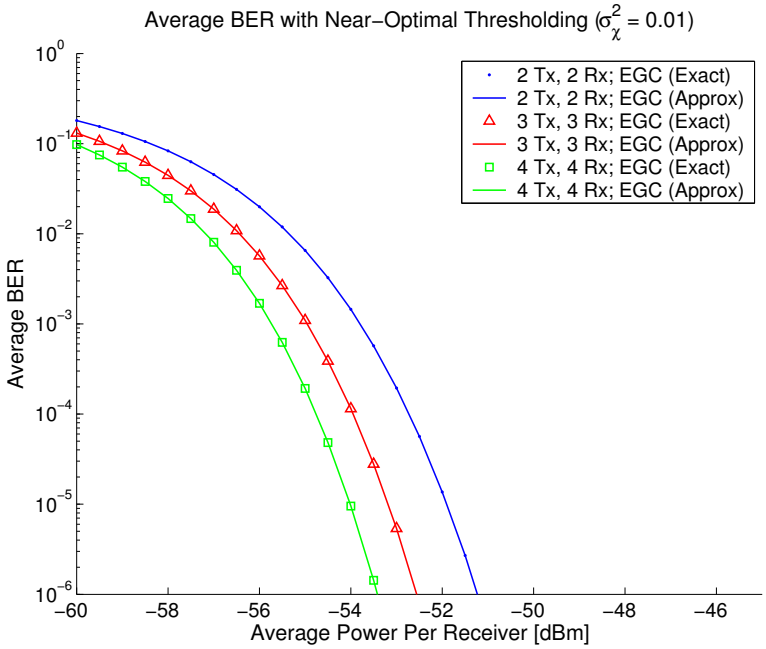

Figure 5-11: A comparison of the exact and approximate average bit error rate for equal gain combining receivers with near-optimal thresholding (5.14) is shown as a function of average power per receiver in mild fading ($\sigma_\chi^2 = 0.01$).

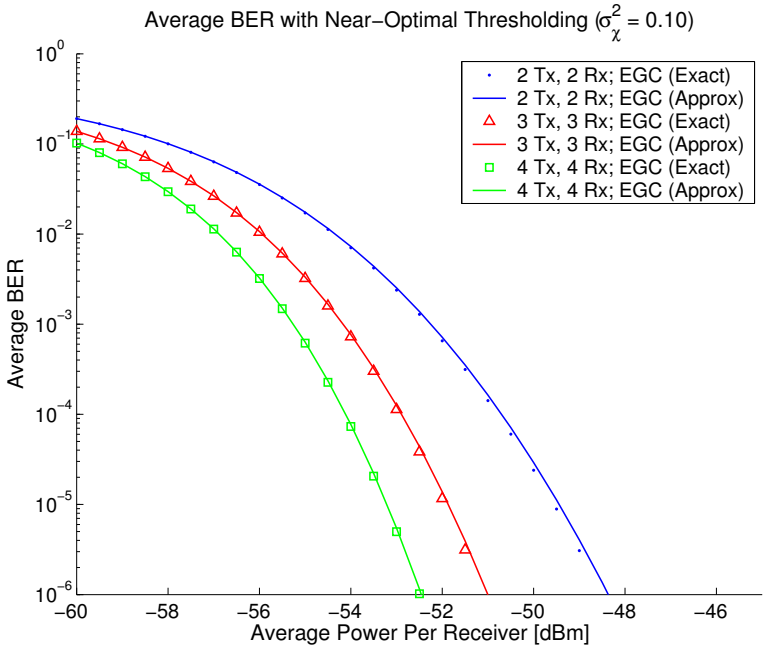

Figure 5-12: A comparison of the exact and approximate average bit error rate for equal gain combining receivers with near-optimal thresholding (5.14) is shown as a function of average power per receiver in moderate fading ($\sigma_\chi^2 = 0.1$).

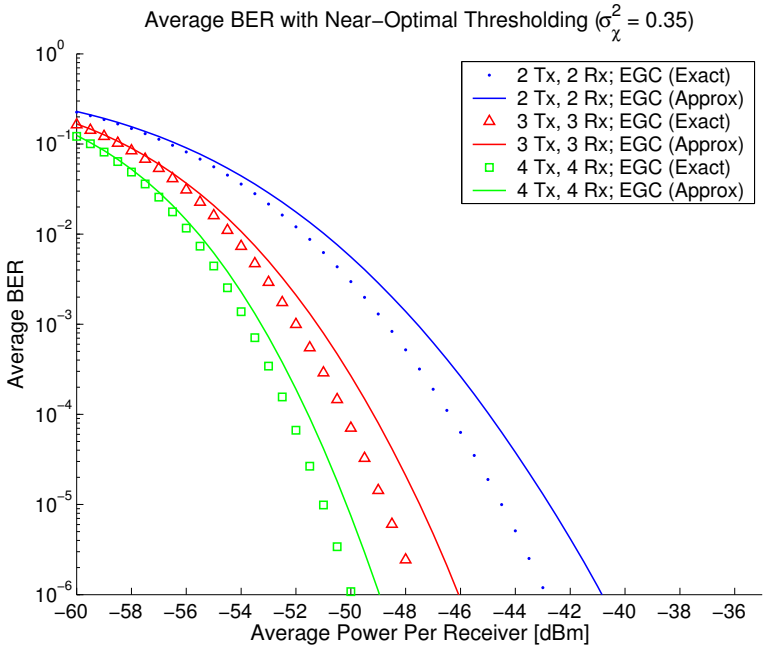

Figure 5-13: A comparison of the exact and approximate average bit error rate for equal gain combining receivers with near-optimal thresholding (5.14) is shown as a function of average power per receiver in strong fading ($\sigma_\chi^2 = 0.35$). In strong fading, our lognormal approximation provides a conservative estimate of error probability.

dB in power at 10^{-5} error rates as fading worsens from mild to severe. To improve the performance we could

- Increase the transmit power,
- Increase the number of transmit apertures,
- Increase the number of receive apertures,

or any combination, thereof. In this section, we address the advantages and disadvantages of each choice.

To make a fair comparison between these choices, we will consider the average total receive power, PM, required to achieve a particular average bit error rate. In other words, for combining receivers, if we double the number of apertures, we must halve the transmit power. For selection diversity at both the transmitter and receiver, the average total receiver power is just P.

Figures 5-14 and 5-15 show the average bit error rate for different diversity schemes involving one and two apertures. The average bit error rate is plotted against the average power per receiver, P. We will examine the power reduction in total average receive power, PM, gained through diversity at 10^{-5} average bit error rates in severe fading ($\sigma_\chi^2 = 0.35$). For example, from Figure 5-14 using one transmit and two receive apertures with maximal ratio combining and midpoint thresholding requires -40 dBm average power per receiver, or -37 dBm average total receive power, to achieve 10^{-5} average bit error rates. Figures 5-7 and 5-10 show the average bit error rates for two transmit and two receive apertures.

Table 5.4 shows the average total receive power required for 10^{-5} average bit error rates for the different diversity schemes. Notice that transmitter and receiver selection diversity perform the same for one transmit, two receive apertures and two transmit, one receive apertures, and we only list receiver selection diversity.

The power saved from using a near-optimal threshold over a midpoint threshold is at most a decibel in severe fading. Note that in mild and moderate fading, the power savings are greater, approximately one to two decibels; for example, compare Figures 5-5 and 5-8, and Figures 5-6 and 5-9. The relatively small threshold gain indicates that in severe fading, the average bit error rate is dominated by deep channel fades. During these deep fades, the fixed component of the receiver noise dominates, and midpoint thresholds are optimal. Furthermore, maximal-ratio combining tends to be slightly better (about a tenth of a decibel) than equal-gain combining because it also performs better when the fixed component dominates.

We now examine the incremental improvement and trade-offs of adding either a receive or transmit aperture over the single aperture system. As seen in Table 5.4, selection diversity saves approximately 7 dB in power compared to a single transmit, single receive aperture system. Selection diversity, however, requires path gain knowledge at either the transmitter or receiver. A momentary application of power on each path is needed to determine the path gains. This power required for channel estimation is not included in our comparison, however.

Transmitter diversity saves approximately 6 dB in power, and does not require path gain knowledge. Similarly, both equal gain and maximal ratio combining save about 5 dB, with the former not requiring path gain knowledge or complicated non-uniform combining. Transmitter diversity has a further advantage in that, presently, lasers are much less expensive than optical amplifiers.

Transmitter diversity, however, requires either angle or frequency separation to get the incoherent addition of optical fields. Angle separation requires coupling multiple spatial modes into the receiver, which is not compatible with single-mode optical amplifiers.

Although, frequency separation will create the desired addition, it also faces a similar problem. To obtain uncorrelated fades, transmitters must be sufficiently separated in distance. If this distance creates too large of an angle between transmitters at the receiver, then coupling into a single-mode fiber might be difficult.

In terms of complexity and performance, an on-off keying, repetition spatial transmitter and/or an equal-gain combining receiver with midpoint thresholding is a very cost-effective method for combatting atmospheric turbulence. In severe fading, this combination can save about five to six decibels in total receive power in a dual versus single aperture system. We could further increase these savings by more transmit or receive apertures without having to increase the power; e.g., two transmit and two receive apertures save around nine decibels.

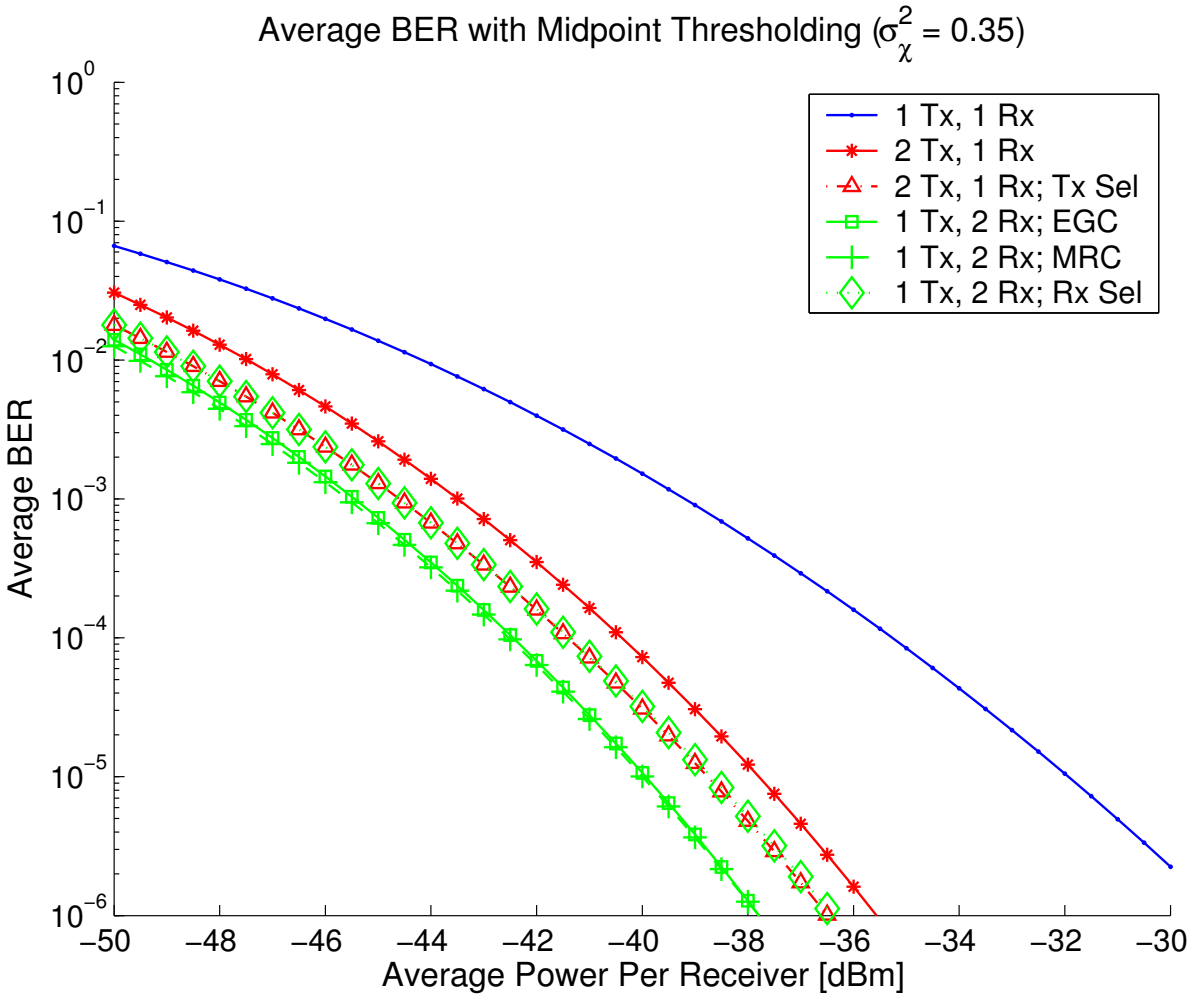

Figure 5-14: The average (over one million channel realizations) bit error rate for different diversity techniques with midpoint thresholding (5.12) is shown as a function of average power per receiver in strong fading ($\sigma_\chi^2 = 0.35$).

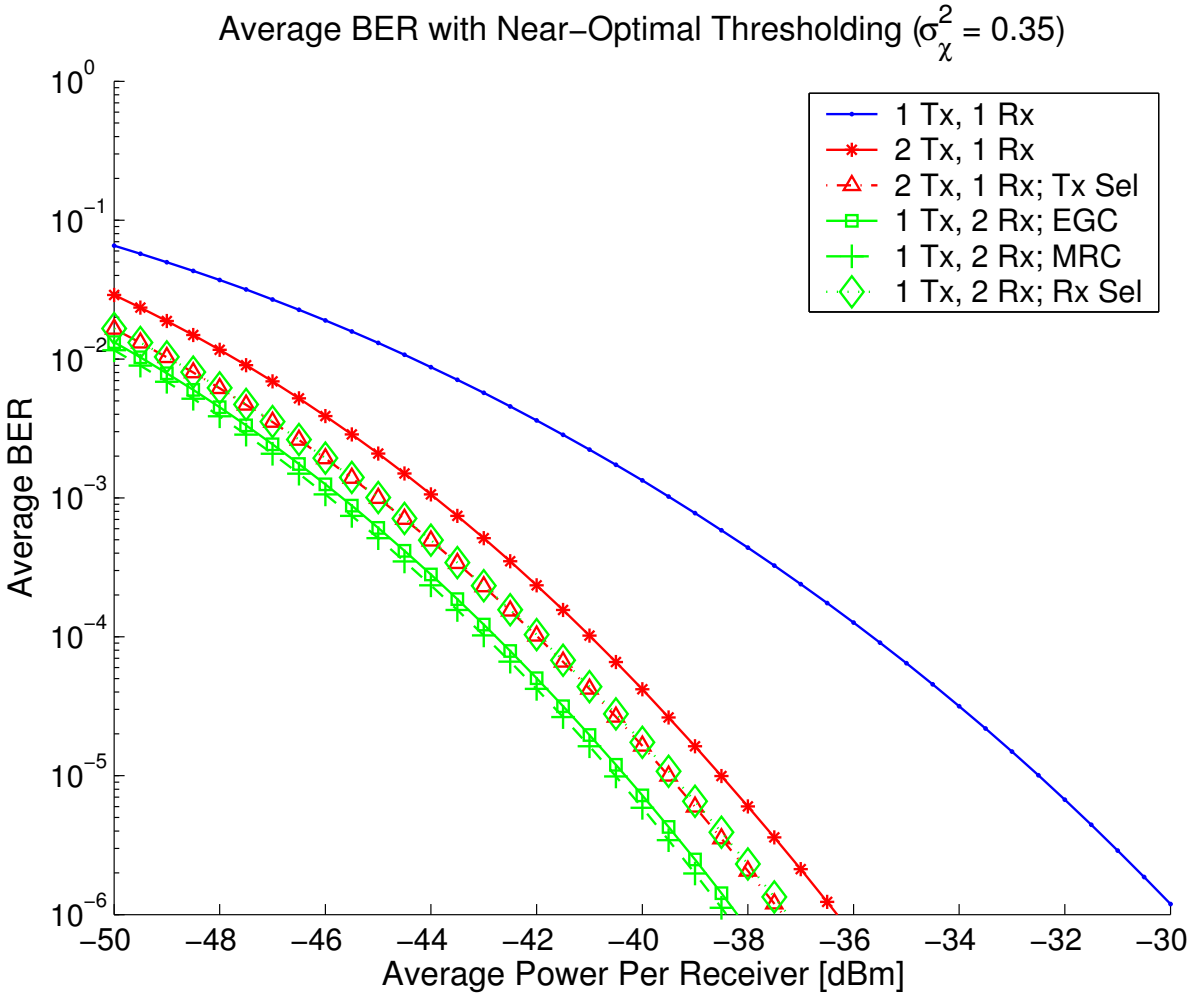

Figure 5-15: The average (over one million channel realizations) bit error rate for different diversity techniques with near-optimal thresholding (5.14) is shown as a function of average power per receiver in strong fading ($\sigma_\chi^2 = 0.35$).

	Midpoint [dBm]	Near-Optimal [dBm]	Thres. Gain [dB]
1 Tx, 1 Rx	-31.9	-32.5	0.6
1 Tx, 2 Rx (EGC)	-36.9	-37.3	0.4
1 Tx, 2 Rx (MRC)	-37.0	-37.5	0.5
2 Tx, 1 Rx	-37.8	-38.5	0.7
1 Tx, 2 Rx (Sel)	-38.7	-39.4	0.7
2 Tx, 2 Rx (EGC)	-40.9	-41.5	0.6
2 Tx, 2 Rx (MRC)	-40.9	-41.6	0.7
2 Tx, 2 Rx (Rx Sel)	-42.4	-43.4	1.0
2 Tx, 2 Rx (Tx & Rx Sel)	-43.6	-44.5	0.9

Table 5.4: The average total receive power (MP for combining schemes, and P for selection diversity) in dBm required for 10^{-5} average bit error rates in severe fading ($\sigma_\chi^2 = 0.35$) is shown for different diversity and thresholding schemes. The accuracy of the power is approximately ± 0.1 dBm. The power gained in decibels from using a near-optimal versus a midpoint threshold is shown in the right most column.

Chapter 6

Experimental Results

In this chapter, we report an experimental study of the performance of a two receive aperture, optically-preamplified receiver. We compare the bit error rate performance with that predicted by the model in Chapter 5. Figure 6-1 shows our experimental configuration for a single transmit and receive aperture system.

A bit error rate tester (BERT) generates a pseudo-random bit sequence that modulates the amplitude of a continuous-wave (CW) laser carrier. The presence of light after the modulator indicates a '1' bit, while its absence, a '0' bit. The optical signal exits the transmitter via a telescope, travels through the atmosphere for 125 meters, reflects off a mirror, travels back another 125 meters, enters the receiver via another telescope, and then couples into a single-mode fiber. An erbium-doped fiber amplifier (EDFA) optically amplifies the received signal, adding amplified spontaneous emission (ASE) noise in the process. This noise is spectrally flat over the optical filter passband.

The resulting filtered signal and noise pass through a photodetector and transimpedance amplifier. A limiting amplifier amplifies the resulting voltage signal, clipping its amplitude to a test-equipment-compatible level. The BERT makes a hard decision as to whether the received signal is a '1' or a '0', and compares the transmitted sequence to this received bit sequence. The BERT then computes the ratio of bits received in error to the total bits transmitted. A digital communication analyzer (DCA) also samples the received waveform, and plots the eye diagram. Table 6.1

summarizes the parameters of the experiment configuration.

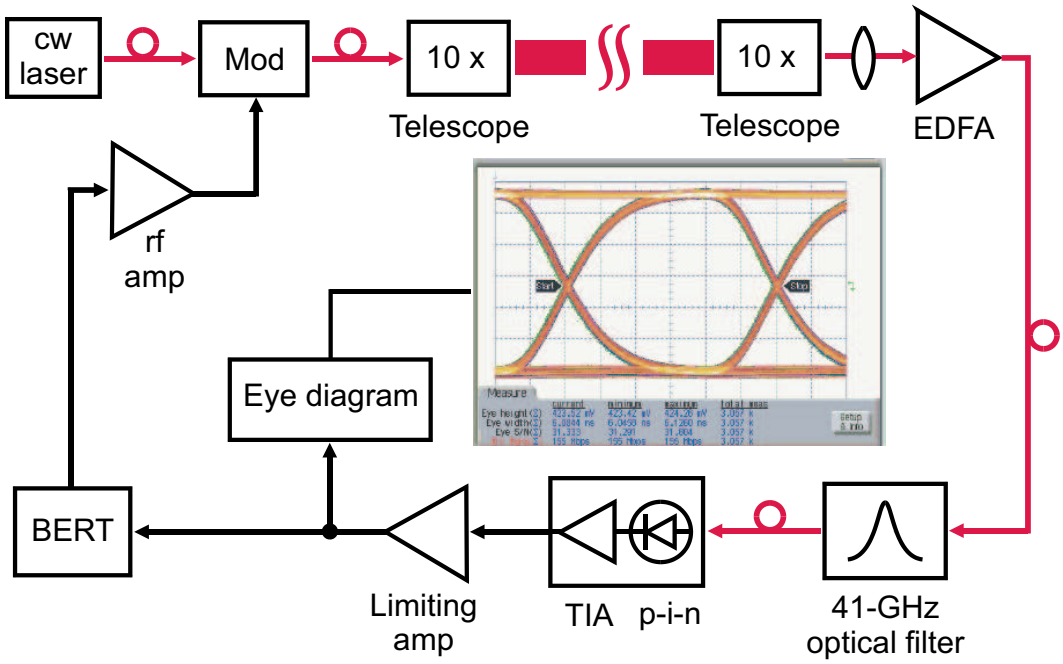

Figure 6-1: The experimental configuration for a single transmit and receive aperture system consists of an externally modulated laser transmitter and optically-preamplified, direct-detection receiver. Test equipment such as the BERT and DCA analyze the communication system performance, such as the eye diagram shown here.

Surprisingly, despite all the approximations we used to arrive at our optically-preamplified, direct-detection channel in (5.1), the model captures the experimental data behavior well in mild and moderate fading. Figures 6-2 and 6-3 shows the bit error rate (BER) as a function of average optical power per receiver branch for 1.25 Gbps (OC-24, or gigabit ethernet) data rates through an optical fiber using midpoint thresholding and equal gain combining. The theoretical curves assume ideal filters and no intersymbol interference, and seem to differ from the measured curve by at most 0.5 dB. The model uses the parameters in Table 6.1. Figure 6-2 shows the bit error rate through optical fiber of a single-transmit, single-receive aperture and single-transmit, dual-receive aperture with equal-gain combining and midpoint thresholding. Figure 6-3 shows the performance of the same two systems in mild fading ($\sigma_\chi^2 \approx 0.02$). These measurements were taken during a clear evening. From these figures we see that this mild fading increases the average power required to maintain 10^{-7} error

Parameter	Symbol	Value	Units
Opt. BW	B_{O}	41	GHz
Elect. BW	B_{E}	750	MHz
ASE Factor	n_{sp}	1.09	
Opt. Gain	G_{O}	40	dB
Opt. Wavelength	λ	1550	nm
Num. Pol. Modes	D_{pol}	2	
Det. Quant. Eff.	η	1	
Dark Power	P_{D}	n/a	W
TIA PSD	N_{TIA}	n/a	A^2/Hz
Backgrd. Power	N_{B}	n/a	W/Hz
Link Distance		250	m
Transmit Power		≈ 2	mW
Receiver Separation		17.8	cm
Receiver Pupil Diameter		2.5	cm
Fixed Noise Comp.	σ^2	2.38×10^{-18}	W^2
Sig. Dep. Noise Factor	γ	4.19×10^{-10}	W

Table 6.1: This table summarizes the parameters of the OC-24 experimental testbed. The dark power, transimpedance amplifier thermal noise, and background power were negligible compared to the amplified spontaneous emission noise; hence, these entries are marked not appreciable (n/a).

rates by about a decibel.

Figures 6-4 and 6-5 show the empirical log-amplitude distribution for the two receivers. The log-amplitude is defined as

$$\text{Log-Amplitude} \equiv \frac{1}{2}\log(\text{Received Faded Signal Power in mW}). \tag{6.1}$$

Writing the received signal power as $e^{2\chi}P$, where P is the average power per receiver, the log-amplitude is

$$\text{Log-Amplitude} = \frac{1}{2}\log P + \chi. \tag{6.2}$$

The received signal power, as measured through a detector monitor, was sampled every 0.5 ms for 20,000 samples. The log-amplitude variance (σ_χ^2) for the first and second receiver was 0.012 and 0.0098, respectively, and the correlation between channels was 0.11. These measurements were again taken during a clear evening. From these figures, we see that the lognormal model is a decent description of the fading around the density's mode, but the measured density tends to be skewed in the tails. Although not shown, we have observed that this skewness is very sensitive to the alignment of the single-mode fiber coupler. Regardless, Figure 6-3 suggests that a lognormal model is appropriate for determining the average bit error rate performance in mild fading.

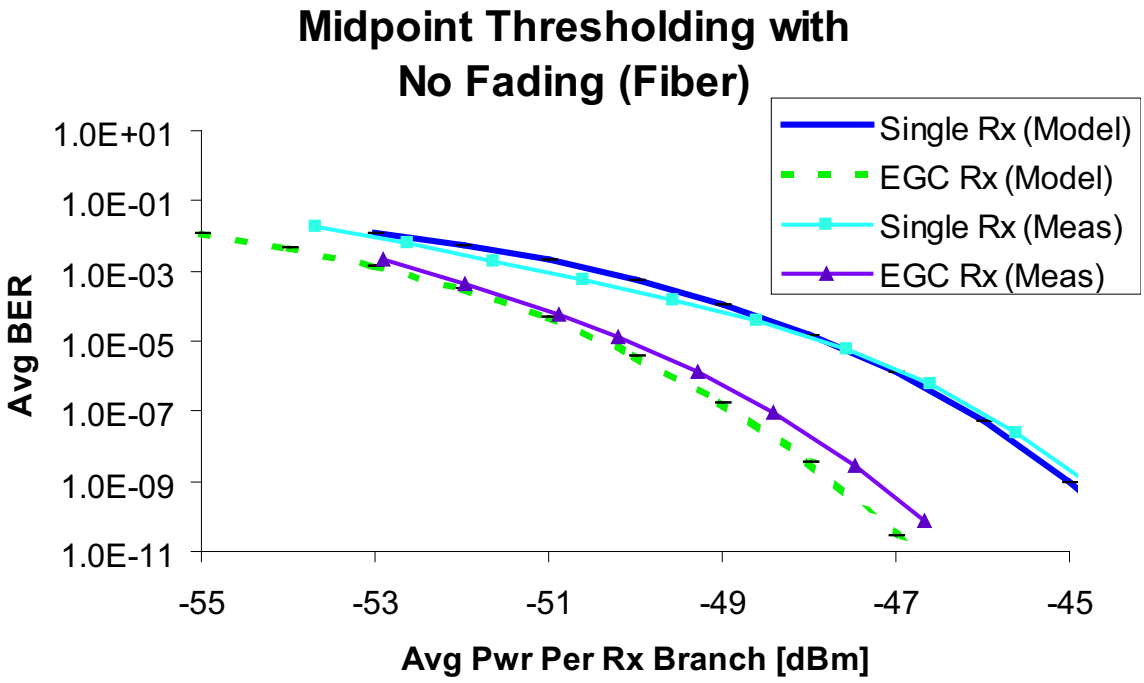

Figure 6-2: This figure plots the measured and theoretical BERs versus the average received optical power for 1.25 Gbps data rates using midpoint thresholding (5.12) in the absence of fading (fiber transmission) for single and dual receiver configurations.

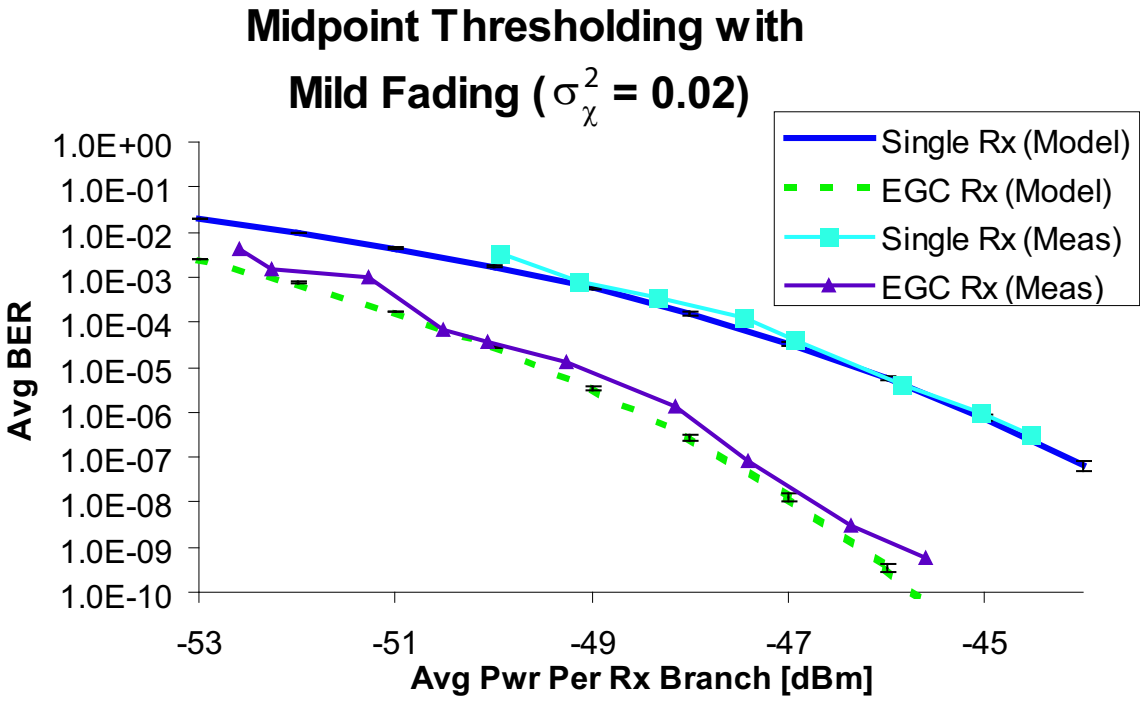

Figure 6-3: This figure plots the measured and theoretical BERs versus the average received optical power for 1.25 Gbps data rates using midpoint thresholding (5.12) in mild fading ($\sigma_\chi^2 \approx 0.02$) for single and dual receiver configurations.

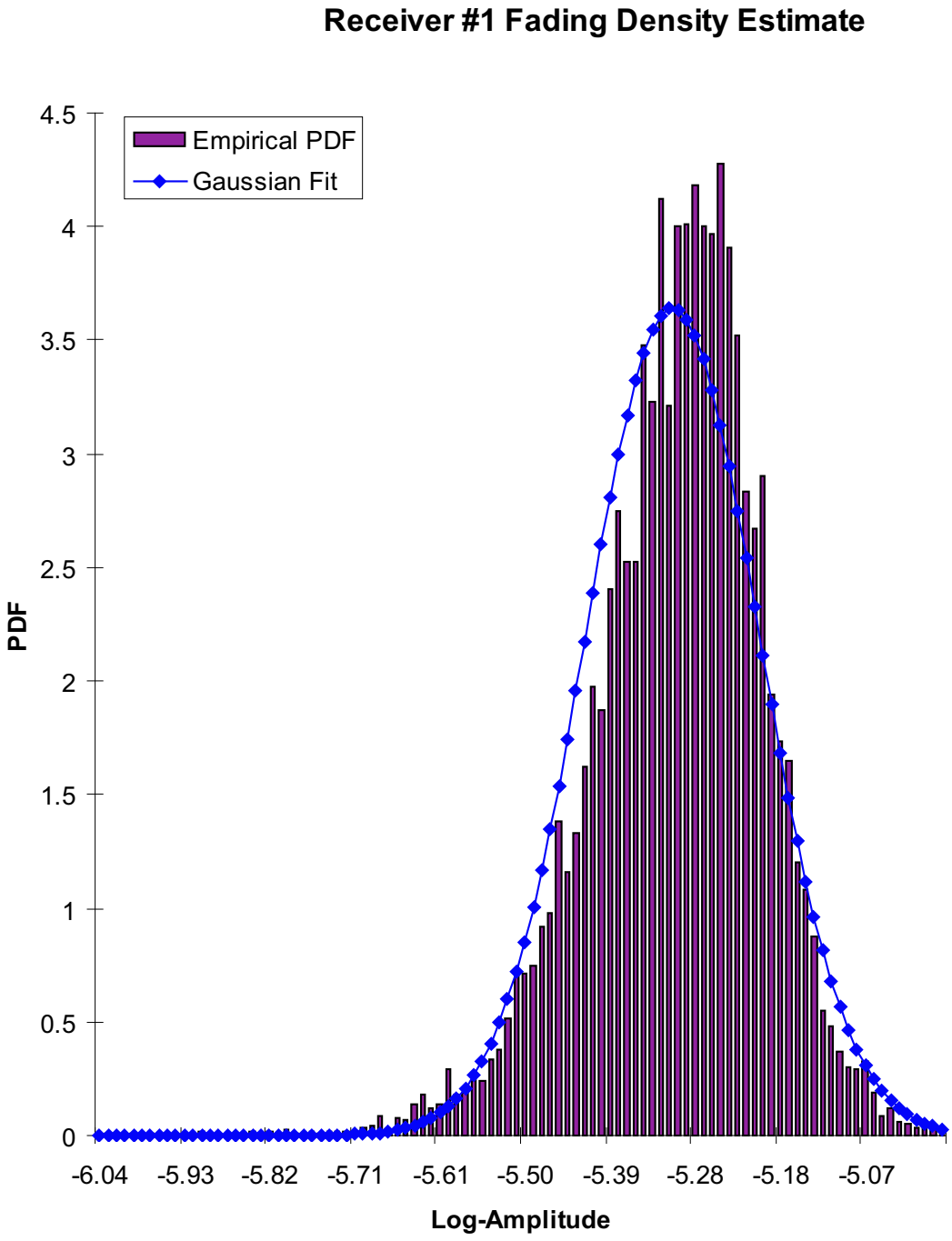

Figure 6-4: The empirical probability density function of the log-amplitude, i.e., $0.5 \log($ Signal Power in mW $)$, for the first receiver is shown with its Gaussian fit.

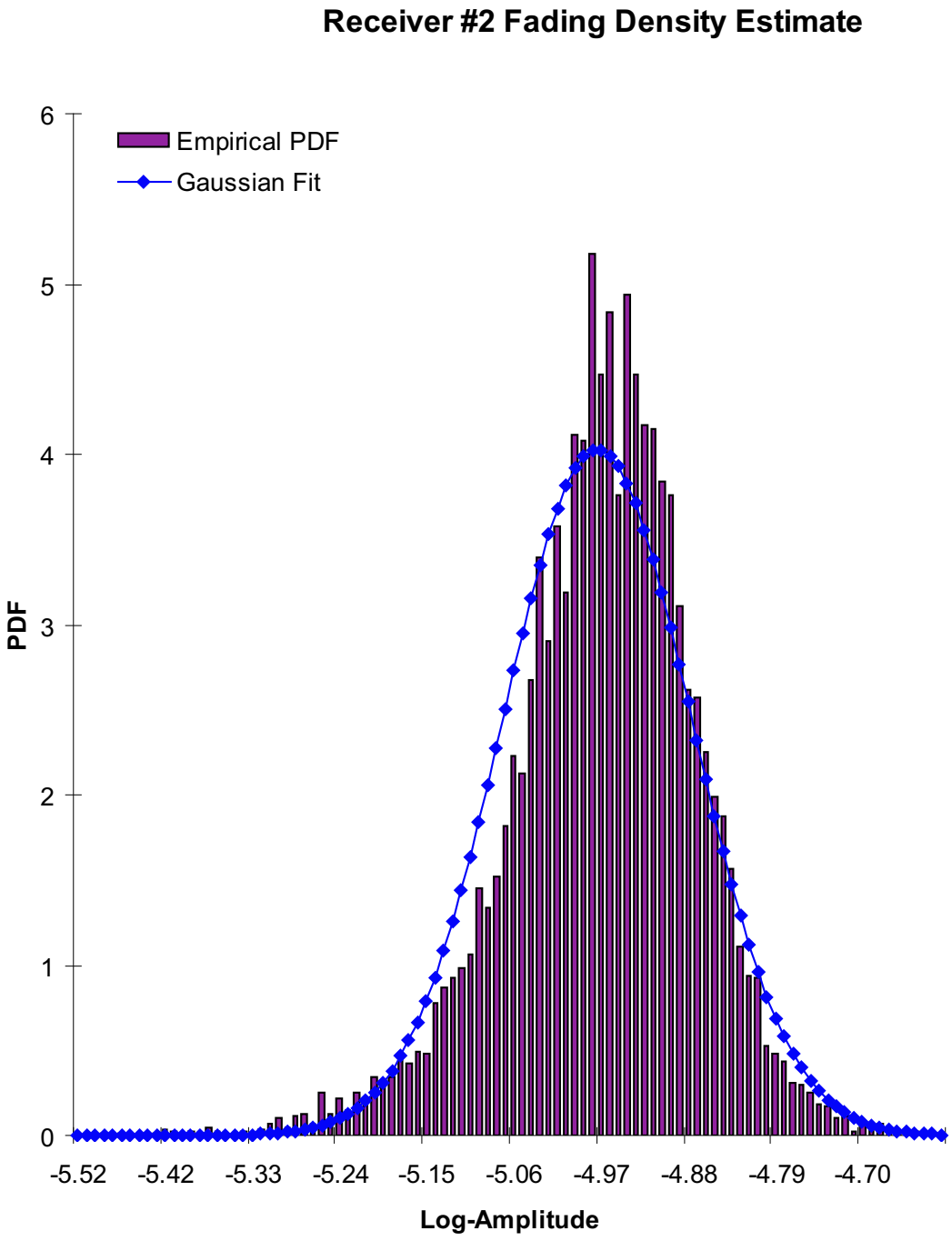

Figure 6-5: The empirical probability density function of the log-amplitude, i.e., $0.5 \log($ Signal Power in mW $)$, for the second receiver is shown with its Gaussian fit.

Chapter 7

Conclusions

Atmospheric turbulence causes random power fluctuations in wireless optical communication systems. Because the coherence time of this fading channel is on the order of milliseconds, a single fade can obliterate millions of bits at gigabit per second data rates. This thesis examined methods to improve the reliability of communication through the turbulent atmosphere.

In Chapter 3, we presented a framework for developing space-time codes for an atmospheric optical heterodyne communication system. We introduced a normalized fading strength, which acts as a "figure of merit" for space-time codes, based on the variance-to-mean-squared ratio of codeword energy difference. Through a central limit theorem approximation, we related this figure of merit to the pairwise error probability. Using this approximation, we found that a diagonal design matrix minimized the pairwise probability of codeword error. Although developed for lognormal fading, this method generalized to other fading distributions in which the fades are zero-mean and independent.

Our design criterion also satisfied the rank and determinant criteria presented in [61] for Rayleigh channels. Furthermore, orthogonal designs provided a method of constructing space-time codes that satisfied our criterion, and required only linear processing at the receiver [60]. We demonstrated the improvement gained in code error rate for a popular orthogonal design, the Alamouti scheme.

In Chapter 4, we introduced the MIMO Poisson channel with peak and average

transmit power constraints for a given set of path gains. We derived upper and lower bounds on the channel capacity that were equal in a number of special cases, such as for channels with low or high background noise, a single receiver (the MISO channel), or low average input power constraints. The lower bound was equal to the MIMO capacity for the single transmitter case (the SIMO channel). When the bounds coincided, the capacity achieving distribution was on-off keying with all transmitters turning on and off in unison.

We developed formulas for the ergodic and outage capacities of photon-counting receivers suffering from turbulence-induced fading. Fading did not reduce capacity at high signal-to-background ratio, and actually increased capacity at low signal-to-background ratio. This improvement was at most 7.4 for a single transmit aperture system in severe fading, and diminished with increasing numbers of transmit apertures.

We showed that path gain knowledge at the transmitter did not increase ergodic capacity at high or low signal-to-background ratios. Path gain knowledge at the receiver increased the ergodic capacity in the high-noise regime by at most a factor of 7.4, attained by using a single transmit aperture and an infinite number of receive apertures. However, for a more realistic case of two transmit and two receive apertures in mild to moderate fading, the use of an optimal receiver provided only one to ten percent capacity improvement over that attained with a simple photon-bucket receiver. In other words, photon-bucket receivers are capacity achieving at high signal-to-noise ratio and nearly optimal at low signal-to-noise ratio for moderate numbers of transmit apertures.

When delay constraints prevent coding over many channel realizations, the distribution of the instantaneous capacity is a more appropriate measure of reliable communication rates. We developed approximations to the capacity-versus-outage probability in the high and low noise regimes for photon-counting receivers. We confirmed that these approximations are quite good for moderate numbers of apertures in mild and moderate fading using Monte Carlo simulations.

In Chapter 5 we examined optically-preamplified receivers, and developed discrete

memoryless channel representations using linear combining receivers and OOK spatial repetition transmitters. Optical preamplification can increase receiver sensitivity one-hundred fold at 10^{-9} error probabilities for 1.25 Gbps data rates. We showed that equal-gain combining is the best linear combining strategy when using much more than -56 dBm average receive power. For low receive power, maximal ratio combining was best. We also developed approximations to the bit-error-rate-versus-outage probability, and verified their efficacy through Monte Carlo simulations.

We built a single-transmit, dual-receive aperture, equal-gain combining, optically-preamplified receiver with midpoint thresholding to examine the validity of our models. We found good agreement between our model and this experimental 1.25 Gbps testbed. Although, the measured fading was not lognormal deep into the tails, the lognormal model provided a good description of the measured average bit error.

For both direct detection channels, we have shown the merits of equal-gain combining. For photon-counting receivers, equal-gain combining is a capacity-achieving receiver in low noise. In fact, for OOK repetition spatial coding, it is the minimum probability of error receiver at high signal-to-noise ratio. We have also seen that for optically-preamplified channels, when the fixed component of the variance is much less than the signal-dependent component, equal-gain combining is the best linear combining strategy. These findings provide theoretical reinforcement to the results of [47, 51] and to common industry practices [16, 10].

Equal-gain combining and repetition spatial coding are attractive architectures for direct detection channels, both in terms of performance and simplicity. For the nominal parameters of the 1.25 Gbps testbed, a single-transmit and dual-receive aperture, equal-gain combining receiver with midpoint threshold saves about five decibels in average total receive power at 10^{-5} average error rates in severe fading. Using dual-transmit, and single-receive apertures saves around six decibels. Neither system requires path gain knowledge, and both have simple implementations.

Appendix A

Derivation Details

A.1 High and Low Noise Duty Cycles

The duty cycle p that maximizes $\mathcal{I}(p, s)$ for a given background noise-to-signal ratio s has the form (see (4.59) and (4.29))

$$p^{\max} = \frac{(1+s)^{(1+s)}}{es^s} - s. \tag{A.1}$$

We first examine the case in which the background noise-to-signal ratio approaches zero, i.e., $s \to 0$. By continuity of the righthand side, and noting that $s^s \to 1$ as $s \to 0$ we have

$$\lim_{s\to 0} p^{\max} = \frac{1}{e \lim_{s\to 0} s^s} = \frac{1}{e}\ . \tag{A.2}$$

The asymptotic expression for the duty cycle in the high noise regime is derived in [67, 17] using the power series expansions $\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \cdots$

and $\exp(x) = 1 + x + x^2/2! + x^3/3! + \cdots$

$$\begin{aligned} p^{\max} &= \frac{1+s}{e}\left(1+\frac{1}{s}\right)^s - s \\ &= \frac{1+s}{e}\exp\left[s\log\left(1+\frac{1}{s}\right)\right] - s \\ &= \frac{1+s}{e}\exp\left[s\left\{\frac{1}{s} - \frac{1}{2s^2} + O\left(\frac{1}{s^3}\right)\right\}\right] - s \\ &= \frac{1+s}{e}\exp\left[1 - \frac{1}{2s} + O\left(\frac{1}{s^2}\right)\right] - s \\ &= \frac{1+s}{e}e\left[1 - \frac{1}{2s} + O\left(\frac{1}{s^2}\right)\right] - s \\ &= \frac{1}{2} + O\left(\frac{1}{s}\right), \end{aligned} \tag{A.3}$$

where the notation $f = O(g)$ means that $f(x) \leq Cg(x)$ for some constant C and all x in the domain.

A.2 High and Low Noise Information Functions

We now examine the behavior of the information function (4.35)

$$\mathcal{I}\,(p,r) \equiv p\,(1+r)\log(1+r) + (1-p\,)\,r\log r - (\,p+r\,)\log(\,p+r\,), \tag{A.4}$$

for small and large values of the background noise-to-signal ratio r. Noting that $r\log r \to 0$ as $r \to 0$, we have

$$\lim_{r\to 0}\mathcal{I}\,(p,r) = -p\log p\ . \tag{A.5}$$

For large r, we again use the expansion $\log(1+x) = x - x^2/2 + x^3/3 - x^4/4 + \cdots$ to obtain [67]

$$\begin{aligned}
\mathcal{I}\,(p,r) &= p\,(1+r)\log\left(1+\frac{1}{r}\right) - (\,p+r\,)\log\left(1+\frac{p}{r}\right) \\
&= p\,(1+r)\left[\frac{1}{r} - \frac{1}{2r^2} + O\left(\frac{1}{r^3}\right)\right] - (\,p+r\,)\left[\frac{p}{r} - \frac{p^2}{2r^2} + O\left(\frac{1}{r^3}\right)\right] \\
&= \frac{p(1-p)}{2r} + O\left(\frac{1}{r^2}\right). \qquad \text{(A.6)}
\end{aligned}$$

A.3 Lognormal Moment-Matching

In this appendix we derive the first- and second-order moment-matching approximation to the sum of real, correlated, lognormal random variables, see (4.89) through (4.92). Utilizing the Gaussian moment-generating function, the first and second moments of Z are

$$\begin{aligned}
E[Z] &= E[e^u] = \exp(\mu + \nu^2/2) \qquad \text{(A.7)} \\
E[Z^2] &= E[e^{2u}] = \exp(2\mu + 2\nu^2). \qquad \text{(A.8)}
\end{aligned}$$

Solving for μ and ν^2 in terms of $E[Z]$ and $E[Z^2]$ gives

$$\begin{aligned}
\nu^2 &= \log\left(\frac{E[Z^2]}{E[Z]^2}\right) = \log\left(1 + \frac{\text{var}[Z]}{E[Z]^2}\right) \qquad \text{(A.9)} \\
\mu &= \log\left(\frac{E[Z]^2}{\sqrt{E[Z^2]}}\right) = \log\left(\frac{E[Z]^2}{\sqrt{\text{var}[Z] + E[Z]^2}}\right). \qquad \text{(A.10)}
\end{aligned}$$

Substituting $m \equiv E[Z] = E[S]$ and $v^2 \equiv \text{var}[Z] = \text{var}[S]$ gives (4.89) and (4.90).

The moments of the sum are

$$E[S] = \sum_{k=1}^{K} E[e^{u_k}] = \sum_{k=1}^{K} e^{\mu_k + \nu_{kk}/2} \tag{A.11}$$

$$\begin{aligned} E[S^2] &= \sum_{k=1}^{K}\sum_{l=1}^{K} E[e^{u_k+u_l}] \\ &= \sum_{k=1}^{K}\sum_{l=1}^{K} e^{\mu_k+\mu_l+[\nu_{kk}+\nu_{ll}+2\nu_{kl}]/2}. \end{aligned} \tag{A.12}$$

Consequently,

$$\begin{aligned} \mathrm{var}[S] &= E[S^2] - E[S]^2 \\ &= \sum_{k=1}^{K}\sum_{l=1}^{K} e^{\mu_k+\mu_l+[\nu_{kk}+\nu_{ll}]/2}\left(e^{\nu_{kl}}-1\right), \end{aligned} \tag{A.13}$$

as desired.

A.4 Low Noise Regime Lognormal Sum Moments

In this appendix we derive the mean and variance of the sum $S(\alpha)$ in (4.96). Each term in the sum is an independent lognormal random variable with log-mean and log-covariances,

$$\begin{aligned} \mu_{nm} &\equiv E[\log(A_n\alpha_{nm})] = \log A_n - 2\sigma_\chi^2 \\ \nu_{nm,\bar{n}\bar{m}} &\equiv \mathrm{cov}[\log(A_n\alpha_{nm}), \log(A_{\bar{n}}\alpha_{\bar{n}\bar{m}})] = \begin{cases} 4\sigma_\chi^2 & n=\bar{n} \text{ and } m=\bar{m} \\ 0 & \text{otherwise} \end{cases}. \end{aligned}$$

Substituting these into (4.91) and (4.92) we get (4.100) and (4.101), respectively,

$$\begin{aligned} m &\equiv E[S(\alpha)] = \sum_{m=1}^{M}\sum_{n=1}^{N} e^{\mu_{nm}+\nu_{nm,nm}/2} \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N} A_n e^{-2\sigma_\chi^2+4\sigma_\chi^2/2} = MP_1, \end{aligned} \tag{A.14}$$

and

$$\begin{aligned} v^2 &\equiv \text{var}[S] \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{\bar{m}=1}^{M}\sum_{\bar{n}=1}^{N} \exp(\mu_{nm}+\mu_{\bar{n}\bar{m}}+[\nu_{nm,nm}+\nu_{\bar{n}\bar{m},\bar{n}\bar{m}}]/2)\,(e^{\nu_{nm,\bar{n}\bar{m}}}-1) \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N} \exp(2\mu_{nm}+\nu_{nm,nm})\,(e^{\nu_{nm,nm}}-1) \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N} A_n^2 \exp[2(-2\sigma_\chi^2)+4\sigma_\chi^2]\left(e^{4\sigma_\chi^2}-1\right) = MS_4P_2. \end{aligned} \tag{A.15}$$

A.5 High Noise Regime Lognormal Sum Moments

A.5.1 Mean and Variance of the Lognormal Sum

Calculating the mean (4.102) and variance (4.104) in the high noise regime is slightly more involved than in the low noise regime. The terms of the sum (4.99) are

$$\frac{A_nA_k}{\lambda_m}\,\alpha_{nm}\alpha_{km} \equiv e^{u_{mnk}}, \tag{A.16}$$

where

$$u_{mnk} \equiv \log\left(\frac{A_nA_k}{\lambda_m}\right) + 2(\chi_{nm}+\chi_{km}). \tag{A.17}$$

The log-mean and log-covariances are

$$\begin{aligned} \mu_{mnk} &\equiv E[u_{mnk}] = \log\left(\frac{A_nA_k}{\lambda_m}\right) - 4\sigma_\chi^2 \\ \nu_{mnk,\bar{m}\bar{n}\bar{k}} &\equiv \text{cov}[u_{mnk},u_{\bar{m}\bar{n}\bar{k}}] = 4\,\text{cov}[\chi_{nm}+\chi_{km},\chi_{\bar{n}\bar{m}}+\chi_{\bar{k}\bar{m}}] \\ &= 4\,(\text{cov}[\chi_{nm},\chi_{\bar{n}\bar{m}}]+\text{cov}[\chi_{nm},\chi_{\bar{k}\bar{m}}]+\text{cov}[\chi_{km},\chi_{\bar{n}\bar{m}}]+\text{cov}[\chi_{km},\chi_{\bar{k}\bar{m}}]) \\ &= 4\sigma_\chi^2(\delta_{n\bar{n}}+\delta_{n\bar{k}}+\delta_{k\bar{n}}+\delta_{k\bar{k}})\delta_{m\bar{m}}, \end{aligned} \tag{A.18}$$

where $\delta_{ij} = 1$ if $i = j$ and zero otherwise. Substituting these moments into (4.91) we have

$$\begin{aligned} m &\equiv E[S] = \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{k=1}^{N} \exp(\mu_{mnk} + \nu_{mnk,mnk}/2) \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N}\left(\exp(\mu_{mnn} + \nu_{mnn,mnn}/2) + \sum_{\substack{k=1 \\ k\neq n}}^{N} \exp(\mu_{mnk} + \nu_{mnk,mnk}/2)\right) \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N}\left(\frac{A_n^2}{\lambda_m}\exp(-4\sigma_\chi^2 + 16\sigma_\chi^2/2) + \sum_{\substack{k=1 \\ k\neq n}}^{N} \frac{A_n A_k}{\lambda_m}\exp(-4\sigma_\chi^2 + 8\sigma_\chi^2/2)\right) \\ &= \left(\sum_{m=1}^{M}\frac{1}{\lambda_m}\right)\left[\left(e^{4\sigma_\chi^2} - 1\right)\sum_{n=1}^{N} A_n^2 + \left(\sum_{n=1}^{N} A_n\right)^2\right] = L_{-1}[S_4 P_2 + P_1^2]. \quad \text{(A.19)} \end{aligned}$$

Calculating the variance is an exercise in counting. The variance is (4.92)

$$\begin{aligned} v^2 &\equiv \mathrm{var}[S] = \sum_{\substack{m,n,k \\ \bar{m},\bar{n},\bar{k}}} \exp(\mu_{mnk} + \mu_{\bar{m}\bar{n}\bar{k}} + [\nu_{mnk,mnk} + \nu_{\bar{m}\bar{n}\bar{k},\bar{m}\bar{n}\bar{k}}]/2)\,(e^{\nu_{mnk,\bar{m}\bar{n}\bar{k}}} - 1) \\ &\equiv T_1 + T_2 + T_3 + T_4, \quad \text{(A.20)} \end{aligned}$$

where the terms T_1, T_2, T_3, and T_4 are defined as follows. Using (A.18), we can set $m = \bar{m}$ because the $m \neq \bar{m}$ portion of (A.20) is zero. Next, we break the remaining sum into four terms based on the structure of the log-mean and log-covariance. The first term is the case when $n = k$ and $\bar{n} = \bar{k}$

$$\begin{aligned} T_1 &\equiv \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{\bar{n}=1}^{N} \frac{A_n^2 A_{\bar{n}}^2}{\lambda_m^2} e^{-8\sigma_\chi^2} \exp([16\sigma_\chi^2 + 16\sigma_\chi^2]/2)\left(e^{4\sigma_\chi^2(4\delta_{n\bar{n}})} - 1\right) \\ &= \sum_{m=1}^{M}\sum_{n=1}^{N} \frac{A_n^4}{\lambda_m^2} e^{8\sigma_\chi^2}\left(e^{16\sigma_\chi^2} - 1\right) \\ &= L_{-2}(S_8 + 1)S_{16}P_4. \quad \text{(A.21)} \end{aligned}$$

The second term is the case when $n = k$ and $\bar{n} \neq \bar{k}$,

$$\begin{aligned}
T_2 &= \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{\bar{n}=1}^{N}\sum_{\substack{\bar{k}=1\\ \bar{k}\neq\bar{n}}}^{N} \frac{A_n^2 A_{\bar{n}} A_{\bar{k}}}{\lambda_m^2} e^{-8\sigma_\chi^2} \exp([16\sigma_\chi^2 + 8\sigma_\chi^2]/2) \left(e^{4\sigma_\chi^2(2\delta_{n\bar{n}}+2\delta_{n\bar{k}})} - 1\right) \\
&= \sum_{m=1}^{M}\sum_{n=1}^{N}\left[\sum_{\substack{\bar{n}=1\\ \bar{n}\neq n}}^{N} \frac{A_n^3 A_{\bar{n}}}{\lambda_m^2} e^{4\sigma_\chi^2}\left(e^{8\sigma_\chi^2} - 1\right) + \sum_{\substack{\bar{k}=1\\ \bar{k}\neq n}}^{N} \frac{A_n^3 A_{\bar{n}}}{\lambda_m^2} e^{4\sigma_\chi^2}\left(e^{8\sigma_\chi^2} - 1\right)\right] \\
&= 2e^{4\sigma_\chi^2}\left(e^{8\sigma_\chi^2} - 1\right)\sum_{m=1}^{M}\lambda_m^{-2}\sum_{n=1}^{N}\left[A_n^3\left(\sum_{\bar{n}=1}^{N} A_{\bar{n}} - A_n\right)\right] \\
&= 2L_{-2}(S_4 + 1)S_8(P_1 P_3 - P_4). \qquad \text{(A.22)}
\end{aligned}$$

The third term is the case when $n \neq k$ and $\bar{n} = \bar{k}$. By symmetry, this term is equal to the term that we just calculated, i.e., $T_3 = T_2$.

The last term is the hardest, and is the case when $n \neq k$ and $\bar{n} \neq \bar{k}$,

$$\begin{aligned}
T_4 &\equiv \sum_{m=1}^{M}\sum_{\substack{n,k=1\\ n\neq k}}^{N}\sum_{\substack{\bar{n},\bar{k}=1\\ \bar{n}\neq\bar{k}}}^{N} \frac{A_n A_k A_{\bar{n}} A_{\bar{k}}}{\lambda_m^2} e^{-8\sigma_\chi^2} \exp([8\sigma_\chi^2 + 8\sigma_\chi^2]/2) \left(e^{4\sigma_\chi^2(\delta_{n\bar{n}}+\delta_{n\bar{k}}+\delta_{k\bar{n}}+\delta_{k\bar{k}})} - 1\right) \\
&= \sum_{m=1}^{M}\sum_{\substack{n,k=1\\ n\neq k}}^{N}\sum_{\substack{\bar{n},\bar{k}=1\\ \bar{n}\neq\bar{k}}}^{N} \frac{A_n A_k A_{\bar{n}} A_{\bar{k}}}{\lambda_m^2}\left(e^{4\sigma_\chi^2(\delta_{n\bar{n}}+\delta_{n\bar{k}}+\delta_{k\bar{n}}+\delta_{k\bar{k}})} - 1\right). \qquad \text{(A.23)}
\end{aligned}$$

There are 16 possible equality and inequality relationships in the pairs of variables $(n, \bar{n})$, $(n, \bar{k})$, $(k, \bar{n})$, $(k, \bar{k})$ that appear in the Kronecker delta terms of the exponent. The possibilities that $n = k = \bar{n} = \bar{k}$ and that equality exists in exactly three of the four pairs, e.g. $n = \bar{n}$, $n = \bar{k}$, $k = \bar{n}$, and $k \neq \bar{k}$, are precluded in this case by our assumption that $n \neq k$ and $\bar{n} \neq \bar{k}$. The case when inequality exists in all four pairs, i.e., $n \neq \bar{n}$, $n \neq \bar{k}$, $k \neq \bar{n}$, and $k \neq \bar{k}$, is not of interest to us because these terms in T_4 are zero. This leaves the two cases in which equality exists in exactly one and two pairs.

Because of our assumption in this last term that $n \neq k$ and $\bar{n} \neq \bar{k}$, equality can exist exactly twice in these four pairs only when $n = \bar{n} \neq k = \bar{k}$ or $n = \bar{k} \neq k = \bar{n}$.

For example, the assignment $n \neq \bar{n}$, $n = \bar{k}$, $k \neq \bar{n}$, and $k = \bar{k}$ is not possible because $n = \bar{k}$ and $k = \bar{k}$ implies $n = k$. Let's examine the case when $n = \bar{n} \neq k = \bar{k}$,

$$\begin{aligned} T_{42} &\equiv \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{\substack{k=1 \\ k\neq n}}^{N} \frac{A_n^2 A_k^2}{\lambda_m^2}\left(e^{8\sigma_\chi^2}-1\right) \\ &= \left(e^{8\sigma_\chi^2}-1\right)\left(\sum_{m=1}^{M}\lambda_m^{-2}\right)\sum_{n=1}^{N}A_n^2\left[\sum_{k=1}^{N}A_k^2 - A_n^2\right] \\ &= L_{-2}S_8(P_2^2 - P_4). \end{aligned} \tag{A.24}$$

By symmetry, the case when $n = \bar{k} \neq k = \bar{n}$ is the same.

Now we examine the T_4 case when equality exists in exactly one pair $(n, \bar{n})$, $(n, \bar{k})$, $(k, \bar{n})$, $(k, \bar{k})$ and inequality exists in the others. For example, consider the case when $n = \bar{n}$, $n \neq \bar{k}$, $k \neq \bar{n}$, and $k \neq \bar{k}$. Notice that $n \neq \bar{k}$ and $k \neq \bar{n}$ are redundant because we have by assumption that $n \neq k$ and $\bar{n} \neq \bar{k}$ for the term T_4. Evaluating this case,

$$\begin{aligned} T_{41} &\equiv \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{\substack{k=1 \\ k\neq n}}^{N}\sum_{\substack{\bar{k}=1 \\ \bar{k}\neq k \\ \bar{k}\neq n}}^{N} \frac{A_n^2 A_k A_{\bar{k}}}{\lambda_m^2}\left(e^{4\sigma_\chi^2}-1\right) \\ &= \left(e^{4\sigma_\chi^2}-1\right)\sum_{m=1}^{M}\lambda_m^{-2}\sum_{n=1}^{N}A_n^2\sum_{\substack{k=1 \\ k\neq n}}^{N}A_k\sum_{\substack{\bar{k}=1 \\ \bar{k}\neq k \\ \bar{k}\neq n}}^{N}A_{\bar{k}} \\ &= L_{-2}S_4\sum_{n=1}^{N}A_n^2\sum_{\substack{k=1 \\ k\neq n}}^{N}A_k\left(\sum_{\bar{k}=1}^{N}A_{\bar{k}} - A_n - A_k\right) \\ &= L_{-2}S_4\sum_{n=1}^{N}A_n^2\left[\sum_{k=1}^{N}A_k\left(P_1 - A_n - A_k\right) - A_n\left(P_1 - 2A_n\right)\right] \\ &= L_{-2}S_4\sum_{n=1}^{N}A_n^2\left[P_1^2 - P_2 - 2A_nP_1 + 2A_n^2\right] \\ &= L_{-2}S_4(P_1^2P_2 - P_2^2 - 2P_1P_3 + 2P_4). \end{aligned} \tag{A.25}$$

By symmetry, the other three cases in which exactly one pair is equal are the same.

So the last term is

$$T_4 = 2T_{42} + 4T_{41} = L_{-2}[2S_8(P_2^2 - P_4) + S_4(P_1^2 P_2 - P_2^2 - 2P_1 P_3 + 2P_4)]. \quad \text{(A.26)}$$

Putting these four terms together, $T = T_1 + T_2 + T_3 + T_4$, we get (4.104).

A.5.2 Approximate Mean and Variance of the Lognormal Sum

In this appendix, we calculate an approximation to the first and second moments of the lognormal sum $S(\alpha)$ in the high noise regime (4.99). We start by using a lognormal approximation to the sum R_m whose mean and variance are

$$E[R_m] = \sum_{n=1}^{N} E[\alpha_{nm}]A_n = P_1 \quad \text{(A.27)}$$

$$\text{var}[R_m] = \sum_{n=1}^{N} \text{var}[\alpha_{nm}]A_n^2 = S_4 P_2 \quad \text{(A.28)}$$

We can then approximate $R_m \approx e^{\bar{u}_m}$ where $\bar{u}_m$ are independent, identically distributed Gaussian random variables with moments given by (4.89) and (4.90)

$$\bar{\mu} \equiv E[\bar{u}_m] = \log\left(\frac{P_1}{\sqrt{1 + \frac{S_4 P_2}{P_1^2}}}\right) = \log\left(\frac{NA}{\sqrt{1 + \frac{e^{4\sigma_\chi^2} - 1}{N}}}\right) \quad \text{(A.29)}$$

$$\bar{\nu}^2 \equiv \text{var}[\bar{u}_m] = \log\left(1 + \frac{S_4 P_2}{P_1^2}\right) = \log\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right), \quad \text{(A.30)}$$

where the last equalities are for identical transmitters. Next, we approximate the sum (4.99) as

$$S(\alpha) = \sum_{m=1}^{M} R_m^2/\lambda_m \approx \sum_{m=1}^{M} e^{2\bar{u}_m}/\lambda_m, \quad \text{(A.31)}$$

with moments

$$\begin{aligned} m &\equiv E[S(\alpha)] \approx \sum_{m=1}^{M} E\left[e^{2\bar{u}_m}\right]/\lambda_m \\ &= L_{-1}e^{2(\bar{\mu}+\bar{\nu}^2)} = L_{-1}\left[S_4P_2 + P_1^2\right] \\ &= \frac{M(NA)^2}{\lambda}\left(1 + \frac{e^{4\sigma_\chi^2}-1}{N}\right) \qquad \text{(A.32)} \\ v^2 &\equiv \mathrm{var}[S(\alpha)] \approx \sum_{m=1}^{M} \mathrm{var}\left[e^{2\bar{u}_m}\right]/\lambda_m^2 \\ &= L_{-2}\left(E\left[e^{4\bar{u}_m}\right] - \left\{E\left[e^{2\bar{u}_m}\right]\right\}^2\right) \\ &= L_{-2}\left(e^{4\bar{\mu}+8\bar{\nu}^2} - e^{4\bar{\mu}+4\bar{\nu}}\right) \\ &= L_{-2}\, e^{4(\bar{\mu}+\bar{\nu}^2)}\left(e^{4\bar{\nu}^2}-1\right) \\ &= L_{-2}P_1^4\left(1+\frac{S_4P_2}{P_1^2}\right)^2\left[\left(1+\frac{S_4P_2}{P_1^2}\right)^4 - 1\right] \\ &= \frac{M(NA)^4}{\lambda^2}\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right)^2\left(\left[1+\frac{e^{4\sigma_\chi^2}-1}{N}\right]^4 - 1\right), \qquad \text{(A.33)} \end{aligned}$$

where again the last equalities are for identical transmitters and receivers.

Bibliography

[1] S.M. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE Journal on Select Areas in Communications*, 16(8):1451–1458, October 1998. 32, 106, 161

[2] R. Barakat. Sums of independent lognormally distributed random variables. *J. Opt. Soc. Am.*, 66(3):211–215, March 1976. 152

[3] N.C. Beaulieu, A.A. Abu-Dayya, and P.J. McLane. Estimating the distribution of a sum of independent lognormal random variables. *IEEE Trans. on Comm.*, 43(12):2869–2873, December 1995. 154

[4] D. Ben-Eli, Y.E. Dallal, and S. Shamai (Shitz). Performance bounds and cutoff rates of quantum limited OOK with optical amplification. *IEEE Journal on Selected Areas in Communications.*, 13(3):510–530, April 1995. 184

[5] E. Biglieri, J. Proakis, and S. Shamai (Shitz). Fading channels: Information-theoretic and communications aspects. *IEEE Trans. on Inform. Theory*, 44(6):2619–2692, October 1998. 31, 83, 152

[6] R. Boel, P. Varaiya, and E. Wong. Martingales on jump processes. II: Applications. *SIAM J. Control*, 13(5):1022–1061, 1975. 63, 120

[7] M. Born and E. Wolf. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light.* Pergamon Press, 6th edition, 1980. 58

[8] P. Bremaud. An averaging principle for filtering a jump process with point observations. *IEEE Trans. Inform. Theory*, 34(3):582–568, May 1988. 128

[9] S.I. Bross, M.V. Brunashev, and S. Shamai (Shitz). Error exponents for the two-user Poisson multiple-access channel. *IEEE Trans. on Inform. Theory*, 47(5):1999–2016, July 2001. 118

[10] W.M. Bruno, R. Mangual, and R.F. Zampolin. Diode laser spatial diversity transmitter. *SPIE Optomechanical Design of Laser Transmitters and Receivers*, 1044:187–194, 1989. 168, 231

[11] E.A. Bucher. Error performance bounds for two receivers for optical communication and detection. *Appl. Opt.*, 11, April 1972. 172

[12] V.W.S. Chan. Coding for the turbulent atmosphere optical channel. *IEEE Trans. on Comm.*, 30(1):269–275, January 1982. 116, 161, 172

[13] V.W.S. Chan, J.H. Shapiro, and F.N.C. Wong. Architecture, system design and concept demonstration of a battlefield agile-beam optical communications system. Steered agile beams project proposal, Massachusetts Institute of Technology, 1999. 36

[14] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications. John Wiley & Sons, Inc., 1991. 30, 80, 119, 127, 140, 187

[15] F.M. Davidson and Y.T. Koh. Interleaved convolutional coding for the turbulent atmospheric optical communication channel. *IEEE Trans. on Comm.*, 39(9):993–1003, September 1988. 161, 172

[16] C.C. Davis, I.I. Smolyaninov, and S.D. Milner. Flexible optical wireless links and networks. *IEEE Communications Magazine*, 41(3):51–57, March 2003. 231

[17] M.A. Davis. Capacity and cutoff rate for Poisson-type channels. *IEEE Trans. on Inform. Theory*, IT-26(6):710–715, November 1980. 117, 122, 123, 128, 130, 133, 139, 233

[18] G.D. Forney. Chapter 5: Signal spaces and optimum detection theory. MIT 6.451 Course Notes, Spring 2000. 86

[19] M.R. Frey. Information capacity of the Poisson channel. *IEEE Trans. on Inform. Theory*, 37(2):244–256, March 1991. 117

[20] R.M. Gagliardi and S. Karp. *Optical Communications*. Wiley, New York, 1995. 45, 46, 50, 64, 77

[21] R.G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968. 30, 135, 170, 183

[22] A.J. Goldsmith and P.P. Varaiya. Capacity of fading channels with side information. *IEEE Trans. on Inform. Theory*, 43(6):1986–1992, November 1997. 36, 83

[23] J. Gowar. *Optical Communication Systems*. Prentice-Hall International, Inc., London, 1984. 45

[24] R.M. Gray. *Entropy and Information*. Springer-Verlag, New York, 1990. 31

[25] S.M. Haas and J.H. Shapiro. Capacity of the multiple-input, multiple-output Poisson channel. In Bozenna Pasik-Duncan, editor, *Proceedings of the Kansas Workshop on Stochastic Theory and Control*, Lecture Notes in Control and Information Sciences, Lawrence, KS, October 18-20, 2001. Springer-Verlag. 19, 37, 117, 153

[26] S.M. Haas and J.H. Shapiro. Capacity of wireless optical communications. *IEEE Journal on Selected Areas in Communications*, 2003. Submitted for publication. 38, 147

[27] S.M. Haas, J.H. Shapiro, and V. Tarokh. Space-time codes for wireless optical communications. *EURASIP Journal on Applied Signal Processing*, 2002(3):1–11, March 2002. Special Issue: "Space-Time Coding and Its Applications – Part I". 37, 83

[28] S.J. Halme, B.K. Levitt, and R.S. Orr. Bounds and approximations for some integral expression involving lognormal statistics. Technical report, MIT Res. Lab. Electron. Quart. Prog. Rept., 1969. 102, 173

[29] R. A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge Univ. Press, New York, 1988. 92

[30] E.V. Hoversten, R.O. Harger, and S.J. Halme. Communication theory for the turbulent atmosphere. *Proc. of the IEEE*, 58(10):1626–1650, October 1970. 168

[31] A. Ishimaru. *Wave Propagation and Scattering in Random Media*. Academic, New York, 2nd edition, 1978. 29

[32] W.C. Jakes. *Microwave Mobile Communications*. IEEE Press, Piscataway, NJ, 1974. 77, 135, 175, 190

[33] Y.M. Kabanov. The capacity of a channel of the Poisson type. *Theory Probab. Appl.*, 23:143–147, 1978. 117, 123, 128, 139

[34] S. Karp, R.M. Gagliardi, S.E. Moran, and L.B. Stotts. *Optical Channels: Fibers, Clouds, Water, and the Atmosphere*. Plenum Press, New York, 1988. 47, 50

[35] R.S. Kennedy. Communication through optical scattering channels: an introduction. *Proc. of the IEEE*, 58(10):1651–1665, October 1970. 50

[36] M.A. Khalighi, J.-M. Brossier, and K. Raoof. Water filling capacity of Rayleigh MIMO channels. In *Proc. of 2001 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, volume 1, pages A–155 – A–158, Sept 2001. 83

[37] I.I. Kim, H. Hakakha, P. Adhikari, E. Korevaar, and A.K. Majumdar. Scintillation reduction using multiple transmitters. *Free-Space Laser Communication Technologies IX, Proc. SPIE*, 2990:102–113, March 1997. 168

[38] I.I. Kim, J. Koontz, H. Hakakha, P. Adhikari, R. Stieger, C. Moursund, M. Barclay, A. Stanford, R. Ruigrok, J. Schuster, and E. Korevaar. Measurement of

scintillation and link margin for the TerraLink laser communication system. *Proc. of SPIE*, 3232:100–118, 1998. Wireless Technologies and Systems: Millimeter-Wave and Optical. 50

[39] Y.T. Koh and F. Davidson. Interleaved concatenated coding for the turbulent atmospheric direct detection optical communication channel. *IEEE Trans. on Comm.*, 37(6):648–651, June 1989. 168

[40] A. Lapidoth and S. Shamai (Shitz). The Poisson multiple-access channel. *IEEE Trans. on Inform. Theory*, 44(2):488–501, March 1998. 115, 118, 127

[41] A. Lapidoth, I.E. Telatar, and R. Urbanke. On wideband broadcast channels. In *IEEE Intl. Symp. on Inform. Theory*, page 188, August 16-21 1998. Cambridge, MA. 118

[42] R.S. Lipster and A.N. Shiryaev. *Statistics of random processes*, volume 2. Springer, Berlin, Heidelberg, New York, 2nd edition, 2000. 120

[43] R.L. Mitchell. Permanence of the log-normal distribution. *J. Opt. Soc. Am.*, 58:1267–1272, September 1968. 152

[44] L.H. Ozarow, A.D. Wyner, and J. Ziv. Information-theoretic considerations for cellular mobile radio. *IEEE Trans. on Inform. Theory*, 43(2):359–378, May 1994. 31

[45] W.K. Pratt. *Laser Communication Systems*. J. Wiley & Sons, New York, 1969. 50

[46] R. Ramaswami and K.N. Sivarajan. *Optical Networks: A Practical Perspective*. Academic Press, London, 1998. 45, 64, 68, 71, 73, 185, 186

[47] M. Razavi and J.H. Shapiro. Wireless optical communications via diversity reception and optical preamplification. *Submitted to IEEE Trans. on Wireless. Commun.*, 2002. 231

[48] S.C. Schwartz and Y.S. Yeh. On the distribution function and moments of power sums with lognormal components. *Bell Syst. Tech. J.*, 61:1441–1462, September 1982. 154

[49] J.H. Shapiro. Imaging and optical communication through atmospheric turbulence. In J.W. Strohbehn, editor, *Laser Beam Propagation in the Atmosphere*, chapter 6. Springer, Berlin, 1978. 29, 46, 47, 49, 51, 56, 134

[50] J.H. Shapiro and R.C. Harney. Simple algorithms for calculating optical communication performance through turbulence. *SPIE Contr. and Commun. Tech. in Laser Sys.*, 295:41–54, 1981. 31, 49, 51, 168

[51] E.J. Shin. Diversity optical communication over the turbulent atmospheric channel. S.M. thesis, Dept. of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 2003. 231

[52] S. Shamai (Shitz). Capacity of a pulse-amplitude modulated direct detection photon channel. *IEEE Proc.*, 137(6):424–436, December 1990. 117

[53] S. Shamai (Shitz). On the capacity of a direct-detection photon channel with intertransition-constrained binary input. *IEEE Trans. on Inform. Theory*, 37(6):1540–1550, November 1991. 117, 141

[54] S. Shamai (Shitz) and I. Bar-David. The capacity of average and peak-power-limited quadrature Gaussian channels. *IEEE Trans. on Inform. Theory*, 41(4):1060–1070, July 1995. 183, 184

[55] S. Shamai (Shitz) and A. Lapidoth. Bounds on the capacity of a spectrally constrained Poisson channel. *IEEE Trans. on Inform. Theory*, 39(1):19–29, January 1993. 118, 141

[56] S. Shamai (Shitz) and A.D. Wyner. Information-theoretic considerations for symmetric, cellular, multiple-access fading channels – Parts I&II. *IEEE Trans. on Inform. Theory*, 43(6):1877–1911, November 1997. 31

[57] J.G. Smith. The information capacity of amplitude- and variance-constrained scalar Gaussian channels. *Information and Control*, 18:203–219, 1971. 183, 184

[58] D.L. Snyder and M.I. Miller. *Random Point Processes in Time and Space*. Springer-Verlag, New York, 2nd edition, 1991. 63, 64, 76, 115, 120, 122

[59] L.G. Tallini, S. Al-Bassam, and B. Bose. On the capacity and codes for the Z-channel. In *Proc. of ISIT*, page 422, Lausanne, Switzerland, June 2002. IEEE. 188

[60] V. Tarokh, H. Jafarkhani, and A.R. Calderbank. Space-time block codes from orthogonal designs. *IEEE Trans. on Inform. Theory*, 45(5):1456–1467, July 1999. 32, 86, 92, 107, 229

[61] V. Tarokh, N. Seshadri, and A.R. Calderbank. Space-time codes for high data rate wireless communication: performance criterion and code construction. *IEEE Trans. on Inform. Theory*, 44(2):744–765, March 1998. 32, 83, 92, 98, 229

[62] V.I. Tatarski. *Wave Propagation in a Turbulent Medium*. Dover, New York, 1961. 29

[63] E. Telatar. Capacity of the multiple-antenna Gaussian channels. Internal tech. memo, AT&T Bell Labs, June 1995. 79, 83

[64] S. Verdu. *The Electrical Engineering Handbook*, chapter "Channel Capacity", pages 1671–1678. IEEE and CRC Press, 1997. 188

[65] H. A. Willebrand and B.S. Ghuman. Fiber optics without fiber. *IEEE Spectrum*, pages 41–45, August 2001. 29

[66] J. Wolfowtiz. *Coding Theorems of Information Theory*. Springer-Verlag, Berlin, Germany, 3rd edition, 1978. 31, 81

[67] A.D. Wyner. Capacity and error exponent for the direct detection photon channel–Parts 1 & 2. *IEEE Trans. Inform. on Theory*, 34(6):1449–1471, November 1988. 117, 130, 133, 135, 139, 233, 235