Free-Space Optical Networks

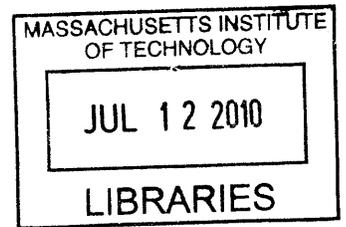## - Fade and Interference Mitigation and Network Congestion Control

by

Etty J. Lee

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 and Computer Science

at the

Massachusetts Institute of Technology

June 2010

Signature of
Author.......................................
Department of Electrical Engineering and Computer Science
February 11, 2010

Certified by...
Vincent W.S. Chan
Joan and Irwin M. Jacobs Professor of Electrical Engineering & Computer Science
Thesis Supervisor

Accepted
by...
Terry P. Orlando
Chair, Department Committee on Graduate Students

1

Free-Space Optical Networks
-Fade and Interference Mitigation and
Network Congestion Control

by

Etty J. Lee

ABSTRACT

Optical communication through the atmospheric channel is commonly known as free-space optical (FSO) communication. When communicating through a clear FSO channel, not only is there atmospheric turbulence which results in fading of the received signal, but there may also be interference that scatters into the receiver and deteriorates performance. In this thesis, we consider mitigating the fading and interference with diversity coherent and diversity incoherent detection. We derive the performance of diversity coherent and diversity incoherent receivers in the presence of fading and various worst case interference types. Diversity coherent detection provides significant power gain over diversity direct detection, and most of the benefit can be achieved with a moderate amount of diversity. Moreover, diversity always improves the performance of coherent detection, whereas diversity improves the performance of direct detection only until an optimal diversity value, beyond which it degrades the performance. We also derive the improvement in expected outage length with diversity, and quantify the amount of interference that the system can handle while still achieving a given outage probability.

Although signal fades or 'outages' in an FSO link can be mitigated on the Physical Layer, they cannot be completely eliminated. In a free-space optical network, these outages affect the performance and design of the Transport Layer. The effect of outages on the TCP sender is to diminish its throughput significantly due to drastic reduction of its rate when its packets do not get received through the outage. We consider a class of TCP-based protocols that is better suited for free-space optical networks. In particular, the protocols in this class have the sender distinguish whether a packet loss is due to an outage or congestion and not reduce its rate if the

loss was due to an outage. We analyze, using an approximate channel model for FSO links, the maximum performance that can be achieved by a sender in this class, and compare the performance against a TCP sender's performance. The protocols in this class can gain back the performance loss in TCP due to link outages and they are particularly beneficial when the path has FSO links with strong turbulence and large bandwidth-delay product. We discuss a possible way to implement the distinguishing of packet loss due to congestion from packet loss due to link outage.

Thesis Supervisor: Vincent W.S. Chan
Title: Joan and Irwin M. Jacobs Professor of Electrical Engineering & Computer Science

Acknowledgments

First, and most importantly, I would like to thank my thesis advisor Vincent Chan for his guidance and support. He has taught me about the many aspects of being a researcher, presenter, and practical engineer. His wisdom, knowledge, and energy are admirable and I am fortunate to have had him as my thesis advisor.

I would also like to thank my thesis committee members John Chapin and Jeffrey H. Shapiro who have also helped guide my research, and who have read my thesis and provided useful feedback.

Thanks also to my husband Dennis Lee, parents, and family for their ongoing and constant support of my academic endeavors.

Table of Contents

Chapter 1

Introduction

Atmospheric optical communication, commonly known as free-space optical communication (FSO), is communication using optical waves (produced by lasers) through the atmosphere. FSO communication provides a means to communicate at high data rates (Gb/s), over distances of tens of meters to tens of thousands of kilometers (even up to a geosynchronous satellite) without being bound to fibers or cables. See Figure 1.1. Since there is no need for fibers to be laid into the ground, faster setup and lower cost is achieved in some applications. Additionally, for long enough distances, the inverse square law of free-space propagation loss [18] has an advantage over fiber which has exponential losses and non-linearities that limit launched optical power [40]. Another appealing feature of FSO communication is that it is physically directed; the transmitter is pointed at the receiver and the signal has a much smaller divergence angle compared to radio frequency communication.
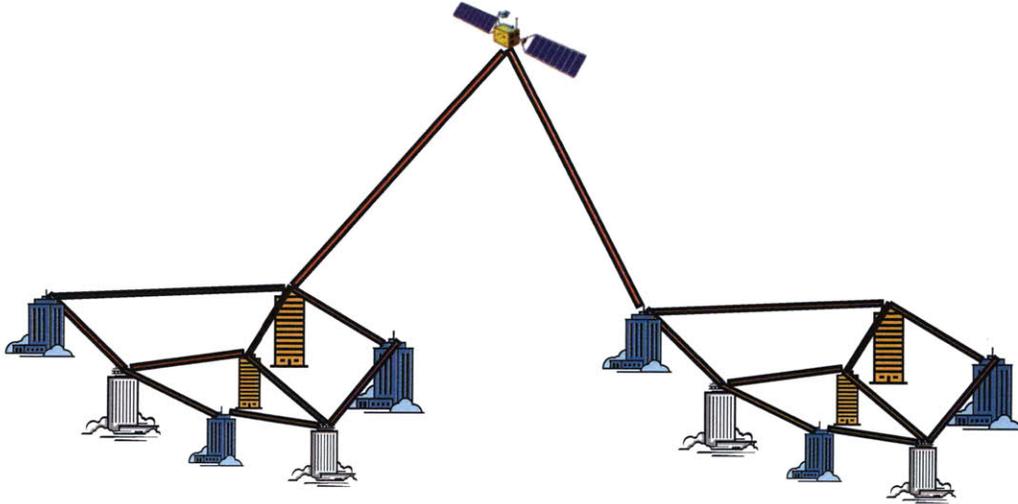
Figure 1.1: Example Free-Space Optical Communication Scenario

Possible applications for FSO communication include remote areas where it is too expensive to lay down cables, urban settings such as between two office buildings where quick setup times are required, and communication to satellites or between aircraft or moving vehicles where it is infeasible to have wired communication.

However, when communicating through the free-space optical channel, the received signal is susceptible to attenuation due to weather conditions. Even in the absence of rain, fog, hail, or snow, the signal undergoes fading due to atmospheric turbulence [19,27,44]. The cause of the fades is the mixing of eddies of air that have slightly different temperatures and hence slightly different refractive indices. As predicted by the Kolmogorov turbulence model, temperature changes in the air on the order of 1° Kelvin cause slight refractive index changes on the order of 10^{-6}. Although these fluctuations may seem small in magnitude, they have an impact on the optical wave that passes through the air (due to the wave's short wavelength) by causing the atmosphere to act as small lenses. The phase front of the transmitted wave undergoes changes as it propagates to the receiver(s), resulting in constructive and destructive interference at the receiver which results in fluctuations in the received signal

amplitude and phase. The durations of the fades roughly equal the time it takes for crosswinds or thermally induced air moments to move the air turbules across the laser beam [44]. It is typical for the fades to be fairly deep (sometimes deeper than 10 dB) and to last anywhere from a few milliseconds to a tenth of a second. When communicating at high data rates (Gb/s), the fades can lead to a loss of a large number of consecutive bits. For example, when communicating at a data rate of 10 Gb/s, a single fade, as seemingly benign as 1 millisecond, can lead to the loss of 10 million consecutive bits (10^{10} Gb/s x 10^{-3} s=10^7 bits).

When communicating through the free-space optical channel, in addition to atmospheric turbulence, one may also need to deal with interference that degrades communication. The interference may be from a multi-access user or an intentional interferer and, even if off-axis, the interference can couple into the receiver through scattering by the receiving equipment. Moreover, this interference may be able to impose a worst case setup for our system and thus have a great detrimental impact on the communication.

A seemingly attractive technique to mitigate fading and interference is to use interleaving with error correction. However, this requires large interleavers (on the order of 1 Gb for a 1 Gb/s channel) and results in long link delay (>1s) which seriously affects the performance of protocols in the Network and Transport Layers of the network. Thus, there is a definite need to mitigate the effects of atmospheric turbulence and possible interference at the Link Layer. Although increasing the transmit power by amounts on the order of 10-20 dB is a simple solution to mitigate the fading and interference, this method requires very expensive optical amplifiers, and may not be suitable for cost-sensitive applications. A sensible technique to mitigate the fades and interference is to use spatial diversity (multiple transmitters or

receivers) and coherent detection. In order to properly evaluate the type of receiver, incoherent versus coherent, and the amount of diversity to use, it is important to understand how diversity direct detection and diversity coherent detection perform in the presence of interference.

Although fades and interference can be mitigated using diversity detection, they cannot be completely eliminated. In a free-space optical *network*, atmospheric turbulence and interference affect not only the performance of the Physical Layer, but also hurt upper layer performance as well. If not designed to deal with the FSO link properties, the upper layers will perform poorly. For example, we show later in this thesis that TCP's congestion control does not perform well in high bandwidth-delay product FSO networks. It is important to re-consider the Transport Layer's congestion control mechanism and how it can be designed to be suited for FSO networks.

This thesis has two main areas of study of free-space optical networks: the Physical Layer and the Transport Layer. Regarding the Physical Layer, we analyze the performance of diversity coherent detection and diversity incoherent detection for communication through the clear atmospheric channel in the presence of various interference types, and quantify the gain of diversity coherent detection over diversity direct detection (both in the presence of and absence of interference and if we mistakenly assume interference is present or absent). We also derive the expected outage length when diversity direct detection is used and when diversity coherent detection is used. On the Transport Layer, we consider a modified TCP congestion control scheme for better sender throughput in free-space optical networks. We analyze the throughput, in steady state and prior to steady state, of TCP's and the

modified TCP's congestion control when operated over free-space optical links with various atmospheric turbulence strengths and congestion loss probabilities.

Chapter 2

Preliminaries

In this Chapter, we describe the fading model and channel model used in this thesis. We also give a brief review of TCP's congestion control algorithm.

2.1   Fading Model

If an optical wave propagates through a vacuum, there is no attenuation or fading. When it propagates through the free-space optical channel within the Earth's atmosphere, however, weather impairments such as rain, snow, hail, and fog cause absorption (irretrievable loss of energy which is absorbed by the particles) and scattering (redirection of the energy) of the optical wave; this is seen as signal attenuation at the receiver [44]. This attenuation is generally modeled as non-random, and a function of the amount of precipitation or fog [26]. Even in the absence of particles in the air, atmospheric turbulence is present, which results in random fading of the optical wave at the receiver. A communication channel that has only atmospheric turbulence without rain, fog, snow, hail or other absorbing particles is called the *clear* atmospheric channel. In this thesis, we consider optical

communication through the clear atmospheric channel. As described in the previous chapter, in clear atmospheric turbulence, the mixing of eddies of air that have slightly different temperatures results in refractive index changes in the air that change the phase front of the wave and cause constructive and destructive addition of the wave fluctuations at the receiver. This phenomenon results in random fading seen at the receiver. Atmospheric turbulence is what causes twinkling of stars and shimmering above pavement or sand on a hot sunny day.

The Extended Huygens-Fresnel Principle [44] describes the propagation of an optical wave through atmospheric turbulence. Before describing this principle, we first state the non-extended Huygens-Fresnel Principle. The Huygens-Fresnel Principle is based on the scalar wave equation and allows us to represent an optical wave after it has traveled through a vacuum from one plane to another parallel plane a distance L away. Given a quasi-monochromatic optical field of complex envelope $U_i(\bar{p}, t)$ in the plane $z=0$, the field, after it has propagated to the plane $z=L$, is given by

$$U_O(\bar{p}', t) = \frac{1}{j\lambda L} \int_{R_1} \left(U_i(\bar{p}, t - \frac{L}{c}) \exp[jk(L + \frac{|\bar{p}' - \bar{p}|^2}{2L})] \right) d\bar{p}$$

(2.1)

where $\bar{p}$ and $\bar{p}'$ are the coordinate vectors in the $z=0$ and $z=L$ planes respectively, λ is the wavelength, t is time, c is the speed of light, k is the wave number ($k=2\pi/\lambda$) and R_1 is the transmitting pupil area as shown in Figure 2.1.
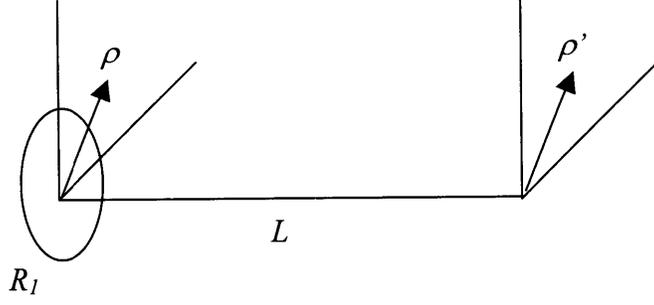
Figure 2.1: Physical Setup for Huygens-Fresnel Principle

The Extended Huygens-Fresnel Principle extends the Huygens-Fresnel Principle to take into account atmospheric turbulence. It states that the field at $z=L$ is given by

$$U_O(\overline{\rho}',t) = \frac{1}{j\lambda L} \int_{R_1} \left(U_i(\overline{\rho},t-\frac{L}{c})\exp[jk(L+\frac{|\overline{\rho}'-\overline{\rho}|^2}{2L})]\exp[\chi(\overline{\rho}',\overline{\rho},t)+j\theta(\overline{\rho}',\overline{\rho},t)] \right) d\overline{\rho} \quad (2.2)$$

where χ and θ are random variables that model the amplitude and phase fluctuation as the field travels through atmospheric turbulence. If we consider a point source or source that is much smaller than the atmospheric coherence length, then the turbulence factor $exp(\chi+j\theta)$ can be factored out of the integral in (2.2). Thus, the output field can be written as

$$U_O(\overline{\rho}',t) = \frac{1}{j\lambda L} \exp[\chi(\overline{\rho}',t)+j\theta(\overline{\rho}',t)] \int_{R_1} \left(U_i(\overline{\rho},t-\frac{L}{c})\exp[jk(L+\frac{|\overline{\rho}'-\overline{\rho}|^2}{2L})] \right) d\overline{\rho} \quad (2.3)$$

and the fading reduces to a multiplicative amplitude and phase factor.

In our multiple receiver atmospheric optical communication system, which we describe in more detail in Chapter 3, we assume that each receiving pupil is separated by more than an intensity coherence length so that the intensity fading seen by each

receiver is approximately independent. This spatial receiver separation can be realistically achieved since the coherence length for communication across distances on the order of tens of kilometers is on the order of centimeters [44] i.e. the multiple receivers only need to be placed centimeters apart to see approximately independent channel fades. Note that for space/ground systems, spatial diversity may be available only at the ground segment due to large coherence lengths in space. Provided that each receiver's pupil size is less than a coherence length, the amplitude fading experienced by each receiver can be modeled as log-normal distributed [44], and the random phase can be modeled as Gaussian distributed [34] with variance so large that the distribution can be practically modeled as uniform over $[0, 2\pi)$. We assume the phase at each receiver is approximately uniform and can be tracked by the coherent receivers.

Let us denote the random multiplicative fading factor seen by receiver i by $e^{\chi_i + j\theta_i}$ where the amplitude fading portion e^{χ_i} is log-normal distributed, χ_i is distributed as $N(m_\chi, \sigma_\chi^2)$, and the random phases θ_i are not necessarily independent and are tracked in the coherent detection system. σ_χ^2 is the log-amplitude variance and is a measure of the amount of atmospheric turbulence. For a horizontal path, the log-amplitude variance is approximately given by

$$\sigma_\chi^2 = \min\{0.124 k^{7/6} C_n^2 L^{11/6}, 0.5\} \tag{2.4}$$

[42] where k is the wave number, C_n^2 is the refractive index structure constant and usually lies between 10^{-16} m$^{-2/3}$ (weak turbulence) and 10^{-12} m$^{-2/3}$ (strong turbulence), and L is the link distance in meters. The variance σ_χ^2 saturates at approximately 0.5 (which is why there is a min{ } and 0.5 in (2.4)). When the variance is this high, we

18

are said to be in the strong fluctuation regime and the log-normal model becomes suspect because the detailed mathematics from which the log-normal distribution is derived does not apply [2,44]. See Figure 2.2 for a plot of this log-amplitude variance for various turbulence levels.
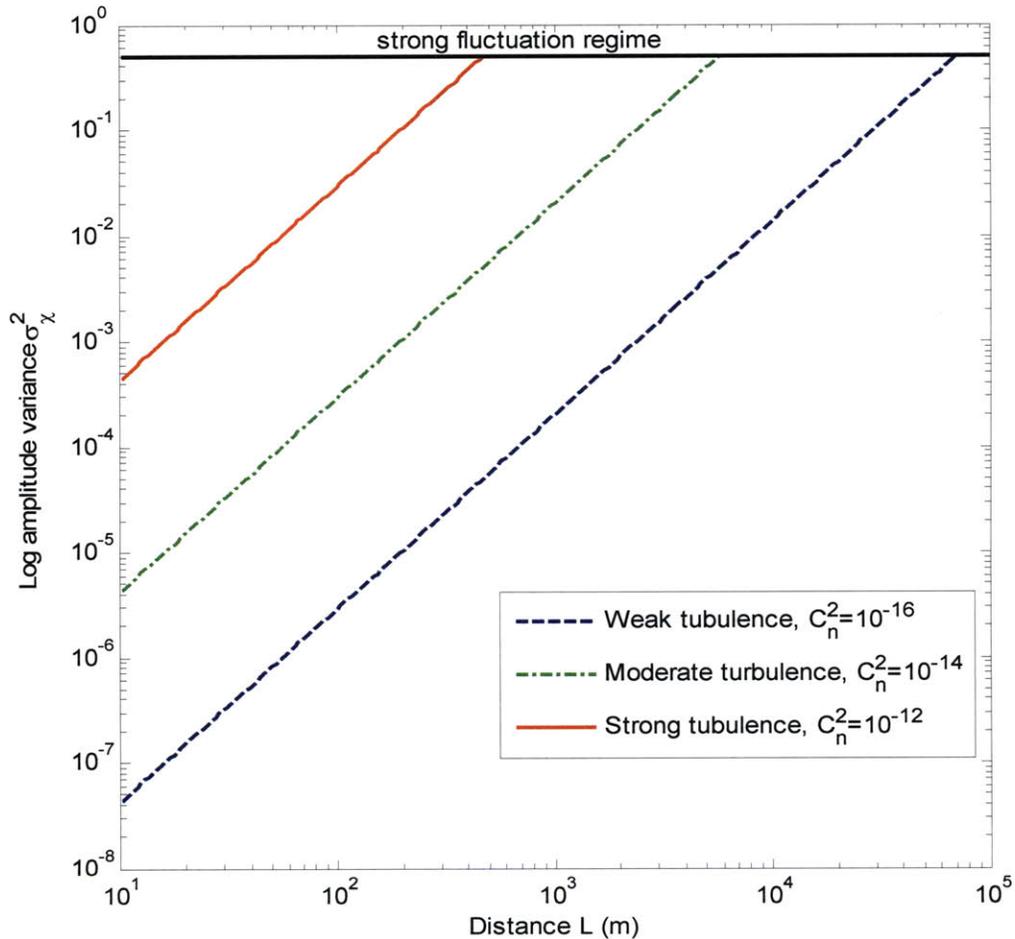
Figure 2.2: Log Amplitude Variance for a horizontal path that uses an optical wave with wavelength 1550 nm

The power fading factor is the square of the amplitude fading factor i.e. the power fading factor is $e^{2\chi_i}$. Thus it is also log-normal distributed where the exponent $2\chi_i$ has Gaussian distribution $N\left(2m_\chi, 4\sigma_\chi^2\right)$. Due to energy conservation, the received

signal is not amplified or attenuated on average i.e. $E[e^{2\chi_i}]$=1. This leads to $m_\chi = -\sigma_\chi^2$ [34]. For use in later chapters, we define $\alpha_i = e^{2\chi_i}$ to be the time-varying log-normal power fading factor seen by receiver i.

In this thesis, we make the assumption that the fading coherence time is much longer than the symbol interval time (realistic for Gb/s communication). Thus, the fading factors α_i can be accurately modeled as constant over each symbol interval time T.

2.2   Outages and Channel Model

In this section, we define an outage and describe our channel model[1].

The usual performance metric in analyzing communication systems is the probability of bit error. In the absence of fading, this is a fine performance metric, and we use it when considering the performance of diversity direct and diversity coherent detection systems in the absence of fading. However, when communicating through atmospheric turbulence at high data rates, error probability is not the best performance metric. Even with a respectable average error probability, the received signal can suffer long and deep fades and cause a large number of consecutive bits to be corrupted. Bit errors due to fades are clearly not independent. This channel memory is not captured by the performance metric of error probability. Since fade lengths (and non-fade lengths) are typically many orders of magnitude longer than bit times for high data rate communication, it makes sense to use the concept of an outage. An outage is defined as follows: if we have a maximum tolerable error probability of $P_e^{thresh} = e^{-\theta_{thresh}}$ (see Figure 2.3), an outage occurs when the short term

---

[1] We described the channel model in our paper [23].

bit error rate over a duration less than the channel coherence time (the operating

point) goes above P_e^{thresh}.
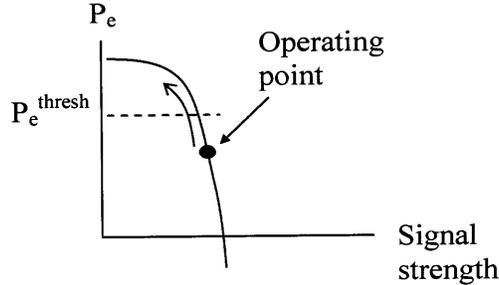
Figure 2.3: Diagram showing that outage occurs when operating point moves above an error probability threshold

Outage probability is the probability that the short-term bit error rate is above the required value P_e^{thresh} i.e.

$$\Pr(outage) = \Pr\left(P_e > P_e^{thresh}\right) \tag{2.5}$$

Outage probability describes the fraction of time during which the system is below a performance threshold. We use outage probability as the performance metric when analyzing diversity direct and diversity coherent detection in atmospheric turbulence.

Describing the channel as being either in the state of an outage or non-outage is a reasonable description of the channel. Consider a communication system that uses a forward error correction code and has additive white Gaussian noise at the receiver. As seen in Figure 2.4, the probability of error curve shifts from sloping gradually, to sloping steeply since the code can correct almost all bits in error if the received signal is above threshold. The communication link can accurately be viewed as out-of-service if the signal is below the threshold (since the error probability is so high), and

in-service with perfect transmission when the signal is above the threshold (since the error probability is so low). The concept of an outage is important as outages and non-outages can have associated time durations (which we call outage length and non-outage length). These durations allow the channel memory to be captured. Outage length and non-outage length are captured in the channel model which we discuss next.
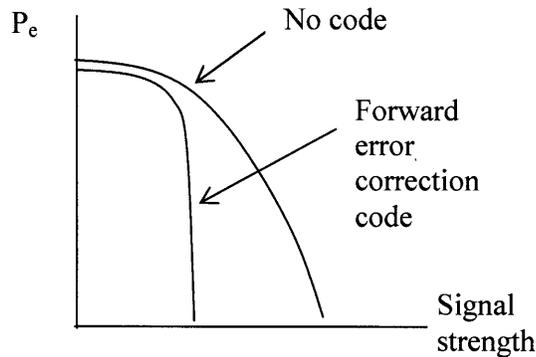
Figure 2.4: Probability of error curve with and without Forward Error Correction (as shown in our paper [28])

We model the channel as a two-state continuous time Markov process, as shown in Figure 2.5, where the two states represent whether we are in an outage or non-outage respectively. The 2-state Markov channel model is justified as an extension of the RF case, in which the channel can be modeled as Markov since the spectrum of the log amplitude factor can be approximated by a 1-pole filter [9]. In the optical case, the log-amplitude fading factor spectrum has a slope of -8/3 [27]. Since a 1-pole filter has slope -2, the log-amplitude fading process in the optical case can be approximated with a 1-pole filter. The 2-state Markov channel model is experimentally confirmed as a reasonable model in the next sub-section.
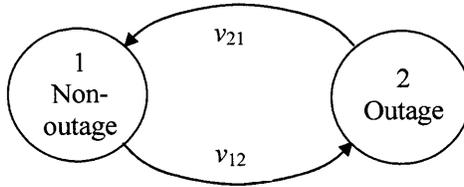
Figure 2.5: Two-state continuous-time Markov channel model

In the two-state channel model, the length of time spent in states 1 and 2, namely the non-outage and outage lengths respectively, are exponentially distributed (as a direct consequence of Markov processes). Letting Y and Z be the outage and non-outage lengths respectively, their probability density functions are given by

$$f_Y(y) = v_{21} \exp(-v_{21}y), \quad y \geq 0$$
$$f_Z(z) = v_{12} \exp(-v_{12}z), \quad z \geq 0$$

(2.6)

where

$$v_{21} = \frac{1}{E[\text{outage length}]}$$

(2.7)

and

$$v_{12} = \frac{1}{E[non - \text{outage length}]}$$

(2.8)

The symbols received during an outage are assumed to be lost, and symbols received during a non-outage are assumed to be received correctly. We use this channel model when analyzing the performance of the Transport Layer. This treatment of the outages is especially important when the link is part of a larger network running network protocols because it captures the channel's memory of the current channel state.

Letting P_{outage} and $1-P_{outage}$ represent the probabilities of being in an outage and non-outage respectively, we obtain the following global balance equation.

$$P_{outage}v_{21} = \left(1 - P_{outage}\right)v_{12}$$

(2.9)

Substituting (2.7) and (2.8) into (2.9) gives

$$\frac{P_{outage}}{E[\text{outage length}]} = \frac{\left(1 - P_{outage}\right)}{E[\text{non - outage length}]}$$

(2.10)

This equation provides a relationship between the expected non-outage and outage lengths in terms of the outage probability. In Chapter 3, we derive expressions that can be used to calculate outage probability and expected outage length for a diversity direct detection or diversity coherent detection system.

Rearranging (2.9), we can express the outage probability and non-outage probability as a function of the transition rates.

$$P_{outage} = \frac{v_{12}}{v_{12} + v_{21}}$$

(2.11)

Moreover,

$$P_{non-outage} = 1 - P_{outage}$$
$$= \frac{v_{21}}{v_{12} + v_{21}}$$

(2.12)

As derived in Appendix A using the Kolmogorov Backward Differential Equations [17], the probability of going from state 1 at time 0 to state 1 at time t (with any combination of transitions during time $(0,t)$) is

$$P_{11}(t) = \frac{\nu_{21}}{\nu_{12} + \nu_{21}} + \frac{\nu_{12}}{\nu_{12} + \nu_{21}} \cdot \exp\{-(\nu_{12} + \nu_{21})t\} \qquad (2.12)$$

and the probability of going from state 1 at time 0 to state 2 at time t (with any combination of transitions during time $(0,t)$) is

$$P_{12}(t) = \frac{\nu_{12}}{\nu_{12} + \nu_{21}} - \frac{\nu_{12}}{\nu_{12} + \nu_{21}} \cdot \exp\{-(\nu_{12} + \nu_{21})t\} \qquad (2.13)$$

These probabilities are used in deriving the throughput of TCP in Chapter 6.

2.3   Experimental Confirmation of Channel Model

We experimentally confirmed [29] that the two-state continuous-time Markov model is a reasonable channel model in the setup shown in Figure 2.6.
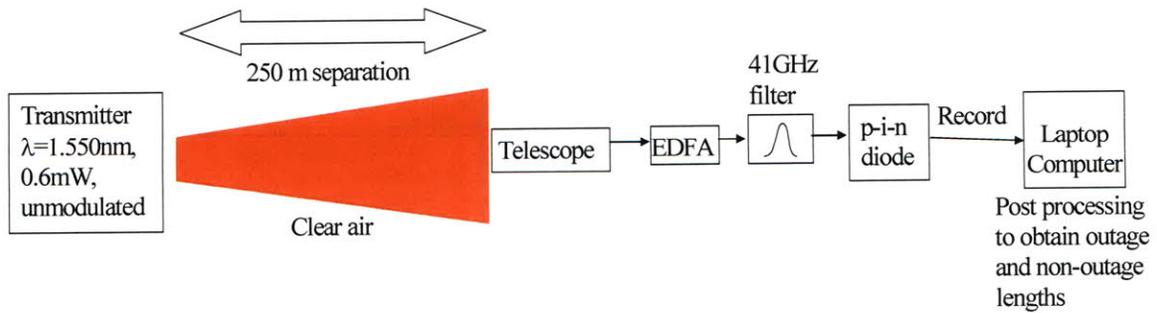
Figure 2.6: Experimental system used to verify two-state channel model

The transmitted optical wave was directed through a telescope and over an outdoor 125 meter path through clear air. The beam was then reflected off a mirror, traveled back 125 meters, and entered the receiving telescope which was coupled into a single

mode fiber. The received signal was optically amplified, filtered, and photo-detected. Since turbulence changes over time, in order to take samples under the same turbulence conditions, the data needed to be taken for limited time segments. Samples of the PIN diodes' output current were taken every millisecond for three minutes, and were used with the power fading factor threshold set at 0.5 to obtain histograms of the outage and non-outage lengths (given in Figure 2.7) [29]. The histograms are plotted together with an exponential probability density function with the same mean. The log amplitude variance was derived from the data to be $\sigma_\chi^2=0.15$. As seen in this figure, the outage and non-outage lengths are approximately exponential.

(a)   (b)
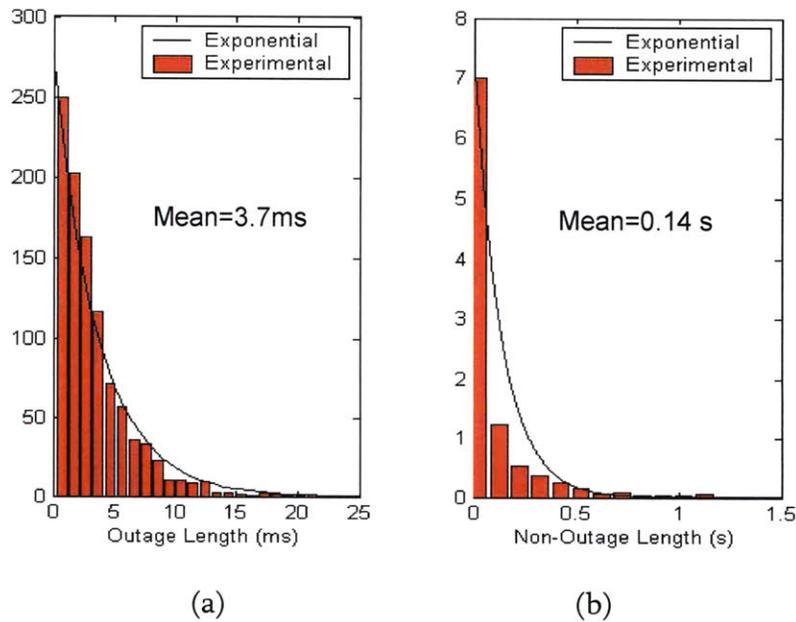
Figure 2.7: Histogram of measured a) outage length and b) non-outage length plotted together with exponential distribution with same mean (as shown in our paper [29])

2.4 Transmission Control Protocol (TCP)'s Congestion Control

In this section, we briefly review TCP's congestion control algorithm [1]. The TCP source and destination nodes are the main active players in controlling the number of packets the senders release into the network. In order to allow the sender to determine if its packets are received by the destination node, TCP uses end-to-end packet acknowledgements (ACKs). The TCP sender includes a packet sequence number (SN) in each packet that it sends to the destination node. If the destination node receives a packet with SN n, it returns an ACK for packet n. If the sender receives this ACK for packet n, it knows that packet n was received. If the next packet the destination node receives is a packet with SN $n+b$, where $b>1$, the destination node returns an ACK for packet n again (thereby signaling that it did not receive packet $(n+1)$). This ACK with the same SN as the previous ACK is called a duplicate ACK. In addition to tracking ACKs, the TCP sender also maintains a value called a window. This window, which we also call "window size", is the maximum number of packets the sender releases into the network for which it has not yet received acknowledgements. The TCP sender dynamically changes its window size and uses the window size to control how many packets it sends into the network.

Before we provide the details of how the TCP sender changes its window size, let us first discuss the basic idea that drives the algorithm. The TCP sender, instead of blasting all of its data at its highest possible transmission rate and possibly causing severe congestion in the network, acts more conservatively. It gradually increases its window as it learns that the packets it sent got received by the destination. If the TCP sender believes there is congestion in the network, it throttles back the amount of data it will release into the network. There are many proposed "variants" of TCP, each of which has a different window update scheme. Since TCP Reno has been

widely employed, we provide on a high level, the details of the TCP Reno's congestion control protocol below:

The TCP sender operates in one of two phases: the Slow Start phase during which the sender increases its window size exponentially, and the Congestion Avoidance phase during which the sender increases its window size linearly. In the Slow Start phase, every time the sender sends a window worth of packets and receives the corresponding ACKs, the sender doubles its window size. In the Congestion Avoidance phase, every time the sender sends a window worth of packets and receives the corresponding ACKs, the sender increases its window size by one packet. The TCP sender starts in the Slow Start phase until the window reaches a threshold (*ssthresh*). After this point, it enters congestion avoidance phase. A TCP sender discerns that a packet was lost in one of two ways: 1) Regardless of what phase the sender is in, if the sender receives three duplicate ACKs (with the same SN), then it assumes that a packet was lost due to congestion and it reduces its window by half (we say the window is closed by a factor half), and goes into or stays in the Congestion Avoidance phase. 2) The other way the TCP sender discerns a packet loss is by setting a timer for each packet, called the retransmission timer (RTO). If the ACK is not received by the time the RTO expires, the sender assumes there must be severe congestion and reduces its window to one packet (we say the window is closed to one packet). It also sets *ssthresh* to half the current window size before the window is reduced, and enters the Slow Start phase. The RTO value is set to be the round-trip time estimate plus four times the standard deviation of the round-trip time estimate. If the RTO expires, we say that a "timeout" occurs, or that the sender "times out".

TCP which has served the users of the Internet well for roughly two decades has been successful in preventing congestion collapse of the network. However, TCP's efficiency decreases as the path's round-trip time increases because after a packet or many packets are lost due to congestion or link errors, TCP window closing is triggered and it takes the TCP sender a long time to ramp up the window size due to the long round-trip times. This causes the throughput to decrease by significantly more than just the bits that are lost. If FSO links are used in the network, they add a different dynamic to the network over that of communication with fiber optics or radio frequency: outages on the links cause a large number of consecutive packet losses rather than occasional single packet losses. In Chapter 6, we discuss the negative effect of outages on TCP and consider a class of TCP-based congestion control protocols that improve sender throughput over high bandwidth-delay product FSO paths.

Chapter 3

Diversity Direct Detection and Diversity Coherent Detection in Absence of Interference

In this chapter, we derive the performance of diversity direct detection and diversity coherent detection in the absence of interference (we published these results in [30] and [33]). We start out by describing the setup of the diversity direct detection and diversity coherent detection systems. Then we derive the outage probability, expected outage length, and effective log-amplitude fading variance of the two systems and provide discussions of each.

One of the original motivations for using coherent detection was to detect a weak signal in the presence of thermal noise [16]. In coherent detection, the mixing of a local oscillator (LO) with the signal raises the signal far above the thermal noise of the electronics and thus the detection process is limited by quantum effects. We will see in this chapter that diversity coherent detection provides additional benefit over

diversity direct detection (that does not use adaptive optics) by selectively filtering out interference and background noise.

3.1 Overview of Diversity Systems

See Figures 3.1 and 3.2 for block diagrams of the diversity direct detection and diversity coherent detection systems that we consider in this thesis. In both diversity systems, each of the N receivers has pupil area A/N so that the total receiving pupil area is A and the total average received signal power is the same regardless of the amount of diversity used.
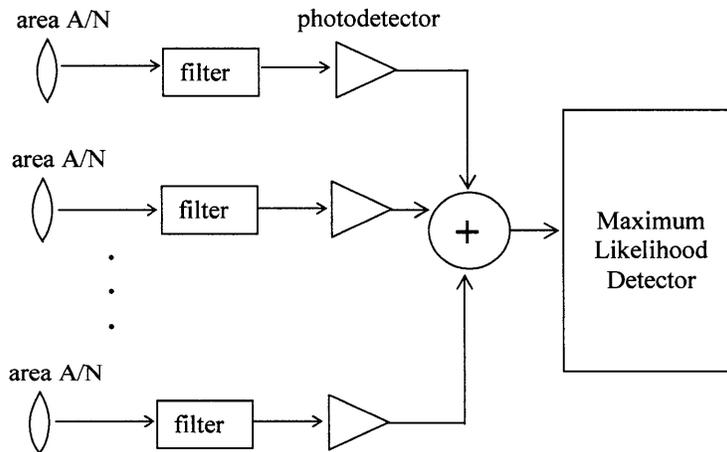
Figure 3.1: Block diagram of multi-aperture incoherent detection (direct detection) with N diversity receivers

photodetector

area A/N

LO <-- | Phase & <---
(γ₁)   fading
          estimator

area A/N

LO <-- | Phase & <---
(γ₂)   fading
          estimator

area A/N

LO <-- | Phase & <---
(γ_N)   fading
          estimator

+ --> filter

Maximum
Likelihood
Detector
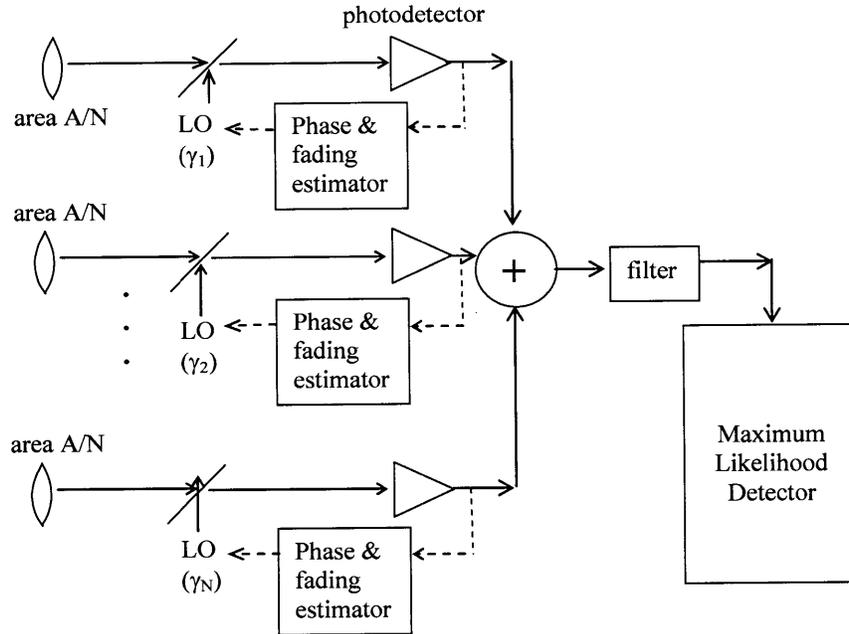
Figure 3.2: Block diagram of multi-aperture coherent detection with N diversity receivers

The diversity direct detection system does not use any adaptive optics such as a rubber mirror to achieve pre-detection coherent combining. In the diversity direct detection system, we assume that the optical filter bandwidth equals the signal bandwidth and that diffraction-limited receivers are used so the minimum amount of interference and background noise is detected. For diffraction-limited receivers, only one of the interference and background noise spatial modes is in the diffraction-limited field of view. Larger field-of-view receivers would not be attractive in the presence of large and multi-spatial-mode interference because all background noise modes in the receiver's field-of-view are detected. Nowadays, the use of diffraction-limited receivers is quite realistic since spatial tracking of the signal is well established. Note that the diffraction-limited solid angle of each receiver in the diversity-N system is N times as large as that of a single receiver system (since the diffraction-limited solid angle is inversely proportional to pupil area). Since the interference and background noise detected is proportional to the pupil area times the

diffraction-limited solid angle [16], each of the N diffraction-limited receivers detects the same average interference and background noise power as the receiver in the single diffraction-limited receiver system. Thus, the N-receiver system detects a total of N times more average interference and background noise than the single receiver system.

In the coherent detection system, each receiver's local oscillator (LO) signal is matched to the incoming signal spatial mode. The local oscillators' angular frequency is denoted by $\omega_L = \omega_0 + \omega_{IF}$ where ω_0 is the optical carrier's angular frequency and ω_{IF} is the intermediate angular frequency (IF). We consider the two types of coherent detection, heterodyne detection and homodyne detection. In heterodyne detection, the IF is non-zero and the filter is band-pass. In homodyne detection, the IF is equal to zero and the filter is low-pass. The total power of the N local oscillators is P_L, regardless of the diversity value. The exact value will not affect performance as long as the power is high enough. Receiver i, 1≤i≤N, has local oscillator power $\gamma_i^2 P_L$ where γ_i^2 is receiver i's local oscillator power scaling factor, and where

$$\sum_{i=1}^{N} \gamma_i^2 = 1 \qquad (3.1)$$

The sum of the γ_i^2's is set equal to a constant as a way to set the relative local oscillator powers. The signal phase and amplitude are tracked by each coherent receiver and we assume this tracking can be done even if there is interference. The coherent detection system coherently combines the received signals with optimal weighting. As we will see later in this chapter and in the next chapter, the total average interference and background noise detected in diversity coherent detection is the same regardless of the amount of diversity (due to the single spatial mode

34

detectors tracking the signal phase and amplitude and coherently combining the N signals with optimal weighting). The diversity-N coherent detection system detects essentially a single spatial mode and thus detects 1/N times less total interference and background noise than does the diversity-N direct detection system.

We will use binary modulation for the direct detection and coherent detection systems. In the absence of background noise, the best performing binary modulation scheme for coherent detection is Binary Phase Shift Keying (BPSK) [7] and for direct detection is On-Off Keying [7]. However, On-Off Keying requires a time-varying threshold[2]. We will consider Binary Pulse Position Modulation (BPPM) for direct detection due to the simple optimum receiver which compares the detector output in the first and second half symbol intervals.

3.2   Derivation of Diversity Direct Detection Performance

We now derive the outage probability of the diversity direct detection system when the signal propagates through the atmospheric channel and there is no interference. In [43] we derived the outage probability of the diversity direct detection system in the absence of interference without explicitly including an error probability threshold value and where all noise was modeled as Gaussian. Modeling the noise as Gaussian is applicable when the noise is large. In this thesis, we find the outage probability as an explicit function of error probability threshold and use the Poisson detection model, which does not make an assumption of large noise.

---

[2] Note that for very high data rates, it is possible to track the optimum threshold of an OOK system with direction detection accurately.

3.2.1 Single Direct Detection Receiver

Consider first the single receiver direct detection system. Appendix B shows (with plots) that the Poisson detection model [16] for the photodetector is a good approximate model for small background noise (for an average number of detected background noise photons per symbol of less than one). Using the Poisson detection model for the photodetector, conditioned on the power fading factor α_i, we have that

$$p\left(N_0 = n_0 \mid H_0, \alpha_1\right) = \frac{\left(\alpha_1 N_S + \frac{N_n}{2}\right)^{n_0} e^{-\left(\alpha_1 N_S + \frac{N_n}{2}\right)}}{n_0}$$

$$p\left(N_1 = n_1 \mid H_0, \alpha_1\right) = \frac{\left(\frac{N_n}{2}\right)^{n_1} e^{-\frac{N_n}{2}}}{n_1}$$

$$p\left(N_0 = n_0 \mid H_1, \alpha_1\right) = \frac{\left(\frac{N_n}{2}\right)^{n_0} e^{-\frac{N_n}{2}}}{n_0}$$

$$p\left(N_1 = n_1 \mid H_1, \alpha_1\right) = \frac{\left(\alpha_1 N_S + \frac{N_n}{2}\right)^{n_1} e^{-\left(\alpha_1 N_S + \frac{N_n}{2}\right)}}{n_1} \tag{3.2}$$

for $n_0 \geq 0$ and $n_1 \geq 0$ where N_0 and N_1 are the number of photons received in the first and second half symbol intervals, N_S and N_n are the average number of detected signal and noise photons per symbol, and the hypothesis H_0 and H_0 represent symbol values of '0' and '1' being sent by the sender corresponding to transmission occurring in the first or second half of the symbol interval. It can be derived [28] that the Maximum Likelihood detector decides a '0' was sent if $N_0 \geq N_1$ and that a '1' was sent otherwise. If a '0' and a '1' being sent are equally likely, the error probability conditional on the power fading factor α_l is

$$P_{DD,noInterference,N=1}(e \mid \alpha_1) = \frac{1}{2}P(e \mid H_0, \alpha_1) + \frac{1}{2}P(e \mid H_1, \alpha_1)$$

$$= P(e \mid H_0, \alpha_1)$$

$$= P(N_1 > N_0 \mid H_0, \alpha_1)$$

$$\leq \min_{s \geq 0} E\left[e^{s(N_1 - N_0)} \mid H_0, \alpha_1\right]$$

$$= \min_{s \geq 0} E\left[E\left[e^{s(N_1 - N_0)} \mid H_0, \alpha_1, N_0\right] \mid H_0, \alpha_1\right] \qquad (3.3)$$

$$= \min_{s \geq 0} E\left[e^{-sN_0} E\left[e^{sN_1} \mid H_0, \alpha_1\right] \mid H_0, \alpha_1\right]$$

$$= \min_{s \geq 0} \exp\left\{\frac{N_n}{2}\left(e^s - 1\right)\right\} E\left[e^{-sN_0} \mid H_0, \alpha_1\right]$$

$$= \min_{s \geq 0} \exp\left\{\left(\alpha_1 N_s + \frac{N_n}{2}\right)\left(e^{-s} - 1\right) + \frac{N_n}{2}\left(e^s - 1\right)\right\}$$

where the inequality is the Chernoff bound and the two last equalities are true because for a Poisson random variable with rate parameter λ, $E\left[e^{sX}\right] = \exp\{\lambda(e^s - 1)\}$. Optimizing over s (by taking the first derivative of the exp{ } expression in the last line in (3.3) and setting equal to zero) gives

$$e^s = \sqrt{\frac{\alpha_1 N_s + \frac{N_n}{2}}{\frac{N_n}{2}}}. \qquad (3.4)$$

Substituting this into the last line of (3.3) and taking the error probability to be well approximated by the Chernoff bound, the conditional error probability is

$$P_{DD,noInterference,N=1}(e \mid \alpha_1) \cong \exp\left\{-\left(\sqrt{\alpha_1 N_s + \frac{N_n}{2}} - \sqrt{\frac{N_n}{2}}\right)^2\right\} \qquad (3.5)$$

(3.5) is the same as in [8] but where N_n for us denotes the total average received background noise photons in both half symbol periods.

3.2.2 Multiple Direct Detection Receivers

Consider the N–receiver direct detection system, where the receivers are separated by more than one coherence length. The sum of N independent Poisson random variables with rate parameters λ_1, λ_2,...,λ_N is a Poisson random variable with rate parameter $\lambda_1+\lambda_2+...+\lambda_N$. Thus, in deriving the error probability conditioned on the fading for multiple receiver direct detection, we can use the same analysis that we did for the single receiver system, but where we replace $\alpha_1 N_s$ with $\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)N_s$ and $\frac{N_n}{2}$ with $\frac{NN_n}{2}$. The conditional error probability of the diffraction-limited N-receiver system is then

$$P_{DD,noInterference}(e\,|\,\underline{\alpha})=\exp\left\{-\left(\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)N_s+\frac{NN_n}{2}}-\sqrt{\frac{NN_n}{2}}\right)^2\right\} \tag{3.6}$$

where $\underline{\alpha}=(\alpha_1,\alpha_2,...\alpha_N)$. If the error probability threshold is $P_e^{thresh}=e^{-\theta_{thresh}}$, using (3.6), the outage probability is

$$
\begin{aligned}
P_{outage,DD} &= \Pr\!\left(P_{DD,noInterference}(e\,|\,\underline{\alpha})>e^{-\theta_{thresh}}\right)\\[4pt]
&= \Pr\!\left(\exp\left\{-\left(\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)N_S+\frac{NN_n}{2}}-\sqrt{\frac{NN_n}{2}}\right)^2\right\}>e^{-\theta_{thresh}}\right)\\[4pt]
&= \Pr\!\left(\left(\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)N_S+\frac{NN_n}{2}}-\sqrt{\frac{NN_n}{2}}\right)^2<\theta_{thresh}\right)\\[4pt]
&= \Pr\!\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i<\frac{1}{N_S}\left[\theta_{thresh}+\sqrt{2\theta_{thresh}NN_n}\right]\right)
\end{aligned}
\tag{3.7}
$$

38

The expression $\frac{1}{N}\sum_{i=1}^{N}\alpha_i$ is the sum of log-normal random variables. For a moderate number (tens or less) of these log-normal random variables, the sum is better approximated by a log-normal random variable [35,43], than by a Gaussian, as we would do by applying the Central Limit Theorem. This is due to the asymmetry of the log-normal probability distribution which requires a large number of random variables to converge to a Gaussian. We approximate the sum of our log-normal random variables to be log-normal, i.e. we let

$$\frac{1}{N}\sum_{i=1}^{N}\alpha_i = e^{U} \tag{3.8}$$

where U is a Gaussian random variable. In [43], we derived the mean and variance of U to be

$$m_U = -0.5\ln\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right) \text{ and } \sigma_U^2 = \ln\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right) \tag{3.9}$$

respectively. Substituting (3.8) into (3.7), the outage probability becomes

$$
\begin{aligned}
P_{outage,DD,noInterference} &= \Pr\left(e^{U} < \frac{1}{N_S}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]\right) \\
&= \Pr\left(U < \ln\left(\frac{1}{N_S}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]\right)\right) \\
&= Q\left(\frac{m_U - \ln\left(\frac{1}{N_S}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]\right)}{\sigma_U}\right) \\
&\cong \frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]}{N_S}\right)\right]^2\right\}
\end{aligned}
\tag{3.10}
$$

where $Q(x) = \int\limits_{x}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ and we have approximated $Q(x)$ by its upper bound $\frac{1}{2} e^{-x^2/2}$.

3.3 Derivation of Diversity Coherent Detection Performance

In this section, we derive the outage probability of diversity coherent detection when the signal propagates through the atmospheric channel and there is no interference. In doing so, we first describe the received field and spectrum as in [16] but where random fading and phase factors due to propagation through atmospheric turbulence are included.

Heterodyne detection is analyzed first (for single and multiple receiver systems) followed by analysis of homodyne detection.

3.3.1 Single Heterodyne Detection Receiver

In a single heterodyne receiver system, the intermediate frequency (IF) is non-zero and the filter in Figure 3.2 is band-pass. The received signal in the detector plane is given by

$$U_d(t,\vec{r}) = \left(\sqrt{\alpha_1} a_s(t) e^{j(\theta_s(t)+\theta_1(t))} \phi_1(\vec{r}) + b_1(t)\phi_1(\vec{r}) \right) e^{j\omega_0 t} \tag{3.11}$$

where $s(t) = a_s(t) e^{j\theta_s(t)}$ is the signal in the absence of fading, $b_1(t)$ is the complex noise envelope of the background noise of the ith mode, $\phi_1(\vec{r})$ is the Airy pattern of the ith mode on the detector plane. Let us denote the LO field on the detector plane by $U_L(t,\vec{r}) = a_L e^{j(\omega_L t + \theta_L)} \phi_L(\vec{r})$ where $\phi_L(\vec{r})$ is the LO field Airy pattern in the detector

plane. We assume the LO field Airy pattern is exactly matched to the first Airy pattern ($\phi_L = \phi_1$), that it is spatially orthogonal to the other Airy patterns (ϕ_2, ϕ_3,...), and that $\int_{A_d} |\phi_i(\vec{r})|^2 d\vec{r} = A$ (where A_d is the photodetector area). The average count rate $n(t)$ out of the photodetector [16] is given by

$$
\begin{aligned}
n(t) &= \frac{\eta}{h\nu} \int_{A_d} |U_d(t,\vec{r}) + U_L(t,\vec{r})|^2 \, d\vec{r} \\
&= \frac{\eta}{h\nu} \left[\int_{A_d} |U_d(t,\vec{r})|^2 \, d\vec{r} + \int_{A_d} |U_L(t,\vec{r})|^2 \, d\vec{r} + 2\operatorname{Re}\left\{ \int_{A_d} U_d(t,\vec{r})U_L^*(t,\vec{r}) d\vec{r} \right\} \right] \\
&= \frac{\eta}{h\nu} A \left[\left| \sqrt{\alpha_1} a_S(t) e^{j(\theta_s(t)+\theta_1(t))} + b_1(t) \right|^2 + a_L^2 + 2\operatorname{Re}\left\{ \left(\sqrt{\alpha_1} a_s(t) e^{j(\theta_s(t)+\theta_1(t))} + b_1(t) \right) a_L e^{j(\omega_0 - \omega_L t + \theta_{L_1}(t))} \right\} \right] \\
&= \frac{\eta}{h\nu} A \left[\begin{aligned} &\left| \sqrt{\alpha_1} a_s(t) e^{j(\theta_s(t)+\theta_1(t))} + b_1(t) \right|^2 + a_L^2 + 2 a_L \sqrt{\alpha_1} a_s(t) \cos\left((\omega_0 - \omega_L)t + \theta_s(t) + \theta_1(t) - \theta_{L_1}(t) \right) \\ &+ 2 a_L \operatorname{Re}\left\{ b_1(t) e^{j(\omega_0 - \omega_L)t - \theta_L} \right\} \end{aligned} \right]
\end{aligned}
$$

(3.12)

where η is the photodetector's quantum efficiency, h is Planck's constant and ν is the frequency of the optical wave. The random fading and phase factors α_1 and θ_1 are included in (3.12). Note that because the LO field is matched to one Airy pattern, only one mode of background noise is actually detected (which is desired). The time average of $n(t)$ is

$$
\begin{aligned}
\bar{n} &= \frac{\eta}{h\nu} A \left[\alpha_1 \overline{|s(t)|^2} + \overline{|b_1(t)|^2} + a_L^2 \right] \\
&= \frac{\eta}{h\nu} \left[\alpha_1 P_S + P_{b0} + P_L \right]
\end{aligned}
$$

(3.13)

where the average signal, background noise, and local oscillator powers are given by $P_s = A\overline{|s(t)|^2}$, $P_{b0} = A\overline{|b_1(t)|^2}$, and $P_L = A a_L^2$ respectively. Let

41

$$y(t) = a_s(t)\cos\left((\omega_0 - \omega_L)t + \theta_s(t) + \theta_1(t) - \theta_{L_1}(t)\right). \tag{3.14}$$

The average intensity of $y(t)$, when $a_s(t) = a_s$ is $I_y = a_s^2 / 2$. The two-sided background noise spectrum of $\text{Re}\left\{b_i(t)e^{j(\omega_0 - \omega_L)t - \theta_L}\sqrt{A}\right\}$ is $N_{0b}/4$ (the b in the subscript represents background noise). Ignoring dark current and thermal noise (valid if the local oscillator power is large enough), for the average photon count rate of (3.12), the spectral density of the detector output current [16] is given by

$$\begin{aligned}
S_c(\omega) &= q^2 S_n(\omega) \\
&= \alpha_1 \left(\frac{2q\eta}{h\nu} A a_L\right)^2 S_y(\omega) + \left(\frac{2q\eta}{h\nu} a_L\right)^2 A \frac{N_{0b}}{4} + \frac{q^2\eta}{h\nu}\left[\alpha_1 P_s + P_{b0} + P_L\right] + (\text{terms at 0 frequency}) \\
&= \alpha_1 \left(\frac{2q\eta}{h\nu}\right)^2 P_L A S_y(\omega) + \left(\frac{q\eta}{h\nu}\right)^2 P_L N_{0b} + \frac{q^2\eta}{h\nu}\left[\alpha_1 P_s + P_{b0} + P_L\right] + (\text{terms at 0 frequency})
\end{aligned}$$

$$(3.15)$$

The first three terms in the last equality correspond to the signal-LO cross signal, background-LO cross noise, and shot noise. The background-only noise term and signal-only term are omitted since they are negligible compared to the signal-LO and background-LO terms. Note that the shot noise is dominated by the local oscillator contribution (i.e. $P_L \gg \alpha_1 P_s + P_{b0}$) and the zero frequency terms are filtered out. Thus, the spectral density is

$$S_c(\omega) \cong \alpha_1 \left(\frac{2q\eta}{h\nu}\right)^2 P_L A S_y(\omega) + \left(\frac{q\eta}{h\nu}\right)^2 P_L N_{0b} + \frac{q^2\eta}{h\nu} P_L \tag{3.16}$$

From (3.16), the squared output signal strength conditioned on the fading is given by

$$P_{s,output} = \alpha_1 \left(\frac{2q\eta}{h\nu} \right)^2 P_L A I_y$$

$$= \alpha_1 \left(\frac{2q\eta}{h\nu} \right)^2 \frac{P_L P_s}{2} \tag{3.17}$$

$$= 2\alpha_1 \left(\frac{q\eta}{h\nu} \right)^2 P_L P_s$$

We now find the probability of symbol error when BPSK is used. Using detection theory in Additive White Gaussian Noise (AWGN), the optimal detector uses a matched filter [38]. The corresponding probability of symbol error for BPSK is

$$P(e) = Q\left(\sqrt{\frac{2E}{N_0}} \right) \cong \frac{1}{2} \exp\left(-\frac{E}{N_0} \right) \tag{3.18}$$

where the signal has energy E and the AWGN has variance $N_0/2$. In our case,

$$E = P_{s,output} T = 2\alpha_1 \left(\frac{q\eta}{h\nu} \right)^2 P_L P_s T \quad \text{and} \tag{3.19}$$

$$\frac{N_0}{2} = \frac{q^2\eta}{h\nu} P_L + \left(\frac{q\eta}{h\nu} \right)^2 P_L N_{0b} \tag{3.20}$$

where T is the symbol time and (3.20) reflects the noise terms from (3.16). Thus, given the fading factor α_1, the error probability is

$$P_{Het,noInterference,N=1}(e \mid \alpha_1) \cong \frac{1}{2}\exp\left(-\frac{\alpha_1\left(\frac{q\eta}{h\nu}\right)^2 P_L P_s T}{\frac{q^2\eta}{h\nu}P_L + \left(\frac{q\eta}{h\nu}\right)^2 P_L N_{0b}}\right)$$

$$= \frac{1}{2}\exp\left(-\frac{\alpha_1\frac{\eta}{h\nu}P_s T}{1 + \frac{\eta}{h\nu}N_{0b}}\right) \qquad (3.21)$$

$$= \frac{1}{2}\exp\left(-\frac{\alpha_1 N_s}{1 + N_n}\right)$$

The last equality is true because the average number of detected signal and background noise photons per symbol is given by

$$N_s = \frac{\eta}{h\nu}P_s T \qquad (3.22)$$

and

$$N_n = \frac{\eta}{h\nu}P_{b0}T = \frac{\eta}{h\nu}N_{0b}B_0 T \qquad (3.23)$$

respectively, where B_0 is the receiver's optical bandwidth and where we assume $B_0 T = 1$ so there is only one temporal mode.

Note that the error probability for heterodyne detection is the same whether diffraction-limited or fixed field-of-view receivers are used. This is because the local oscillator is matched to just one spatial mode and thus the receiver ignores all other spatial noise modes.

44

3.3.2 Multiple Heterodyne Detection Receivers

Consider the N-receiver heterodyne system where the receivers are separated by more than an intensity coherence length. Recall that the average received signal power in each of N receivers is 1/N times the received signal power in the single receiver system and that the ith LO amplitude is weighted by factor γ_i. The ith receiver's average photodetector count rate is the same as the single receiver system's average count rate (3.12) but where we replace $a_S(t)$ with $\frac{a_S(t)}{\sqrt{N}}$, a_L with $\gamma_i a_L$, and where α_1, θ_1 and θ_{L_1} are replaced with α_i, θ_i and θ_{L_i} (to represent the ith receiver in the N-receiver system rather than the one receiver in the single receiver system). Thus, the ith receiver's average photodetector count rate is

$$
\begin{aligned}
n_i(t) &= \frac{\eta}{h\nu} \int_{A_d} \left| f_d(t,\vec{r}) + f_L(t,\vec{r}) \right|^2 d\vec{r} \\
&= \frac{\eta}{h\nu} \int_{A_d} \left| f_d(t,\vec{r}) \right|^2 d\vec{r} + \frac{q\eta}{h\nu} \int_{A_d} \left| f_L(t,\vec{r}) \right|^2 d\vec{r} + \frac{2q\eta}{h\nu} \mathrm{Re}\left\{ \int_{A_d} f_d(t,\vec{r}) f_L^*(t,\vec{r}) d\vec{r} \right\} \\
&= \frac{\eta}{h\nu} A \left[\begin{array}{l} \left| \frac{\sqrt{\alpha_i}\, a_s(t)}{\sqrt{N}} e^{j(\theta_s(t)+\theta_i(t))} + b_1(t) \right|^2 + \gamma_i^2 a_L^2 + \frac{2\gamma_i \sqrt{\alpha_i}\, a_L a_s(t)}{\sqrt{N}} \cos\left((\omega_0 - \omega_L)t + \theta_s(t) + \theta_i(t) - \theta_{L_i}(t)\right) \\ + 2\gamma_i a_L\, \mathrm{Re}\left\{ b_1(t) e^{j(\omega_0 - \omega_L)t - \theta_{L_i}} \right\} \end{array} \right]
\end{aligned}
$$

$$(3.24)$$

which includes random fading and phase factors α_i and θ_i. The total average count rate is the sum of the count rates of each of the photodetectors, namely

$$n_{tot}(t) = \frac{\eta}{h\nu} A \sum_{i=1}^{N} \left[\begin{array}{l} \left| \frac{\sqrt{\alpha_i} a_s(t)}{\sqrt{N}} e^{j(\theta_s(t)+\theta_i(t))} + b_1(t) \right|^2 + \gamma_i^2 a_L^2 + \frac{2\gamma_i \sqrt{\alpha_i} a_L a_s(t)}{\sqrt{N}} \cos\left((\omega_0 - \omega_L)t + \theta_s(t) + \theta_i(t) - \theta_{L_i}(t)\right) \\ + 2\gamma_i a_L \operatorname{Re}\left\{ b_1(t) e^{j(\omega_0-\omega_L)t - \theta_{L_i}} \right\} \end{array} \right]$$

$$= \frac{\eta}{h\nu} A \left[\begin{array}{l} \sum_{i=1}^{N} \left| \frac{\sqrt{\alpha_i} a_s(t)}{\sqrt{N}} e^{j(\theta_s(t)+\theta_i(t))} + b_1(t) \right|^2 + \left(\sum_{i=1}^{N} \gamma_i^2\right) a_L^2 + \left(\sum_{i=1}^{N} \gamma_i \sqrt{\alpha_i}\right) \frac{2 a_L a_s(t)}{\sqrt{N}} \cos\left((\omega_0 - \omega_L)t + \theta_s(t)\right) \\ + \sum_{i=1}^{N} 2\gamma_i a_L \operatorname{Re}\left\{ b_1(t) e^{j(\omega_0-\omega_L)t - \theta_{L_i}(t)} \right\} \end{array} \right]$$

$$= \frac{\eta}{h\nu} A \left[\begin{array}{l} \sum_{i=1}^{N} \left| \frac{\sqrt{\alpha_i} a_s(t)}{\sqrt{N}} e^{j(\theta_s(t)+\theta_i(t))} + b_1(t) \right|^2 + a_L^2 + \left(\sum_{i=1}^{N} \gamma_i \sqrt{\alpha_i}\right) \frac{2 a_L a_s(t)}{\sqrt{N}} \cos\left((\omega_0 - \omega_L)t + \theta_s(t)\right) \\ + \sum_{i=1}^{N} 2\gamma_i a_L \operatorname{Re}\left\{ b_1(t) e^{j(\omega_0-\omega_L)t - \theta_{L_i}(t)} \right\} \end{array} \right]$$

(3.25)

Similar to how the spectral density (3.15) was found from the count rate in (3.12) in the single receiver system, the total spectral density for the output average photon count rate of (3.25) given the fading factors α_i, $1 \le i \le N$, is

$$S_c(\omega) = \frac{\left(\sum_{i=1}^{N} \gamma_i \sqrt{\alpha_i}\right)^2}{N} \left(\frac{2q\eta}{h\nu} A a_L\right)^2 S_y(\omega) + \left(\frac{2q\eta}{h\nu} a_L\right)^2 A \frac{N_{0b}}{4}$$
$$+ \frac{q^2 \eta}{h\nu} \left[\frac{\left(\sum_{i=1}^{N} \gamma_i \sqrt{\alpha_i}\right)^2}{N} P_s + N P_{b0} + P_L \right] + (\text{terms at 0 frequency})$$

(3.26)

$$= \frac{\left(\sum_{i=1}^{N} \gamma_i \sqrt{\alpha_i}\right)^2}{N} \left(\frac{2q\eta}{h\nu}\right)^2 P_L A S_y(\omega) + \left(\frac{q\eta}{h\nu} a_L\right)^2 A N_{0b}$$
$$+ \frac{q^2 \eta}{h\nu} \left[\frac{\left(\sum_{i=1}^{N} \gamma_i \sqrt{\alpha_i}\right)^2}{N} P_s + N P_{b0} + P_L \right] + (\text{terms at 0 frequency})$$

where the first three terms correspond to the signal-LO cross signal, background-LO cross noise, and shot noise. Again, the background-only noise terms and signal-only terms are omitted since they are negligible compared to the signal-LO and background-LO terms. Compared to the single receiver case, the multiple receiver noise level remains the same for the following reasons:

1) The background-LO cross noise remains the same. Although each local oscillator is matched to one spatial mode so that each receiver sees one diffraction angle worth of background noise, each receiver's noise variance is then reduced by a factor γ_i^2. Since the sum of these N factors is 1, the overall background-LO cross noise level remains the same.

2) The shot noise level is approximately the same. This is true when the LO is large enough that

$$P_s \ll P_L \text{ and } NP_{b0} \ll P_L \tag{3.27}$$

such that the shot noise due to the LO dominates the shot noise due to the signal and background noise.

One may want to apply a similar thought process as that of point one above to the signal and say that in the absence of fading, since the signal power received by each receiver is Ps/N, and since each receiver's power is reduced by a factor γ_i^2 and the sum of these N factors is 1, that the overall signal power is only Ps/N ($1/N$ times less the signal power received by the single heterodyne receiver system). However, this reasoning is not correct because the signal portion of the detector output current is proportional to the square root of the received power by that receiver (not

47

proportional to the received power). In the absence of fading, $\alpha_i = 1$ and the optimal LO weights are $\gamma_i = 1/\sqrt{N}$ (as we will see later). Substituting these values of α_i and γ_i into the first term in (3.26), we see that the total signal power of the multiple heterodyne receiver system is indeed the same as the single heterodyne receiver system.

In Section 3.3.1, we derived the error probability for a single receiver heterodyne detection system that uses Maximum Likelihood detection. Following the same development, but using the spectrum for the diversity-N system (3.26) instead of the spectrum for the single receiver system (3.15), the error probability for the multiple receiver system, given the fading factors $(\alpha_1, \alpha_2, ..., \alpha_N) = \underline{\alpha}$ is

$$P_{e,Het,noInterference}(e \mid \underline{\alpha}) = \frac{1}{2}\exp\left(-\frac{\dfrac{\left(\sum\limits_{i=1}^{N}\gamma_i\sqrt{\alpha_i}\right)^2}{N}\left(\dfrac{q\eta}{h\nu}\right)^2 P_L P_s T}{\dfrac{q^2\eta}{h\nu}P_L + \left(\dfrac{q\eta}{h\nu}\right)^2 P_L N_{0b}}\right)$$

$$= \frac{1}{2}\exp\left(-\frac{\dfrac{\left(\sum\limits_{i=1}^{N}\gamma_i\sqrt{\alpha_i}\right)^2}{N}\dfrac{\eta}{h\nu}P_s T}{1+\dfrac{\eta}{h\nu}N_{0b}}\right)$$

$$= \frac{1}{2}\exp\left(-\frac{\dfrac{\left(\sum\limits_{i=1}^{N}\gamma_i\sqrt{\alpha_i}\right)^2}{N}N_s}{1+N_n}\right)$$

(3.28)

where again, the last equality is true because the average number of detected signal and background noise photons per symbol are given by (3.22) and (3.23).

The value of the ith local oscillator's scaling factor γ_i that minimizes the error probability is

$$\gamma_i = \frac{\sqrt{\alpha_i}}{\sqrt{\sum_{k=1}^{N} \alpha_k}} \tag{3.29}$$

(see Appendix C for derivation). When there is no fading, $\alpha_i = 1$ for all i and $\gamma_i = \frac{1}{\sqrt{N}}$ (found by substituting $\alpha_i = 1$ into (3.29)). In other words, when there is no fading, the error probability is minimized when all the local oscillators are weighted equally. Substituting the optimal weighting (3.29) of the N local oscillator powers, into the spectral density (3.26) and error probability (3.28), the spectral density becomes

$$S_c(\omega) = \left(\frac{1}{N} \sum_{i=1}^{N} \alpha_i \right) \left(\frac{2q\eta}{h\nu} A a_L \right)^2 S_v(\omega) + \left(\frac{2q\eta}{h\nu} a_L \right)^2 A \frac{N_{0b}}{4}$$
$$+ \frac{q^2\eta}{h\nu} \left[\left(\frac{1}{N} \sum_{i=1}^{N} \alpha_i \right) P_s + N P_{b0} + P_L \right] + (\text{terms at 0 frequency}) \tag{3.30}$$

and the error probability becomes

$$P_{Het,noInterference}(e \mid \underline{\alpha}) = \frac{1}{2} \exp \left(-\frac{\left(\frac{1}{N} \sum_{i=1}^{N} \alpha_i \right) N_s}{1 + N_n} \right) \tag{3.31}$$

Just as in the single heterodyne receiver case, this error probability is the same whether diffraction-limited or fixed field-of-view receivers are used because the local oscillators are each matched to just one spatial mode. Note also that the total amount of background noise detected is the same regardless of the amount of diversity.

If the error probability threshold is $P_e^{thresh} = e^{-\theta_{thresh}}$, using (3.31), the outage probability for the heterodyne detection diversity system is given by

$$
\begin{aligned}
P_{outage,Het,noInterference} &= \Pr\left(P_{Het,noInterference}\left(e \mid \underline{\alpha}\right) > e^{-\theta_{thresh}}\right) \\
&= \Pr\left(\frac{1}{2}\exp\left(-\frac{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)N_s}{1+N_n}\right) > e^{-\theta_{thresh}}\right) \\
&= \Pr\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i < \frac{\left(1+N_n\right)\left(-\ln 2 + \theta_{thresh}\right)}{N_s}\right)
\end{aligned}
\tag{3.32}
$$

Again, approximating $\frac{1}{N}\sum_{i=1}^{N}\alpha_i$ to be log-normal, and substituting $\frac{1}{N}\sum_{i=1}^{N}\alpha_i = e^{U}$ into (3.32), the outage probability becomes

$$
\begin{aligned}
P_{outage,Het,noInterference} &= \Pr\left(e^{U} < \frac{\left(1+N_n\right)\left(-\ln 2 + \theta_{thresh}\right)}{N_s}\right) \\
&= \Pr\left(U < \ln\left(\frac{\left(1+N_n\right)\left(-\ln 2 + \theta_{thresh}\right)}{N_s}\right)\right) \\
&= Q\left(\frac{m_U - \ln\left(\frac{\left(1+N_n\right)\left(-\ln 2 + \theta_{thresh}\right)}{N_s}\right)}{\sigma_U}\right)
\end{aligned}
\tag{3.33}
$$

3.3.3 Multiple Homodyne Detection Receivers

Consider an N-receiver homodyne receiver system. The local oscillator frequency is ω_0 (i.e. $\omega_{IF}=0$) and the filter is low-pass rather than band-pass.

The analysis for homodyne detection is the same as for heterodyne detection except the IF is 0. As a result, the signal power after the filter is twice that in heterodyne detection (seen by setting $\omega_L=\omega_0$ in (3.14)), while the received noise power is the same. Equivalently, if we scale the signal and noise power after the filter by a factor of 1/2, the signal power is the same as for heterodyne, but the noise power half that of heterodyne. All the results of heterodyne detection remain the same for homodyne detection except the noise power is halved. Thus, from (3.31), the conditional error probability of diversity homodyne detection is

$$P_{Homo,noInterference}\left(e \mid \underline{\alpha}\right)=\frac{1}{2}\exp\left[-\frac{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)2N_s}{1+N_n}\right] \qquad (3.34)$$

and from (3.33) the outage probability is

$$P_{outage,\,Homo,noInterference}=Q\left(\frac{m_U-\ln\left(\frac{\left(1+N_n\right)\left(-\ln 2+\theta_{thresh}\right)}{2N_s}\right)}{\sigma_U}\right)$$

$$\approx\frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U-\ln\left(\frac{\left(1+N_n\right)\left(-\ln 2+\theta_{thresh}\right)}{2N_s}\right)\right]^2\right\} \qquad (3.35)$$

Since homodyne detection performs better than heterodyne detection, we will assume that the coherent system uses homodyne detection in the remainder of this thesis.

3.4   Performance of Diversity Direct and Homodyne Detection

See Figure 3.3 for a plot of the error probability of diversity direct detection and diversity homodyne detection in the absence of any interference and fading. The curves correspond to (3.6) and (3.34) with the fading factors α_i set equal to 1. In homodyne detection, the error probability curve is the same regardless of how much diversity is used. This is due to the same amount of background noise detected being regardless of diversity value. In direct detection, however, the curves shift right as diversity increases due to the extra background noise detected.
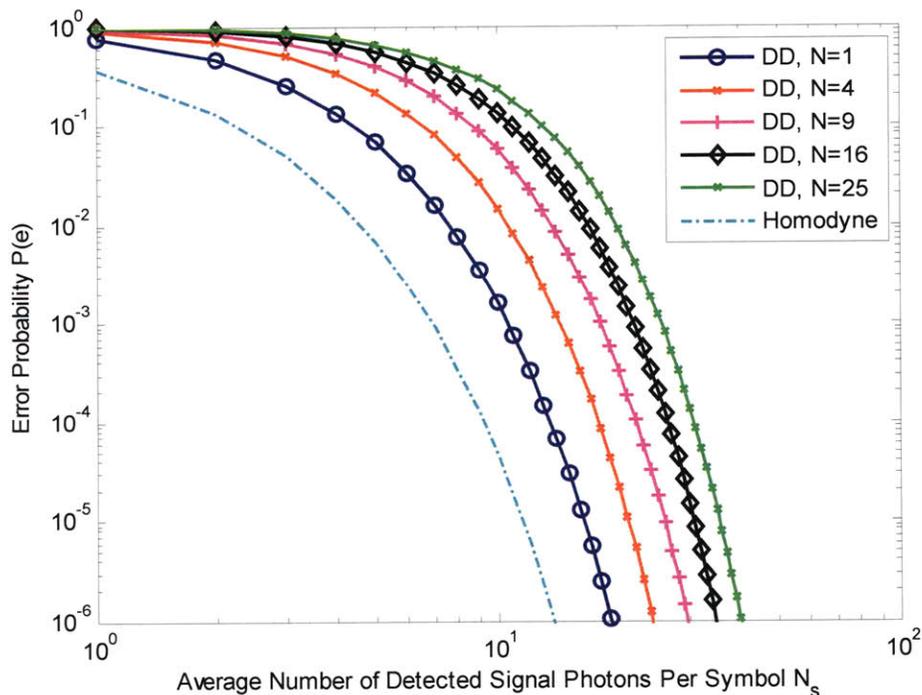
Figure 3.3: Error probability in the absence of interference and fading when $N_n=1$

52

See Figures 3.4 and 3.5 for plots of the outage probability of diversity direct detection and diversity homodyne detection when P_e^{thresh} =0.1 and P_e^{thresh} =10^{-4} respectively. The curves correspond to (3.10) and (3.35). Consider an outage probability of 10^{-2}. Adding diversity in direct detection initially helps to reduce the amount of signal power needed to achieve this outage probability, but then begins to increase the power needed as diversity is increased further. This is because the diversity improves the fading statistics but also adds more background noise. For direct detection, there is a threshold diversity value at which the improvement in outage statistics with diversity begins to be overpowered by the added background noise.
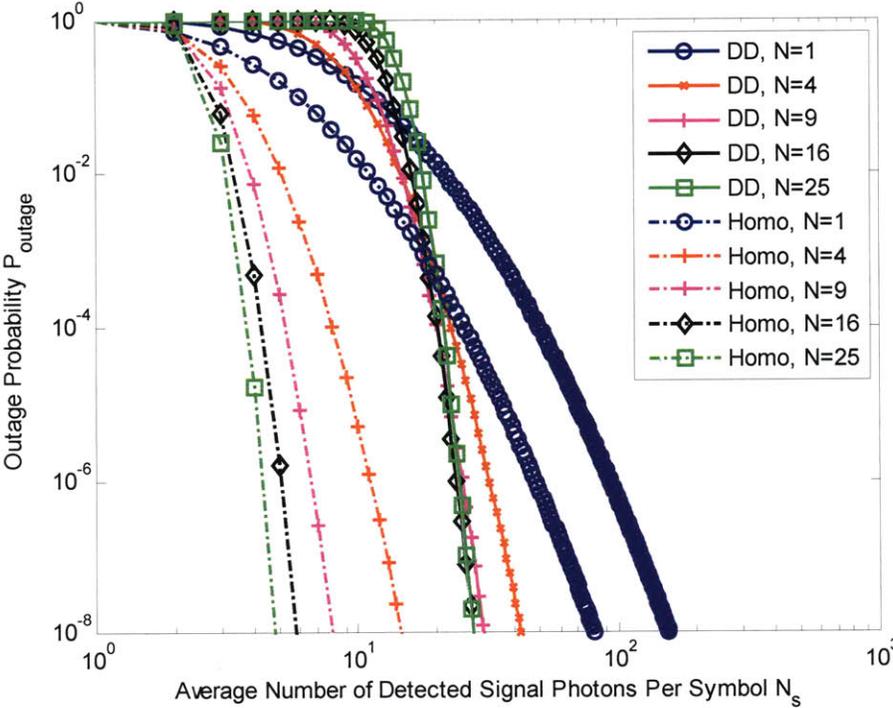
Figure 3.4: Outage probability in the absence of interference when P_e^{thresh} =0.1, σ_χ =0.3, N_n=1
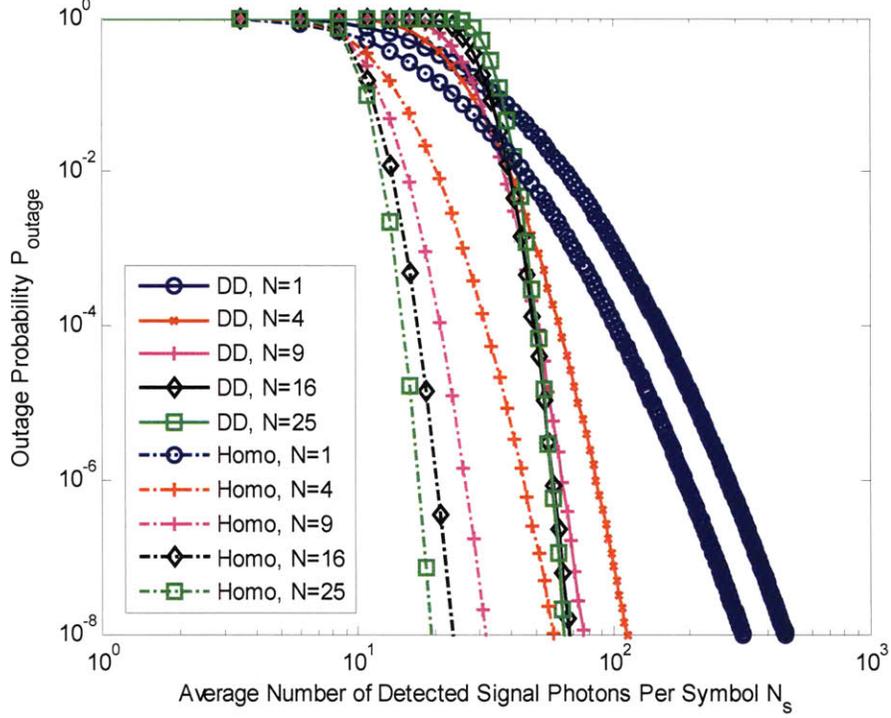
Figure 3.5: Outage probability in the absence of interference when $P_e^{thresh} = 10^{-4}$, $\sigma_\chi = 0.3$, $N_n = 1$

We derive in Appendix D, the diversity value for direct detection in the absence of interference that minimizes the required sender signal level for a desired outage probability. This optimal direct detection diversity value, which we denote by N_{opt}, is the root (that is ≥ 1) of the equation

$$0 = a_3 N_{opt}^{3/2} + a_2 N_{opt} + a_1 N_{opt}^{1/2} + a_0 \qquad (3.36)$$

where

$$
\begin{aligned}
a_3 &= \sqrt{2N_n} \\
a_2 &= -2\sqrt{-N_n\left(e^{4\sigma_\chi^2} - 1\right)\ln\left(2P_{outage}\right)} \\
a_1 &= -\sqrt{-\theta^{thresh}\left(e^{4\sigma_\chi^2} - 1\right)\ln\left(2P_{outage}\right)} \\
a_0 &= -\sqrt{\frac{\theta^{thresh}}{2}\left(e^{4\sigma_\chi^2} - 1\right)}
\end{aligned}
\qquad (3.37)
$$

and where we assumed $\dfrac{e^{4\sigma_\chi^2}-1}{N} \ll 1$ (reasonable for mild turbulence). In cases where N_{opt} is not an integer, one should check if rounding up or down yields the best diversity value.

As seen in Figures 3.4 and 3.5, for homodyne detection, increasing diversity always improves the outage probability. This is because diversity improves the fading statistics while the amount of background noise stays the same. In fact, for large diversity, the improvement in outage probability with diversity goes as

$$P_{outage,\,Homo,\,noInterference} \sim \frac{1}{2}\exp\{-cN\} \qquad (3.38)$$

$$\text{where} \quad c = \frac{\left[\ln\left(\dfrac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_S}\right)\right]^2}{2\left(e^{4\sigma_\chi^2}-1\right)}$$

(found in Appendix E using (3.35) and (3.9)). We can also see from Figures 3.4 and 3.5 that for any given diversity N, homodyne detection has lower outage probability than direct detection. Moreover, the difference increases with increasing diversity.

3.5   Effect of Diversity on Average Outage Length

The motivation for using diversity is that it improves fading statistics. Intuitively, if we average many diversity branches, it is unlikely that the sum is faded as often as a single branch. But how much statistical improvement does diversity provide? It would be beneficial to understand how the average outage length improves with

diversity. In Appendix F, we derive, using Gaussian level crossing theory, the expected outage length for the diversity direct detection system and for the diversity homodyne detection system. We now state the results of Appendix F and then provide plots and discussion of the expected outage length. Approximating the sum of N log-normal random variables as log-normal, i.e. letting

$$e^{2W} = \frac{1}{N}\left(e^{2\chi_1} + e^{2\chi_2} + ... e^{2\chi_N}\right),$$

$$(3.39)$$

the expected outage length is given by

$$E[outage \ length] = \left(P_{outage}\pi\right)\sqrt{\frac{\lambda_0^W}{\lambda_2^W}}\exp\left\{\frac{(\zeta - E[W])^2}{2\lambda_0^W}\right\}.$$

$$(3.40)$$

where we now summarize each variable (detailed derivations are in Appendix F). P_{outage} is the outage probability and is given by (3.10) for direct detection and (3.35) for homodyne detection. $E[W]$ is the expected value of W and is given by

$$E[W] = -\frac{1}{4}\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right).$$

$$(3.41)$$

ζ is the level that W goes below to be considered in an outage and is given by

$$\zeta = \frac{1}{2}\ln\left(\frac{\left(\sqrt{-\ln\left(P_e^{thresh}\right)} + \sqrt{\frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{-m\frac{(1+N_n)}{2}\ln\left(2P_e^{thresh}\right)}\right)$$

$$(3.42)$$

for homodyne detection and

$$\zeta = \frac{1}{2}\ln\left(\frac{1}{m}\right)$$ (3.43)

for direct detection where m is the power link margin the system provides beyond the signal power of a single receiver homodyne system that achieves P_e^{thresh}. λ_0^W and λ_2^W are the 0th and 2nd moment of the spectral density of W and are given by

$$\lambda_0^W = \frac{1}{4}\ln\left(\frac{1}{N}\exp(4\lambda_0^\chi) + \frac{N-1}{N}\right)$$ (3.44)

and

$$\lambda_2^W = \frac{\lambda_2^\chi \exp(4\lambda_0^\chi)}{e^{4\lambda_0^\chi} + N - 1}$$ (3.45)

where

$$\lambda_0^\chi = 2(0.15)\langle\chi^2\rangle\left[1 + \frac{3(0.48)}{7}\right] - \frac{6}{5}(1.14)\langle\chi^2\rangle\left[(g)^{-5/3} - 1\right]$$ (3.46)

and

$$\lambda_2^\chi = \frac{v_\perp^2}{2\pi\lambda L}\left\{2(2\pi)^2(0.15)\langle\chi^2\rangle\left[\frac{1}{3} + \frac{3(0.48)}{13}\right] + 6(2\pi)^2(1.14)\langle\chi^2\rangle\left[g^{1/3} - 1\right]\right\}$$ (3.47)

and $v_\perp$ is the transverse wind speed, λ is the wavelength, L is the path distance, and $f_{cutoff} = gf_0$ (where g>>1).

The expected outage length is inversely proportional to the transverse wind speed. This is because none of the variables in the expected outage length expression (3.40) are a function of $v_\perp$ except for λ_2^W which is proportional to $v_\perp^2$.

The expected non-outage length is given by

$$E[non-outage \quad length] = \left(\frac{1 - P_{outage}}{P_{outage}} \right) E[outage \quad length].$$ (3.48)

See Figure 3.6 for a plot of the expected outage length versus link margin for various amounts of diversity and link margin. The curves are calculated using (3.40). The link margin corresponds to additional power over that needed by a one-receiver homodyne detection system that achieves the threshold error probability P_e^{thresh} in absence of fading. As seen in Figure 3.6a, the expected outage length decreases as we increase diversity or link margin. In Figure 3.6b, we see that if we fix the link margin to the reasonable amount of 5 dB, increasing diversity from 1 to a reasonable value of 25 reduces the expected outage length in the homodyne system from 10 milliseconds to 0.5 milliseconds. However, the smaller expected outage length still results in a large number of consecutive bits to be lost when operating at high data rates. Increasing the diversity to larger values reduces the outage length further as seen Figure 3.6c. However, such large diversity values may not be practical or possible due to the limited footprint of the optical signal and the need to put receivers more than a coherence length apart. The exact values in Figure 3.6c may not be accurate since we used the approximation that the sum of the N fading factors is log-normal, and this may not apply for such large N. In Figure 3.6b where the link margin is fixed to be 5 dB, the expected outage length for direct detection decreases gently with diversity compared to homodyne. This is because the error probability curve for direct detection is higher than for homodyne detection and with a limited link margin of 5 dB, the operating point (in the absence of fading) for direct detection is closer to the error probability threshold that for homodyne detection. If the link margin is fixed to a higher value (8 dB for example), direct detection's expected outage length decreases

58

more steeply with diversity (as does homodyne detection's expected outage length). This can be seen in Figure 3.6a and also in Figure 3.6d which shows the expected outage length when 8 dB of link margin is used. Using a higher link margin puts the error probability operating point (in the absence of fading) of direct detection and homodyne detection at a lower value. Thus, this operating point is farther away from the error probability threshold and results in reduced outage lengths.
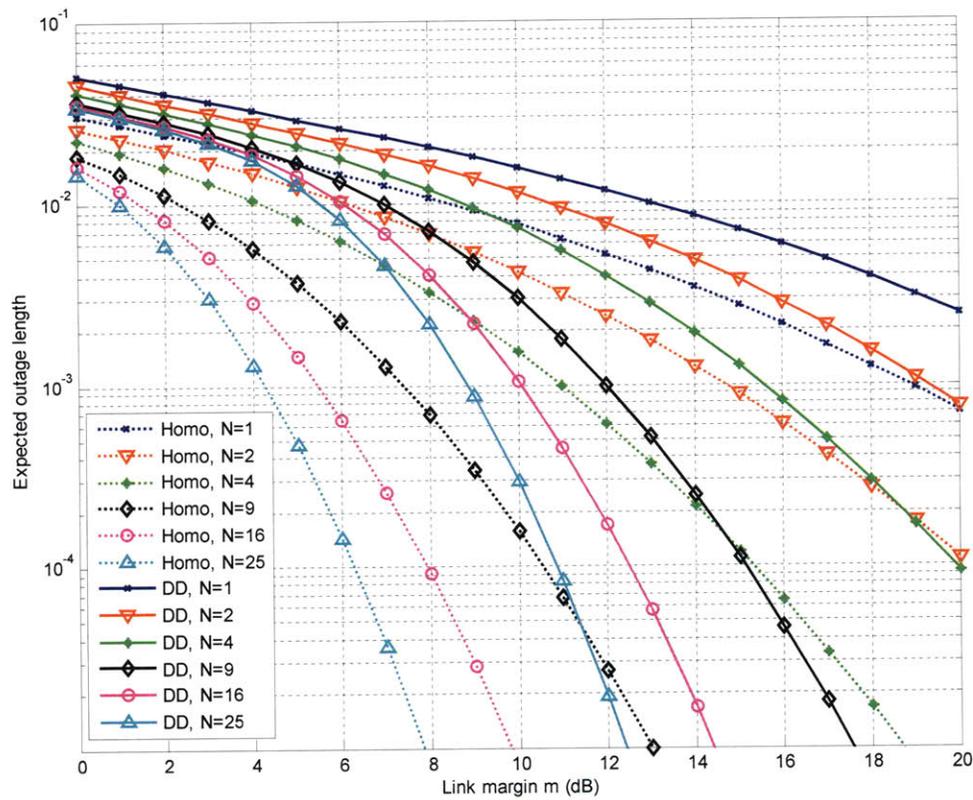
Figure 3.6 (a) link margin on the x-axis
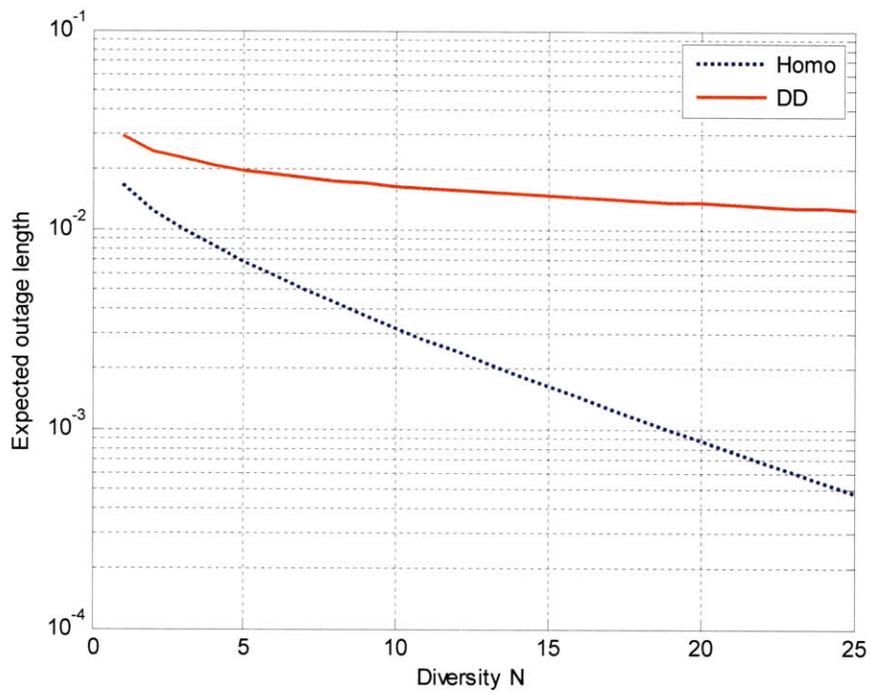
Figure 3.6 (b) fixed link margin of 5 dB
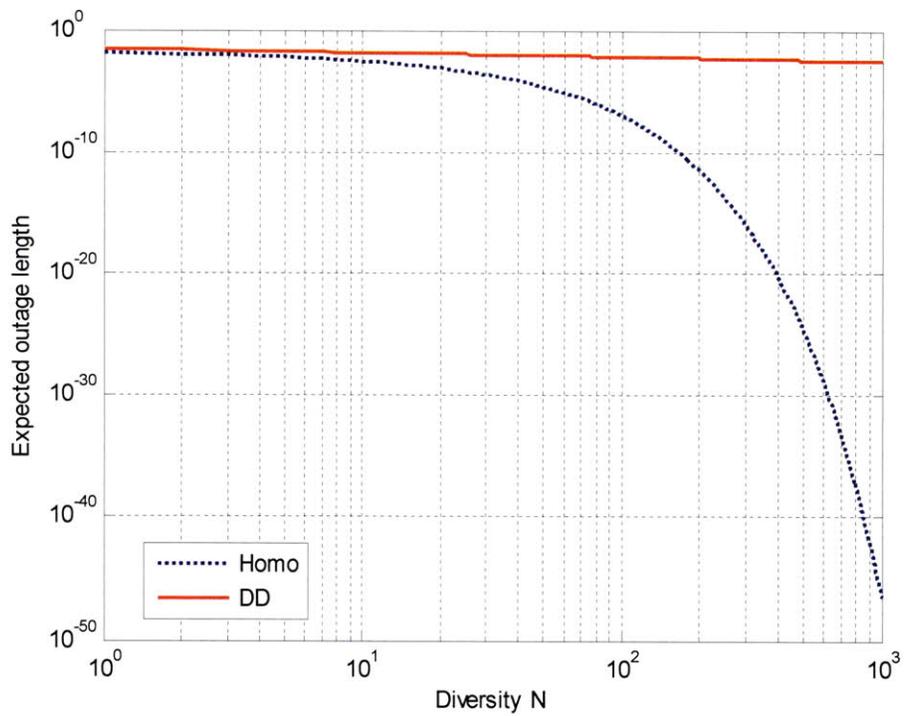
Figure 3.6 (c) fixed link margin of 5 dB
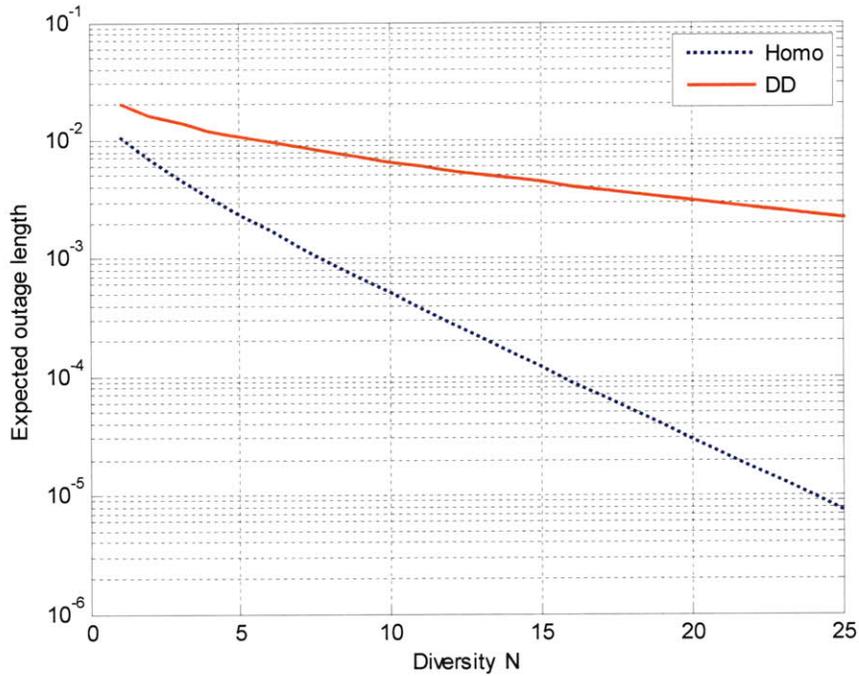
Figure 3.6 (d) fixed link margin of 8 dB

Figure 3.6: Expected outage length a) versus link margin for various amounts of diversity, b), c) versus diversity for fixed link margin of 5 dB, d) versus diversity for fixed link margin of 8 dB. For all three plots, σ_χ^2=0.5, P_e^{thresh}=0.1, N_n=10^{-6}, transverse wind speed is 10 km/hr, path distance=20km.

See Figure 3.7 for plots of the expected outage length versus diversity for large background noise. The direct detection curve actually arcs upward as diversity gets large due to the large amount of background noise that is detected with increasing diversity.

Figure 3.7: Expected outage length versus diversity for fixed link margin of 5 dB when $\sigma_\chi^2=0.5$, $P_e^{thresh}=0.1$, $N_n=1$, transverse wind speed is 10 km/hr, path distance=20km

As we summarized in the mathematics of this section, the expected outage length is inversely proportional to the transverse wind speed. See Figures 3.8a and 3.8b for plots of the expected outage length when the wind speed is increased to 20 km/hr and 40 km/hr respectively (with all other parameters kept constant compared to those in Figure 3.6b). Comparing Figure 3.6b with Figure 3.8, one can see that increasing the wind speed from 10 km/hr to 20 km/hr and 10 km/hr to 40 km/hr reduces the expected outage length by factor 1/2 and 1/4. This is consistent with the expected outage length being proportional to the transverse wind speed.
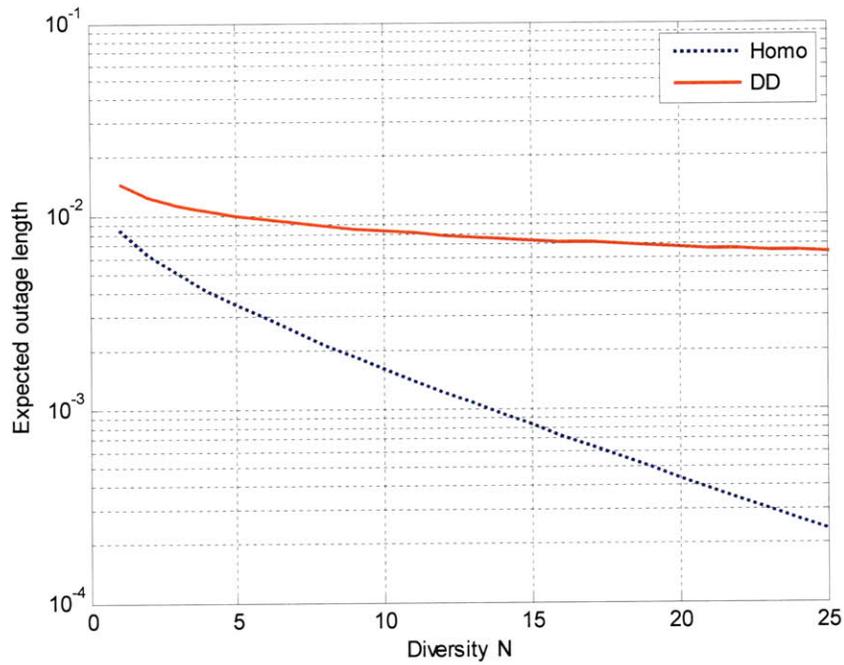
Figure 3.8 (a)
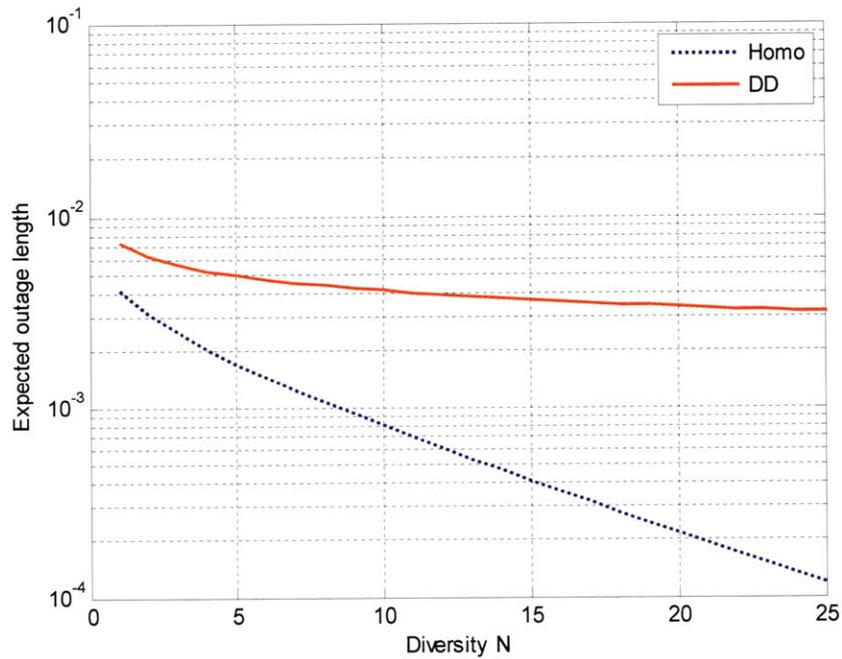
Figure 3.8 (b)

Figure 3.8: Expected outage length versus diversity for fixed link margin of 5 dB when $\sigma_\chi^2=0.5$, $P_e^{thresh}=0.1$, $N_n=10^{-6}$, path distance=20km and a) transverse wind speed is 20 km/hr b) transverse wind speed is 40 km/hr

See Figure 3.9 for a plot of the expected non-outage length versus link margin for various amounts of diversity and link margin (plotted using (3.48)). Recall that 0 dB of link margin corresponds to the power needed by a single homodyne detection receiver to achieve the error probability threshold in the absence of fading. We see that for homodyne detection, the expected non-outage length increases as we increase link margin or diversity. This is because increasing link margin moves the operating point on the error probability curve, in the absence of fading, further away from the error probability threshold and diversity tightens the fluctuation due to fading around that point. For direct detection, if enough link margin is not provided, the expected non-outage length increases with diversity. This because for small link margins the operating point on the error probability curve in the absence of fading is above the error probability threshold, and increasing diversity tightens the fluctuation around this point. In direct detection, provided enough link margin is used, the non-outage length also increases with diversity and link margin. For both homodyne detection and direct detection, significant amounts of link margin and diversity are required to increase the expected outage length substantially (for the large log-amplitude variance of $\sigma_\chi^2 = 0.5$). For example, for homodyne detection, in order to increase the expected outage length by an order of magnitude, 11 dB of link margin and a diversity of 25 are required.
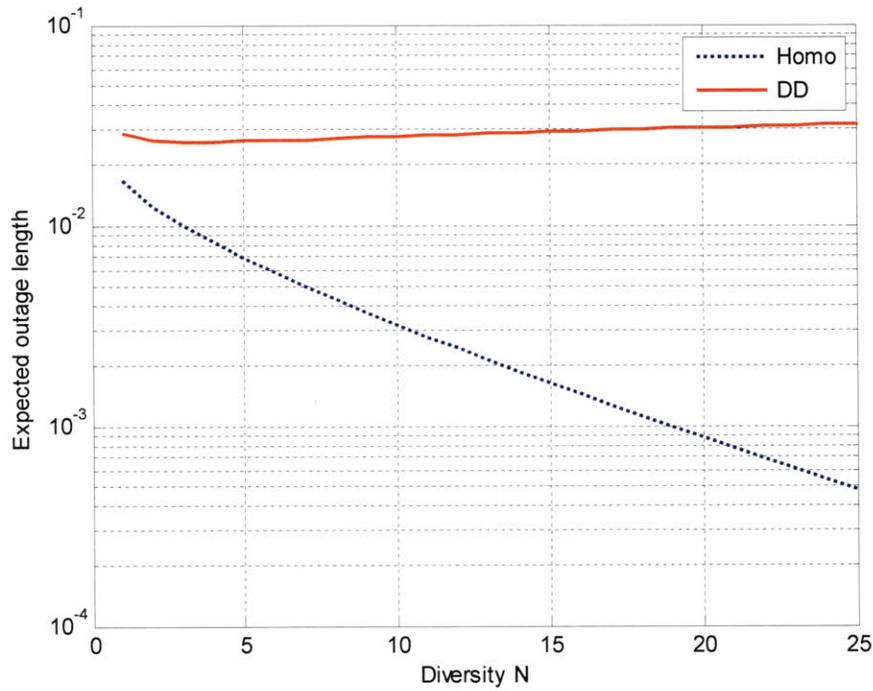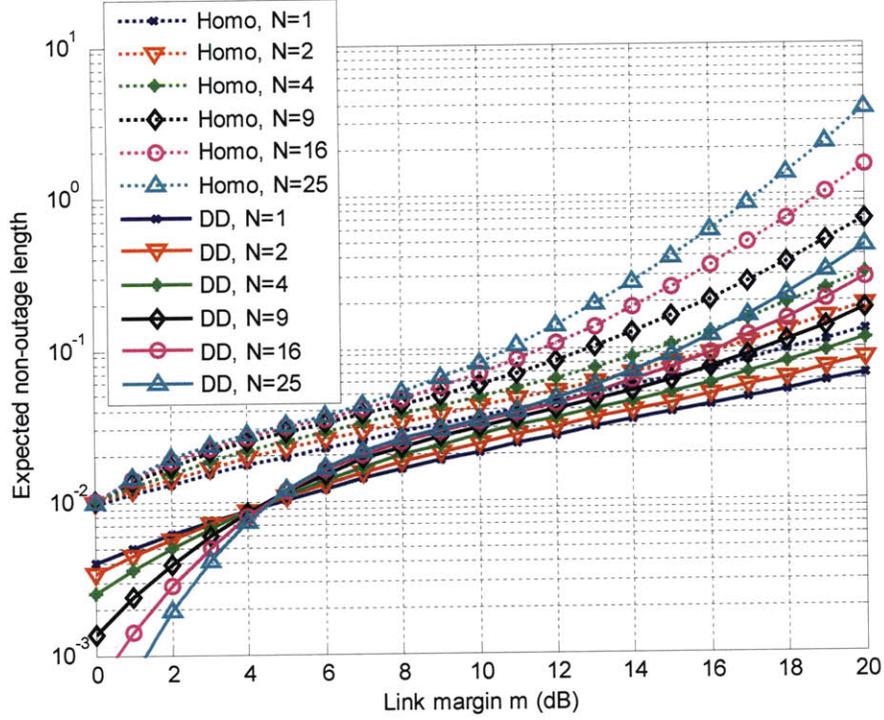
Figure 3.9: Expected outage length versus link margin for various amounts of diversity. $\sigma_\chi^2 = 0.5$, $P_e^{\text{thresh}} = 0.1$, $N_n = 10^{-6}$, transverse wind speed is 10 km/hr, path distance=20km

3.6 Intuitive Understanding of Improvement with Diversity

One way of looking at the improvement in fading statistics with diversity is that diversity has the net effect of reducing the average or *pseudo* log-amplitude variance experienced i.e. diversity provides a net improvement in link conditions. Specifically, for a diversity-N system with a given amount of turbulence in the atmospheric path to each receiver (represented by log-amplitude variance σ_χ^2), the same outage probability is achieved by a single receiver system with log-amplitude variance $\sigma_{\chi,pseudo}^2$ where $\sigma_{\chi,pseudo}^2 < \sigma_\chi^2$. We now find, for coherent detection and direct detection, the value of $\sigma_{\chi,pseudo}^2$ as a function of the amount of diversity N and the log-amplitude variance σ_χ^2 seen by each of the N diversity receivers.

Earlier in this chapter, we found the outage probability of a diversity-N homodyne detection system to be given by (3.35), namely

$$P_{outage, Homo, noInterference} \cong \frac{1}{2} \exp \left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln \left(\frac{(1 + N_n)(-\ln 2 + \theta_{thresh})}{2N_s} \right) \right]^2 \right\}$$

The outage probability of a diversity-N homodyne system (where each receiver sees a log-amplitude variance of σ_χ^2) equals that of a system with no diversity (where the receiver sees a log-amplitude variance of $\sigma_{\chi, pseudo}^2$) if

$$\frac{1}{2\sigma_U^2} \left[m_U - \ln \left(\frac{(1 + N_n)(-\ln 2 + \theta_{thresh})}{2N_s} \right) \right]^2 = \frac{1}{2(4\sigma_{\chi, pseudo, Homo}^2)} \left[m_{2\chi pseudo} - \ln \left(\frac{(1 + N_n)(-\ln 2 + \theta_{thresh})}{2N_s} \right) \right]^2$$

$$(3.49)$$

The left-hand side of this equation is the outage probability exponent of a diversity-N homodyne detection system and the right-hand side is the outage probability exponent of a "pseudo" single receiver system. As we mentioned in Section 2.1, due to conservation of energy, the power fading factor $e^{2\chi}$ has $m_\chi = -\sigma_\chi^2$. Thus, $m_{2\chi} = -2\sigma_\chi^2$. Substituting $m_{2\chi pseudo} = -2\sigma_{\chi, pseudo}^2$ into (3.49) gives

$$\frac{1}{\sigma_U^2} \left[m_U - \ln \left(\frac{(1 + N_n)(-\ln 2 + \theta_{thresh})}{2N_s} \right) \right]^2 = \frac{1}{(4\sigma_{\chi, pseudo}^2)} \left[-2\sigma_{\chi, pseudo, Homo}^2 - \ln \left(\frac{(1 + N_n)(-\ln 2 + \theta_{thresh})}{2N_s} \right) \right]^2$$

$$(3.50)$$

Isolating for $\sigma_{\chi, pseudo}^2$ gives

$$\sigma^2_{\chi,pseudo,Homo} = -\frac{A}{2} - \frac{1}{2}\sqrt{A^2 - \left(\ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right)^2} \qquad (3.51)$$

where $A = \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right) + \frac{1}{\sigma^2_U}\left[m_U - \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right]^2$.

For a diversity-N direct detection system, the outage probability is given by (3.10), namely

$$P_{outage,DD,noInterference} \cong \frac{1}{2}\exp\left\{-\frac{1}{2\sigma^2_U}\left[m_U - \ln\left(\frac{\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]}{N_s}\right)\right]^2\right\}$$

Similar to how we found $\sigma^2_{\chi,pseudo}$ above for homodyne detection, we find $\sigma^2_{\chi,pseudo}$ for direct detection to be

$$\sigma^2_{\chi,pseudo,DD} = -\frac{B}{2} - \frac{1}{2}\sqrt{B^2 - \left(\ln\left(\frac{\theta_{thresh} + \sqrt{2\theta_{thresh}N_n}}{N_s}\right)\right)^2} \qquad (3.52)$$

where $B = \ln\left(\frac{\theta_{thresh} + \sqrt{2\theta_{thresh}N_n}}{N_s}\right) + \frac{1}{\sigma^2_U}\left[m_U - \ln\left(\frac{\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}}{N_s}\right)\right]^2$.

We show in Figure 3.10 the pseudo log-amplitude variance for homodyne and direct detection for σ_χ^2 values of 0.5, 0.3 and 0.1. Diversity improves the net amount of turbulence and homodyne detection is more effective at reducing it than direct detection is. This is due to the extra background noise detected as diversity is increased.
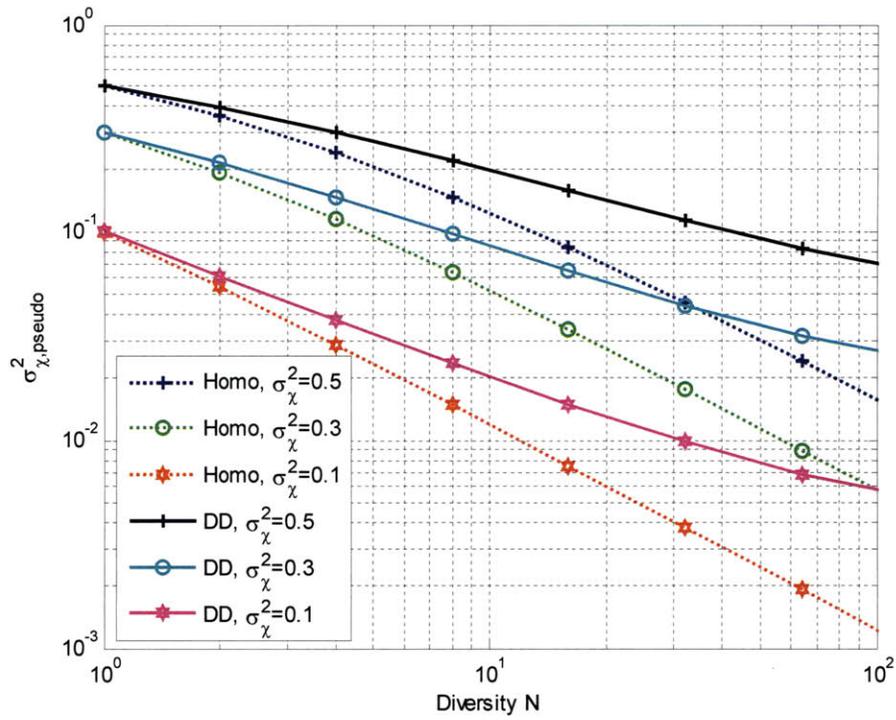
Figure 3.10: The pseudo log-amplitude variance of a diversity-N homodyne and diversity direct detection system where $N_n=1$, $N_S=100$, $P_e^{thresh}=0.1$

Chapter 4

Diversity Direct Detection and Diversity Coherent Detection in Presence of Interference

In this chapter, we consider various types of interference and derive the performance of the diversity direct detection and the diversity coherent detection systems in the presence of each of these interference types (we published much of these results in [31]-[33]). Specifically, for each interference type, we derive the interference duty cycle that causes the highest error probability in the absence of fading, the interference duty cycle that causes the highest outage probability in the presence of fading, and the corresponding error and outage probabilities when the interference uses the worst case duty cycle. We also quantify the amount of interference that the diversity system can tolerate and discuss a sensible way to choose the amount of diversity and link margin.

4.1 Interference Types

The interference we consider in this thesis arrives off-axis[3] from the receiver and is scattered by the receiving optical equipment so that the interference impinging on the receiving telescope is multi-spatial-mode. We assume the interference is idealized and worst-case in that it has the worst possible setting for the intended user's communication system (including the user's modulation scheme, transmit power, receiver type, amount of receiver diversity, and receiver location). Worst-case interference is useful to analyze as it provides an upper bound on system performance. We consider five types of interference in the diversity direct detection system, and two types of interference in the diversity coherent detection system. These interference types are described in the next section and are representative of those that might cause the maximum performance degradations. Some of the interference types are symbol synchronized with the communication signal. This assumption is not unreasonable because the interference could sweep its symbol boundary slowly over time compared with the sender's symbol time to eventually be aligned with the sender's symbol boundary. Some of the interference types cancel the communication signal. This is not unreasonable because the interference could sweep its phase gradually over 2π phase change so that the mode in the same spatial mode as the communication signal eventually cancels the communication signal. This would dislodge the communication system synchronization and the communication system would have to re-acquire time synchronization and likely also spatial tracking.

For all the interference types, we assume that the interference is average power constrained. We allow the interference to have a partial duty cycle, where duty cycle

---

[3] On-axis interference which is not the focus of this thesis can be much stronger and must be treated separately using other techniques such as frequency hopping. The performance in the presence of on-axis interference is expected to be very poor and will not be of interest as an operating scenario.

is the fraction of symbols for which the interference is on. The duty cycle is denoted as β, where $0 < \beta \leq 1$. The interference can cause more damage if it is in a burst mode, and has partial duty cycle and we derive the worst case duty cycle. We denote the average number of received interference photons per symbol per receiver as N_I. If the interference has a duty cycle of β, then during the symbols for which the interference is on, the average number of received interference photons per symbol per receiver is N_I / β.

The communication system should, at the very least, be functional in the absence of interference (in the presence of only background noise). We consider the effect of interference on performance when the interference is larger than the background noise i.e. we assume that when the interference is on, the average received background noise photons per symbol is much less than the average received interference photons per symbol $\left(N_n \ll N_I / \beta \right)$.

4.1.1 Interference Types in Direct Detection (Incoherent Detection)

Recall that the modulation scheme used in the direct detection system is BPPM. In BPPM, if the sender sends a '0', its signal is on during the first half of the symbol and if it sends a '1', its signal is on during the second half of the symbol. The first of five types of interference that we consider in the diversity direct detection system is one that masks the communication signal when the sender sends a '1': the interference is on as a constant signal only during the first half of the symbol for fraction β of the symbols. See Figure 4.1 for a visualization of the communication and interference signals. The communication and interference signals are shown in white and grey respectively. If the sender sends a '1', then the interference masks the

71

communication signal, but if the sender sends a '0', then the interference actually helps by adding to the communication signal.
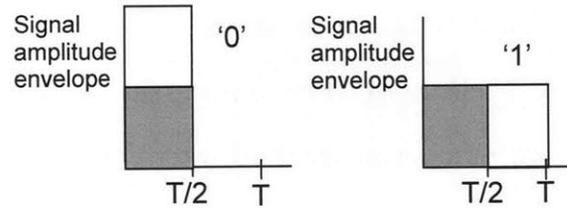
Figure 4.1: Visualization of signals in direct detection when the interference is on as a constant signal for only the first half of the symbol. The communication and interference signals are shown in white and grey respectively. 'T' is the symbol duration.

When the interference and communication signals are received in the same half symbol, we assume their powers add. (This assumes that either the interference is 90 degrees out of phase from the communication signal or that the interference phase fluctuates rapidly relative to the communication signal's phase.) This type of interference has symbol synchronization with the communication system since its signal is time-aligned to be in the first half of the symbol.

In some cases, a more 'random' interference may inflict more damage to the communication performance. The second type of interference that we consider for diversity direct detection has random amplitude with (2-sided) Gaussian distribution and is on during the entire symbol. The third type of interference also has random amplitude with Gaussian distribution but is on only during the first half of the symbol. This type of interference has symbol synchronization with the communication system since its signal is time-aligned to be in the first half of the symbol. For the second and third interference types, we assume for a worst case analysis that the interference at the N receivers are in phase.

The fourth type of interference we consider is on for the entire symbol and is received with opposite phase from the communication signal such that the communication signal is cancelled. See Figure 4.2 for a visualization of the communication and interference signals and the net signal after cancellation. Again, the communication and interference signals are shown in white and grey respectively. This type of interference causes errors by canceling the communication signal when either a '0' or '1' is sent. This type of interference has phase synchronization with the communication system since the interference cancels the communication signal.

(a) communication and interference signals
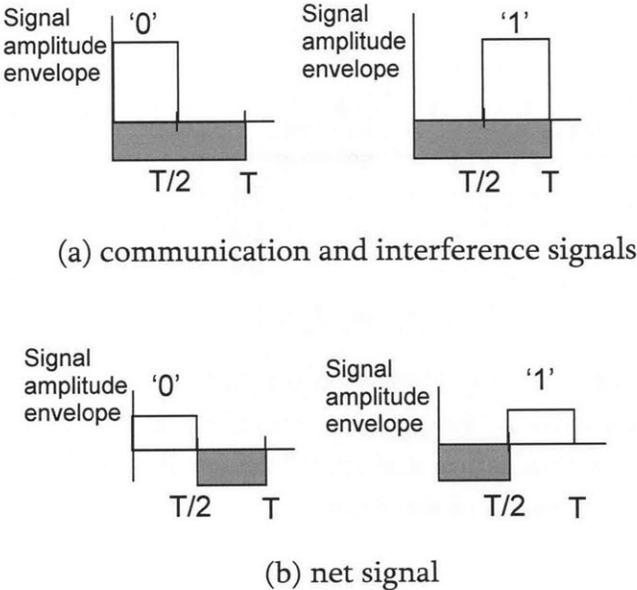
(b) net signal

Figure 4.2: Visualization of a) communication and interference signals and b) net signal in direct detection when the canceling interference is on for the entire symbol. The communication and interference signals are shown in white and grey respectively. 'T' is the symbol duration.

The fifth type of interference, like the fourth type, is received with opposite phase from the communication signal, but it is on only during the first half of the symbol. It is also a canceling interference type, but only cancels the communication signals

that correspond to a '0'. See Figure 4.3 for a visualization of the communication and interference signals. This type of interference has symbol and phase synchronization with the communication system since it is time-aligned to be in the first half of the symbol and its signal cancels the communication signal.

(a) communication and interference signals
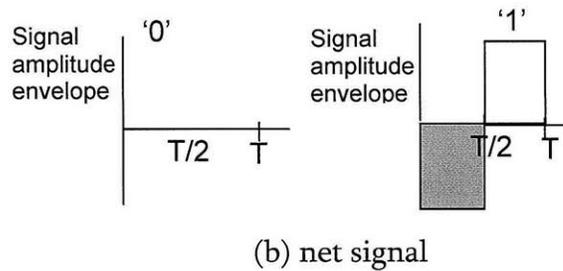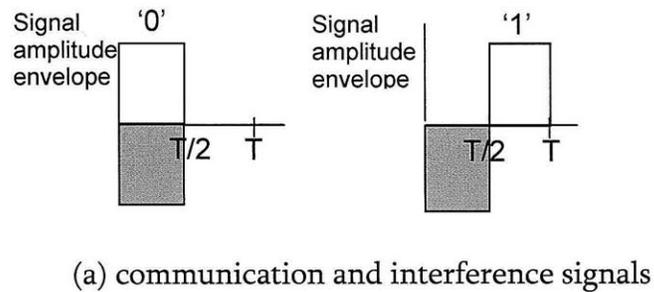
(b) net signal

Figure 4.3: Visualization of a) communication and interference signals and b) net signal in direct detection when the canceling interference is on for only the first half of the symbol. The communication and interference signals are shown in white and grey respectively. 'T' is the symbol duration.

4.1.2 Interference Types in Coherent Detection

Recall that the modulation scheme used in coherent detection is BPSK. In BPSK, if the sender sends a '0', there is a 180 degree phase difference in its signal compared to when the sender sends a '1'. The first of two types of interference that we consider in diversity coherent detection is one that is Gaussian noise.

The second type of interference that we consider for coherent detection is one that is on as a '0' signal and cancels the communication signal when the sender sends a '1'. See Figure 4.4 for a visualization of the communication and interference signals. If the sender sends a '1', then the interference cancels the communication signal, but if the sender sends a '0', then the interference actually helps by adding to the communication signal. This type of interference is phase synchronized with the communication system since the interference cancels the communication signal. We did not consider the other three interference types one, three and five as in direct detection since these others correspond to transmission in only half of the symbol and are not as performance degrading to coherent detection with BPSK.
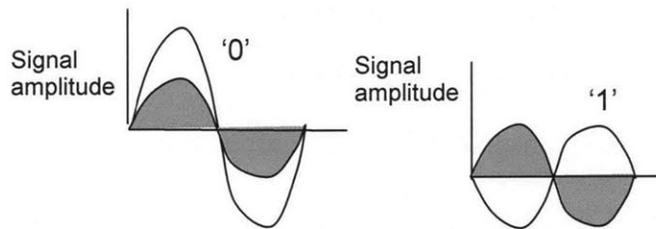
Figure 4.4: Visualization of communication and signals in coherent detection when the interference is on as a '0'. The communication and interference signals are shown in white and grey respectively.

4.2 Power Margin to Mitigate Interference

Consider that we have a given outage probability requirement, $P_{outage,required} = p$. In the *absence* of interference, the necessary average received signal photons per symbol that homodyne detection requires to achieve outage probability p (given the turbulence level, background noise level, error probability threshold, and diversity N) is given by

$$N_S^* = \frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2} \exp\left\{ \sqrt{-2\ln(2p)\ln\left(1 + \frac{e^{4\sigma_\chi^2}-1}{N}\right)} + \frac{1}{2}\ln\left(1 + \frac{e^{4\sigma_\chi^2}-1}{N}\right) \right\} \qquad (4.1)$$

This is found by rearranging (3.35), the outage probability of homodyne detection in the absence of interference. (4.1) gives the baseline minimum signal power needed to achieve the performance requirement. We use a homodyne detection system for the baseline because it requires less power than direct detection to achieve any given outage probability.

In the *presence* of interference, we must provide extra power margin $m > 1$ such that we can meet the outage probability requirement despite the presence of interference. Thus, in the presence of interference, we will denote the average received signal photons per symbol by mN_s^* instead of N_s to emphasize that a power margin is needed to mitigate the interference.

We discuss the amount of link margin required by the system to achieve the given outage probability requirement in the presence of interference in Section 4.8.

4.3 Quantities Derived for Diversity Direct Detection and Diversity Coherent Detection in the Presence of Interference

In the next two sections, we derive the following quantities for diversity direct detection and diversity homodyne detection in the presence of the various types of interference:

a) the error probability in the absence of fading

b) the interference worst case duty cycle in the absence of fading (that maximizes the error probability)

c) the outage probability in the presence of fading

d) the interference worst case interference duty cycle in the presence of fading (that maximizes the outage probability)

In deriving the error probability, P(e), and outage probability, P_{outage}, we make the following simplifying approximations:

$$P(e) = (1 - \beta)P(e \mid \text{interference is not on}) + \beta P(e \mid \text{interference is on})$$
$$\cong \beta P(e \mid \text{interference is on})$$

(4.2)

and

$$P_{outage} = (1 - \beta)P(\text{outage} \mid \text{inteference is not on}) + \beta P(\text{outage} \mid \text{interference is on})$$
$$\cong \beta P(\text{outage} \mid \text{interference is on})$$

(4.3)

These assumptions are reasonable if the contribution to error and outage probability when the interference is on is much larger than the contribution when the interference is not on i.e. when the average received interference power is not much smaller than the average received signal power. When deriving the worst case interference duty cycle, we assume the duty cycle is a positive number not greater than one (this is true by definition of a duty cycle that is non-zero). We assume that the sender sends a '0' or a '1' symbol with equal probability.

As we did in Chapter 3, we will assume that the sum of the N log-normal random variables $\alpha_1, \alpha_2, ..., \alpha_N$ is log-normal i.e. $\frac{1}{N}\sum_{i=1}^{N}\alpha_i = e^U$ where U is a Gaussian random variable with mean and variance given in (3.9).

If the interference is assumed to experience independent fading from the sender, the analysis of outage probability becomes unwieldy and it does not lend itself to closed form solutions or a form that provides intuition. Thus, when it simplifies the analysis of diversity coherent and incoherent detection in the presence of interference, we assume that the interference signal does not go through fading (or goes through the same fading as the communication signal). This assumption does not have an appreciable effect on outage probability as we will explain. Consider the following four scenarios:

(1) the communication signal is not faded and the interference is faded

(2) the communication signal is not faded and the interference is not faded

(3) the communication signal is faded and the interference is faded

(4) the communication signal is faded and the interference is not faded

The system is not in an outage in scenario (1). Outages will occur in the other three scenarios (2), (3) and (4). Assuming the interference does not go through fading (or goes through the same fading as the communication signal) has no effect on the sum probability of scenarios (3) and (4), and has little effect on the probability of scenario (2) provided the probability of the interference being faded is small. Thus, the assumption that the interference does not go through fading (or goes through the same fading as the communication signal) does not have an appreciable effect on outage probability.

4.4 Derivation of Diversity Direct Detection Performance

A block diagram of the diversity direct detection system is shown in the previous chapter in Figure 3.1.

Let us define the quantities used in this section. N_0 and N_1 represent the total number of received photons in the first and second half symbol intervals. N_n and N_I represent the average number of received background noise photons and interference photons per symbol per receiver. The hypothesis H_0 and H_1 represent symbol values of '0' and '1' being sent by the sender. $\underline{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_N)$ is a vector of power fading factors. N_S^* is the average number of received signal photons per symbol that homodyne detection requires to achieve a required outage probability (see Section 4.2). m is the power link margin.

4.4.1 Direct Detection in Presence of Constant Interference That is On for First Half of Symbol

Let us first consider the performance of diversity direct detection in the presence of interference that is on for only half the symbol as a constant signal. As mentioned previously, when the communication signal and interference are on during the same half symbol, their powers are assumed to add. If the interference is on and the sender sends a '0', the average photon counts in the first half and second half symbol intervals are

$$N_0 = mN_S^* + \frac{NN_I}{\beta} + \frac{NN_n}{2} \text{ and} \tag{4.4}$$

$$N_1 = \frac{NN_n}{2} \tag{4.5}$$

respectively. If the interference is on and the sender sends a '1', the total average photon counts in the first and second half symbol intervals are

$$N_0 = \frac{NN_I}{\beta} + \frac{NN_n}{2} \quad \text{and} \tag{4.6}$$

$$N_1 = mN_S^* + \frac{NN_n}{2} \tag{4.7}$$

respectively. The error probability in the absence of fading is given by

$$
\begin{aligned}
P(e) &\cong \beta P\big(e \,|\, \text{interference is on}\big) \\
&= \frac{\beta}{2} P\big(e \,|\, \text{interference is on, } H_0\big) + \frac{\beta}{2} P\big(e \,|\, \text{interference is on, } H_1\big) \\
&= \frac{\beta}{2} P\big(N_1 \geq N_0 \,|\, \text{interference is on, } H_0\big) + \frac{\beta}{2} P\big(N_0 \geq N_1 \,|\, \text{interference is on, } H_1\big) \\
&\leq \frac{\beta}{2} \min_{s \geq 0} E\Big[e^{s(N_1 - N_0)}\Big|\text{interference is on, } H_0\Big] + \frac{\beta}{2} \min_{t \geq 0} E\Big[e^{t(N_0 - N_1)}\Big|\,\big|\, \text{interference is on, } H_1\Big]
\end{aligned} \tag{4.8}
$$

The fourth line is the Chernoff Bound where the value of s and t are to be optimized to get the tightest bound. We find the tightest upper bound in Appendix G. Taking the error probability to be well approximated by this tightest bound, the error probability is

$$
P(e) \cong \frac{\beta}{2} \exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_I}{\beta} + \frac{NN_n}{2}} - \sqrt{\frac{NN_n}{2}}\right)^2\right\} + \frac{\beta}{2} \exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\}
$$

$$\tag{4.9}$$

Notice that if we increase the receiver diversity from 1 to N, the total average received signal photons does not change from the one receiver system but the total received average background noise and interference increase by a factor of N. Taking the derivative of (4.9) with respect to β and assuming that errors when the communication and interference signals add are negligible, $N_n \ll \dfrac{N_I}{\beta}$, and

$mN_S^* \gg NN_n$ (reasonable since the received signal should be much larger than the received background noise for decent operation of the system), the interference duty cycle that maximizes the error probability is

$$\beta_{wc} = \frac{NN_I}{mN_S^*} \qquad (4.10)$$

See Appendix G for the derivation. With this duty cycle, when the interference is on and the sender sends a '1', the average received interference photons in the first half symbol is $\dfrac{NN_I}{\beta_{wc}} = mN_S^*$ and the average received signal photons in the second half symbol is also mN_S^*. Thus, when the interference is on and the sender sends a '1', the average received energy is constant across the symbol. In other words, the interference worst case duty cycle is such that the interference puts just enough energy into the symbols (during which it is on) to cause the average received energy across the symbol to be constant. The error probability is dominated by errors that occur when the sender sends a '1', and the average error probability when the sender sends a '1' is sent is ½. Thus, the resulting error probability is

$$\begin{aligned} P(e) &\cong \beta_{wc} \cdot \Pr(H_1) \cdot \frac{1}{2} \\ &= \frac{\beta_{wc}}{2} \cdot \frac{1}{2} \\ &= \frac{NN_I}{4mN_S^*} \end{aligned} \qquad (4.11)$$

In the presence of fading of only the communication signal, the error probability when the interference is on is

$$P(e \mid \underline{\alpha}, \text{interference is on}) \cong \frac{1}{2} \exp\left\{ -\left(\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)mN_S^* + \frac{NN_I}{\beta} + \frac{NN_n}{2}} - \sqrt{\frac{NN_n}{2}} \right)^2 \right\}$$
$$+ \frac{1}{2}\exp\left\{ -\left(\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}} \right)^2 \right\} \qquad (4.12)$$

This is simply (4.9) with a multiplying fading factor for the communication signal included and the β factor in front of the two terms removed. The outage probability is given by

$$P_{outage} \cong \beta P(\text{outage} \mid \text{interference is on})$$
$$= \beta P\left(P(e \mid \underline{\alpha}, \text{interference is on}) > e^{-\theta_{thresh}} \right)$$
$$\cong \beta P\left(\frac{1}{2}\exp\left\{ -\left(\sqrt{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}} \right)^2 \right\} > e^{-\theta_{thresh}} \right)$$
$$= \beta P\left(\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right) < \frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}} \right)^2 - \frac{NN_n}{2}}{mN_S^*} \right)$$
$$= \beta P\left(e^U < \frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}} \right)^2 - \frac{NN_n}{2}}{mN_S^*} \right)$$
$$= Q\left(m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}} \right)^2 - \frac{NN_n}{2}}{mN_S^*} \right) \right)$$
$$\cong \frac{\beta}{2}\exp\left\{ -\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}} \right)^2 - \frac{NN_n}{2}}{mN_S^*} \right) \right]^2 \right\} \qquad (4.13)$$

where, in the third line, we assumed that errors when the communication and interference are both on during the first half symbol are negligible i.e. that the first term in (4.12) is negligible compared to the second term. In the second last line, $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \exp(-t^2/2) dt$ and in the last line, we have approximated $Q(x)$ by the upper bound $\frac{1}{2}\exp(-x^2/2)$. Taking the derivative of the last line of (4.13) with respect to β and assuming $N_n \ll \frac{N_I}{\beta}$ and

$$\frac{NN_I}{\beta} \gg -\ln 2 + \theta_{thresh} \tag{4.14}$$

(reasonable since $-\ln 2 + \theta_{thresh} = 1.6$ for an error probability threshold of 0.1 and N_I/β is much larger than 1 in scenarios where the interference is much larger than worst case background noise of $N_n = 1$), the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{NN_I}{mN_S^*}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^{3/2} \tag{4.15}$$

See Appendix G for the derivation. Substituting this worst case duty cycle into the outage probability expression in the last line of (4.13), the outage probability when the interference uses this worst case duty cycle is approximately

$$P_{outage} = \frac{NN_I}{2mN_S^*}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right) \tag{4.16}$$

The worst case duty cycle and corresponding outage probability expressions are valid when (4.14) is true for the worst case duty cycle i.e. when the power margin is large enough that $mN_S^* \gg (-\ln 2 + \theta_{thresh}) \left(1 + \dfrac{e^{4\sigma_\chi^2} - 1}{N} \right)^{3/2}$. For example, for $N = 4$, $P_e^{thresh} = 0.1$, and $\sigma_\chi^2 = 0.3$, any power margin greater than 3.2 is large enough to satisfy the approximation.

4.4.2 Direct Detection in Presence of Gaussian Interference That is On for Entire Symbol

Next, let us consider the performance of diversity direct detection in the presence of interference that is on for fraction β of the symbols as a randomly Gaussian signal. Because the field impinging on the receiving pupils is not constant, the output of each photodetector is a doubly stochastic Poisson process where the rate parameter varies with the impinging Gaussian field. For a high power incoming field, the photodetector output can be modeled as Gaussian. Assuming background noise is negligible compared to the interference, we derive in Appendix G, the mean and variance of the sum of the N photodetector outputs in each of the first and second half symbol intervals (when the interference is on) to be

$$m = \frac{N N_I}{2\beta} \text{ and} \tag{4.17}$$

$$\sigma^2 = \frac{N N_I}{2\beta} + 2\left(\frac{N N_I}{2\beta} \right)^2 \tag{4.18}$$

respectively. The last term in the variance is due to the added randomness of the varying rate parameter in the Poisson detection. Since each diversity receiver detects

the same average interference power as the receiver detects in the single receiver system, the sum of the N photodetector outputs has N times the mean of the photodetector output of a single receiver system. The variance of the N-receiver system is more than N times that of the single receiver system because we assumed, for a worst case analysis, that the received Gaussian fields cause the N rate parameters to vary together. In general, due to independent scattering, the phases of the fields are actually independent. The error probability in the absence of fading is given by

$$
\begin{aligned}
P(e) &\cong P(e \mid \text{interference is on}, H_0) \\
&= \beta P\big(m N_S^* + n_0 < n_1 \mid \text{interference is on}, H_0\big) \\
&= \beta P\big(n_1 - n_0 > m N_S^* \mid \text{interference is on}, H_0\big) \\
&= \beta Q\!\left(\frac{m N_S^*}{\sqrt{2\sigma^2}}\right) \\
&\cong \frac{\beta}{2} \exp\!\left\{-\frac{\left(m N_S^*\right)^2}{4\sigma^2}\right\} \\
&= \frac{\beta}{2} \exp\!\left\{-\frac{\left(m N_S^*\right)^2}{2\left[\dfrac{N N_I}{\beta} + \left(\dfrac{N N_I}{\beta}\right)^2\right]}\right\}
\end{aligned}
\tag{4.19}
$$

where n_0 and n_1 represent the total interference signal received in the first and second half symbol intervals respectively. Assuming $\dfrac{N N_I}{\beta} << \left(\dfrac{N N_I}{\beta}\right)^2$, the value of β that maximizes the error probability is

$$
\beta_{wc} = \frac{N N_I}{m N_S^*}
\tag{4.20}
$$

See Appendix G for the derivation. If the interference uses this worst case duty cycle, the resulting error probability is

$$P(e) = \frac{NN_I}{2mN_S^*} e^{-1/2} \tag{4.21}$$

(found by substituting the worst case duty cycle expression (4.20) into the last line of the error probability expression (4.19) and assuming $mN_S^* \gg 1$). In the presence of fading of only the communication signal, the error probability becomes

$$P(e \mid \underline{\alpha}, \text{interference is on}) = \frac{1}{2}\exp\left\{-\frac{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)\left(mN_S^*\right)^2}{2\left[\frac{NN_I}{\beta} + \left(\frac{NN_I}{\beta}\right)^2\right]}\right\} \tag{4.22}$$

This is (4.19) with a multiplying fading factor for the communication signal included and the β factor in front of the term removed. The outage probability of the diversity direct detection system in the presence of the Gaussian interference is given by

$$P_{outage} \cong \beta \Pr(\text{outage} \mid \text{interference is on})$$

$$= P\left(P(e \mid \underline{\alpha}, \text{interference is on}) > e^{-\theta_{thresh}} \mid \text{interference is on}\right)$$

$$= \beta P\left(\frac{1}{2}\exp\left\{-\frac{\left[\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)mN_S^*\right]^2}{2\left[\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right]}\right\} > e^{-\theta_{thresh}}\right)$$

$$= \beta P\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i < \frac{1}{mN_S^*}\cdot\sqrt{2\left[\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right]\left(-\ln 2 + \theta_{thresh}\right)}\right)$$

$$= \beta P\left(e^U < \frac{1}{mN_S^*}\cdot\sqrt{2\left[\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right]\left(-\ln 2 + \theta_{thresh}\right)}\right)$$

$$= \beta Q\left(\frac{1}{\sigma_U}\left[m_U - \ln\left(\frac{\sqrt{2\left[\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right]\left(-\ln 2 + \theta_{thresh}\right)}}{mN_S^*}\right)\right]\right) \tag{4.23}$$

$$\cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\sqrt{2\left[\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right]\left(-\ln 2 + \theta_{thresh}\right)}}{mN_S^*}\right)\right]^2\right\}$$

Assuming again that $\frac{NN_I}{\beta} \ll \left(\frac{NN_I}{\beta}\right)^2$, the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{NN_I}{mN_S^*}\sqrt{2\left(-\ln 2 + \theta_{thresh}\right)}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^{3/2} \tag{4.24}$$

See Appendix G for the derivation. The outage probability when the interference uses this worst case duty cycle is

$$P_{outage} = \frac{NN_I}{2mN_S^*} \sqrt{2\left(-\ln 2 + \theta_{thresh}\right)} \left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right)$$ (4.25)

(found by substituting the worst case duty cycle expression (4.24) into the last line of outage probability expression (4.23)).

4.4.3 Direct Detection in Presence of Gaussian Interference That is On for First Half of Symbol

Now we consider diversity direct detection in the presence of interference that is on for the first half of the symbol for fraction β of the symbols as a randomly Gaussian signal. Again, for a worst case analysis, we assume that the N Poisson rate parameters vary together. Modeling the sum of the N photodetector outputs as Gaussian and assuming background noise is negligible compared to the interference, the mean and variance of this Gaussian signal in the half symbol interval that the interference is on is

$$m = \frac{NN_I}{\beta} \quad \text{and}$$ (4.26)

$$\sigma^2 = \frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2$$ (4.27)

respectively. (This is the same as the mean and variance in (4.17) and (4.18) but where $\frac{NN_I}{2\beta}$ is replaced with $\frac{NN_I}{\beta}$.) The error probability in the absence of fading is given by

$$P(e) \cong \beta P(e \mid \text{interference is on})$$

$$\cong \frac{\beta}{2} P(e \mid \text{interference is on}, H_1)$$

$$= \frac{\beta}{2} \Pr(n_1 > mN_S^* \mid \text{interference is on})$$

$$= \frac{\beta}{2} Q\left(\frac{mN_S^* - \frac{NN_I}{\beta}}{\sqrt{\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2}} \right)$$

$$\cong \frac{\beta}{2} \exp\left\{ - \frac{\left(mN_S^* - \frac{NN_I}{\beta}\right)^2}{2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)} \right\} \tag{4.28}$$

where in the second line, we assumed the error probability is dominated by errors when sender sends a '1'. Assuming $\frac{NN_I}{\beta} \ll \left(\frac{NN_I}{\beta}\right)^2$, the value of β that maximizes the error probability is

$$\beta_{wc} = \frac{2NN_I}{mN_S^*} \tag{4.29}$$

See Appendix G for the derivation. If the interference uses this worst case duty cycle, the resulting error probability is

$$P(e) = \frac{NN_I}{mN_S^*} \exp\left(-\frac{1}{4}\right) \tag{4.30}$$

(found by substituting the worst case duty cycle expression (4.29) into the last line of the error probability expression (4.28) and assuming $\frac{mN_S^*}{2} \gg 1$). In the presence of fading of the communication signal, the error probability when the interference is on

89

is given by

$$P(e \mid \underline{\alpha}, \text{interference is on}) = \frac{1}{2}\exp\left\{-\frac{\left[\left(\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)mN_S^* - \frac{NN_I}{\beta}\right)^2\right]}{2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)}\right\} \qquad (4.31)$$

This is the last line of (4.28) with a multiplying fading factor for the communication signal included and the β factor in front of the term removed. The outage probability is given by

$$P_{outage} \cong \beta P(\text{outage} \mid \text{interference is on})$$

$$= \beta P\left(P(e \mid \underline{\alpha}, \text{interference is on}) > e^{-\theta_{thresh}}\right)$$

$$= \beta P\left(\frac{1}{2}\exp\left\{-\frac{\left[\left(\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)mN_S^* - \frac{NN_I}{\beta}\right)^2\right]}{2\left(\frac{NN_I}{\beta} + 2\left[\frac{NN_I}{\beta}\right]^2\right)}\right\} > e^{-\theta_{thresh}}\right)$$

$$= \beta P\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i < \frac{\sqrt{2\left(\frac{NN_I}{\beta} + 2\left[\frac{NN_I}{\beta}\right]^2\right)\left(-\ln 2 + \theta_{thresh}\right)} + \frac{NN_I}{\beta}}{mN_S^*}\right)$$

$$= \beta P\left(e^U < \frac{\sqrt{2\left(\frac{NN_I}{\beta} + 2\left[\frac{NN_I}{\beta}\right]^2\right)\left(-\ln 2 + \theta_{thresh}\right)} + \frac{NN_I}{\beta}}{mN_S^*}\right)$$

$$\cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\sqrt{2\left(\frac{NN_I}{\beta} + 2\left[\frac{NN_I}{\beta}\right]^2\right)\left(-\ln 2 + \theta_{thresh}\right)} + \frac{NN_I}{\beta}}{mN_S^*}\right)\right]^2\right\} \qquad (4.32)$$

Assuming $\frac{NN_I}{\beta} \ll \left(\frac{NN_I}{\beta}\right)^2$, the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{NN_I}{mN_S^*}\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^{3/2} \tag{4.33}$$

See Appendix G for the derivation. The outage probability when the interference uses this worst case duty cycle is

$$P_{outage} = \frac{NN_I\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)}{2mN_S^*}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right) \tag{4.34}$$

(found by substituting the worst case duty cycle expression (4.33) into the last line of the outage probability expression (4.32)).

4.4.4 Direct Detection in Presence of Canceling Interference That is On for Entire Symbol

Let us consider diversity direct detection in the presence of interference that is on for the entire symbol of fraction β of the symbols, during which it cancels the communication signal. As discussed previously, it may be difficult for interference to be aligned to exactly cancel the communication signal, so this type of canceling interference is a worst case type of interference. If the sender sends a '0', as we show in Appendix G, the average number of received photons in the first half and second half symbol times are given by

$$N_0 = \left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{2\beta}}\right)^2 + \frac{NN_n}{2} = mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2mN_S^*NN_I}{\beta}} + \frac{NN_n}{2} \quad \text{and} \quad (4.35)$$

$$N_1 = \frac{NN_I}{2\beta} + \frac{NN_n}{2} \tag{4.36}$$

respectively. If the sender sends a '1', the average number of received photons in the first half and second half symbol intervals are reversed from (4.35) and (4.36). If the interference is on for fraction β of the symbols, the error probability in the absence of fading is given by

$$
\begin{aligned}
P(e) &\cong \beta P(e \mid \text{interferer is on}) \\
&= \beta P(N_1 \geq N_0 \mid \text{interferer is on}, H_0) \\
&\leq \min_{s \geq 0} \beta E\left[e^{s(N_1 - N_0)} \mid \text{interferer is on}, H_0\right] \\
&= \beta \exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2NN_I mN_S^*}{\beta}} + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{2\beta} + \frac{NN_n}{2}}\right)^2\right\}
\end{aligned}
\tag{4.37}
$$

where the third line is the Chernoff Bound and the last line uses the optimal s to get the tightest bound (derived in Appendix G). Assuming $N_n \ll \frac{N_I}{2\beta}$, and taking the error probability to be well approximated by the tightest Chernoff bound, the error probability is

$$
\begin{aligned}
P(e) &\cong \beta \exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2NN_I mN_S^*}{\beta}}} - \sqrt{\frac{NN_I}{2\beta}}\right)^2\right\} \\
&= \beta \exp\left\{-\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{2\beta}} - \sqrt{\frac{NN_I}{2\beta}}\right)^2\right\} \\
&= \beta \exp\left\{-\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2\right\}
\end{aligned}
\tag{4.38}
$$

Assuming $mN_S^* >> 4$, the value of β that maximizes the error probability is

$$\beta_{wc} = \frac{2NN_I}{mN_S^*} \qquad (4.39)$$

See Appendix G for the derivation. With this worst case duty cycle, when the interference is on, the average number of received photons in the two half symbol intervals are equal. This is seen by substituting (4.39) into (4.35) and (4.36). The average received photon count in the half symbol interval with the communication signal is

$$\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{2\beta_{sc}}} \right)^2 + \frac{NN_n}{2} = \frac{mN_S^*}{4} + \frac{NN_n}{2} \qquad (4.40)$$

and in the other half symbol interval is

$$\frac{NN_I}{2\beta_{wc}} + \frac{NN_n}{2} = \frac{mN_S^*}{4} + \frac{NN_n}{2}. \qquad (4.41)$$

Thus, the interference worst case duty cycle is such that the interference puts just enough energy into the symbols (during which it is on) to cause the average received energy across the symbol to be constant. The resulting error probability when the interference uses the worst case duty cycle is thus

$$P(e) = \beta_{wc} \cdot \frac{1}{2} = \frac{NN_I}{mN_S^*} \qquad (4.42)$$

(found by substituting (4.39) into (4.38)). Assuming that the interference goes

through the same fading as the communication signal (as we described near the start of Section 4.3, this assumption does not have an appreciable effect on performance), the error probability when the interference is on is becomes

$$P(e \mid \underline{\alpha}, \text{interference is on}) = \exp\left\{-\left(\sum_{i=1}^{N}\alpha_i\right)\left(\sqrt{\frac{mN_S^*}{N}} - \sqrt{\frac{2N_I}{\beta}}\right)^2\right\}$$

$$= \exp\left\{-\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2\right\}$$

(4.43)

This is (4.38) with the same multiplying fading factor included for the communication signal and interference and the β factor in front of the term removed. The outage probability is

$$P_{outage} \cong \beta P\left(\text{outage} \mid \text{interference is on}\right)$$

$$= \beta \Pr\left(P(e \mid \underline{\alpha}, \text{interference is on}) > e^{-\theta_{thresh}}\right)$$

$$= \beta \Pr\left(\exp\left\{-\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2\right\} > e^{-\theta_{thresh}}\right)$$

$$= \beta \Pr\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i < \frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2}\right)$$

$$= \beta \Pr\left(e^U < \frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2}\right)$$

$$= \beta \Pr\left(U < \ln\left(\frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2}\right)\right)$$

$$= \beta Q\left(\frac{1}{\sigma_U}\left[m_U - \ln\left(\frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2}\right)\right]\right) \qquad (4.44)$$

$$\cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2}\right)\right]^2\right\}$$

It is difficult to find a closed form expression for the duty cycle that maximizes this outage probability, so we assume the worst case duty cycle is the same as the duty cycle that maximizes error probability i.e.

$$\beta_{wc} = \frac{2NN_I}{mN_S^*} \qquad (4.45)$$

This assumption is reasonable for large error probability thresholds because it is the detection noise, background noise and fading that push the system into outage. If the interferer spreads its energy across much more than this fraction β_{wc} of symbols, the efficacy of the noise and fading in causing outages will be reduced (due to moving the error probability operating point further away from the error probability threshold). However, this duty cycle is an approximation since it does not include an atmospheric turbulence parameter.

When the interference uses this value of duty cycle, the resulting outage probability is

$$P_{outage} = \beta_{wc} = \frac{2NN_I}{mN_S^*} \tag{4.46}$$

4.4.5 Direct Detection in Presence of Canceling Interference That is On for First Half of Symbol

Now consider diversity direct detection in the presence of interference that is on for only the first half of the symbol for fraction β of the symbols, during which it cancels the communication signal. Again, it would be difficult for the interference to be aligned to exactly cancel the communication signal, so this type of interference is also a worst case type of interference. If the sender sends a '0', as we show in Appendix G, the average number received photons in the first and second half symbol times are given by

$$N_0 = mN_S^* + \frac{NN_I}{\beta} - \sqrt{\frac{2mN_S^*NN_I}{\beta}} + \frac{NN_n}{2} = \left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}} \right)^2 + \frac{NN_n}{2} \quad \text{and} \tag{4.47}$$

96

$$N_1 = \frac{NN_n}{2} \tag{4.48}$$

If the sender sends a '1', the average number of received photons in the first and second half symbol intervals are given by

$$N_0 = \frac{NN_I}{\beta} + \frac{NN_n}{2} \quad \text{and} \tag{4.49}$$

$$N_1 = mN_S^* + \frac{NN_n}{2} \tag{4.50}$$

The error probability in the absence of fading is given by

$$
\begin{aligned}
P(e) &\cong \beta P(e \,|\, \text{interference is on}) \\
&= \frac{\beta}{2} P(e \,|\, \text{interference is on, H}_0) + \frac{\beta}{2} P(e \,|\, \text{interference is on, H}_1) \\
&\cong \frac{\beta}{2} \exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_I}{\beta} - 2\sqrt{\frac{NN_I mN_S^*}{\beta}} + \frac{NN_n}{2}} - \sqrt{\frac{NN_n}{2}}\right)^2\right\} \\
&\quad + \frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\}
\end{aligned}
\tag{4.51}
$$

where we again used the tightest Chernoff Bound. Assuming $\frac{N_n}{2} \ll \frac{N_I}{\beta}$ and $\frac{NN_n}{2} \ll mN_S^*$, this error probability reduces to

$$P(e) = \beta \exp\left\{-\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2\right\} \tag{4.52}$$

This error probability is the same as the error probability for diversity direct detection

in the presence of canceling interference that is on for the entire symbol except that $2NN_I / \beta$ is replaced with NN_I / β. Maximizing this error probability over β, the interference worst case duty cycle is

$$\beta_{wc} = \frac{NN_I}{mN_S^*} \qquad (4.53)$$

See Appendix G for the derivation. Substituting this worst case duty cycle into (4.47)-(4.50), the average number of received signal photons in the first and second half symbol intervals are equal when the interference is on. Specifically, when the sender sends a '0', the average number of received photons in both half symbols is $\frac{NN_n}{2}$ and when the sender sends a '1', the average received photons in both half symbol intervals is $mN_S^* + \frac{NN_n}{2}$. The resulting error probability when the interference uses this worst case duty cycle is thus

$$P(e) = \beta_{wc} \cdot \frac{1}{2} = \frac{NN_I}{2mN_S^*} \qquad (4.54)$$

when the average number of background noise photons is much smaller than the average number of interference photons. Assuming that the interference experiences the same fading as the communication signal (as we described near the start of Section 4.3, this assumption does not have an appreciable effect on performance), the error probability becomes

$$P(e) = \exp\left\{-\left(\sum_{i=1}^{N}\alpha_i\right)\left(\sqrt{\frac{mN_S^*}{N}} - \sqrt{\frac{N_I}{\beta}}\right)^2\right\}$$

$$= \exp\left\{-\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2\right\}$$

(4.55)

This is (4.52) with the same multiplying fading factor included for the communication signal and interference and the β factor in front of the term removed. Assuming that the interference goes through the same fading as the communication signal, the outage probability is

$$P_{outage} \cong \beta P(\text{outage} \,|\, \text{interference is on})$$

$$= \beta \Pr\left(\exp\left\{-\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2\right\} > e^{-\theta_{thresh}}\right)$$

$$= \beta \Pr\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i < \frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2}\right)$$

$$= \beta \Pr\left(e^U < \frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2}\right)$$

$$= \beta Q\left(\frac{1}{\sigma_U}\left[m_U - \ln\left(\frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2}\right)\right]\right)$$

(4.56)

$$\cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\theta_{thresh}}{\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2}\right)\right]^2\right\}$$

This is the same as the outage probability of diversity direct detection in the presence

of canceling interference that is on for the entire symbol except that $2NN_I/\beta$ is replaced with NN_I/β. Again, it is difficult to find a closed form expression for the duty cycle that maximizes this outage probability, so we assume that the worst case duty cycle is the same as the duty cycle that maximizes error probability i.e.

$$\beta_{wc} = \frac{NN_I}{mN_S^*} \tag{4.57}$$

Again, as we explained in the previous section, this assumption is reasonable for large P_e^{thresh}. When the interference uses this value of duty cycle, the resulting outage probability is

$$P_{outage} = \beta_{wc} = \frac{NN_I}{mN_S^*} \tag{4.58}$$

(found by substituting (4.57) into the third line of (4.56)).

4.5 Derivation of Diversity Coherent Detection Performance

A block diagram of the diversity coherent detection system is shown in the previous chapter in Figure 3.2. As discussed in Chapter 3, homodyne detection results in lower error probability than heterodyne detection. Thus, we will assume that the coherent detection system uses homodyne detection in our analysis in this section.

4.5.1 Homodyne Detection in Presence of Gaussian Interference

Let us consider the performance of diversity homodyne detection in the presence of

interference that transmits Gaussian noise for fraction β of the symbols. In Section 3.3.3, we found the error probability of diversity homodyne detection in the absence of interference to given by (3.34), namely

$$P_{Homo,noInterference}\left(e \mid \underline{\alpha}\right) = \frac{1}{2}\exp\left[-\frac{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)2N_s}{1+N_n}\right]$$

In the presence of Gaussian interference, assuming the background noise is negligible compared to the interference, the error probability (when the interference is on) is the above error probability where N_n is replaced with N_I / β. Thus, in the presence of Gaussian interference and the absence of fading, the error probability is given by

$$P(e) \cong \beta P\left(e \mid \text{interference is on}\right)$$

$$\cong \frac{\beta}{2}\exp\left\{-\frac{2mN_S^*}{1+\frac{N_I}{\beta}}\right\} \qquad (4.59)$$

As we discussed in Chapters 2 and 3, the total detected interference does not increase with diversity due to single spatial mode detectors coherently combining the signals with optimal weights. In essence, diversity coherent detection only sees a total of one spatial mode. Assuming $1 \ll \frac{N_I}{\beta}$, the value of β that maximizes the error probability is

$$\beta_{wc} = \frac{N_I}{2mN_S^*} \qquad (4.60)$$

See Appendix G for the derivation. The error probability when the interference uses

101

this worst case duty cycle is approximately

$$P(e) = \frac{N_I}{4mN_S^*} e^{-1} \tag{4.61}$$

(found by substituting (4.60) into (4.59) and assuming $2mN_S^* \gg 1$). In the presence of fading of only the communication signal, the error probability when the interference is on is

$$P_{Homo,noInterference}(e \mid \underline{\alpha}, \text{interference is on}) = \frac{1}{2} \exp\left[-\frac{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right) 2N_s}{1+N_n} \right] \tag{4.62}$$

This is the error probability in the absence of interference (3.34) but where N_n is replaced with N_I / β. The outage probability is given by

$$P_{outage} \cong \beta P(\text{outage} \mid \text{interference is on})$$

$$= \beta P\left(P(e \mid \underline{\alpha}, \text{interference is on}) > e^{-\theta_{thresh}}\right)$$

$$= \beta P\left(\frac{1}{2}\exp\left\{-\frac{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)2mN_S^*}{1+\frac{N_I}{\beta}}\right\} > e^{-\theta_{thresh}}\right)$$

$$= \beta P\left(-\frac{\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right)2mN_S^*}{1+\frac{N_I}{\beta}} > \ln\left(2e^{-\theta_{thresh}}\right)\right)$$

$$= \beta P\left(\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right) < \frac{\left(1+\frac{N_I}{\beta}\right)\left(-\ln 2 + \theta_{thresh}\right)}{2mN_S^*}\right)$$

$$\cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\left(1+\frac{N_I}{\beta}\right)\left(-\ln 2 + \theta_{thresh}\right)}{2mN_S^*}\right)\right]^2\right\} \tag{4.63}$$

Assuming $1 \ll \frac{N_I}{\beta}$, the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{N_I\left(-\ln 2 + \theta_{thresh}\right)}{2mN_S^*}\left(1 + \frac{e^{4o_\chi^2}-1}{N}\right)^{3/2} \tag{4.64}$$

See Appendix G for the derivation. The outage probability when the interference uses this worst case duty cycle is

$$P_{outage} = \frac{N_I\left(-\ln 2 + \theta_{thresh}\right)}{4mN_S^*}\left(1 + \frac{e^{4o_\chi^2}-1}{N}\right) \tag{4.65}$$

(found by substituting the worst case duty cycle expression (4.64) into the last line of

outage probability expression (4.63)).

4.5.2 Homodyne Detection in Presence of Canceling Interference

Finally, we consider the performance of diversity homodyne detection in the presence of interference that has symbol and phase synchronization with our system, and whose signal cancels the communication signal. Recall that symbol and phase synchronization may occur if the interference slowly sweeps its phase and symbol boundary so that the interference signal is symbol and phase aligned with the communication signal. This is a worst case interference analysis. Let the received signal amplitude (after photodetection and filtering) in the absence of interference be denoted by a_s if the sender sends a '0', and $-a_s$ if the sender sends a '1' (see Figure 4.5a). Since the interference always sends a '0', its signal subtracts from the sender's signal if the sender sends a '1' (and adds to the sender's signal if the sender sends a '0'). We denote the amplitude (after the filter) of the canceling interference by a_I. As depicted in Figure 4.5b, the total received signal amplitude is a_s+a_I if the sender sends a '0' and $-a_s+a_I$ if the sender sends a '1'.
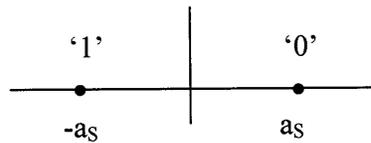
(a) in the absence of any interference

(b) in the presence of a canceling interference

Figure 4.5: Received signal constellation

The error probability is given by

$$
\begin{aligned}
P(e) &\cong \beta P\big(e \mid \text{interference is on}\big) \\
&= \beta\left[\frac{1}{2}\Pr\big(n > (a_S - a_I)\mid H_1\big) + \frac{1}{2}\Pr\big(n < -(a_S + a_I)\mid H_0\big)\right] \\
&= \frac{\beta}{2}Q\left(\frac{a_S - a_I}{\sqrt{N_0/2}}\right) + \frac{\beta}{2}Q\left(\frac{a_S + a_I}{\sqrt{N_0/2}}\right) \qquad\qquad (4.66)\\
&\cong \frac{\beta}{4}\exp\left\{-\frac{(a_S - a_I)^2}{N_0}\right\} + \frac{\beta}{4}\exp\left\{-\frac{(a_S + a_I)^2}{N_0}\right\}
\end{aligned}
$$

where n is the total local oscillator and background noise at the output of the filter, and has variance $N_0/2$. Using (3.19) in the analysis of heterodyne detection but scaling the noise by factor ½ to adjust for homodyne detection, the signal and interference amplitude and noise variance are given by

$$
a_S = \sqrt{E_S} = \sqrt{2\left(\frac{q\eta}{h\nu}\right)^2 P_L P_S T} \qquad\qquad (4.67)
$$

$$
a_I = \sqrt{E_I} = \sqrt{2\left(\frac{q\eta}{h\nu}\right)^2 P_L \frac{P_I}{\beta} T} \qquad\qquad (4.68)
$$

$$
\frac{N_0}{2} = \frac{q^2\eta}{2h\nu}P_L + \frac{1}{2}\left(\frac{q\eta}{h\nu}\right)^2 P_L N_{0,background} \qquad\qquad (4.69)
$$

where E_S is the average energy of the communication signal after the filter, E_I is the average energy of the interference signal after the filter, P_S is the average received signal power, P_I is the average received interference power, T is the symbol time, q is the charge of an electron, $N_{0,background}$ is the one sided spectral density of the background noise, η is the photodetector's quantum efficiency, h is Plank's constant, and ν is the frequency of the optical wave. Substituting (4.67)-(4.69) into (4.66), the

error probability reduces to

$$P(e) = \frac{\beta}{4}\left[\exp\left\{ \frac{-2\left(\sqrt{mN_S^*} - \sqrt{\frac{N_I}{\beta}} \right)^2}{1 + N_n} \right\} + \exp\left\{ \frac{-2\left(\sqrt{mN_S^*} + \sqrt{\frac{N_I}{\beta}} \right)^2}{1 + N_n} \right\} \right] \quad (4.70)$$

Assuming that the second term is negligible (reasonable since the second term represents errors when the signal and interference add constructively) and that $2(1 + N_n) \ll mN_S^*$ (reasonable since the received signal should be much larger than the worst case received background noise for decent operation of the system), we find the value of β that maximizes the error probability to be

$$\beta_{wc} = \frac{N_I}{mN_S^*} \quad (4.71)$$

See Appendix G for the derivation. The error probability when the interference uses this worst case duty cycle is

$$P(e) = \frac{N_I}{4mN_S^*} \quad (4.72)$$

(found by substituting (4.71) into (4.70) and again taking the second term to be negligible compared to the first term). In the presence of fading of the communication signal, the error probability when the interference is on is

106

$$P(e \mid \underline{\alpha}, \text{interference is on}) = \frac{1}{4}\left[\exp\left\{\frac{-2\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i mN_S^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1+N_n}\right\} + \exp\left\{\frac{-2\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i mN_S^*} + \sqrt{\frac{N_I}{\beta}}\right)^2}{1+N_n}\right\}\right]$$

(4.73)

This is the error probability (4.70) with a multiplying fading factor for the communication signal included and the β factor in front of the term removed. The outage probability is given by

$$
\begin{aligned}
P_{outage} &\cong \beta P\left(\text{outage} \mid \text{interference is on}\right)\\
&= \beta P\left(P(e \mid \underline{\alpha}, \text{interference is on}) > e^{-\theta_{thresh}}\right)\\
&= \beta \Pr\left(\frac{1}{4}\left[\exp\left\{\frac{-2\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i mN_S^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1+N_n}\right\} + \exp\left\{\frac{-2\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i mN_S^*} + \sqrt{\frac{N_I}{\beta}}\right)^2}{1+N_n}\right\}\right] > e^{-\theta_{thresh}}\right)\\
&\approx \beta \Pr\left(\frac{1}{4}\exp\left\{\frac{-2\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\alpha_i mN_S^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1+N_n}\right\} > e^{-\theta_{thresh}}\right)\\
&\cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left\{\frac{1}{mN_S^*}\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^2\right\}\right]^2\right\}
\end{aligned}
$$

(4.74)

where in the fourth line, we assumed that the second term in the error probability (due to a symbol error when the sender sends a '0' i.e. when the sender's and interference signal add constructively) is negligible. This is a reasonable assumption since most errors will occur due to the communication and interference signals canceling. Assuming

$$\frac{N_I}{\beta} \gg \frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2} \tag{4.75}$$

(reasonable since $-\ln 4 + \theta_{thresh} = 0.9$ for an error probability threshold of 0.1, $N_n = 1$ in worst case scenarios, and $\frac{N_I}{\beta}$ is much larger than N_n in scenarios where the interference is much larger than worst case background noise), the worst case duty cycle that maximizes the outage probability is

$$\beta_{wc} = \frac{N_I}{mN_S^*}\left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right)^{3/2} \tag{4.76}$$

See Appendix G for the derivation. The outage probability when the interference uses this worst case duty cycle is

$$P_{outage} = \frac{N_I}{2mN_S^*}\left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right) \tag{4.77}$$

(found by substituting the worst case duty cycle expression (4.76) into the last line of outage probability expression (4.74)). The worst case duty cycle and corresponding outage probability are valid when (4.75) is true for worst case duty cycle i.e. when the power margin is large enough that $mN_S^* \gg \frac{1}{2}(1 + N_n)(-\ln 4 + \theta_{thresh})\left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right)$.

4.6 Tables Summarizing Worst Case Duty Cycle, Error Probabilities and Outage Probabilities for Various Interference Types

In Sections 4.4 and 4.5, we derived for the various interference types, the worst case duty cycle in the absence and presence of fading and the corresponding error and outage probabilities. The worst case interference duty cycle for the various interference types is summarized in Table 4.1 and the corresponding error and outage probabilities are summarized in Table 4.2. For direct detection, the duty cycle, error probability, and outage probability increase linearly with increasing diversity for all the interference types. With increasing diversity, more interference is detected per symbol, and the interference can thus increase its duty cycle to cause more damage to more symbols. This causes the corresponding error probability and outage probability to increase. For coherent detection, increasing diversity does not cause an increase in duty cycle, error probability or outage probability due to detecting the same amount of interference regardless of the amount of diversity.

Interference type	Worst case duty cycle in absence of turbulence (corresponding to P(e))	Worst case duty cycle in presence of turbulence (corresponding to P_{outage})
Direct Detection:		
Constant on for half symbol	$\dfrac{NN_I}{mN_S^*}$	$\dfrac{NN_I}{mN_S^*}\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)^{3/2}$
Gaussian on for entire symbol	$\dfrac{NN_I}{mN_S^*}$	$\dfrac{NN_I}{mN_S^*}\sqrt{2\left(-\ln 2+\theta_{thresh}\right)}\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)^{3/2}$
Gaussian on for half symbol	$\dfrac{2NN_I}{mN_S^*}$	$\dfrac{NN_I}{mN_S^*}\left(2\sqrt{-\ln 2+\theta_{thresh}}+1\right)\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)^{3/2}$
Canceling on for entire symbol	$\dfrac{2NN_I}{mN_S^*}$	$\dfrac{2NN_I}{mN_S^*}$ (approximated to be the same as in the absence of fading)
Canceling on for half symbol	$\dfrac{NN_I}{mN_S^*}$	$\dfrac{NN_I}{mN_S^*}$ (approximated to be the same as in the absence of fading)
Homodyne Detection:		
Gaussian	$\dfrac{N_I}{2mN_S^*}$	$\dfrac{N_I}{2mN_S^*}\left(-\ln 2+\theta_{thresh}\right)\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)^{3/2}$
Canceling	$\dfrac{N_I}{mN_S^*}$	$\dfrac{N_I}{mN_S^*}\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)^{3/2}$

Table 4.1: Worst case interference duty cycle for various interference types

Interference type	Error probability (in absence of turbulence)	Outage probability (in presence of turbulence)
Direct Detection:		
Constant on for half symbol	$\dfrac{NN_I}{4mN_S^*}$	$\dfrac{NN_I}{2mN_S^*}\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)$
Gaussian on for entire symbol	$\dfrac{NN_I}{2mN_S^*}e^{-1/2}$	$\dfrac{NN_I}{2mN_S^*}\sqrt{2(-\ln 2+\theta_{thresh})}\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)$
Gaussian on for half symbol	$\dfrac{NN_I}{mN_S^*}e^{-1/4}$	$\dfrac{NN_I}{2mN_S^*}\left(2\sqrt{-\ln 2+\theta_{thresh}}+1\right)\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)$
Canceling on for entire symbol	$\dfrac{NN_I}{mN_S^*}$	$\dfrac{2NN_I}{mN_S^*}$
Canceling on for half symbol	$\dfrac{NN_I}{2mN_S^*}$	$\dfrac{NN_I}{mN_S^*}$
Homodyne Detection:		
Gaussian	$\dfrac{N_I}{4mN_S^*}e^{-1}$	$\dfrac{N_I}{4mN_S^*}(-\ln 2+\theta_{thresh})\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)$
Canceling	$\dfrac{N_I}{4mN_S^*}$	$\dfrac{N_I}{2mN_S^*}\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)$

Table 4.2: Error probability and outage probability for various interference types when the interference uses the worst case duty cycle.

4.7 Performance Plots of Diversity Coherent and Incoherent Detection in Presence of Various Interference Types

In this section, we use the expressions that we derived for worst case interference duty cycle, error probability, and outage probability to present various plots and make some observations.

111

4.7.1 Performance in Non-Fading Channel

First, let us discuss the performance of diversity direct detection and diversity homodyne detection in the presence of interference but in the absence of fading. See Figure 4.6 for a plot of the interference worst case duty cycle in a non-fading channel for the various interference types, and the corresponding error probability when the interference uses this worst case duty cycle. As seen in Figure 4.6a, for each of the interference types, the interference worst case duty cycle increases with increasing interference power. In other words, if the interference spreads its larger power over more symbols, it causes more damage; adding more interference to the same fraction of symbols does not increase the error probability as much as adding the additional interference to additional symbols. Also, as expected, we see in Figure 4.6b that for each interference type, the error probability increases as the ratio of interference to communication signal power increases. In direct detection, the interference type that causes the highest error probability is the canceling interference that is on for the entire symbol followed by the Gaussian interference that is on for only the first half of the symbol. For homodyne detection, we have an analogous scenario: the canceling interference causes higher error probability than the Gaussian interference. If we use direct detection, the interference can take on a strategy that causes worse error probability than if we use coherent detection.

Note that for direct detection, canceling interference that is on for the entire symbol causes *higher* error probability than if it is on for just half the symbol, whereas the direct detection Gaussian interference that is on for the entire symbol causes *lower* error probability than if it is on for just half the symbol. This seems inconsistent at first glance, but there is a logical explanation. Recall from Sections 4.4.4 and 4.4.5 that the worst case duty cycle for each of the canceling interference types in direct detection is such that the interference power in each symbol is the amount that

causes the net energy level to be constant across the entire symbol. (See Figures 4.3 and 4.4 for a visualization of the communication and interference signals for the canceling interference types in direct detection and the net signal after cancellation.) It takes the canceling interference that is on for the entire symbol half the energy to make the net signal level across the symbol as compared to the canceling interference that is on for only the first half of the symbol. This is because the canceling interference that is on for the entire symbol only has to use enough energy to subtract half the amplitude in the first half symbol (and use this same amount of energy in the second half symbol interval), whereas the canceling interference that is on for only the first half of the symbol has to subtract the entire amplitude in the first half symbol. Thus, the canceling interference that is on for the entire symbol can spread its energy across more symbols and cause more damage overall.
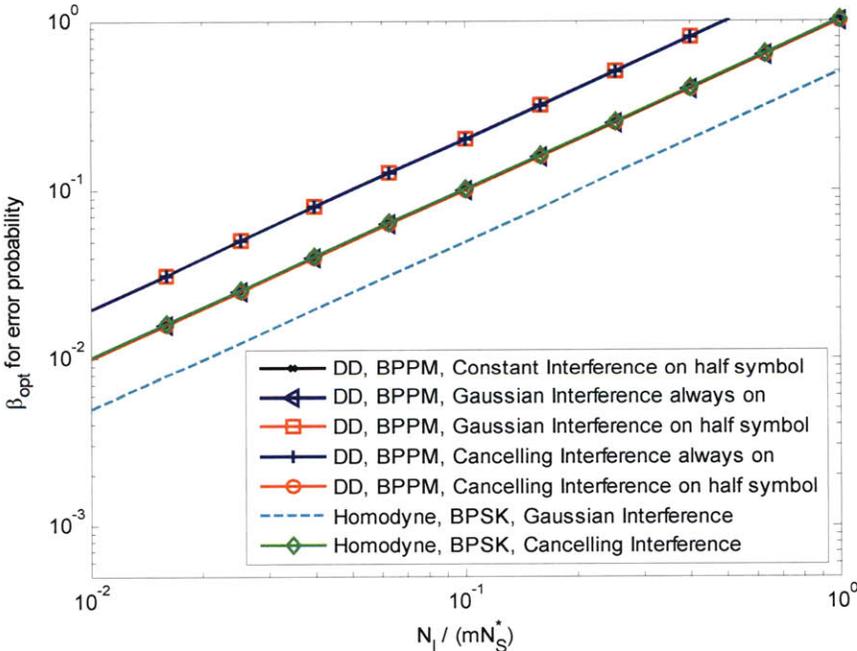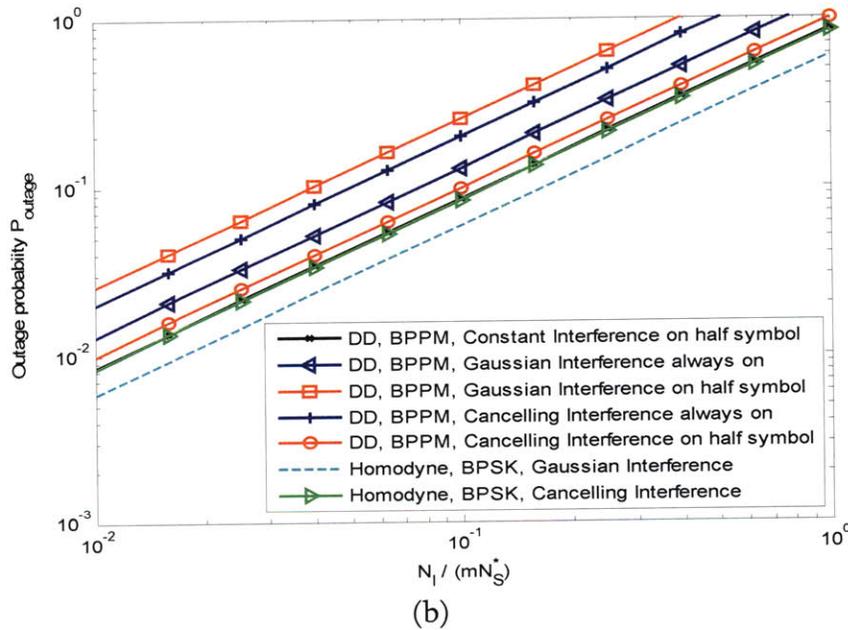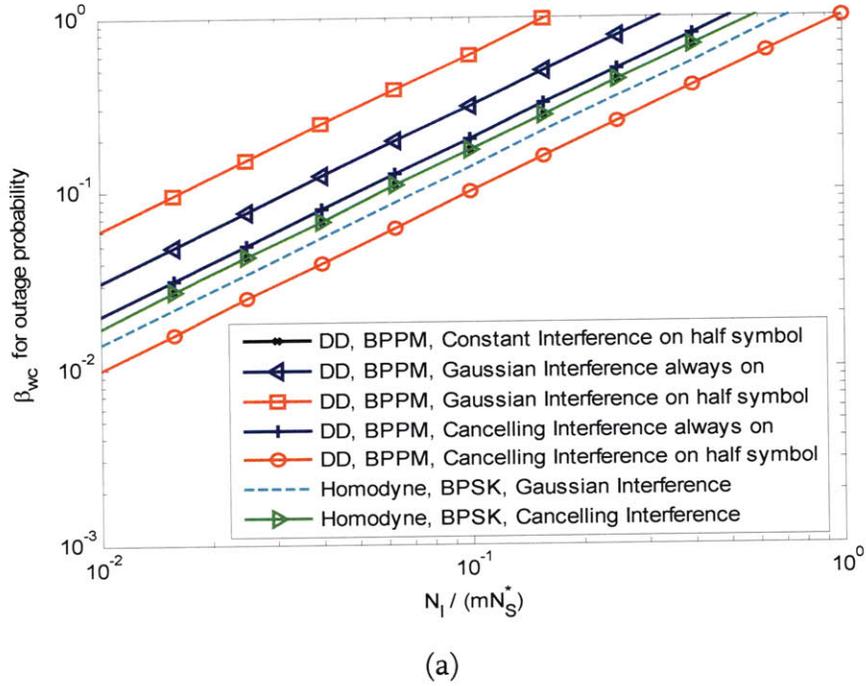

Figure 4.6 (a)
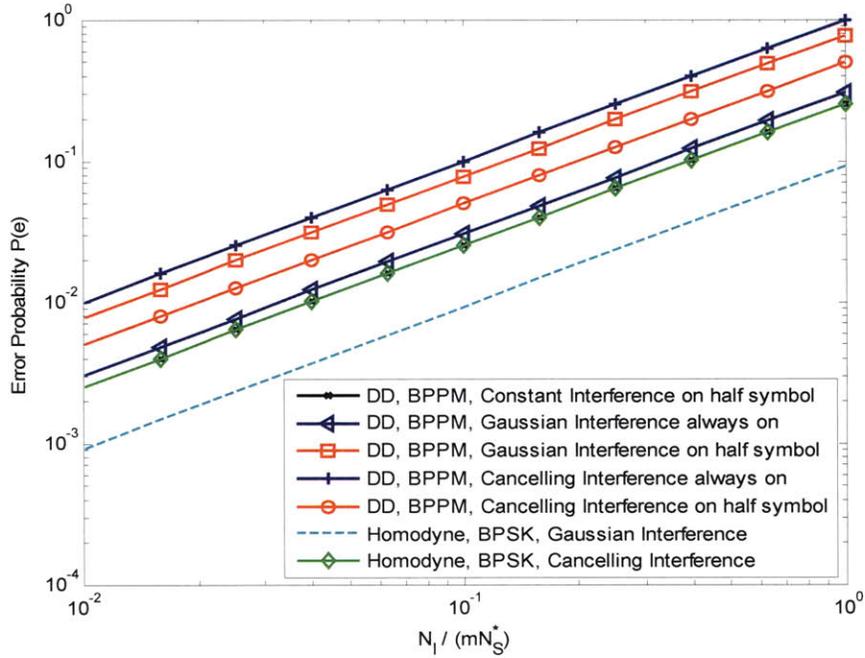
Figure 4.6 (b)

Figure 4.6: a) Interference worst case duty cycle and b) Error probability of direct detection and homodyne detection in the presence of interference that uses the worst case duty cycle in the following conditions: no fading, N=1, N_n=1. mN_S^* is the average number received signal photons per symbol. (Recall from Section 4.2 that N_S^* is the baseline number of average received signal photons per symbol needed by homodyne detection to achieve a given required outage probability and m is the link margin provided beyond N_S^*.) Note that the error probability on the left side of Figure 4.6b where $N_I/(mN_S^*) \cong 10^{-2}$ is not accurate when the average received signal photons mN_S^* is less than roughly seven[4]. This is because the contribution to error probability when the interference is not on is not negligible compared to the contribution when the interference is on, and the approximation (4.2) is not a good one. As $N_I/(mN_S^*)$ becomes smaller, and in the limit when it approaches zero, the error probability would approach the error probability value that corresponds to no interference i.e. the error probability due to just background noise and receiver detection noise.

In the case of the two Gaussian interference types that use their worst case duty cycle,

---

[4] For an average received signal photons per symbol of 7 and average received noise photons per symbol of 1, direct detection's error probability in the absence of interference is 1.6×10^{-2} (by (3.6)) and homodyne detection error probability in the absence of interference is 5×10^{-3} (by (3.34)).

when the interference is on, it turns out that the mean average photon count per half symbol due to the interference is the same for the two Gaussian interference types and the variances are the same also (this can be found by substituting the worst case duty cycle into the expressions for the mean and variance in Sections 4.4.2 and 4.4.3). The error probability for the Gaussian interference that is on for the first half symbol is dominated by errors that occur when the sender sends a '1', during which the average received photon difference between the two half symbol intervals is $N_s/2$ (see Appendix G). In the presence of the Gaussian interference that is on for the entire symbol, the average received energy difference between the two half symbol intervals is N_s (see Appendix G). Moreover, the worst case duty cycle for the Gaussian interference that is on for the first half of the symbol is twice that of the Gaussian interference that is on for the entire symbol. Therefore, the error probability for the Gaussian interference that is on for the first half of the symbol is higher than the error probability for the Gaussian interference that is on for the entire symbol.

4.7.2 Performance in Atmospheric Log-Normal Fading Channel

Next, we discuss the performance of the direct detection and homodyne detection systems in the presence of interference when we have log-normal fading due to clear atmospheric turbulence. As mentioned, the performance metric we use in the presence of fading is outage probability.

See Figure 4.7 for plots of the interference worst case duty cycle and the corresponding outage probability for the various interference types when the signal experiences fading and the error probability threshold is 0.1.

(a)

(b)

Figure 4.7: a) Interference worst case duty cycle b) outage probability of direct detection and homodyne detection in the presence of interference that uses the worst case duty cycle in the following conditions: log-normal fading with $\sigma_\chi = 0.3$, $P_e^{thresh} = 0.1$, N=1, N_n=1

Just as in the absence of fading, the interference can hamper performance more if we use direct detection than if we use coherent detection. For direct detection in the

presence of fading, Gaussian interference that is on for the first half of the symbol causes higher outage probability than canceling interference that is on for the entire symbol (where we approximate the canceling interference worst case duty cycle to be the same as in the absence of fading). For homodyne detection, the canceling interference causes higher outage probability than the Gaussian interference. By comparing the outage probability expressions of homodyne detection in the presence of the Gaussian interference and canceling interference ((4.65) and (4.77)), the canceling interference causes higher outage probability than the Gaussian interference does roughly when

$$1 > \frac{-\ln 2 + \theta_{thresh}}{2} \tag{4.78}$$

which reduces to

$$P_e^{thresh} > \frac{1}{2} e^{-2} = 0.068 \tag{4.79}$$

In general terms, canceling interference, though hard to implement, is the more damaging type of interference when the error probability threshold is high. This is because the canceling interference focuses its energy to reduce the amplitude of the desired signal thereby putting the error probability operating point in the absence of fading close to the error probability threshold (background noise and turbulence take care of causing outages after that). The Gaussian interference moves the operating point in the absence of fading farther from the error threshold (and its variance is what pushes the operating point above threshold).

Conversely, the Gaussian interference is more damaging when the error probability threshold is low. With a lower error probability threshold, the threshold is closer to the operating point in the absence of fading for both interference types. Thus, for the Gaussian interference, its variance pushes the point above threshold often. For the canceling interference, the background noise and atmospheric turbulence do not push the operating point above threshold as often as the Gaussian interference.

The plots shown thus far were for no-diversity direct detection and no-diversity homodyne detection in the presence of various interference types. In Figure 4.8, we show the outage probability when the diversity is increased to 4. Comparing Figure 4.7b and 4.8, we observe that when diversity is increased to 4, direct detection's outage probability curves become higher than those with no diversity. In other words, increasing diversity to 4 made the outage probability worse. In contrast, homodyne detection's outage probability with N=4 is lower than with no diversity. Homodyne detection's outage probability improves with diversity because the fading statistics improve with diversity while the total interference and background noise detected remains the same. Direct detection's outage probability worsens because although the fading statistics improve with diversity, the total interference and background noise detected increases with diversity. In direct detection, the increase in detected interference overpowers the improvement in fading statistics.
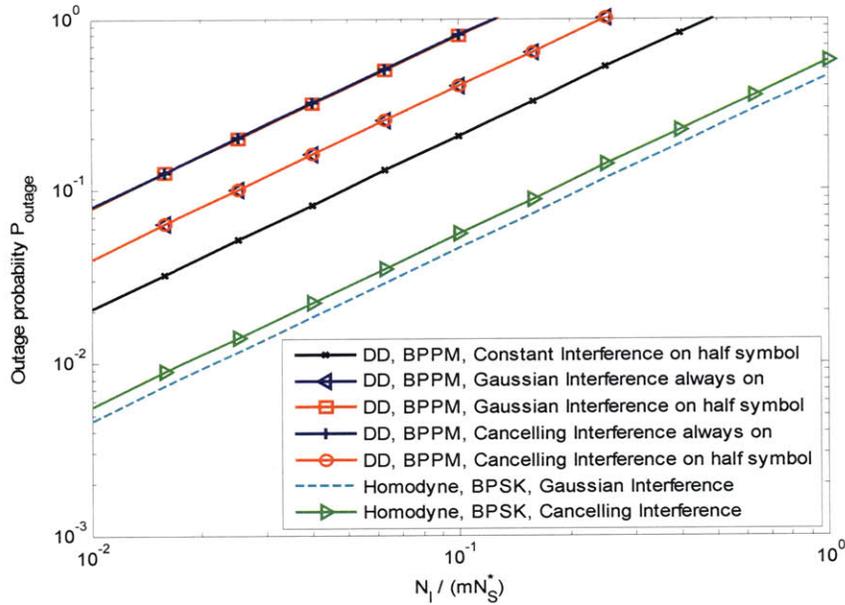
Figure 4.8: Outage probability (in the presence of fading) of direct detection and homodyne detection in the presence of interference that uses the worst case duty cycle in the following conditions: log-normal fading with $\sigma_\chi=0.3$, $P_e^{thresh}=0.1$, N=4, $N_n=1$

Let us consider the effect of increasing diversity in direct detection when the worst interference type is present. Among the interference types we considered, the interference that causes the worst performance in direct detection (for large error probability thresholds) is the Gaussian interference that is on for the first half symbol. We show in Figures 4.9 and 4.10, for various diversity values, the error probability and outage probability of direct detection in the presence of the Gaussian interference that is on for the first half of the symbol. We see that increasing diversity only worsens the performance rather than improving it. We prove in the next section that this is true for diversity direct detection in the presence of a Gaussian interference that is on for the first half of the symbol, provided $\frac{NN_I}{\beta} << \left(\frac{NN_I}{\beta}\right)^2$.
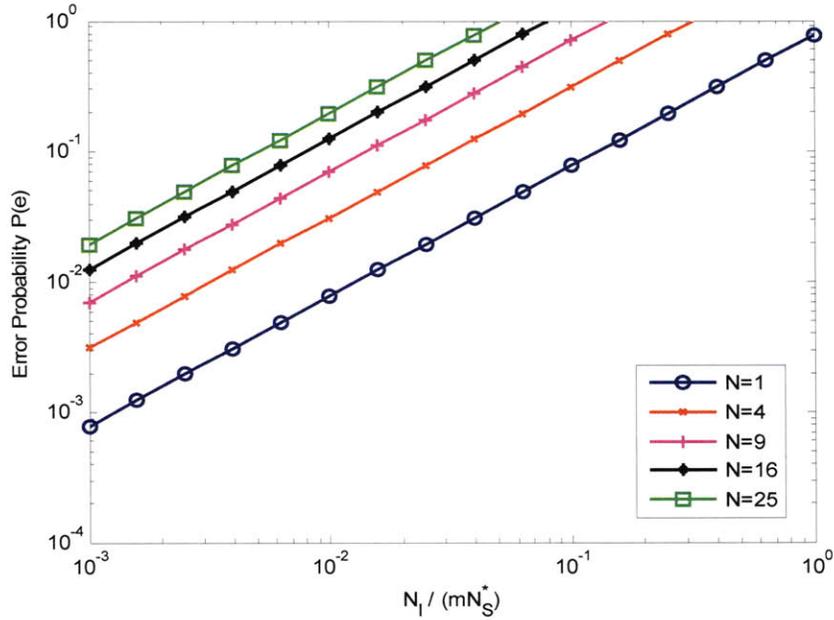
Figure 4.9: Error probability (in the absence of fading) of direct detection in the presence of Gaussian interference that turns on for only the first half symbol, and that uses the worst case duty cycle in the following conditions: $N_n=1$
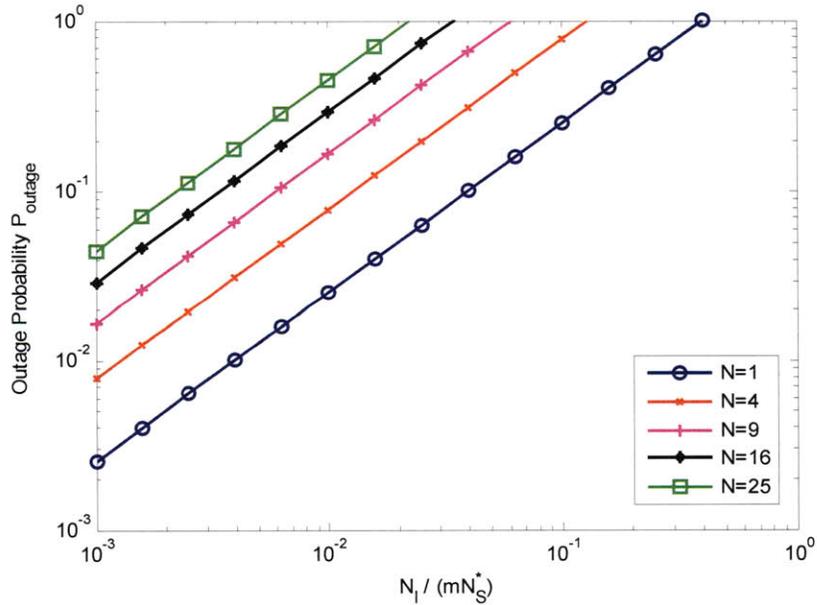
Figure 4.10: Outage probability (in the presence of fading) of direct detection in the presence of interference that turns on for only the first half symbol, and that uses the worst case duty cycle in the following conditions: log-normal fading, $\sigma_\chi=0.3$, $P_e^{thresh}=0.1$, $N_n=1$

4.8 Optimal Diversity of Direct Detection in Presence of Interference

We now show that the optimal diversity value for direct detection in the presence of Gaussian interference that is on for the first half symbol is one when $\frac{NN_I}{\beta} << \left(\frac{NN_I}{\beta}\right)^2$.

The outage probability of direct detection in the presence of Gaussian interference that is on for the first half symbol (and uses the worst case duty cycle) is given by (4.34), namely

$$P_{outage} = \frac{NN_I}{2mN_S^*}\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)\left(1 + \frac{e^{4\sigma_{\chi}^2} - 1}{N}\right) \qquad (4.80)$$

when the background noise is negligible compared to the interference and $\frac{NN_I}{\beta} << \left(\frac{NN_I}{\beta}\right)^2$. Isolating for mN_S^* gives

$$mN_S^* = \frac{N_I\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)}{2P_{outage}}\left(N + e^{4\sigma_{\chi}^2} - 1\right) \qquad (4.81)$$

Since this expression increases with increasing diversity value N, the diversity value that minimizes the amount of required signal power for a desired outage probability is one. Denoting the optimal direct detection diversity value in the presence of the Gaussian interference that is on for the first half of the symbol as $N_{opt,GaussianHalfSymbolInterference}$,

$$N_{opt,GaussianHalfSymbolInterference} = 1 \qquad (4.82)$$

In other words, for direct detection in the presence of Gaussian interference that is on for the first half the symbol, if $\frac{NN_I}{\beta} << \left(\frac{NN_I}{\beta}\right)^2$, adding diversity only degrades the outage probability; the lowest outage probably is achieved if no diversity is used.

4.9 Amount of Interference the System Can Tolerate

Since coherent detection outperforms direct detection, if we are interested in better performance (all other factors aside), we would like to use a coherent detection system. In this section, we discuss the amount of interference the homodyne system can tolerate while still achieving the required outage probability. We also describe a sensible way to select the amount of diversity and power link margin to deal with atmospheric turbulence and interference.

If we provide a factor m of power margin and diversity N, the outage probability of a diversity homodyne system in the presence of canceling interference (the worst of the types we considered) is given by (4.77), namely

$$P_{outage} = \frac{N_I}{2mN_S^*}\left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right)$$

Thus, the amount of interference, N_I, that the system can tolerate to still meet an outage probability requirement of $P_{outage,required}$ is

$$N_I = \frac{2mN_S^* P_{outage,required}}{\left(1 + \dfrac{e^{4o_\chi^2} - 1}{N}\right)} \tag{4.83}$$

(found by re-arranging the outage probability above). Phrased another way, the maximum ratio of average received interference to signal photons that achieves the required outage probability is

$$\frac{N_I}{mN_S^*} = \frac{2P_{outage,required}}{\left(1 + \dfrac{e^{4o_\chi^2} - 1}{N}\right)}$$

(4.84)

For the case when no diversity is used (N=1), in the presence of interference, the minimum link margin that is needed to achieve $P_{outage,required}$ is

$$mN_S^* = \frac{N_I e^{4o_\chi^2}}{2P_{outage,required}}$$

(4.85)

(found by substituting N=1 into (4.84)). So, for example, if we have a required outage probability of 10^{-2} and $\sigma_\chi^2 = 0.5$, we need $mN_S^* = \dfrac{N_I e^2}{2x10^{-2}} = 369N_I$ to tolerate the interference. The power margin reduces outages due to both turbulence and interference. Depending on the level of interference we wish to be able to tolerate, we can increase the power margin appropriately.

Note that if, instead of adding power margin, we increase the diversity from N to N', the amount of N_I we can tolerate increases as well (as we can see in (4.83)). However, as diversity increases to larger values, the benefit of added diversity becomes smaller. Even if it were possible to increase diversity to an infinite value, as we can see by taking (4.83) to the limit as in $N \rightarrow \infty$, the amount of N_I we can tolerate is limited to $2mN_S^* P_{outage,required}$. This is because diversity helps reduce outages due to atmospheric

turbulence (as opposed to outages due to interference), and there is a limit to this benefit. Thus, it makes sense to set the diversity to the value that provides adequate performance in the absence of interference and add enough link margin to tolerate the interference.

Chapter 5

Comparison of Diversity Direct Detection and Diversity Coherent Detection

It is apparent from the previous chapters that diversity homodyne detection provides better performance than diversity direct detection. In this chapter, we will quantify the amount of performance gain (we published these results in [33]). We define the power gain of homodyne detection over direct detection as the factor increase in power required by direct detection compared to homodyne detection to achieve outage probability P_{outage}. The power gain will be plotted as a function of diversity N.

We will first derive the following power gains in the absence of interference:

(1) power gain of homodyne detection with diversity N over direct detection with diversity N

(2) power gain of homodyne detection with diversity N over direct detection with diversity fixed to be N_{opt} (recall from Chapter 3 that N_{opt} is the optimal direct detection diversity in the absence of interference)

(3) power gain of homodyne detection with diversity N over direct detection with diversity fixed to be $N_{opt,GaussianHalfSymbolInterference} = 1$) (recall from Chapter 4 that $N_{opt,GaussianHalfSymbolInterference} = 1$ is the optimal direct detection diversity in the presence of Gaussian interference that is on for the first half of the symbol if $\frac{NN_I}{\beta} << \left(\frac{NN_I}{\beta}\right)^2$).

We fix the direct detection diversity value in (2) and (3) above because in direct detection, it is not optimal to continue to increase the diversity past its optimal value; we want to consider selecting the best possible diversity value for direct detection and compare the performance against diversity homodyne detection. Moreover, we may not know if interference is present or absent, and may select the optimal diversity value to be either in the presence or absence of interference. These different power gains allow us to compare the performance of homodyne detection over direct detection from the different perspectives of when the direct detection system's diversity varies and when it is fixed to the optimal value.

We will also derive the above three power gains in the presence of interference. In deriving the power gain in the presence of interference, we will assume that the diversity direct detection and diversity homodyne detection systems are operating in the presence of their corresponding worst interference type (of the types we considered). We will let direct detection be in the presence of Gaussian interference that is on for the first half of the symbol and will let homodyne detection be in the presence of canceling interference.

5.1 Power Gain in Absence of Interference

5.1.1 Power Gain of Homodyne Detection with Diversity N Over Direct Detection with Diversity N

Direct detection with diversity N achieves the same outage probability as homodyne detection with diversity N when

$$P_{outage,DD,noInterference}(N) = P_{outage,Homo,noInterference}(N) \tag{5.1}$$

$$Q\left(\frac{m_U(N) - \ln\left(\frac{1}{N_{S,DD}(N)}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]\right)}{\sigma_U(N)}\right) = Q\left(\frac{m_U(N) - \ln\left(\frac{1}{N_{S,Homo}(N)}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]\right)}{\sigma_U(N)}\right)$$

$$\frac{1}{N_{S,DD}(N)}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right] = \frac{(1 + N_n)(-\ln 2 + \theta_{thresh})}{2N_{s,Homo}(N)}$$

where in the second line, we used the outage probabilities of diversity direct detection and diversity homodyne detection in the absence of interference given in (3.10) and (3.35) and we explicitly show which variables are a function of N. For an error probability threshold of $P_e^{thresh} = e^{-\theta_{thresh}}$, the power gain is

$$\text{Power Gain} = \frac{N_{s,DD}(N)}{N_{s,Homo}(N)}$$

$$= 2\frac{\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}}{(1 + N_n)(-\ln 2 + \theta_{thresh})} \tag{5.2}$$

(found by re-arranging (5.10)). When N is taken to approach ∞, the first term in the numerator of (5.8) is negligible compared to the second term in the numerator. Thus, the asymptotic power gain, as N becomes large, increases as

$$\text{Power Gain} = c\sqrt{N}$$
$$\scriptstyle N\to\infty$$

$$\text{where } c = \frac{2\sqrt{2\theta_{thresh}N_n}}{\left(1+N_n\right)\left(-\ln 2 + \theta_{thresh}\right)} \qquad (5.3)$$

5.1.2 Power Gain of Homodyne Detection with Diversity N Over Direct Detection with Diversity N_{opt}

Direct detection with diversity N_{opt} achieves the same outage probability as homodyne detection with diversity N when

$$P_{outage,DD,noInterference}\left(N=N_{opt}\right) = P_{outage,Homo,noInterference}\left(N\right)$$

$$Q\left(\frac{m_U\left(N_{opt}\right)-\ln\left(\frac{1}{N_{S,DD}\left(N_{opt}\right)}\left[\theta_{thresh}+\sqrt{2\theta_{thresh}N_{opt}N_n}\right]\right)}{\sigma_U\left(N_{opt}\right)}\right) = Q\left(\frac{m_U\left(N\right)-\ln\left(\frac{1}{N_{S,Homo}\left(N\right)}\left[\theta_{thresh}+\sqrt{2\theta_{thresh}NN_n}\right]\right)}{\sigma_U\left(N\right)}\right)$$

$$\frac{m_U\left(N_{opt}\right)-\ln\left(\frac{1}{N_{S,DD}\left(N_{opt}\right)}\left[\theta_{thresh}+\sqrt{2\theta_{thresh}N_{opt}N_n}\right]\right)}{\sigma_U\left(N_{opt}\right)} = \frac{m_U\left(N\right)-\ln\left(\frac{1}{N_{S,Homo}\left(N\right)}\left[\theta_{thresh}+\sqrt{2\theta_{thresh}NN_n}\right]\right)}{\sigma_U\left(N\right)}$$

$$(5.5)$$

where in the second line, we used the outage probabilities given in (3.10) and (3.35). This is similar to (5.1) except that in the direct detection outage probability expression, we set $N = N_{opt}$. The power gain is given by

128

$$\text{Power Gain} = \frac{N_{s,DD}(N_{opt})}{N_{s,Homo}(N)} \qquad (5.6)$$

Isolating for $N_{S,DD}(N_{opt})/N_{S,Homo}(N)$ in (5.5)[5], the resulting power gain is given by

$$\text{Power Gain} = k_1 \exp\left\{ -\sqrt{-2\ln(2P_{outage})\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)} - \frac{1}{2}\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right) \right\} \qquad (5.7)$$

$$\text{where } k_1 = 2\left(\frac{\theta^{thresh} + \sqrt{2\theta^{thresh}N_{opt}N_n}}{(1 + N_n)(-\ln 2 + \theta^{thresh})}\right)\sqrt{1 + \frac{e^{4\sigma_\chi^2} - 1}{N_{opt}}} \cdot \exp\left\{ \sqrt{-2\ln(2P_{outage})\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N_{opt}}\right)} \right\}$$

We now find the asymptotic power gain as N becomes large. When $N \gg e^{4\sigma_\chi^2} - 1$,

we can use the approximation $\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right) \cong \frac{e^{4\sigma_\chi^2} - 1}{N}$ and (5.7) becomes

$$\text{Power Gain} = k_1 \exp\left\{ -\sqrt{-\frac{1}{N}2\left(e^{4\sigma_\chi^2} - 1\right)\ln(2P_{outage})} - \left(\frac{e^{4\sigma_\chi^2} - 1}{2N}\right) \right\} \qquad (5.8)$$

As $N \to \infty$, the second term in the exponent in (5.8) is much smaller than the first term. Thus, the asymptotic power gain, as $N \to \infty$, increases as

$$\text{Power Gain}_{N \to \infty} = k_1 \exp\left(-k_2 / \sqrt{N}\right)$$

$$\text{where } k_2 = \sqrt{-2\left(e^{4\sigma_\chi^2} - 1\right)\ln(2P_{outage})} \qquad (5.9)$$

and approaches the limit k_1 for large N.

---

[5] We make the resulting expression independent of $N_{S,Homo}(N)$ by substituting (4.1) into our expression (where p is the outage probability P_outage).

5.1.3 Power Gain of Homodyne Detection with Diversity N Over Direct Detection with Diversity $N_{opt,GaussianHalfSymbolInterference}$ (=1)

Direct detection with diversity $N_{opt,GaussianHalfSymbolInterference}=1$ achieves the same outage probability as homodyne detection with diversity N when

$$P_{outage,DD,noInterference}(N=1) = P_{outage,Homo,noInterference}(N)$$

$$Q\left(\frac{m_U(N=1) - \ln\left(\frac{1}{N_{S,DD}(N=1)}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}N_{opt}N_n}\right]\right)}{\sigma_U(N=1)}\right) = Q\left(\frac{m_U(N) - \ln\left(\frac{1}{N_{S,Homo}(N)}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]\right)}{\sigma_U(N)}\right)$$

(5.10)

This is the same as (5.5) except N_{opt} is replaced with 1. Thus, the power gain and the asymptotic power gain are the same as in (5.7) and (5.9) except N_{opt} is replaced with 1 i.e. the power gain is

$$\text{Power Gain} = c_1 \exp\left\{-\sqrt{-2\ln(2P_{outage})\ln\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right)} - \frac{1}{2}\ln\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right)\right\}$$

(5.11)

$$\text{where } c_1 = 2\left(\frac{\theta^{thresh} + \sqrt{2\theta^{thresh}N_n}}{(1+N_n)(-\ln 2 + \theta^{thresh})}\right)e^{2\sigma_\chi^2}\exp\left\{\sqrt{-2\ln(2P_{outage})4\sigma_\chi^2}\right\}$$

and the asymptotic power gain, as N becomes large is

$$\text{Power Gain} \atop {}_{N\to\infty} = c_1\exp\left(-k_2/\sqrt{N}\right)$$

(5.12)

$$\text{where } k_2 = \sqrt{-2\left(e^{4\sigma_\chi^2}-1\right)\ln(2P_{outage})}.$$

5.2 Power Gain in Presence of Interference

We now find the power gain of homodyne detection over direct detection when interference is present. As we discussed at the start of this chapter, in direct detection, the interference type is assumed to be the Gaussian interference that is on for the first half symbol and in homodyne detection, the interference type is assumed to the canceling interference.

5.2.1 Power Gain of Homodyne Detection with Diversity N Over Direct Detection with Diversity N

When both direct detection and homodyne detection have diversity N, they achieve the same outage probability in the presence of interference when

$$P_{outage,DD,GaussHalfSymbolInterference}(N) = P_{outage,Homo,CancellingInterference}(N) \qquad (5.13)$$

$$\frac{NN_I\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)}{2N_S^* m_{DD,GaussHalfSymbolInterference}(N)} = \frac{N_I\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)}{2N_S^* m_{Homo,CancellingInterference}(N)} \qquad (5.14)$$

where in (5.14), we used the outage probability of diversity direct detection in the presence of Gaussian interference that is on for the first half of the symbol given in (4.34) and the outage probability of diversity homodyne detection in presence of canceling interference given in (4.77). Rearranging (5.14), the power gain is

$$\text{Power Gain} = \frac{m_{DD,GaussHalfSymbolInterference}(N)}{m_{Homo,CancellingInterference}(N)} \qquad (5.15)$$

$$= \left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)N \qquad \propto N$$

131

5.2.2 Power Gain of Homodyne Detection with Diversity N Over Direct Detection with Diversity N_{opt}

When direct detection's diversity is N_{opt} instead of N, we find the power gain of homodyne detection with diversity N over direct detection with diversity N_{opt} in the presence of interference by letting $N = N_{opt}$ in the left side of (5.14). Doing so and re-arranging the resulting equation, the power gain is

$$\text{Power Gain} = \frac{m_{DD,GaussHalfSymbolInterference}\left(N_{opt}\right)}{m_{Homo,CancellingInterference}\left(N\right)}$$

$$= \frac{N_{opt}\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)\left(1 + \dfrac{e^{4\sigma_\chi^2} - 1}{N_{opt}}\right)}{\left(1 + \dfrac{e^{4\sigma_\chi^2} - 1}{N}\right)} \qquad \propto \frac{1}{1 + \dfrac{e^{4\sigma_\chi^2} - 1}{N}} \qquad (5.16)$$

5.2.3 Power Gain of Homodyne Detection with Diversity N over Direct Detection with Diversity $N_{opt,GaussianHalfSymbolInterference}$ (=1)

When direct detection's diversity is 1 instead of N, we find the power gain of homodyne detection with diversity N over direct detection with diversity 1 in the presence of interference by letting N=1 in the left side of (5.14). Doing so and re-arranging the resulting equation, the power gain is

$$\text{Power Gain} = \frac{m_{DD,GaussHalfSymbolInterference}\left(N_{optGaussianHalfSymbolInterference} = 1\right)}{m_{Homo,CancellingInterference}\left(N\right)}$$

$$= \frac{\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)e^{4\sigma_\chi^2}}{\left(1 + \dfrac{e^{4\sigma_\chi^2} - 1}{N}\right)} \qquad \propto \frac{1}{1 + \dfrac{e^{4\sigma_\chi^2} - 1}{N}} \qquad (5.17)$$

5.3 Power Gain Summary, Plots, and Observations

Table 5.1 summarizes the power gain of homodyne detection over direct detection both in the absence of and presence of interference. The interference type in direct detection is the Gaussian interference that is on for the first half of the symbol and the interference type in homodyne detection is the canceling interference.

We show in Figures 5.1 and 5.2, the power gain when the outage probability is 0.01 and 0.1 respectively. Consider the power gain in the *absence* of interference where direct detection's diversity blindly increases together with homodyne detection's diversity, even though increasing direct detection's diversity without bound actually hurts its performance. As diversity increases to large values, the power gain increases proportionally to the square root of the amount of diversity. This boundless gain is due to the boundless increase in background noise seen by direct detection as diversity increases. Note that part of the power gain is due to direct detection using BPPM rather than OOK (the amount due to suboptimal modulation is less than the gain seen at a diversity of one – the left most point on the curve). One may argue that since direct detection's performance worsens if diversity is increased beyond an optimal value, and homodyne detection's performance always improves with diversity, a fairer comparison of the two systems occurs if direct detection's diversity is fixed to the optimal value N_{opt} (the value at which direct detection has its best performance in the absence of interference), and homodyne detection's diversity is allowed to vary. This power gain approaches a constant as diversity increases because there is a limit to the amount of outage statistical improvement that can be achieved with diversity. We see, though, that the power gain is significant. For example, at an outage probability of 0.01, the power gain approaches 10 dB. Most of the power gain is attained when homodyne detection uses a moderate amount of diversity. In the

Diversity in direct detection and homodyne detection	Power Gain
Absence of interference	
Direct detection and homodyne detection use diversity N	$2\dfrac{\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}}{(1+N_n)(-\ln 2 + \theta_{thresh})}$
Direct detection uses diversity N, homodyne detection uses diversity N_{opt}	$k_1 \exp\left\{ -\sqrt{-2\ln(2P_{outage})\ln\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)} - \dfrac{1}{2}\ln\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right) \right\}$ where $\quad k_1 = 2\left(\dfrac{\theta^{thresh} + \sqrt{2\theta^{thresh}N_{opt}N_n}}{(1+N_n)(-\ln 2 + \theta^{thresh})}\right)\sqrt{1+\dfrac{e^{4\sigma_\chi^2}-1}{N_{opt}}}$ $\cdot \exp\left\{\sqrt{-2\ln(2P_{outage})\ln\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N_{opt}}\right)}\right\}$
Direct detection uses diversity N, homodyne detection uses diversity $N_{opt,GaussianHalfSymbolInterference}$	$c_1 \exp\left\{ -\sqrt{-2\ln(2P_{outage})\ln\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)} - \dfrac{1}{2}\ln\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right) \right\}$ where $c_1 = 2\left(\dfrac{\theta^{thresh} + \sqrt{2\theta^{thresh}N_n}}{(1+N_n)(-\ln 2 + \theta^{thresh})}\right)e^{2\sigma_\chi^2}\exp\left\{\sqrt{-2\ln(2P_{outage})4\sigma_\chi^2}\right\}$
Presence of Interference	
Direct detection and homodyne detection use diversity N	$\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)N$
Direct detection uses diversity N, homodyne detection uses diversity N_{opt}	$\dfrac{N_{opt}\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N_{opt}}\right)}{\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)}$
Direct detection uses diversity N, homodyne detection uses diversity $N_{opt,GaussianHalfSymbolInterference}$	$\dfrac{\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)e^{4\sigma_\chi^2}}{\left(1+\dfrac{e^{4\sigma_\chi^2}-1}{N}\right)}$

Table 5.1: Power gain of homodyne detection (in the presence of canceling interference) over direct detection (in presence of Gaussian interference that is on for the first half of the symbol)
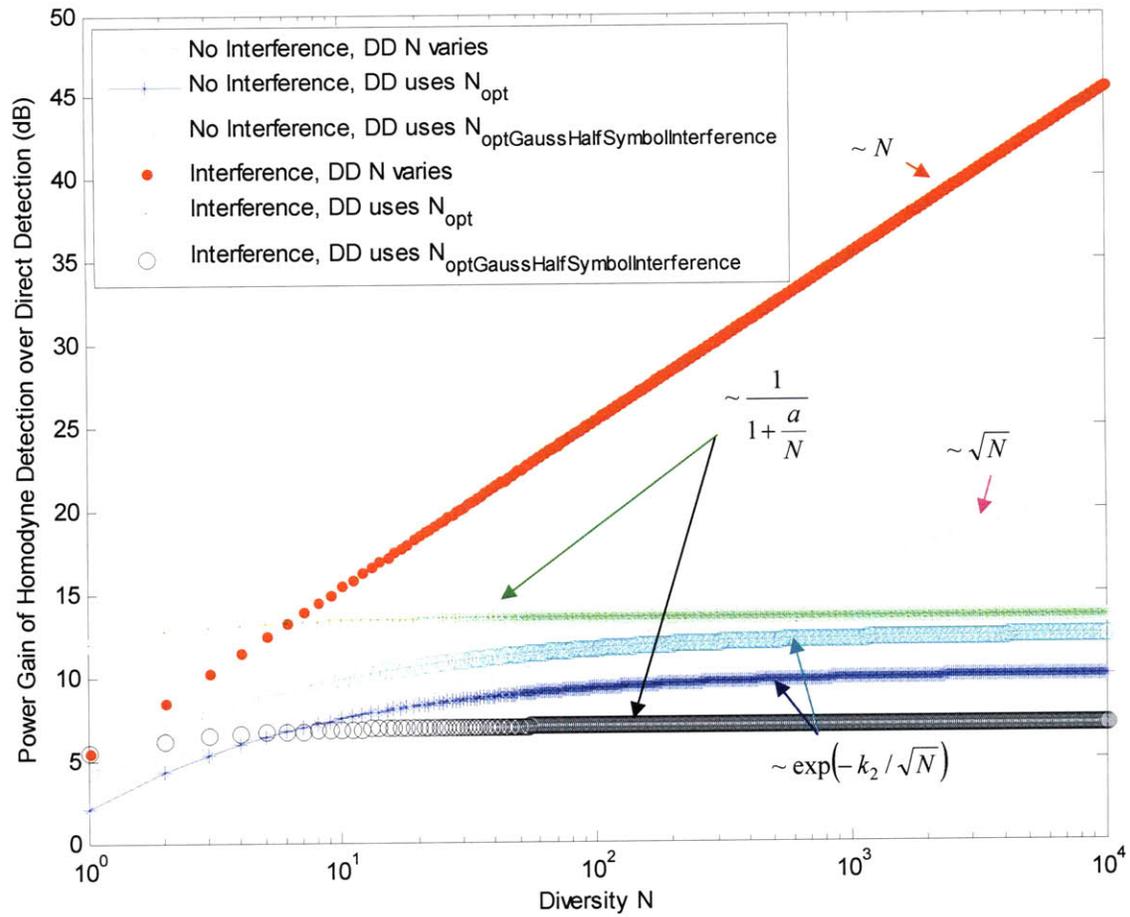
Figure 5.1: Power Gain of homodyne detection over direct detection (if interference is present and it uses the worst case duty cycle) for outage probability of 0.01 when $\sigma_\chi = 0.3$, $P_e^{thresh} = 0.1$, $N_n = 1$. $N_{opt} = 6$.
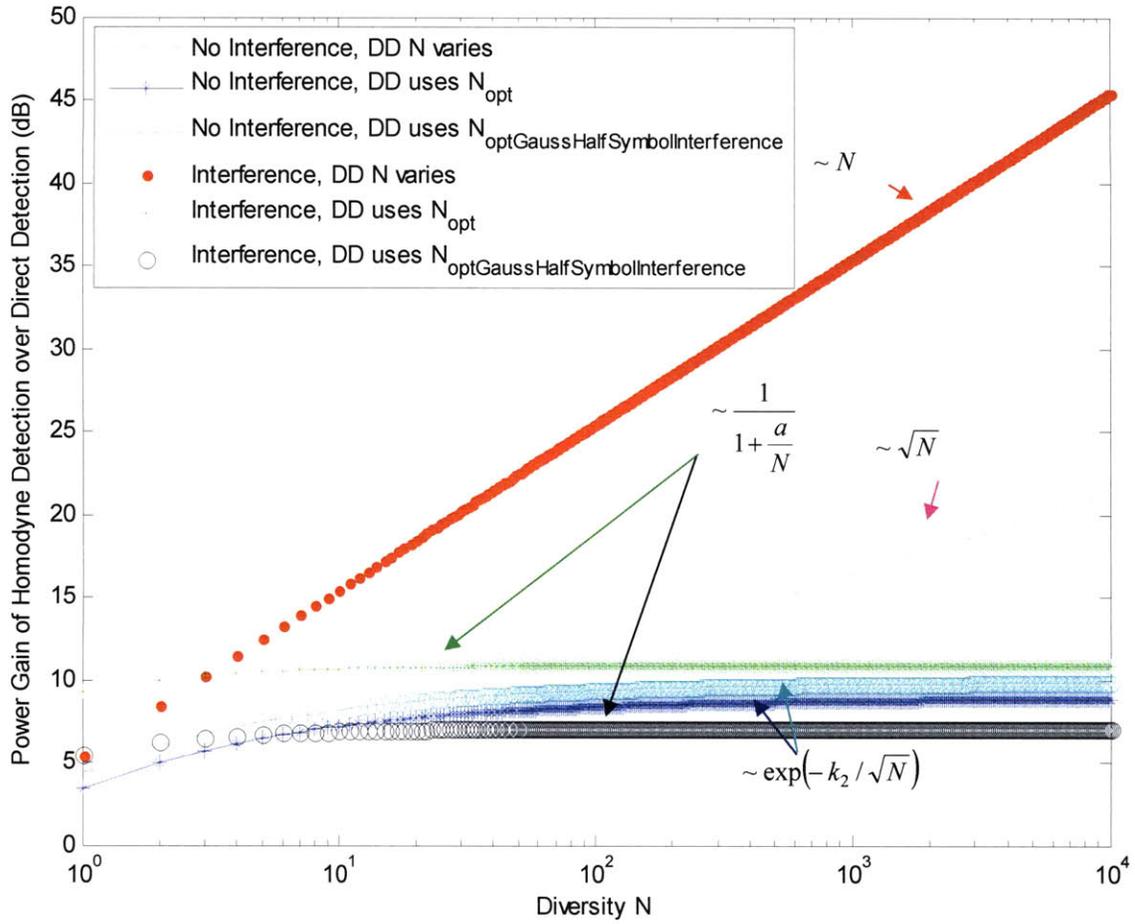
Figure 5.2: Power gain of homodyne detection over direct detection (if interference is present and it uses the worst case duty cycle) for outage probability of 0.1 when σ_χ=0.3, P_e^{thresh}=0.1, N_n=1. N_{opt}=3.

absence of interference, homodyne detection clearly offers better performance over direct detection.

In the *presence* of interference, if direct detection's and homodyne detection's diversities are allowed to increase together, the power gain actually increases proportionally to the diversity. The power gain is even larger than in the absence of interference. The reason for the larger rate of increase in power gain with diversity is due to the increase in unwanted interference detected by direct detection as diversity increases, together with the fact that the interference duty cycle is worst case (and a function of the communication system including diversity). Note that when both systems have a diversity of one, the additional power gain when interference is present over when interference is absent is due to direct detection's Gaussian interference affecting the detected signal variance at the receiver. Consider the case when direct detection's diversity is fixed to be $N_{opt,,GaussianHalfSymbolInterference}=1$ (the value at which direct detection, in the presence of interference, has its best performance), and homodyne detection has arbitrary diversity. Again, the power gain approaches a constant whose value is significant (at an outage probability of 0.01, the power gain approaches 7 dB), and most of that power gain is achieved even when homodyne detection uses a small amount of diversity. In the presence of interference, diversity homodyne detection offers significant performance improvement over diversity direct detection.

Note that if we assume no interference is present and set direct detection's diversity value to be N_{opt}, but interference is actually present, then the power gain of homodyne detection over the fixed diversity direct detection is significantly higher (several dB) than if the actual optimum diversity is used for direct detection in the presence of the interference. On the other hand, if we assume interference exists and

set direct detection's diversity to one, and interference is actually not present, then the power gain is a little higher (a few dB) than if direct detection used N_{opt}. These differences in power gain, if we mistakenly assume no interference exists, or mistakenly assume interference does exist, are due to direct detection having a different optimal diversity value in the presence of and in the absence of interference. Homodyne detection is advantageous in that it is not subject to worse performance if too large of a diversity value is chosen, as is true with direct detection.

Overall, homodyne detection always outperforms direct detection where direct detection was not allowed to use rubber mirror technology to achieve homodyne pre-detection combining. In the presence of worst case interference and turbulence, not only does single receiver homodyne detection perform better than single receiver direct detection, but diversity improves homodyne detection's performance even further, while actually degrading direct detection's performance.

Chapter 6

Transport Layer

In this part of the thesis, we will focus on the Transport Layer's congestion control in a network with free-space optical links.

6.1 TCP Shortcomings in FSO Networks and Motivation for Modified TCP

TCP's congestion control[6] which has been used in the Internet for roughly two decades has been successful in preventing congestion collapse. However, when data rates and geographic spans increase, and communication links without wires or cables (such as satellite and free-space optical links) are added into the network, TCP has performance issues leading to low throughput. This is partly due to the TCP sender's limited rate of window increase. It is also due to the TCP sender unnecessarily reducing its window upon link losses (it assumes every packet loss is due to congestion and that it should thus reduce its window to relieve the congestion).

---

[6] See Chapter 2 for a summary of the TCP sender's congestion control algorithm.

FSO optical links are different from fiber optic and radio frequency (RF) wireless links because they have long fades. It is typical for an outage to last tens of milliseconds even when reasonable amounts of link margin and diversity are used, as discussed in Chapter 3. These long link outages cause a large number of consecutive packets to be dropped and can cause the TCP sender's retransmission timer (RTO) to expire thereby resulting in reduction of the sender's window to one packet in flight without acknowledgement. In fiber optic and RF wireless links, the typical losses are of single packets which usually cause three duplicate ACKs and window halving rather than timeouts. The fact that FSO outages can cause timeouts is an observation that has not been considered before. TCP throughput plots in Section 6.5 show that when high bandwidth-delay product paths are combined with FSO links with atmospheric turbulence, the outages cause severe throughput degradation. The throughput degradation is severe because after a long outage occurs and the sender's window is reduced to one packet in flight, it takes the sender a long time to increase the window to a size large enough to make good use of the available rate in the network.

Others have considered modified versions of TCP to improve throughput. However, none of these variants fix the problem of link losses causing timeouts and drastic window reduction to one packet. Some of the variants try to address link losses that cause only window halving rather than the window being closed to one packet (for example, TCP Westwood [34] which instead of cutting the window in half, cuts the window only to the bandwidth estimate). One variant, TCP Veno [15], monitors the congestion level and reduces the window less drastically if it makes an educated guess that the loss is not due to congestion. TCP Veno will have poor throughput if several outage packet losses occur in a row. Some other TCP variants [6,14,22,25] provide improved additive increase and multiplicative decrease parameters or faster window

140

increase methods to improve the slow window increase and conservative window halving that occurs after a packet loss. None of these variants address outages causing timeouts.

Sending larger packet sizes helps the slow window increase problem by improving the window increase rate by a factor of $b>1$, for packets that are a factor b larger than normal. However, this benefit comes at the cost of possibly more frequent and larger oscillations in the network. Moreover, with larger packet sizes, the sender is also still susceptible to drastic window reduction upon a link outage. The congestion control algorithm Explicit Congestion Protocol (XCP) [23,24] improves the slow window increase problem as we discuss in Chapter 8. However, the XCP sender is also susceptible to window reduction upon timeout, although it has the advantage over other the TCP variant sender in that its window can be increased to larger values within two round-trip times.

The congestion control algorithms proposed in the literature generally assume timeouts are used, just as in TCP, and that the window is closed to one packet upon timeout. Since timeouts cause reduced throughput in FSO networks, let us consider why timeouts are used in congestion control algorithms. Timeouts provide a way for senders to determine if the network has severe congestion. TCP senders assume that all timeouts (and packet losses) are due to congestion and that they should immediately relieve the congestion by reducing their sending rate. Drastic window reduction is an appropriate response to congestion but not to link losses. Thus, in order for senders to achieve a higher throughput, they need to distinguish whether a packet loss is due to link outage or congestion and respond appropriately.

There is a whole class of possible protocols that distinguish whether a packet loss is due to link outage or congestion and thereby help the problem of FSO link outages causing window reduction. In this thesis, we analyze the maximum performance benefit, in the framework of TCP, that can be achieved if the sender distinguishes outage versus congestion losses and reacts appropriately. The analysis gives the best performance that can be attained because it is idealized in the following ways: a) it assumes the sender has perfect knowledge of whether a packet loss is due to outage or congestion, and b) it does not consider the overall impact on network performance caused by any additional packets that are fed back to the sender to help it distinguish outage loss versus congestion loss. In a real network, the performance of a protocol that distinguishes outage loss from congestion loss will be less than the performance shown in the analysis due to imperfect sender knowledge of the cause of a packet loss and the effect of any explicit feedback packets on network congestion. In this thesis, we also consider one possible approach to accomplish the distinguishing of an outage loss from a congestion loss. We discuss this particular design to a moderate level of detail to show what level of complexity is added to the system to support it. It is not a full protocol proposal, but rather is an example of how we may build such a protocol. Our performance analysis of throughput applies not only to the particular design implementation we discuss, but to the entire class of TCP-based protocols that do not reduce their window upon packet loss due to link outage.

Other proposals in the literature that try to improve TCP's performance on wireless or FSO links include Snoop [3], Spoofing [3] and Split-Connection [37]. These schemes have a device in the middle of the path to deceive the TCP sender in a way that prevents the sender from seeing link losses. Snoop does local retransmissions to hide link packet losses from the sender. Snoop would have a problem when used for long distance, high data rate FSO links (for example up to satellite): it would cause

the TCP sender to timeout due to the long retransmission time. Spoofing has a local device in the middle of the path, just prior to the wireless link, send fake acknowledgements to the TCP sender while performing the packet transmissions to the destination itself. Spoofing has the general problem/concern of not having the desired end-to-end semantics of the Transport Layer; they tell the user that packets have been received even if they have not yet been received (and may never be received) by the destination. Split-Connection terminates the TCP sessions before and after the worst error-prone link so as to create consecutive, shorter distance TCP sessions. When a path includes a long distance satellite link, there are devices called performance enhancing proxies (PEPs) that terminate the sessions just prior to and just after the satellite link. However, PEPs have unresolved issues and complexities regarding how to handle/handoff all the incoming and stored packets of the TCP senders if a link coming into or going out of the PEP fails.

6.2 Modification to TCP

Let us consider a modified version of TCP that does not have the problem of reduced throughput due to outages causing window reduction to one packet. Specifically, the Modified TCP does not respond to a large number of outage losses by timing out. Rather, it is able to distinguish whether the packet loss is due to congestion or link outage. One way this can be accomplished is by having the router feed back a 'Congestion Loss' message to the sender for each packet the router drops (and optionally try to guarantee that this feedback gets to the sender by having routers give priority to these feedback messages or by using out-of-band feedback)[7]. If a

---

[7] In-band 'Congestion Loss' feedback in which the router does not do any extra work to give priority to the feedback packets simplifies the amount of processing needed by the router to handle the 'Congestion Loss' feedback. However, it is possible for the network to try to guarantee that the 'Congestion Loss' feedback gets to the sender by having routers give priority to feedback

'Congestion Loss' feedback packet is received, the sender responds to the congestion by reducing its window. If no 'Congestion Loss' feedback packet is received and the RTO expires, the sender assumes the timeout is due to a link outage and the sender does not cut its window. In other words, the sender distinguishes outage loss from congestion loss by explicit feedback for each congestion loss, and assumes any other loss (causing timeout) is due to link outage. For this particular implementation of Modified TCP, see Figure 6.1 for a state diagram of the sender's actions and events that occur to trigger the actions. The events and actions of the Modified TCP that differ from TCP are shown in red.

For the proposed Modified TCP implementation, the reason for considering explicit 'Congestion Loss' feedback packets rather than explicit 'Outage Loss' notification packets to avoid window reduction upon outage is two-fold. Firstly, if outage lengths are long, the 'Outage Loss' feedback may not be generated quickly enough to avoid timeouts. Since outage link losses occur on the link, not at the router, it is only after the outage is over and more packets are getting through that the routers can determine that an outage occurred *and* which senders' packets were lost in the outage. When outages are long, by the time the sender receives the 'Outage Loss' feedback, the sender may have already timed out. Secondly, it is simpler to generate the 'Congestion Loss' feedback because a router can easily determine which sender's packet it is dropping by looking at the address in the header. Although possible, it is practically more challenging for the network to discern outage losses and feed back to the appropriate senders. In order for a router to determine whose packet was dropped in a link outage, the router at the end of the link would need to keep track of

---

packets or by using a dedicated out-of-band channel (with lower rate) for the feedback packets. Giving priority to feedback packets increases the processing required by routers. Using a dedicated out-of-band channel requires a separate channel and processing of that channel and is generally not desirable.

Start

Window starts at 1 packet. No packets have been sent yet.
Number of outstanding (sent but unacknowledged) packets is 0.

(1) Is number of packets in Q < window?

No

Note: For every packet that is sent, it goes into a queue (Q) of outstanding packets. When a packet is removed from the queue, its RTO and Backup-Timer are turned off.

For each packet sent, there are two timers:
$$RTO = RTT + 4\sigma_{RTT}$$
$$Backup\text{-}Timer = RTO + max(E[outage\ length] + \psi\sigma_{outage\ length},\ RTO)$$

Yes

Send a new packet and set its RTO and Backup-Timer

Wait for one of five below events (a,b,c,d,e) to happen then follow that branch

(2) Put packet in queue

Each packet has a timestamp.

Did one of five below events (a,b,c,d,e) happen while sending packet? If yes, follow that branch

No

a

ACK received (for lowest packet number in queue)

Remove packet from Q

Increase window

Go to (1)

b

'Congestion Loss' feedback received

Remove packets in window from Q

Halve window

Go to (1)

c

3 duplicate ACKS received

d

RTO expires

(Packet was dropped in outage or dropped due to congestion but feedback was lost)

Do not cut window

Remove packet from Q and ignore any future duplicate ACKs corresponding to the removed packet

Resend packet. Reset its RTO but do not reset its Backup-Timer (it continues running).

Go to (2)

e

Backup-timer expires

Remove packets in window from Q.

Cut window to 1 packet

Go to (1)

Figure 6.1 Sender State Diagram of the Proposed Implementation of Modified TCP

how many packets were missing in the link and communicate with the router at the sending end of the link which would track the sender's address of each packet[8].

This scheme that we propose of feeding back a 'Congestion Loss' message for each packet that is dropped due to congestion is different from Explicit Congestion Notification (ECN) [39]. ECN is a congestion avoidance scheme in which the router sets the ECN bit in packets if congestion is building at the router. The idea is that the router backs off TCP senders before congestion becomes severe enough that the router drops packets. But when routers actually drop packets, no explicit feedback goes to the senders (unlike the Modified TCP implementation we propose in which explicit feedback is sent to the sender if the router drops a packet). In TCP with ECN, senders react to any loss or ECN-marked packet by cutting their window. If congestion starts mildly then gradually gets severe, a congestion loss is usually preceded by some ECN bits being set in the returned ACKs. However, it has been shown that ECN is not a good predictor of congestion losses [5], and thus trying to differentiate outage loss from congestion loss based on the recent history of ECN bits being set may not provide desired results.

The Modified TCP sender's ability to distinguish outage losses and congestion losses (and respond appropriately) improves sender throughput, as we show in this thesis. For the analysis in Sections 6.3 to 6.5, we assume perfect feedback of the 'Congestion Loss' messages to find the maximum performance that is possible if the sender is always able to correctly distinguish whether a packet loss is due to outage or congestion (and respond appropriately). We will discuss the effect of the loss of the 'Congestion Loss' feedback packets on the protocol design next.

---

[8] Although more complex, a scheme similar to this is discussed in Chapter 7 in the context of deploying only a few specialized devices (with significant complexity) in a heterogeneous network to provide the 'Congestion Loss' feedback rather modifying all routers to provide the 'Congestion Loss' feedback.

Consider what happens if the explicit 'Congestion Loss' feedback is not guaranteed to be received by the sender. In the extreme case that severe congestion causes all of a sender's packets to be dropped, and all of the 'Congestion Loss' feedback packets are lost, the sender will not reduce its window. However, the sender should have a way to reduce its rate drastically in this case. We propose the sender have an additional timer which we call the "Backup-Timer". Upon expiration of the Backup-Timer, the sender assumes there is severe congestion and reduces its window size to one packet. The Backup-Timer duration should be large enough that when a link outage occurs, it does not expire and gives the sender enough time to retransmit the packet, receive an acknowledgement and increase the window (thereby preventing Backup-Timer from expiring). If a link outage causes the loss of a series of packets but subsequent packets get through after the outage, there will be no received packets for a round-trip time plus the outage length. Thus, the Backup-Timer duration should be at least the RTO plus the outage length in order for it to not expire due to the outage. The RTO is used instead of the estimate of RTT in order to allow for variability in the round-trip time, just as the RTO does. In order to allow the sender to retransmit the packet and receive an acknowledgement before Backup-Timer expires, the Backup-Timer duration should be at least 2 RTO. Thus, we should have

Backup-Timer = RTO + max(E[outage length] + $\psi\sigma_{outage\ length}$, RTO)

where $\sigma_{outage\ length}$ is the standard deviation of the outage length and ψ is to be optimized. If the value of ψ is selected appropriately, the expected value of the outage length plus ψ times the outage length standard deviation would capture most of the outage lengths. As discussed, the performance analysis of the Modified TCP in Sections 6.3 to 6.5 assumes a perfect feedback channel in order to show how much

147

performance gain can be achieved by differentiating outage losses versus congestion losses. Thus, in the analysis, the Backup-Timer is not used and ψ is not optimized.

The Backup-Timer will expire if all the 'Congestion Loss' feedback packets and ACKs are lost for the duration of the Backup-Timer. We discuss in Chapter 7, the effect of the Backup-Timer expiring on network congestion. We also discuss in Chapter 7 the effect of providing 'Congestion Loss' feedback packets on network congestion and router design.

The expected outage length and standard deviation used in the Backup-Timer can be estimated by the Link Layer at the routers and reported to the senders. If the expected outage length plus ψ times the standard deviation is larger than the RTO, this estimation allows the Backup-Timer to have a smaller value when the link conditions are such that the expected outage length is smaller. Consider that, for simplicity, instead of having the mean and standard deviation of the outage lengths estimated, the senders set the Backup-Timer to a value on the large end of possible outage lengths for all turbulence conditions (0.5 seconds, for example). This could make the Backup-Timer unnecessarily long and cause longer response times to actual severe congestion in the event that all the 'Congestion Loss' feedback packets and ACKs are lost.

In order to decrease the expected outage length (and thus the length of the Backup-Timer if $E[\text{outage length}] + 4\sigma_{\text{outage length}} > \text{RTO}$), one may add receiver diversity on the free-space optical links. We have shown in Chapter 3 that diversity detection reduces the expected outage length. We show Figure 3.6b again below. In this figure, we plot the expected outage length for a reasonable amount of link margin of 5 dB and for diversity values up to a reasonable value of 25. For coherent detection systems,

increasing diversity to 25 decreases the outage length by more than an order of magnitude.
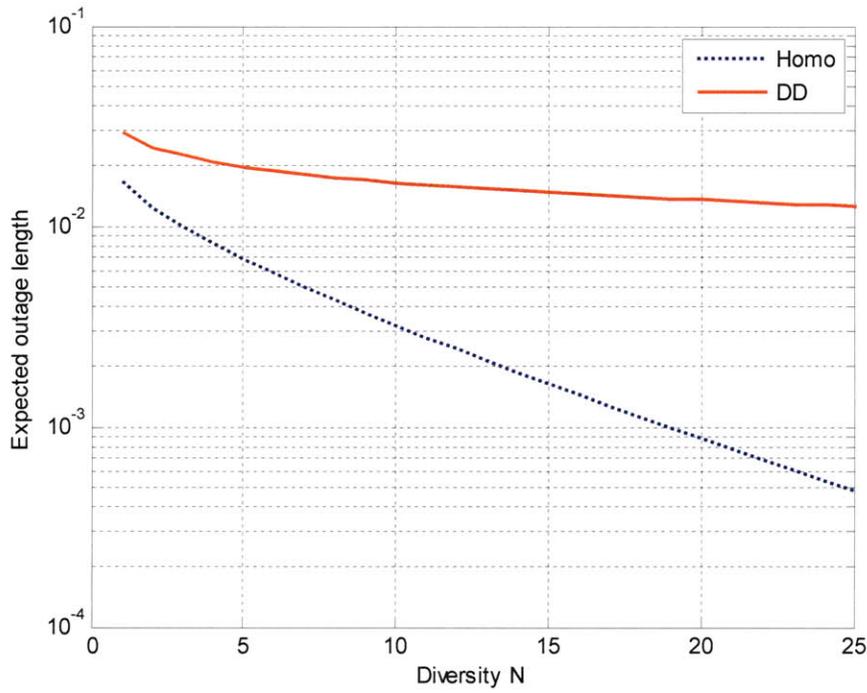
Figure (reproduction of Figure 3.6b): Expected outage length versus diversity for fixed link margin of 5 dB when σ_χ^2=0.5, P_e^{thresh}=0.1, N_n=1, transverse wind speed is 10 km/hr, path distance=20km

If it were possible to increase the link margin to very large values (more than several dB) and/or the transverse wind speed were large enough (perhaps due to non-stationary link endpoints that move at high speeds), the expected outage length may be decreased such that outages typically cause single or a few consecutive packet losses. In this case, outages would typically cause TCP to cut its window in half instead of to one packet (provided subsequent packets result in ACKs to be received). However, TCP would still suffer from reduced throughput due to the window being reduced by half. In this case, the Modified TCP sender could additionally include the functionality of not halving its window upon three duplicate ACKs unless a

'Congestion Loss' message was first received for the missing packet. We do not study this variant in this thesis.

For the Modified TCP implementation we discussed, if packets are dropped in an outage and the window is not reduced, the sender continues to transmit packets (using the same window size) every RTO. However, even though the window is not reduced, the sender is able to respond to congestion and reduce its window if a 'Congestion Loss' feedback packet or three duplicate ACKs are received or the Backup-Timer expires (see Figure 6.1).

In the analysis in Sections 6.3-6.5, we use the words "Modified TCP" to refer to the TCP-based protocol that perfectly distinguishes outage loss from congestion loss and does not reduce its window in response to an outage loss. The analysis gives the best case performance of the class of TCP-based protocols that distinguish outage loss from congestion loss.

6.3 Steady State Analysis (Average Throughput)

In this section, we analyze TCP's and Modified TCP's throughput over atmospheric optical links. In the literature, TCP throughput is analyzed assuming packets are lost with independent probability. However, as discussed in Chapter 2, this is not a good assumption for atmospheric optical links. For atmospheric links operating at high data rates, the loss probabilities of consecutive packets are highly correlated i.e. the channel has memory. In order to obtain more realistic results of TCP's throughput in a network that uses atmospheric optical links, it is essential that the channel model incorporates the channel's memory. In Chapter 2, we described an appropriate FSO channel model (for a no-diversity system), namely a two–state continuous-time

Markov process where the states represent the channel as being in an outage, or not in an outage. For links with diversity, this Markov channel model may only be a crude approximation; an appropriate channel model needs further investigation. Assuming we have a peak power constraint, there is little hope for communicating through an outage. In our analysis, we assume that packets received during an outage are lost, and packets received during a non-outage are received correctly.

We analyze the *throughput* (number of bits or packets sent per unit time) of TCP and Modified TCP over atmospheric optical links rather than analyzing the actual number of correctly received bits per unit time, commonly known as *goodput*. This is in order to analyze the performance of the window closing mechanism in atmospheric optical channels and the loss in throughput that it causes. Although not the focus in this thesis, the loss in goodput due to retransmitted packets may be significant in the Go-Back-N style retransmissions of TCP but may be improved by using select repeat retransmissions.

In order to derive the TCP and Modified TCP throughput bounds, we first define the following quantities:

R_{max} = sender's maximum transmission rate [bits per second]

G = packet size [bits per packet]

τ_{pkt} = time to transmit a packet [s]

RTT   = round-trip time [s]

M   = the maximum possible number of packets in flight

RTO   = TCP's retransmission timeout value

          = estimated RTT + 4 x (estimated standard deviation of RTT)

Then

$$\tau_{pkt} = \frac{G}{R_{\max}} \text{ and} \tag{6.1}$$

$$M = \frac{R_{\max} \cdot RTT}{G} \tag{6.2}$$

For high data rate communication over physically long optical channels, even if there are no packet errors, TCP's standard maximum window size ($W_{\max}$) of 2^{16} bytes severely limits the throughput. This is because the sender is not able to send enough packets to fully use the link while waiting for ACKs. For example, at a data transmission rate of 1 Gb/s over a GEO satellite link, in a vacuum, the average throughput is limited to 2 Mb/s. We assume that a simple fix called TCP window scaling [21] is used so that the maximum window size is not the cause of reduced throughput in TCP over the long links. Moreover, we assume that the maximum window size is set to be the maximum possible number of packets in flight[9].

6.3.1   TCP Throughput

In actual operation, as described in Section 2.4, TCP is sometimes in the Slow Start (exponential increase) phase and sometimes in the Congestion Avoidance (linear increase) phase. Linear window increase allows for fewer packets to be sent per unit time compared to exponential window increase. Thus, letting the window increase be linear yields a lower bound on TCP throughput, and letting the window increase be exponential yields an upper bound.

---

[9] Whether the maximum window size is set to the maximum possible number of packets in flight, M, or to a value larger than this, the throughput bounds we find in this thesis are roughly the same. The difference if we allow the maximum window size to be larger than M is the value to which the window cuts upon window halving when the window is larger than M. For small round-trip times, the throughput does not change much since the window increases back up to M (or higher) quickly. For long and medium RTTs, the throughput does not change much since the window rarely reaches M or beyond due to getting closed as a result of congestion or outage losses.

In this thesis (and as we published in [29]), we model TCP's window evolution by a discrete time Markov chain in which transitions occur every round-trip time[10], and the states represent a measure of the window size. We depict in Figures 6.2 and 6.3, Markov chains that assume linear window increase and exponential window increase respectively. For linear window increase, state n represents a window size of n, and the maximum number of states in the Markov chain is $n_{max}=M$. In the case of exponential window increase, state n represents a window size of 2^{n-1} and the maximum number of states is $n_{max}=\log_2(M)+1$.
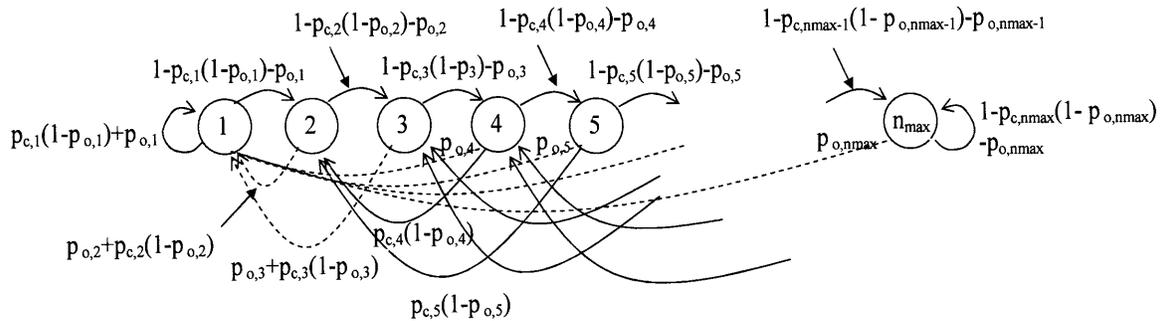
Figure 6.2: TCP linear increase (lower bound) Markov chain. Timeouts (and thus window reduction to one packet) are represented by dotted transitions. The dotted transitions leaving states 2 and 3 also include window halving.
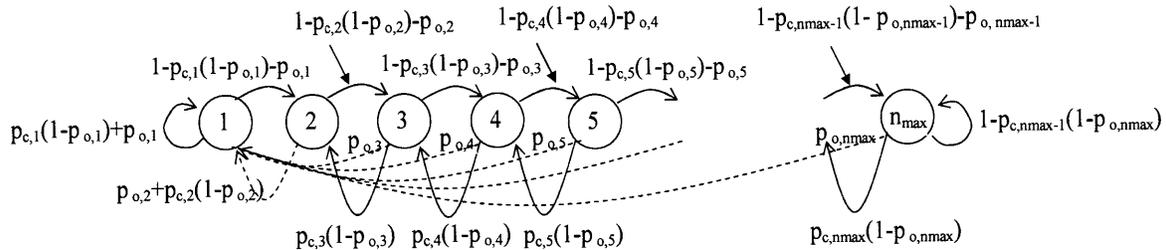
Figure 6.3: TCP exponential increase (upper bound) Markov chain. Timeouts (and thus window reduction to one packet) are represented by dotted transitions. The dotted transition leaving state 2 also includes window halving.

---

[10] We discuss this model of transitions occurring every round-trip time in detail is Section 6.3.1.1.

The probability of congestion packet loss and outage packet loss of any of the transmitted packets while in state n are denoted by $p_{c,n}$ and $p_{o,n}$ respectively. If no outage or congestion loss occurs in the packets sent in any particular window, the window size increases (by one for linear window increase and by a factor of two for exponential window increase). In Figures 6.2 and 6.3, this is represented by the transitions from state n to n+1 for n=1 to $n_{max}-1$. For state n_{max}, if there is no outage or congestion loss in the packets sent in the window, there is a self transition. We assume that if a congestion loss occurs with no outage loss, the window is halved[11], and if an outage loss occurs, the RTO expires and the window is reduced to one packet[12].

There are many different possibilities for modeling congestion losses in the network. In one extreme, we can model each packet as being dropped due to congestion with independent probability. This may be a reasonable model if there are a large number of users in the network. In the other extreme, we can model the probability of congestion loss as 0 below a given window size and 1 above the window size. i.e model congestion losses by a step function. This is a reasonable model if there is only one sender going over links with possibly different rates. Provided there is enough buffer space to queue up the packets due to the different link rates, there will be no packet losses due to congestion while the sender is sending at a window size that the smallest rate link can handle. However, as the sender increases its window to larger values, it will ultimately cause packet drops by the router prior to the smallest rate link. We take these two extremes in modeling congestion losses (where in the latter

---

[11] To include the probability that a congestion loss causes a timeout, one would add an additional probability to all of the transitions that correspond to a timeout.

[12] An outage would cause a timeout if it causes the loss of the last packet(s) in a window or the loss of all packets in a window such that the sender does not receive any ACKS for an RTO. An outage would also cause a timeout if the outage causes the loss of a series of consecutive packets in the beginning (or middle) of the window and either a) the outage is longer than RTO-RTT or b) the subsequent packets encounter delay such that the outage plus delay is longer than RTO-RTT and they cannot return triple duplicate acknowledgements before a timeout occurs. In case b), if the outage is less than RTO-RTT i.e. RTO-RTT is not small, then congestion must be building to cause it to not be small, thereby making it likely for there to be additional delays in the subsequent packets.

case, the threshold window size that causes the congestion packet loss is n_{max}). In the former case where each packet is dropped with independent probability, the probability of congestion loss in state n is given by

$$p_{c,n} = \Pr(\text{at least one of the packets sent in state n is dropped due to congestion})$$
$$= 1 - \Pr(\text{none of the packets sent in stage n is dropped due to congesiton})$$
$$= \begin{cases} 1 - \left(1 - p_{congperpkt}\right)^{n} & \text{for linear window increase Markov chain} \\ 1 - \left(1 - p_{congperpkt}\right)^{2^{n-1}} & \text{for exponential window increase Markov chain} \end{cases}$$

(6.3)

where $p_{congperpkt}$ is the probability that any given packet is dropped due to congestion.

We derive the value of $p_{o,n}$, the probability of outage loss in state n, in Section 6.3.1.1 using the 2-state channel model discussed in Chapter 2. Incorporating the channel's memory into the analysis of the Transport Layer performance is one of the unique aspects of this work.

The average throughput of TCP in steady state is given by

$$\text{steady state TCP throughput} = \frac{E[\text{window size}]}{RTT}$$

(6.4)

Using the linear window increase Markov chain, we obtain a lower bound on TCP throughput at steady state

$$\text{steady state TCP throughput} \geq \frac{\sum_{n=1}^{n_{max}} n \pi_{n}^{linear}}{RTT} \text{ packets/sec}$$

(6.5)

155

where the superscript "linear" denotes the linear window increase Markov chain, and π_n is the steady state probability of being in state n. Using the exponential window increase Markov chain, we obtain an upper bound on TCP throughput at steady state

$$\text{steady state TCP throughput} \leq \frac{\sum_{n=1}^{n_{\max}} 2^{n-1} \pi_n^{\text{exp}}}{RTT} \text{ packets/sec} \qquad (6.6)$$

where the superscript "exp" denotes the exponential window increase Markov chain.

The global balance equations for the (lower bound) TCP linear increase Markov chain of Figure 6.2 are

$$\pi_1^{linear}\left[1-\left(p_{c,1}\right)\!\left(1-p_{o,1}\right)-p_{o,1}\right]=\pi_2^{linear}\left[\left(p_{c,2}\right)\!\left(1-p_{o,2}\right)+p_{o,2}\right]+\pi_3^{linear}\left[\left(p_{c,3}\right)\!\left(1-p_{o,3}\right)+p_{o,3}\right]+\sum_{i=4}^{n_{\max}}\pi_i^{linear}p_{o,i}$$

$$\pi_n^{linear}=\pi_{n-1}^{linear}\left[1-\left(p_{c,n-1}\right)\!\left(1-p_{o,n-1}\right)-p_{o,n-1}\right]+\pi_{2n}^{linear}\left(p_{c,2n}\right)\!\left(1-p_{o,2n}\right)+\pi_{2n+1}^{linear}\left(p_{c,2n+1}\right)\!\left(1-p_{o,2n+1}\right)$$
$$\text{for } n = 2,...,n_{\max}/2-1$$

$$\pi_{n_{\max}/2}^{linear}=\pi_{n_{\max}/2-1}^{linear}\left[1-\left(p_{c,n_{\max}/2-1}\right)\!\left(1-p_{o,n_{\max}/2-1}\right)-p_{o,n\max/2-1}\right]+\pi_{n_{\max}}^{linear}\left(p_{c,n_{\max}}\right)\!\left(1-p_{o,n_{\max}}\right)$$

$$\pi_n^{linear}=\pi_{n-1}^{linear}\left[1-\left(p_{c,n-1}\right)\!\left(1-p_{o,n-1}\right)-p_{o,n-1}\right] \quad \text{for } n = n_{\max}/2+1,...n_{\max}-1$$

$$\pi_{n_{\max}}^{linear}\left[\left(p_{c,n_{\max}}\right)\!\left(1-p_{o,n_{\max}}\right)+p_{o,n_{\max}}\right]=\pi_{n_{\max}-1}^{linear}\left[1-\left(p_{c,n_{\max}-1}\right)\!\left(1-p_{o,n_{\max}-1}\right)-p_{o,n_{\max}-1}\right]$$

$$(6.7)$$

where π_n^{linear} is the steady state probability of being in state n in the Markov chain of Figure 6.2. The sum of the steady probabilities is 1 i.e.

$$1 = \sum_{n=1}^{n_{max}} \pi_n^{linear} \tag{6.8}$$

Solving for the $n_{max}+1$ equations in (6.7) and (6.8) gives the steady state probabilities of the TCP linear increase Markov chain as a function of the outage and congestion loss probabilities.

Similarly, the global balance equations for the (upper bound) TCP exponential increase Markov chain of Figure 6.3 are

$$\pi_1^{exp}\left[1 - \left(p_{c,1}\right)\left(1 - p_{o,1}\right) - p_{o,1}\right] = \pi_2^{exp}\left(p_{c,2}\right)\left(1 - p_{o,2}\right) + \sum_{i=3}^{n_{max}} \pi_i^{exp} p_{o,i}$$

$$\pi_n^{exp} = \pi_{n-1}^{exp}\left[1 - \left(p_{c,n-1}\right)\left(1 - p_{o,n-1}\right) - p_{o,n-1}\right] + \pi_{n+1}^{exp}\left(p_{c,n+1}\right)\left(1 - p_{o,n+1}\right) \qquad \text{for } n = 2,...,n_{max} - 1 \tag{6.9}$$

$$\pi_{n_{max}}^{exp}\left[\left(p_{c,n_{max}}\right)\left(1 - p_{o,n_{max}}\right) + p_{o,n_{max}}\right] = \pi_{n_{max}-1}^{exp}\left[1 - \left(p_{c,n_{max}-1}\right)\left(1 - p_{o,n_{max}-1}\right) - p_{o,n_{max}-1}\right]$$

and

$$1 = \sum_{n=1}^{n_{max}} \pi_n^{exp} \tag{6.10}$$

where π_n^{exp} is the steady state probability of being in state n in the Markov chain of Figure 6.3. Solving for the $n_{max}+1$ equations in (6.9) and (6.10) gives the steady state probabilities of the TCP exponential increase Markov chain as a function of the outage and congestion loss probabilities.

6.3.1.1 Outage Transition Probabilities for the Markov Chains that Model the TCP Sender's Window Progression

In this sub-section, we use the 2-state channel model discussed in Section 2.2, to derive the outage transition probabilities $p_{o,n}$ in the TCP Markov chains in Figure 6.2 and 6.3. These outage transition probabilities are found as a function of channel parameters assuming outages occur only on the forward links (on the links going toward the destination node). If outages also occur on the backward links, the outage transition probabilities would be larger than those derived here, and the throughput bounds would also be lower.

The TCP Markov chain models in Figures 6.2 and 6.3 assume that state transitions occur every RTT i.e. if we think of time as being the concatenation of discrete multiples of round-trip times, all window halving and reductions to one packet occur at the round-trip time boundaries. In other words, the model assumes that the sender has perfect and immediate knowledge of whether its transmitted data is lost (without knowing the cause). So if packet losses occur due to congestion or outage in any RTT time segment, the window is reduced at the start of the next RTT. We believe this is a reasonable approximation of the operation of TCP for purposes of calculating steady state throughput. TCP senders find out about packet losses one round-trip time (or RTO) later, and window halving and timeouts may not occur exactly at round-trip time boundaries, but rather in the middle of the boundaries. Assuming that the window reductions occur at the boundaries means we are ignoring the time from the boundary to the actual detection of packet loss. Provided these timeouts and window halving do not occur every round-trip time, but rather infrequently, the assumption that the window reductions occur at the boundaries does not have an appreciable effect on the calculation of throughput.

Let conceptualize time as the concatenation of round-trip time segments or 'stages' and term the round-trip time segments corresponding to state n as 'stage n'. Let S_n represent whether stage n starts in an outage or non-outage. The outage transition probabilities are given by

$$p_{o,n} = \left(p_{o,n|S_n=outage}\right)\Pr\left(S_n = outage\right) + \left(p_{o,n|S_n=non-outage}\right)\Pr\left(S_n = non-outage\right) \quad \forall n. \quad (6.11)$$

By random incidence, we approximate $\Pr\left(S_1 = non-outage\right) \cong 1 - P_{outage}$. Since we assume that outages are long such that when packets are lost in an outage, a timeout occurs, the probability of a timeout in stage n given that stage n starts in an outage is approximately 1. i.e.

$$\begin{aligned}p_{o,n|S_n=outage} &= \Pr\left(\text{outage in stage n causes timeout} \mid S_n = outage\right) \\ &\cong 1\end{aligned} \quad (6.12)$$

Given that stage n starts in a non-outage, for linear window increase, the probability of a timeout in stage n is given by

$$\begin{aligned}p_{o,n|S_n=non-outage} &= \Pr\left(\text{outage in stage n causes timeout} \mid S_n = \text{non - outage}\right) \\ &= \Pr\left(\text{non - outage length} < n\tau_{pkt}\right) \\ &= \int_0^{n\tau_{pkt}} v_{12} \exp\left(-v_{12}z\right)dz \\ &= 1 - \exp\left(-v_{12}n\tau_{pkt}\right)\end{aligned} \quad (6.13)$$

If stage n starts in a non-outage, from (6.13) we can see that for larger n i.e. larger window sizes, the probability of a timeout increases. Now let us find the probability that stage n starts in a non-outage. If stage n is reached, then the previous stage was stage n-1 and there was no outage in that stage n-1 that caused packet loss and

159

timeout. If there was an outage that caused packet losses in stage n-1, then the next stage would be stage 1 not stage n. Thus, for stage n, we know that at the end of the last packet in the previous stage (stage n-1), the system was in a non-outage. Using the 2-state Markov channel model, we derive in Appendix A that the probability that the channel is in a non-outage (or outage) t time units in the future given that the process is currently in a non-outage is given by

$$P_{11}(t) = 1 - P_{outage} + P_{outage} \cdot \exp\left\{-\left(v_{12} + v_{21}\right)t\right\} \text{ and} \qquad (6.14)$$

$$P_{12}(t) = P_{outage} - P_{outage} \cdot \exp\left\{-\left(v_{12} + v_{21}\right)t\right\} \qquad (6.15)$$

respectively. Thus,

$$
\begin{aligned}
\Pr(S_n = non-outage) &= P_{11}\left(RTT - (n-1)\tau_{pkt}\right) \\
&= 1 - P_{outage} + \left(P_{outage} \cdot \exp\left\{-(v_{12} + v_{21})\left[RTT - (n-1)\tau_{pkt}\right]\right\}\right)
\end{aligned} \qquad (6.16)
$$

and

$$
\begin{aligned}
\Pr(S_n = outage) &= P_{12}\left(RTT - (n-1)\tau_{pkt}\right) \\
&= 1 - P(S_n = non-outage) \\
&= P_{outage} - \left(P_{outage} \cdot \exp\left\{-(v_{12} + v_{21})\left[RTT - (n-1)\tau_{pkt}\right]\right\}\right)
\end{aligned} \qquad (6.17)
$$

for n≥2. Substituting (6.12), (6.13), (6.16) and (6.17) into (6.11), the transition probabilities are given by

$$p_{o,1} = 1 - \left(1 - P_{outage}\right)\exp\left(-v_{12}\tau_{pkt}\right)$$

$$
\begin{aligned}
p_{o,n} = 1 &- \left[\left(1 - P_{outage}\right) \cdot \exp\left(-v_{12}n\tau_{pkt}\right)\right] \\
&- \left[P_{outage} \cdot \exp\left\{-v_{12}\left(RTT + \tau_{pkt}\right) - v_{21}\left[RTT - (n-1)\tau_{pkt}\right]\right\}\right] \quad \text{for n} \geq 2
\end{aligned} \qquad (6.18)
$$

Similarly, for exponential window increase,

$$p_{o,n|S_n=non-outage} = \Pr\left(\text{outage in stage n causes timeout} \,|\, S_n = \text{non - outage}\right)$$

$$= 1 - \exp\left(-v_{12}2^{n-1}\tau_{pkt}\right) \quad \text{for } n = 1,2,\ldots,n_{\max} \tag{6.19}$$

and

$$\Pr\left(S_n = outage\right) = P_{outage}\left(1 - \exp\left\{-(v_{12}+v_{21})\left(RTT - 2^{n-2}\tau_{pkt}\right)\right\}\right) \text{ for } n = 2,3,\ldots,n_{\max} \tag{6.20}$$

So the transition probabilities are given by

$$p_{o,n} = \begin{cases} 1 - \left(1 - P_{outage}\right)\cdot \exp\left(-v_{12}\tau_{pkt}\right) & \text{for } n = 1 \\ 1 - \left[\left(1 - P_{outage}\right)\exp\left(-v_{12}2^{n-1}\tau_{pkt}\right)\right] - \left[P_{outage}\cdot\exp\left\{-(v_{12}+v_{21})\left(RTT + 2^{n-2}\tau_{pkt}\right)\right\}\right] & \text{for } n = 2,\ldots,n_{\max} \end{cases}$$

$$\tag{6.21}$$

6.3.2 Modified TCP Throughput

We depict in Figures 6.4 and 6.5 the Markov chains for Modified TCP assuming linear window increase and exponential window increase respectively. The transitions occur every RTT and the states represent a measure of the window size. Just as for TCP, for the Modified TCP linear window increase chain in Figure 6.4, state n represents a window size of n, and the maximum number of states in the Markov chain is $n_{\max}=M$. Also, for the Modified TCP exponential window increase chain in Figure 6.5, state n represents a window size of 2^{n-1} and the maximum number of states is $n_{\max}=\log_2(M)+1$. The $p_{c,n}$ are the same as in the TCP Markov chains.
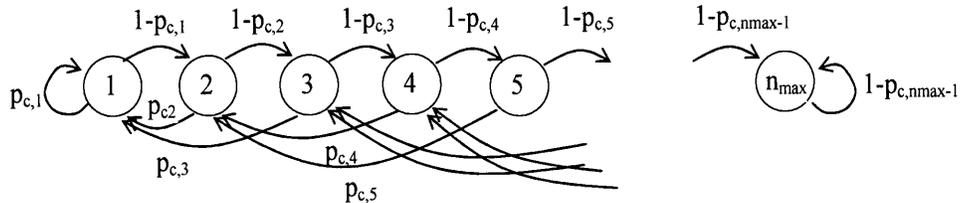
Figure 6.4 Modified TCP linear increase Markov chain
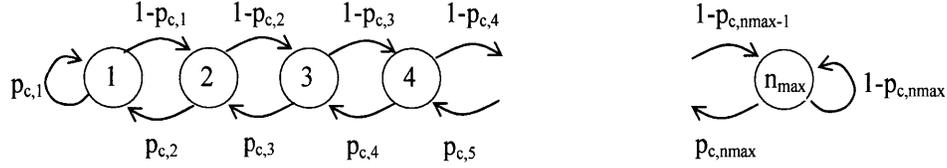
Figure 6.5 Modified TCP upper bound (exponential increase) Markov chain

If no outage or congestion loss occurs in the packets sent in any particular window, the process transitions from state n to n+1 for n=1 to n_{max}-1 and self transitions when in state n_{max}. If a congestion loss occurs, the window is halved. The key difference between the Modified TCP and TCP Markov chains is that if an outage occurs, the window is not reduced at all in the Modified TCP chains since the sender knows the loss is not due to congestion.

The steady state throughput of the Modified TCP is given by

$$\text{steady state Modified TCP throughput} = \frac{E[\text{window size}]}{RTT} \qquad (6.22)$$

Using the Modified TCP linear window increase Markov chain, a lower bound of Modified TCP steady throughput can be calculated using

$$\text{steady state Modified TCP throughput} \geq \frac{\sum_{n=1}^{n_{max}} n\pi_n^{linear}}{RTT} \qquad (6.23)$$

where π_n^{linear} is the steady state probability of being in state n in the Modified TCP Markov chain of Figure 6.4. Without timeouts due to outages, and assuming congestion losses cause window halving rather than RTO timeouts, in steady state,

Modified TCP is always in linear window increase except at the start of the session. Thus, the lower bound is a tight lower bound. The global balance equations for the Modified TCP linear increase Markov chain are

$$\pi_1^{linear} \cdot \left(1 - p_{c,1}\right) = \pi_2^{linear} p_{c,2} + \pi_3^{linear} p_{c,3}$$

$$\pi_n^{linear} = \pi_{n-1}^{linear} \cdot \left(1 - p_{c,n-1}\right) + \pi_{2n}^{linear} \cdot \left(p_{c,2n}\right) + \pi_{2n+1}^{linear} \cdot \left(p_{c,2n+1}\right) \quad \text{for n} = 2,\dots,\left(\frac{n_{max}}{2} - 1\right)$$

$$\pi_{n_{max}/2}^{linear} = \pi_{n_{max}/2-1}^{linear} \cdot \left[1 - p_{c,n_{max}/2-1}\right] + \pi_{n_{max}}^{linear} \cdot \left(p_{c,n_{max}}\right) \qquad (6.24)$$

$$\pi_n^{linear} = \pi_{n-1}^{linear} \cdot \left(1 - p_{c,n-1}\right) \quad \text{for n} = \left(\frac{n_{max}}{2} + 1\right),\dots,\left(n_{max} - 1\right)$$

$$\pi_{n_{max}}^{linear} p_{c,n_{max}} = \pi_{n_{max}-1}^{linear} \cdot \left[1 - p_{c,n_{max}-1}\right]$$

and the sum of the steady state distributions is 1.

$$1 = \sum_{n=1}^{n_{max}} \pi_n^{linear} \qquad (6.25)$$

Solving for the steady state probabilities using the equations in (6.24) and (6.25), we can calculate the lower bound on Modified TCP's steady state throughput using (6.23).

The detailed balance equations for the Modified TCP exponential increase Markov chain are

$$\pi_{n-1}^{exp}\left(1 - p_{c,n-1}\right) = \pi_n^{exp} p_{c,n} \quad \text{for n=2,3,\dots,}n_{max} \qquad (6.26)$$

163

where $\pi_n^{\exp}$ is the steady state probability of being in state n in the Modified TCP Markov chain of Figure 6.5. Again, the sum of the steady state distributions is 1.

$$1 = \sum_{n=1}^{n_{\max}} \pi_n^{\exp} \qquad (6.27)$$

The solution to this system of equations given in (6.26) and (6.27) is found in closed form in Appendix H and gives the following steady state equations

$$\pi_1^{\exp} = \left(1 + \sum_{i=2}^{n_{\max}} \prod_{j=1}^{i-1}\left(\frac{1-p_{c,j}}{p_{c,j+1}}\right)\right)^{-1}$$

$$\pi_n^{\exp} = \pi_1^{\exp} \prod_{j=1}^{n-1}\left(\frac{1-p_{c,j}}{p_{c,j+1}}\right) \qquad (6.28)$$

Using the Modified TCP exponential window increase Markov chain, we can calculate a gross upper bound of the Modified TCP steady state throughput as

$$\text{Modified TCP throughput} \leq \frac{\sum_{n=1}^{n_{\max}} n\pi_n^{\exp}}{RTT} \qquad (6.29)$$

This is a loose upper bound because the Modified TCP sender is almost always in linear window increase when congestion losses and outage losses do not cause timeouts.

In the next section, we continue with deriving the throughputs of TCP and Modified TCP, but we find the throughput prior to steady state rather than in steady state.

Plots and discussion of TCP's and Modified TCP's throughput in steady state and prior to steady state are given in Section 6.5.

6.4 Transient Analysis

When a user sends a small or moderate sized file (for example <1 MB), TCP is not necessarily in steady state for most of the file transfer, particularly if the round-trip distance is long. Thus, steady state throughput does not give an accurate estimate of the time it takes to send the file. In this section, we find the transient throughput of TCP and Modified TCP after session initiation.

The expected number of packets sent in the mth round-trip time is given by

$$E\left[\text{number of packets sent in } m^{th} \text{ RTT}\right] = \begin{cases} \displaystyle\sum_{i=1}^{n_{max}} ip_i(m) & \text{for linear increase(lower bound) of both Modified TCP and TCP} \\[2em] \displaystyle\sum_{i=1}^{n_{max}} 2^{i-1} p_i(m) & \text{for exponential increase (upper bound) of both Modified TCP and TCP} \end{cases}$$

(6.30)

where $p_i(m)$ is the probability of being in state i in the mth RTT. The $p_i(m)$ can be obtained from the following evolution of probability distribution across the states:

$$\begin{aligned} \bar{p}(m) &= \bar{p}(m-1)P \\ &= \bar{p}(1)P^{(m-1)} \end{aligned}$$

(6.31)

where $\bar{p}(m)$ is a row vector of probabilities of being in the n_{max} states in the mth RTT, P is the probability transition matrix for the Markov chain, and $P^{(m-1)}$ is the matrix

product of P with itself (m-1) times and represents the transition matrix from any given RTT to (m-1) RTTs later. TCP starts with an initial window size of 1. Thus, $\vec{p}(1) = [1 \quad 0 \quad 0 \quad \dots \quad 0]$.

The expected number of packets sent in K round-trip times is given by:

$$E[\text{number of packets sent in K RTTs}] = \sum_{m=1}^{K} E[\text{number of packets sent in m}^{\text{th}}\text{ RTT}] \qquad (6.32)$$

Thus, we can calculate the expected number of packets sent in the m$^{\text{th}}$ RTT and in K RTTs if we know the transition matrix of the Markov chain. Using Figures 6.2-6.5, the corresponding transition matrices of the:

1) TCP linear increase Markov chain

2) TCP exponential increase Markov chain

3) Modified TCP linear increase Markov chain

4) Modified TCP exponential increase Markov chain

are given below.

Transition matrix of TCP linear increase Markov Chain:

$$P = \begin{bmatrix}
p_{c,1}(1-p_{o,1})+p_{o,1} & 1-p_{c,1}(1-p_{o,1})-p_{o,1} & 0 & 0 & \cdots & & & & & & 0 \\
p_{c,2}(1-p_{o,2})+p_{o,2} & 0 & 1-p_{c,2}(1-p_{o,2})-p_{o,2} & 0 & \cdots & & & & & & \cdot \\
p_{c,3}(1-p_{o,3})+p_{o,3} & 0 & 0 & 1-p_{c,3}(1-p_{o,3})-p_{o,3} & 0 & & & & & & \cdot \\
p_{o,4} & p_{c,4}(1-p_{o,4}) & 0 & 0 & \cdots & & & & & & \cdot \\
p_{o,5} & p_{c,5}(1-p_{o,5}) & 0 & & & & & & & & \cdot \\
\cdot & 0 & p_{c,6}(1-p_{o,6}) & & & & & & & & \\
& & p_{c,7}(1-p_{o,7}) & & & & & & & & \\
& & 0 & & \cdots & & & & & & \\
\cdot & \cdot & & & & 0 & p_{c,n_{\max}-2}(1-p_{o,n_{\max}-2}) & 0 & & & 0 \\
& \cdot & & & & 0 & p_{c,n_{\max}-1}(1-p_{o,n_{\max}-1}) & 0 & & & 1-p_{c,n_{\max}-1}(1-p_{o,n_{\max}-1})-p_{o,n_{\max}-1} \\
& & & & & & 0 & p_{c,n_{\max}}(1-p_{o,n_{\max}}) & 0\cdots & 0 & 1-p_{c,n_{\max}}(1-p_{o,n_{\max}})-p_{o,n_{\max}} \\
p_{o,n_{\max}} & 0 & & \cdots & & & & & & &
\end{bmatrix}$$

where the $p_{c,n_{\max}}(1-p_{o,n_{\max}})$ entry in the last row is in column $\dfrac{n_{\max}}{2}$.

Transition matrix of TCP exponential increase Markov Chain:

$$
P = \begin{bmatrix}
p_{c,1}(1-p_{o,1})+p_{o,1} & 1-p_{c,1}(1-p_{o,1})-p_{o,1} & 0 & 0 & \cdots & & & 0 \\
p_{c,2}(1-p_{o,2})+p_{o,2} & 0 & 1-p_{c,2}(1-p_{o,2})-p_{o,2} & 0 & \cdots & & & \vdots \\
p_{o,3} & p_{c,3}(1-p_{o,3}) & 0 & 1-p_{c,3}(1-p_{o,3})-p_{o,3} & 0 & & & \vdots \\
p_{o,4} & 0 & p_{c,4}(1-p_{o,4}) & 0 & \cdots & & & \vdots \\
\vdots & & & 0 & \cdots & & & \\
\vdots & & & 0 & & & & 0 \\
& & & & & 0 & 1-p_{c,n_{\max}-1}(1-p_{o,n_{\max}-1})-p_{o,n_{\max}-1} & \\
p_{o,n_{\max}} & 0 & & & & 0 & p_{c,n_{\max}}(1-p_{o,n_{\max}}) & 1-p_{c,n_{\max}}(1-p_{o,n_{\max}})-p_{o,n_{\max}}
\end{bmatrix}
$$

Transition matrix of Modified TCP linear increase Markov Chain:

$$
P = \begin{bmatrix}
p_{c,1} & (1-p_{c,1}) & 0 & \cdots & & & & & 0 \\
p_{c,2} & 0 & (1-p_{c,2}) & 0 & \cdots & & & & \cdot \\
p_{c,3} & 0 & 0 & (1-p_{c,3}) & 0 & \cdots & & & \cdot \\
0 & p_{c,4} & 0 & 0 & \cdots & & & & \cdot \\
\cdot & p_{c,5} & 0 & & & & & & \\
\cdot & 0 & p_{c,6} & & & & & & \\
& & p_{c,7} & & & & & & \\
\cdot & & 0 & \cdots & & 0 & & & \\
\cdot & & & 0 & p_{c,n_{\max}-2} & 0 & & & 0 \\
\cdot & & & 0 & p_{c,n_{\max}-1} & 0 & 0 & \cdots & 0 & (1-p_{c,n_{\max}-1}) \\
0 & \cdots & & & 0 & p_{c,n_{\max}} & 0 & \cdots & 0 & (1-p_{c,n_{\max}})
\end{bmatrix}
$$

where the $p_{c,n_{\max}}$ entry in the last row is in column $\dfrac{n_{\max}}{2}$.

Transition matrix of Modified TCP exponential increase Markov Chain:

$$
P = \begin{bmatrix}
p_{c,1} & (1-p_{c,1}) & 0 & 0 & \cdots & & 0 \\
p_{c,2} & 0 & (1-p_{c,2}) & 0 & \cdots & & \cdot \\
0 & p_{c,3} & 0 & (1-p_{c,3}) & 0 & \cdots & \cdot \\
0 & 0 & p_{c,4} & 0 & & & \cdot \\
\cdot & & 0 & \cdots & & & 0 \\
\cdot & & & & & & 0 \\
\cdot & & & & 0 & (1-p_{c,n\,\max-1}) \\
0 & 0 & \cdots & & 0 & p_{c,n\,\max-1} & (1-p_{c,n\,\max})
\end{bmatrix}
$$

Plots of the expected number of packets sent as time progresses give us an estimate of how long it takes, on average, to send a file of a given size. We show plots of the expected number of packets sent in the Kth RTT and in the first K RTTs in Section 6.5.2.

169

Below are analytic expressions for the approximate expected number of packets sent in the Kth round-trip time and in the K round-trip times.

For Modified TCP linear increase Markov chain:

$$E[\text{number of packets sent in } K^{th} \text{ RTT}] \approx K \prod_{i=1}^{K-1} (1 - p_{c,i}) \tag{6.33}$$

For Modified TCP exponential increase Markov chain:

$$E[\text{number of packets sent in } K^{th} \text{ RTT}] \approx 2^{K-1} \prod_{i=1}^{K-1} (1 - p_{c,i}) \tag{6.34}$$

For TCP linear increase Markov chain:

$$E[\text{number of packets sent in } K^{th} \text{ RTT}] \approx K \prod_{i=1}^{K-1} (1 - p_{c,i}(1 - p_{o,i}) - p_{o,i}) \tag{6.35}$$

For TCP exponential increase Markov chain:

$$E[\text{number of packets sent in } K^{th} \text{ RTT}] \approx 2^{K-1} \prod_{i=1}^{K-1} (1 - p_{c,i}(1 - p_{o,i}) - p_{o,i}) \tag{6.36}$$

where the $p_{c,i}$ and $p_{o,i}$ correspond to the probabilities in the applicable Markov chain.

These approximate expressions represent only the contribution of the probability that transitions from state 1 to state K by progressing from state 1 to each higher state until reaching state K (without window halving or window cutting to 1 in the process). In other words, the expressions omit the contribution of the probability distribution in the states lower than K. As we will see in plots in Section 6.5.2, these approximate expressions are good for the first few time steps for both Modified TCP and TCP particularly for small congestion probabilities and packet loss due to turbulence.

170

6.5 Performance of TCP and Modified TCP

6.5.1 Steady State Throughput of TCP and Modified TCP

In this section, we show plots of and discuss the steady state throughput of TCP and Modified TCP. Using the Markov chain analysis in Section 6.3, we show in Figures 6.6-6.40 the TCP and Modified TCP throughput bounds for a wide range of congestion loss probabilities[13] and atmospheric turbulence levels. The throughput is scaled to have a maximum value of one. i.e. in the figures, throughput efficiency = throughput[packets/s] x G[bits/packet] / R_{max}[bits/s]. The plots use a maximum transmission rate of 10 Gb/s, packet size of 1500 bytes, 8 dB of link margin, error probability threshold of 0.1, transverse wind speed of 10 km/hr and background noise photons per symbol per mode of one[14]. When plotting the throughput lower bounds, we extended the curves for round-trip times longer than 10 milliseconds as the curves could practically not be calculated for large RTTs using the same method described in Section 6.3 (due to computer memory constraints)[15]. The extended parts of the curves are shown in red. In extending the curves, we used the following intuition and assumption: as round-trip times increase to very large values, the change in steady state distribution is small since outage and congestion losses prevent the window from increasing to higher values (or at least make the probability of going to high values

---

[13] The throughput plots where congestion loss probability per packet is 10^{-1} may not be accurate since the model of transitions occurring at the round-trip boundaries may not be a good approximation for calculating throughput.

[14] A transmission rate of 10 Gb/s is a typical current rate, 1.5 kB is the maximum packet size for Ethernet, 8 dB is a significant amount of link margin (a few dB is very realistic), error probability threshold of 0.1 is aggressive for powerful forward error correction codes, and 10 km/hr is a gentle breeze which causes longer outages than a fast wind, and background noise photons of 1 per bit is a large amount. These values are selected to represent a current system with unfavorable link conditions and significant extra transmit power to mitigate the fading.

[15] As round-trip distances increase, the number of states in the linear Markov chains becomes very large. We use the program MATLAB on a 32 bit Windows machine to calculate the steady state distribution of the Markov chain. For a large number of states in the Markov chain, MATLAB runs out of memory (Applications in 32 bit Windows have a memory limit of 2 GB per application) and thus cannot calculate the throughput lower bound.

very small). We confirmed that the change in steady state distribution in fact becomes smaller and smaller as RTT increases (see Figure 6.41). We plotted the change in steady state distribution for every combination of turbulence and congestion loss probability for which we plotted the throughput curves, but we elected to include only one of the figures in the thesis because the others are similar. Since the change in the steady state distribution gets small as RTT increases close to 10 milliseconds, in extending the throughput lower bound curves, we assumed that the steady state distribution does not change for round-trip times above 10 milliseconds (above M=9000). Note that the throughput lower bound is proportional to 1/RTT for long round-trip distances since the numerator of (6.5) and (6.23) becomes a fixed value when the steady state distribution does not change for larger RTTs.
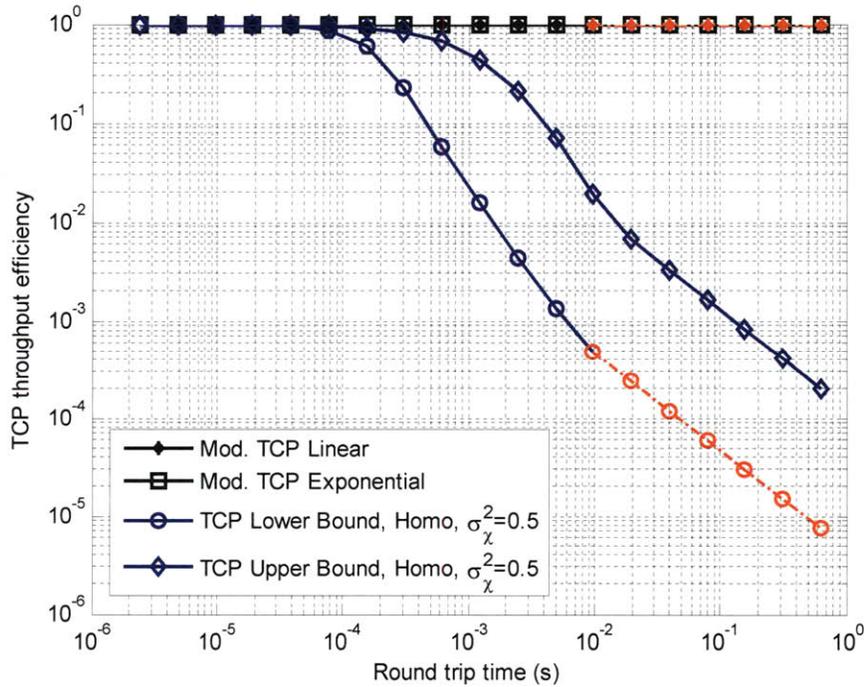
Figure 6.6: Throughput efficiency for congestion loss per packet of 0 and σ_χ^2=0.5
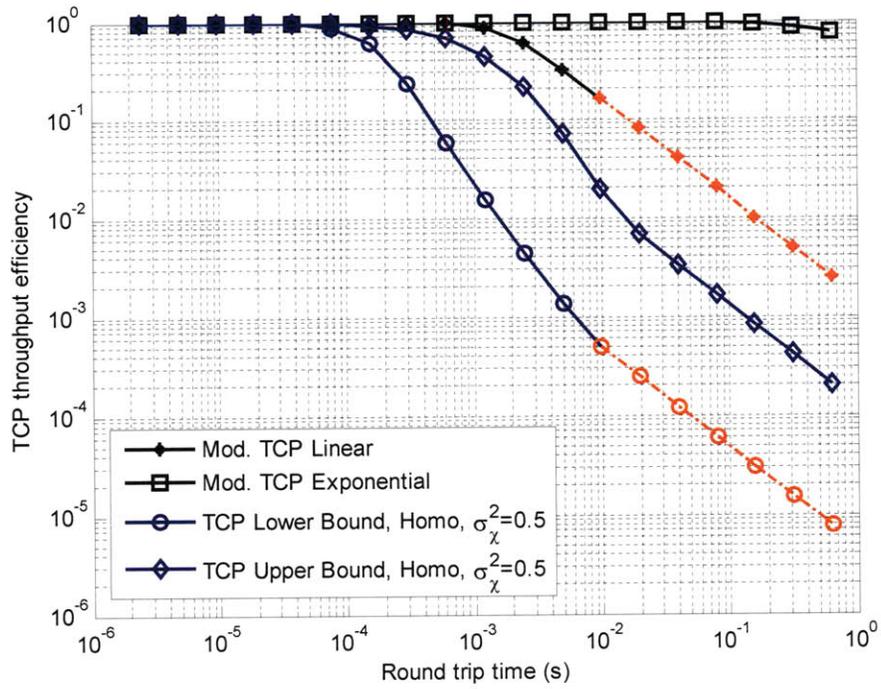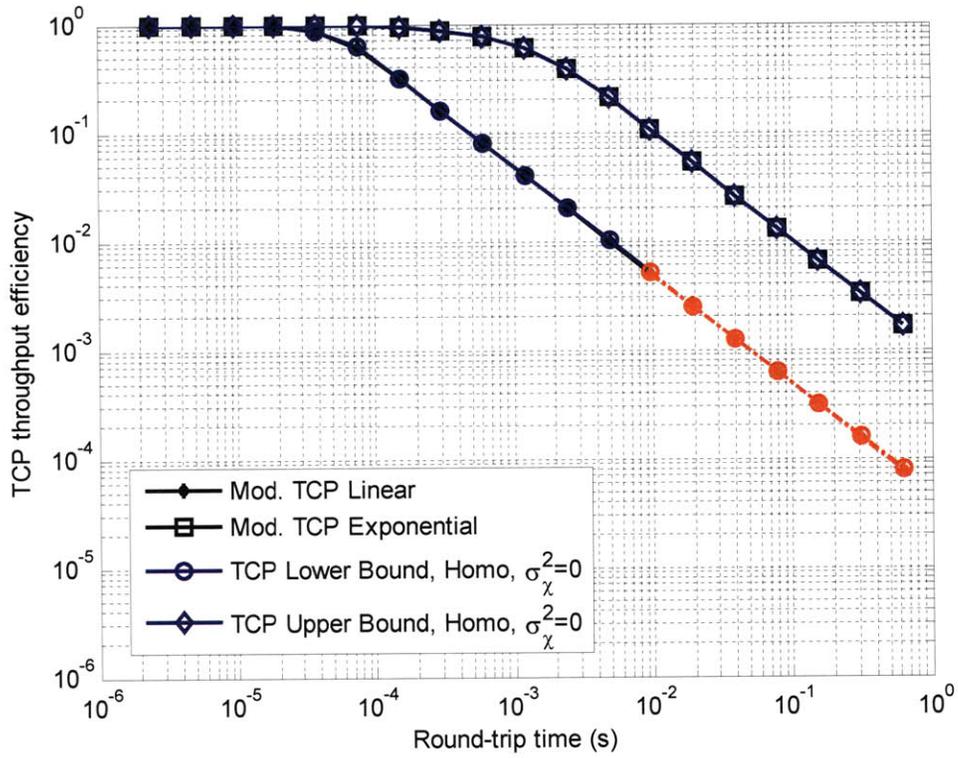
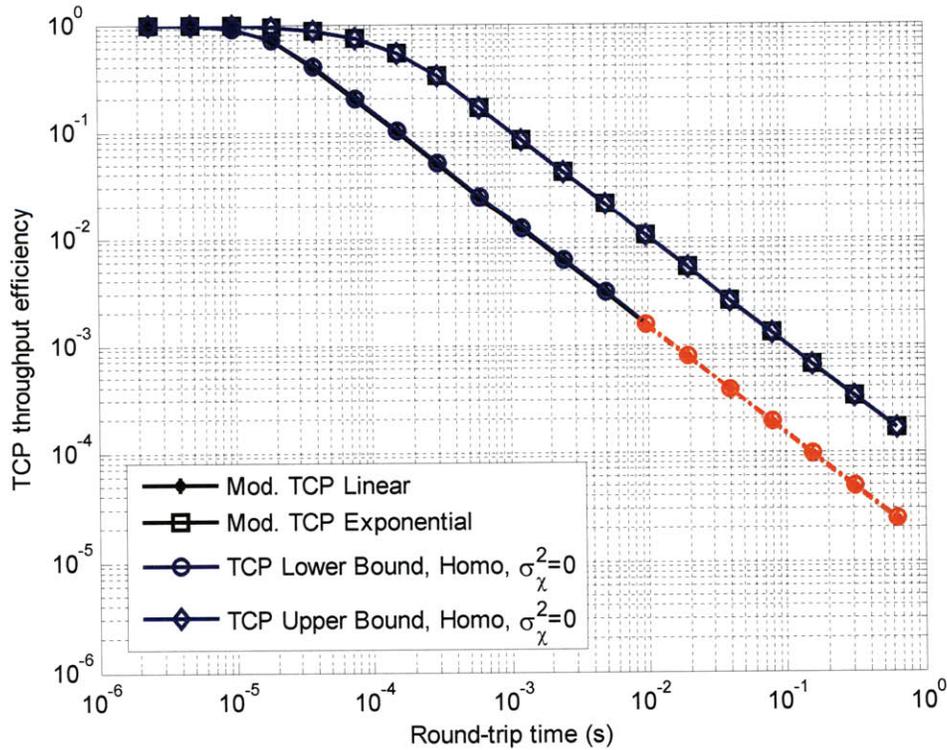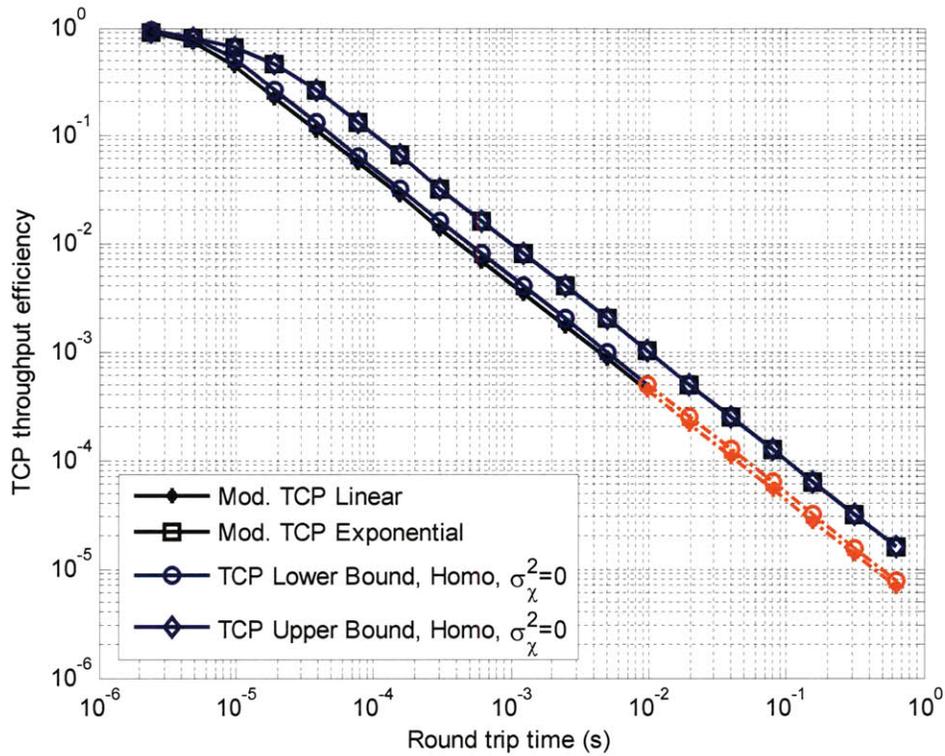Figure 6.7: Throughput efficiency for congestion loss per packet of 10^{-6} and σ_χ^2=0.5
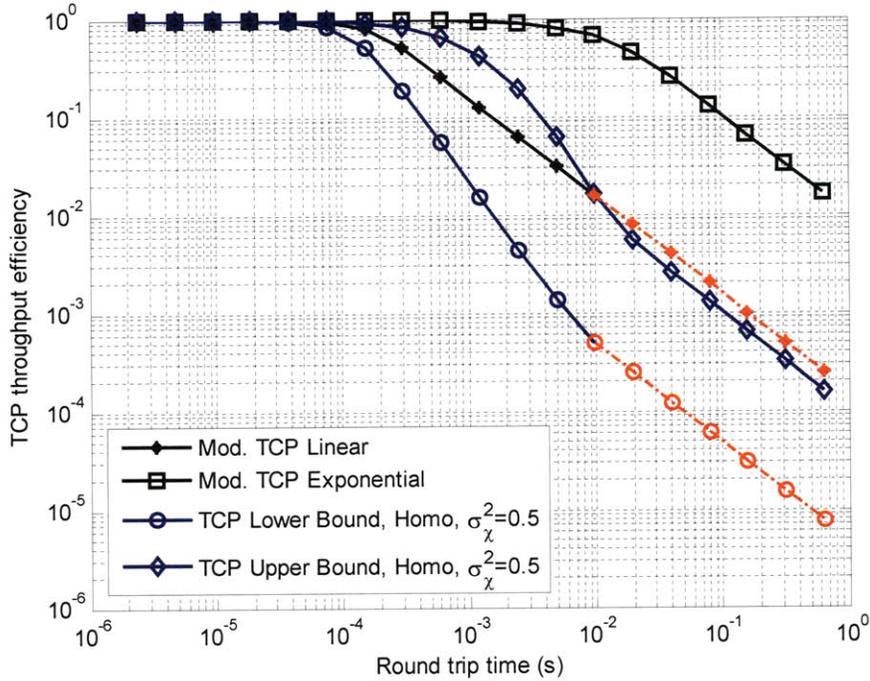
Figure 6.8: Throughput efficiency for congestion loss per packet of 10^{-4} and σ_χ^2=0.5
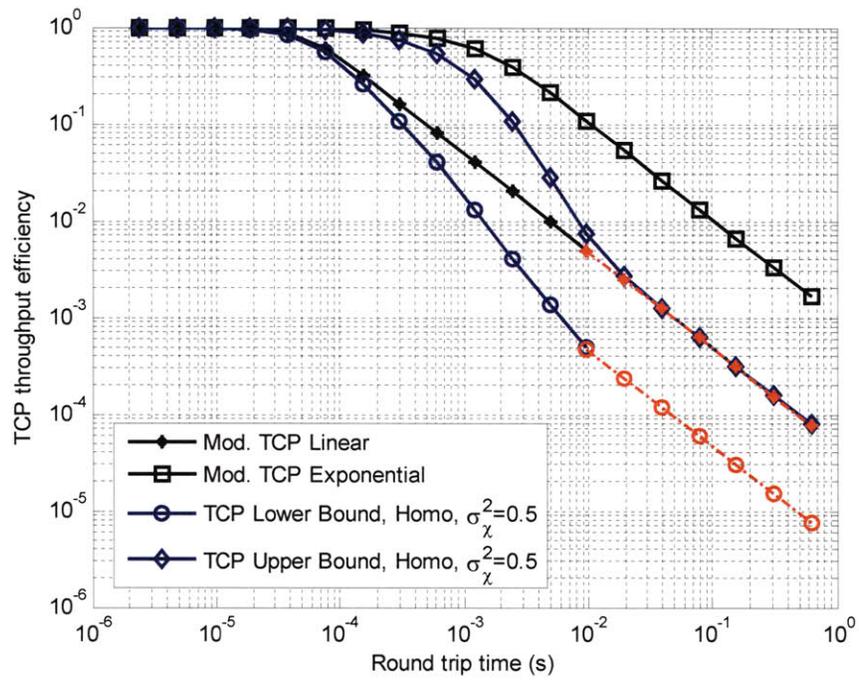
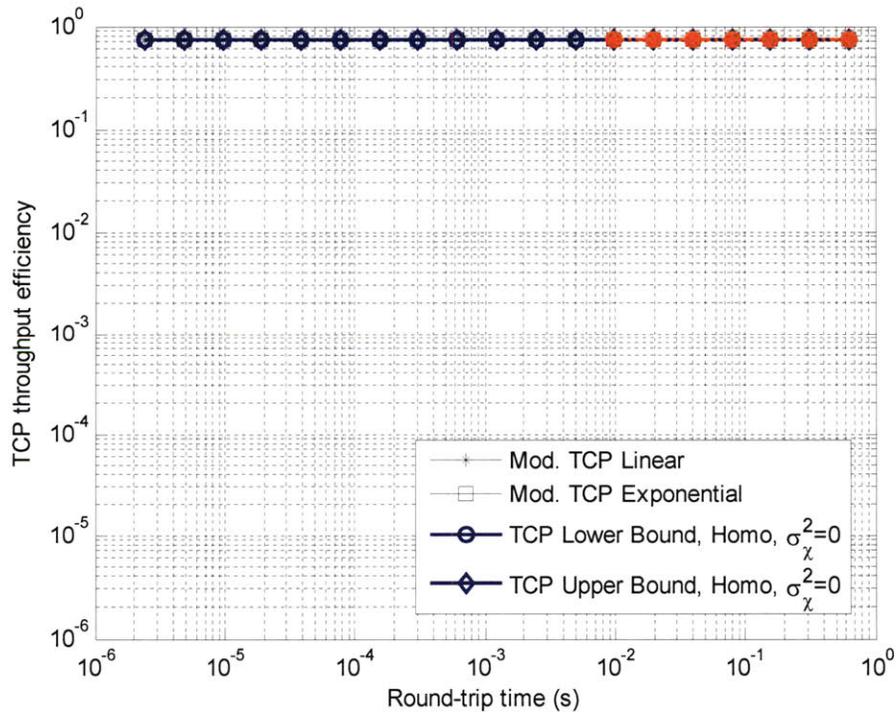Figure 6.9: Throughput efficiency for congestion loss per packet of 10^{-3} and $\sigma_\chi^2=0.5$
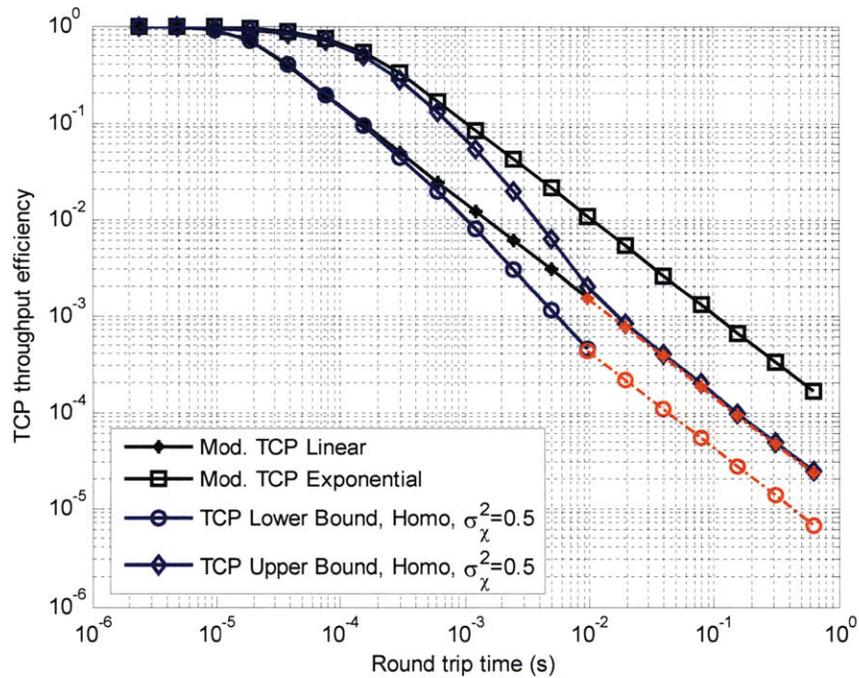
Figure 6.10: Throughput efficiency for congestion loss per packet of 10^{-2} and $\sigma_\chi^2=0.5$
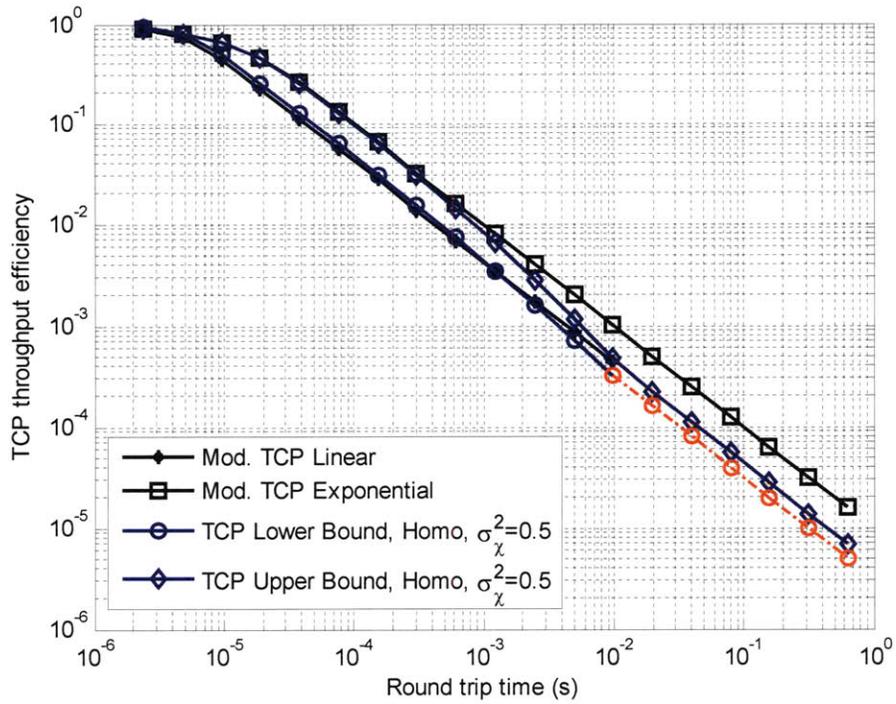
Figure 6.11: Throughput efficiency for congestion loss per packet of 10^{-1} and $\sigma_\chi^2 = 0.5$
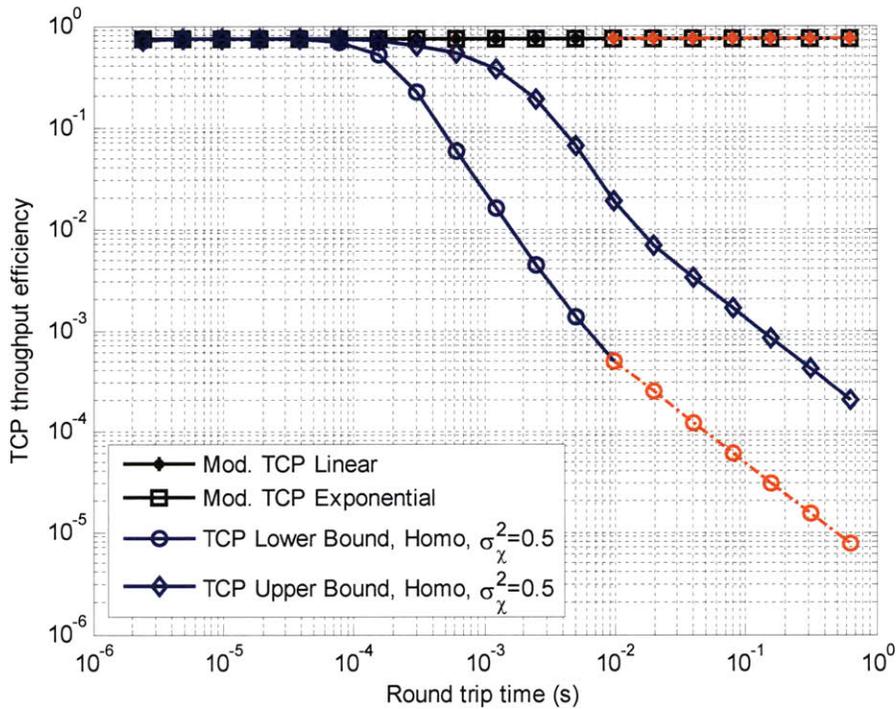
Figure 6.12: Throughput efficiency for congestion loss being modeled as a step function and $\sigma_\chi^2 = 0.5$
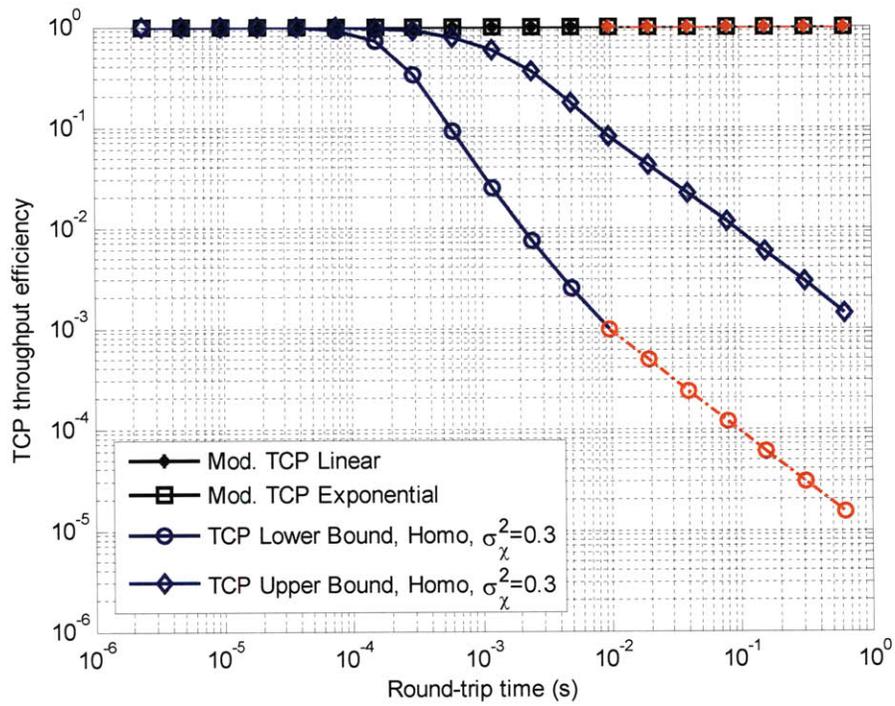
Figure 6.13: Throughput efficiency for congestion loss per packet of 0 and $\sigma_\chi^2 = 0.3$
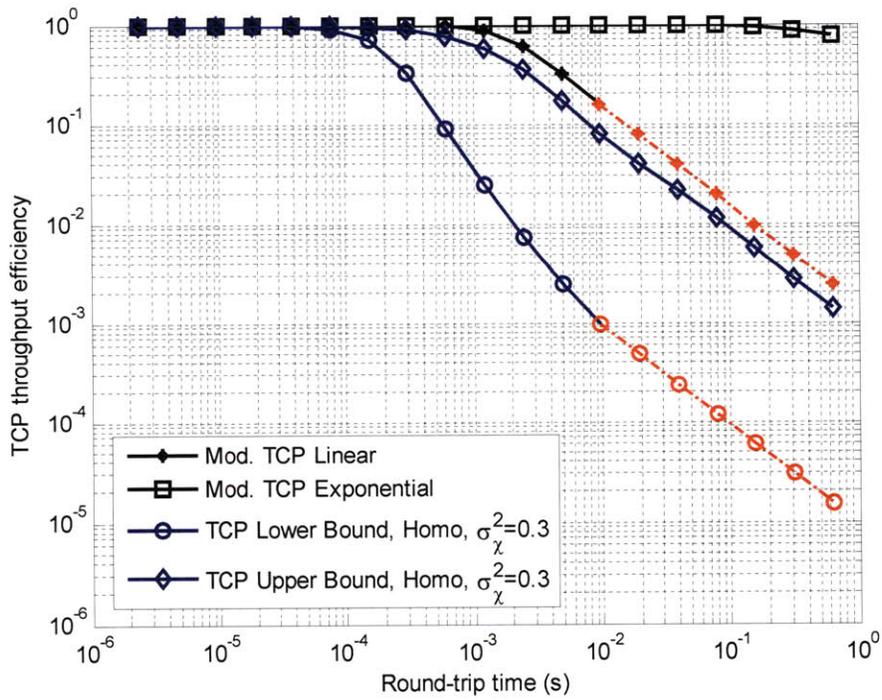
Figure 6.14: Throughput efficiency for congestion loss per packet of 10^{-6} and $\sigma_\chi^2 = 0.3$
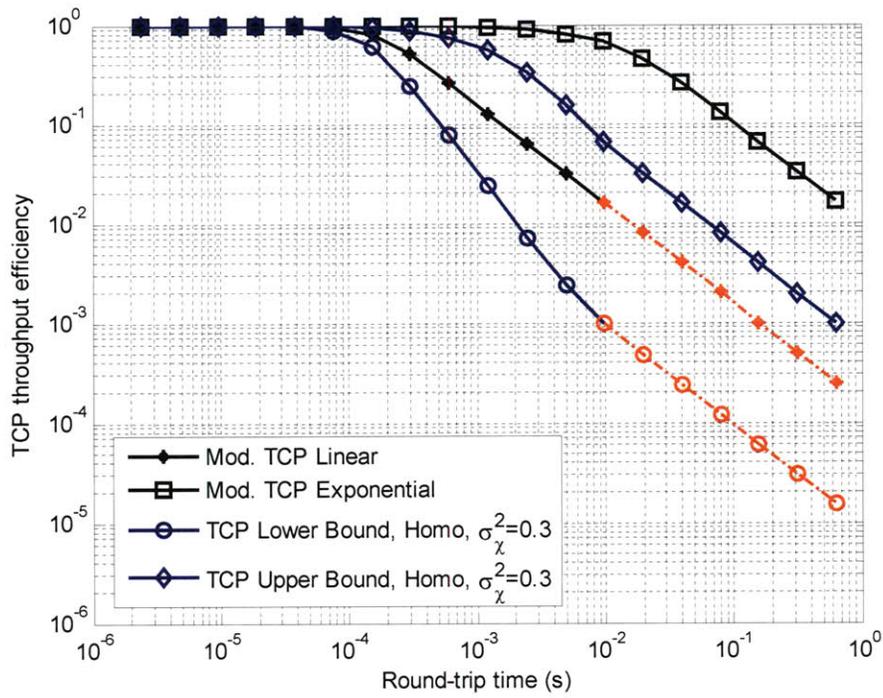
Figure 6.15: Throughput efficiency for congestion loss per packet of 10^{-4} and $\sigma_\chi^2 = 0.3$
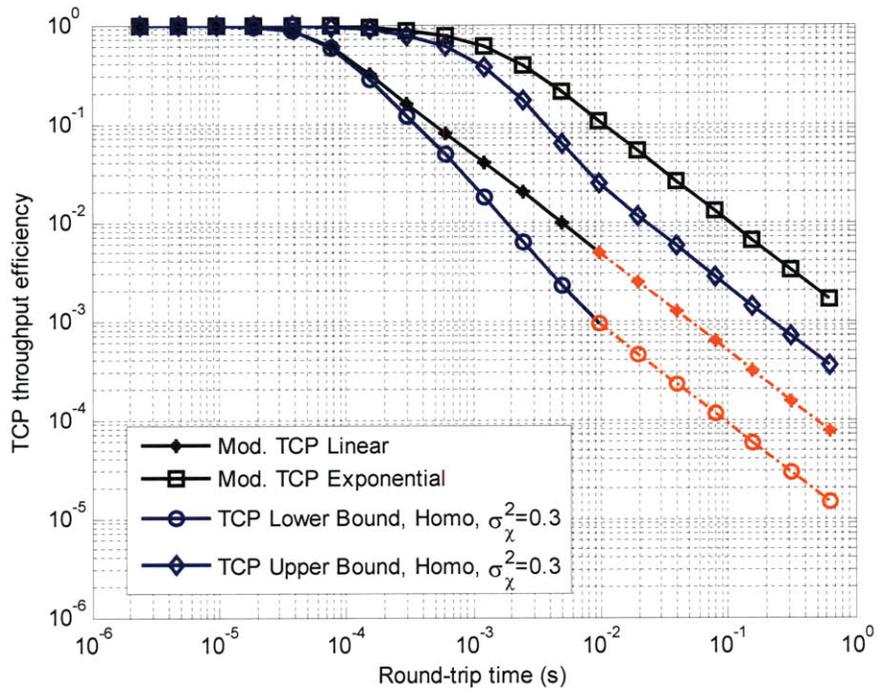
Figure 6.16: Throughput efficiency for congestion loss per packet of 10^{-3} and $\sigma_\chi^2 = 0.3$
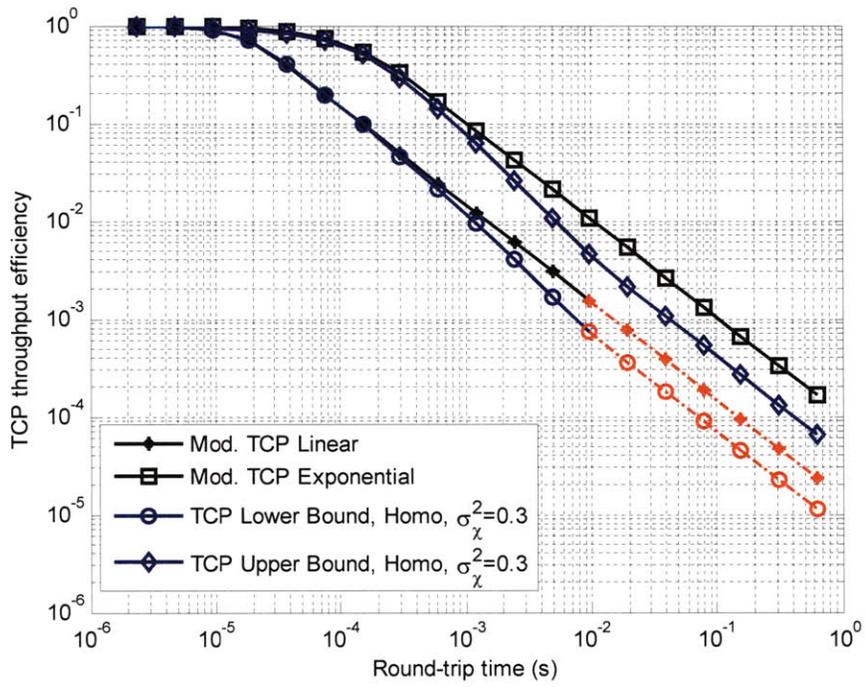
Figure 6.17: Throughput efficiency for congestion loss per packet of 10^{-2} and $\sigma_\chi^2 = 0.3$
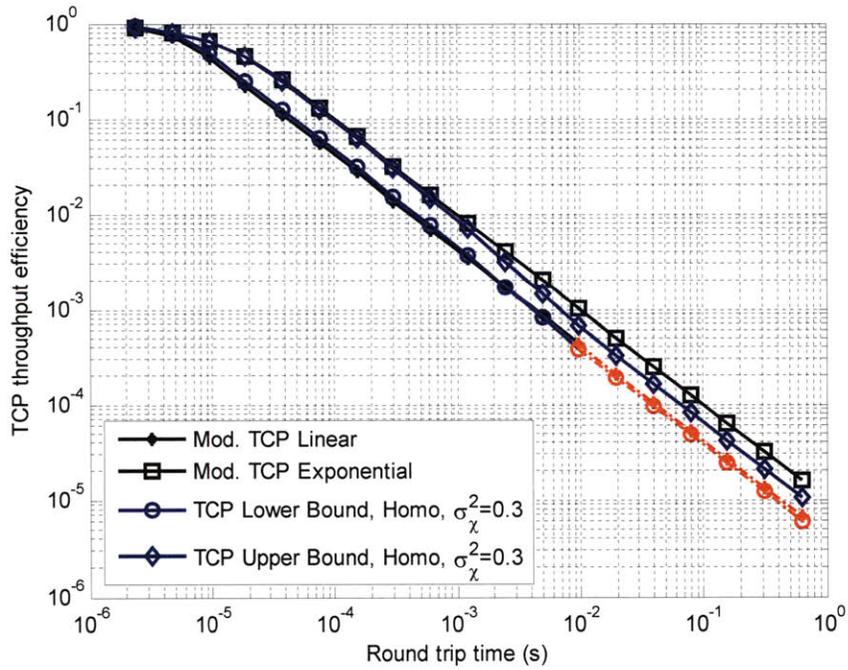
Figure 6.18: Throughput efficiency for congestion loss per packet of 10^{-1} and $\sigma_\chi^2 = 0.3$
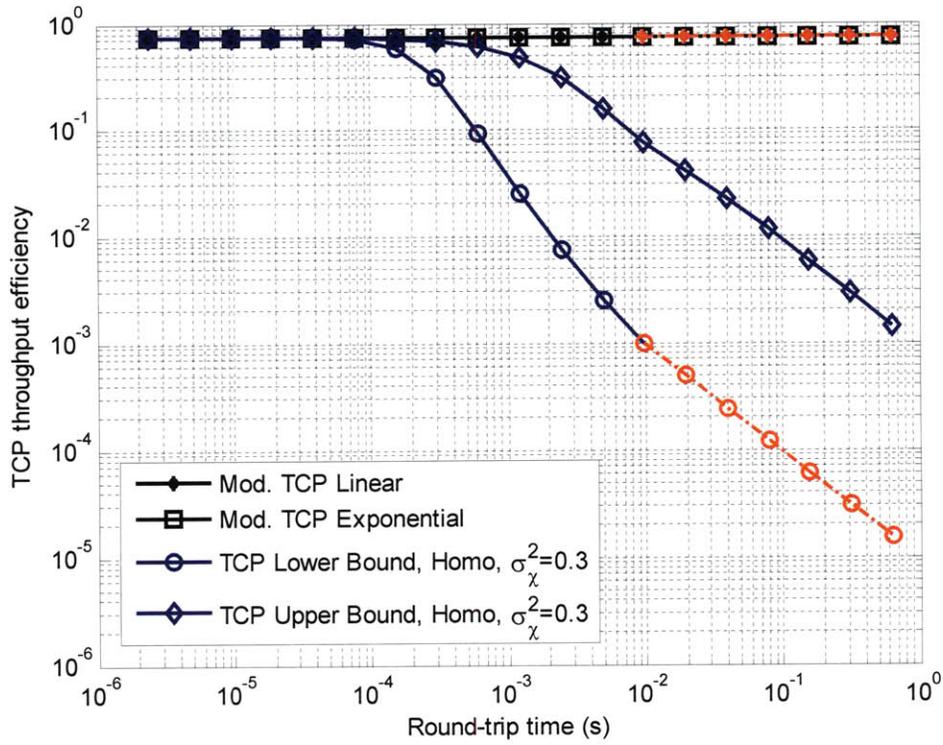
Figure 6.19: Throughput efficiency for congestion loss being modeled as a step function and $\sigma_\chi^2=0.3$
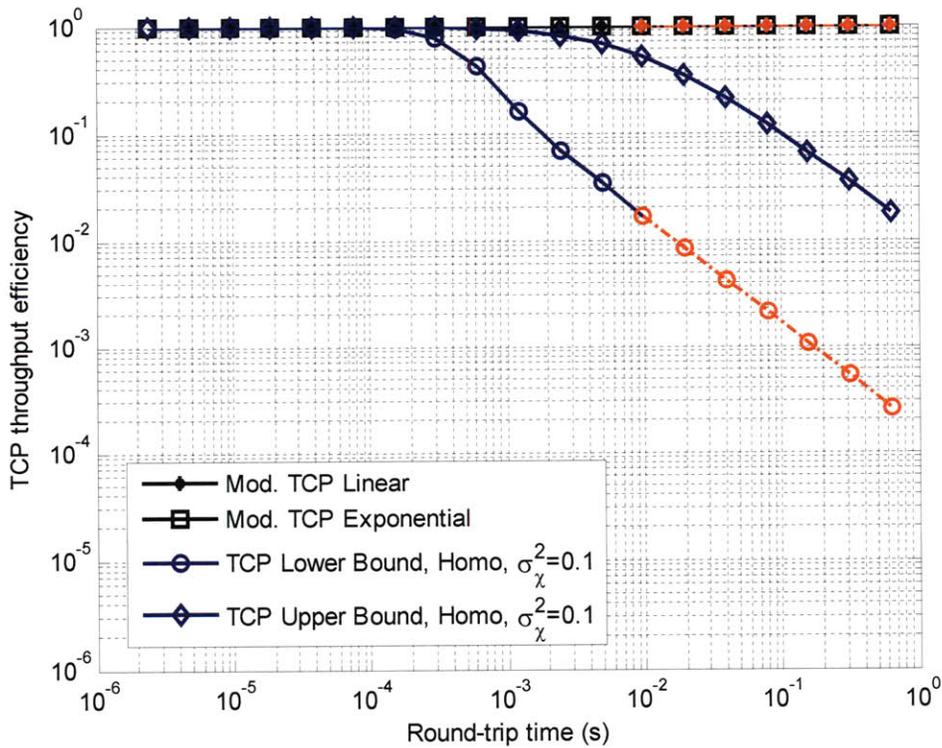
Figure 6.20: Throughput efficiency for congestion loss per packet of 0 and $\sigma_\chi^2=0.1$
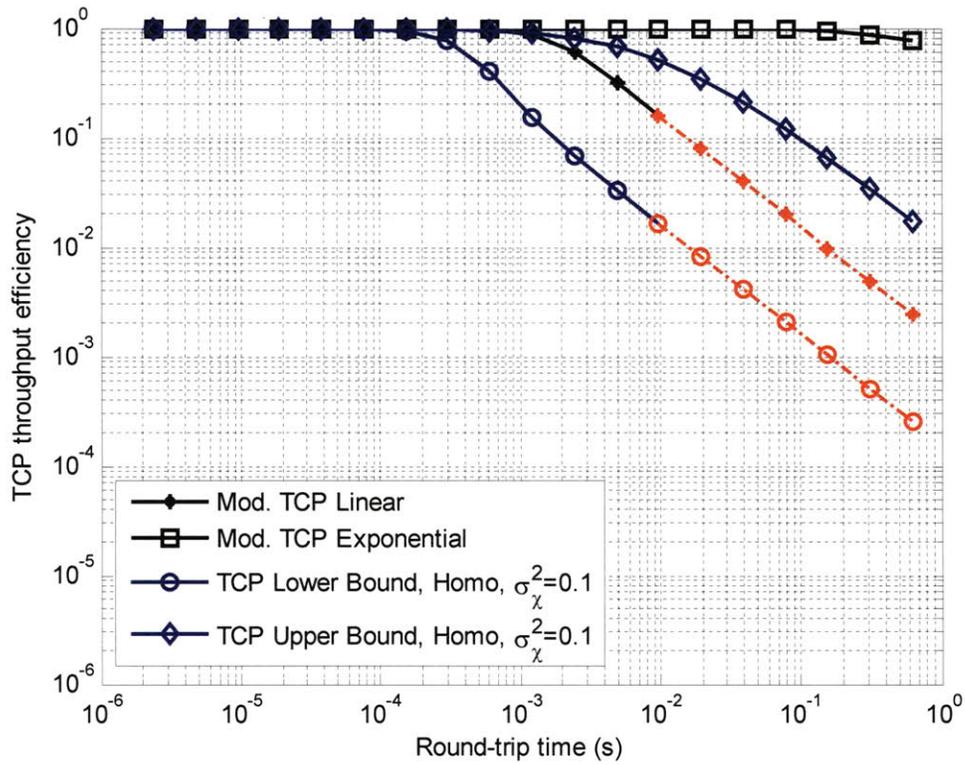
Figure 6.21: Throughput efficiency for congestion loss per packet of 10^{-6} and $\sigma_\chi^2=0.1$
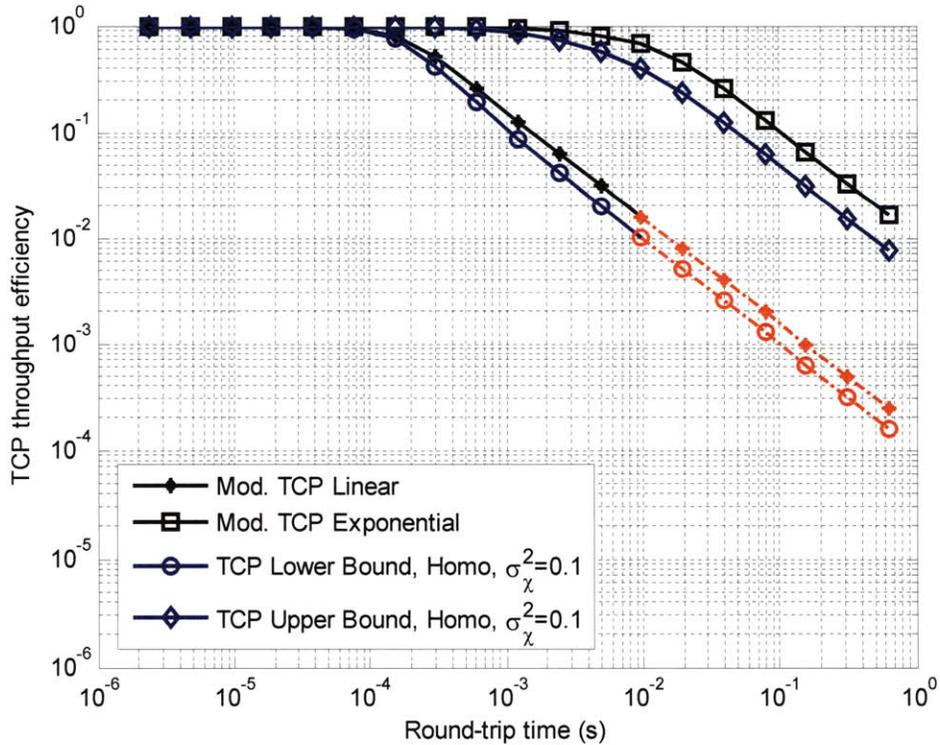
Figure 6.22: Throughput efficiency for congestion loss per packet of 10^{-4} and $\sigma_\chi^2=0.1$
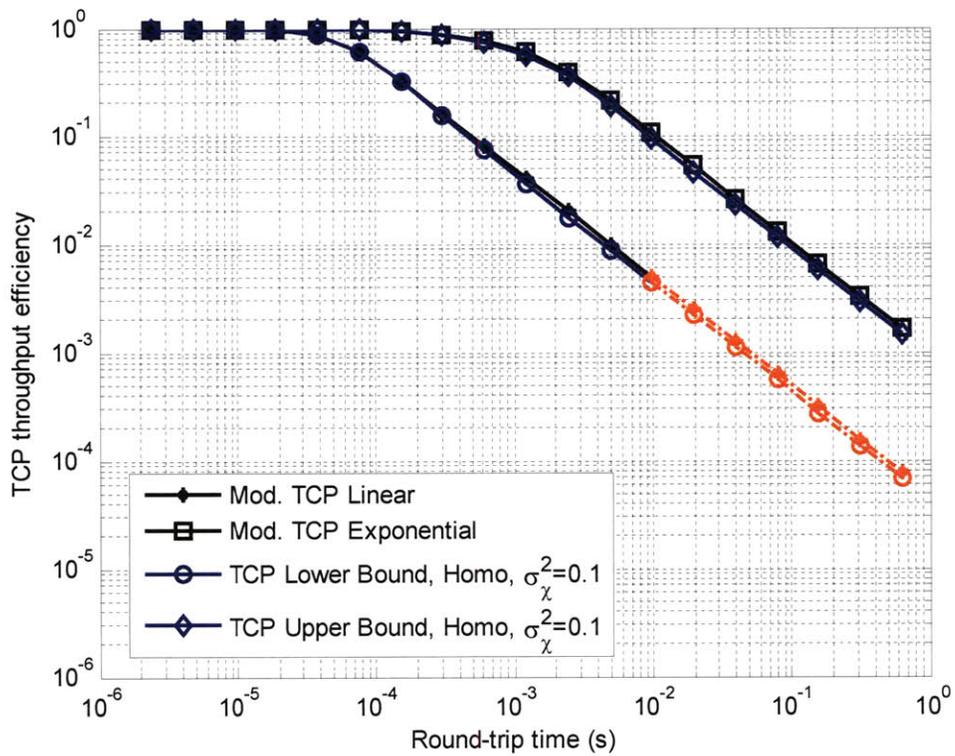
Figure 6.23: Throughput efficiency for congestion loss per packet of 10^{-3} and $\sigma_\chi^2=0.1$
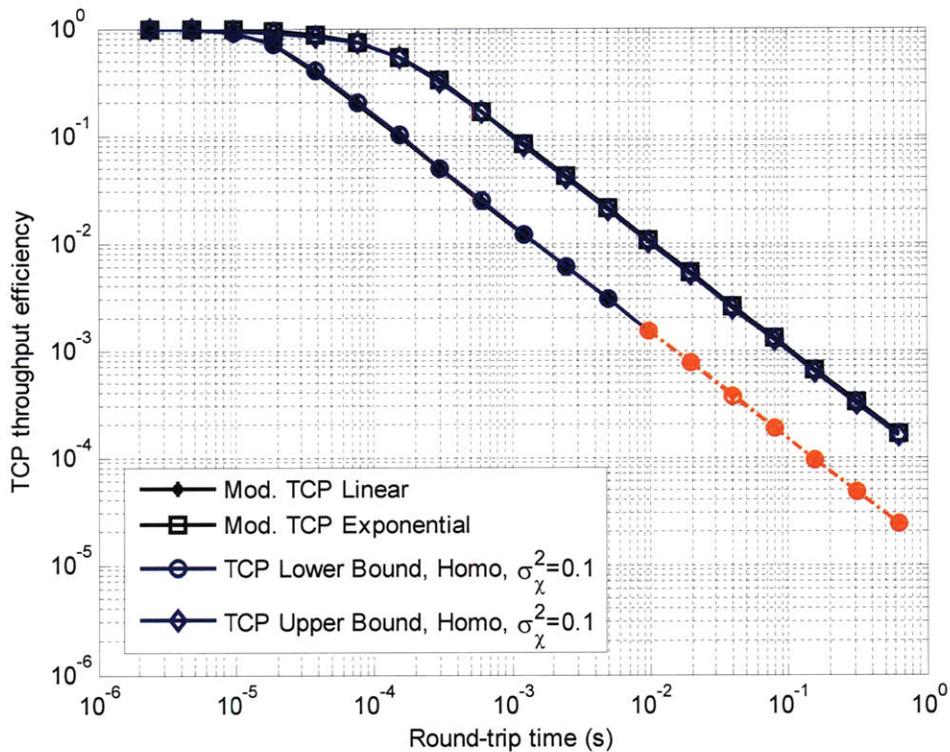
Figure 6.24: Throughput efficiency for congestion loss per packet of 10^{-2} and $\sigma_\chi^2=0.1$
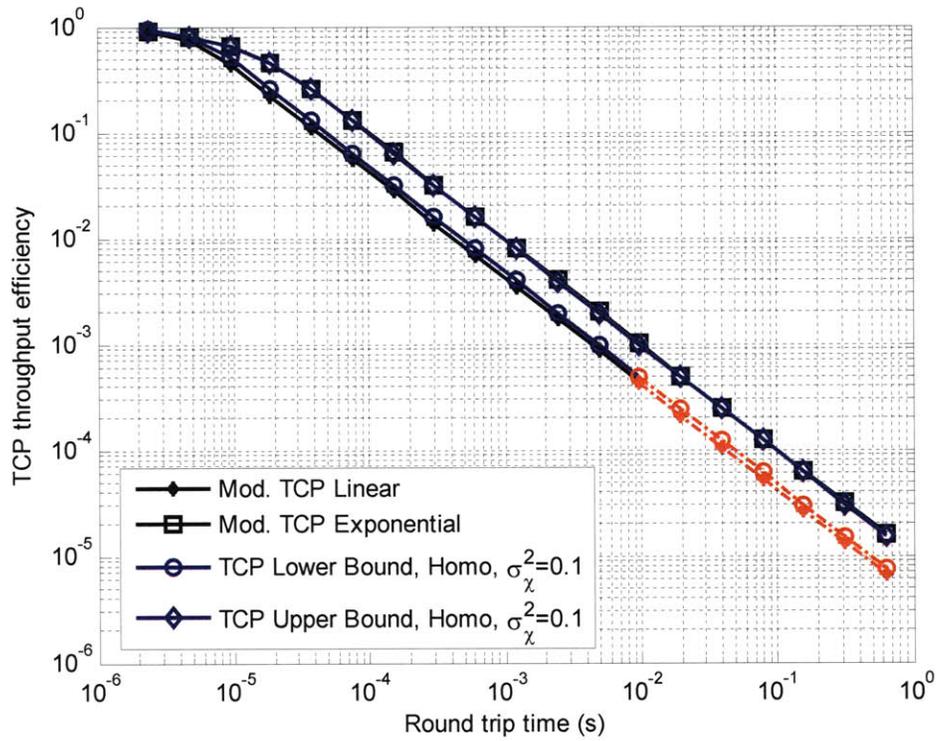
Figure 6.25: Throughput efficiency for congestion loss per packet of 10^{-1} and $\sigma_\chi^2=0.1$
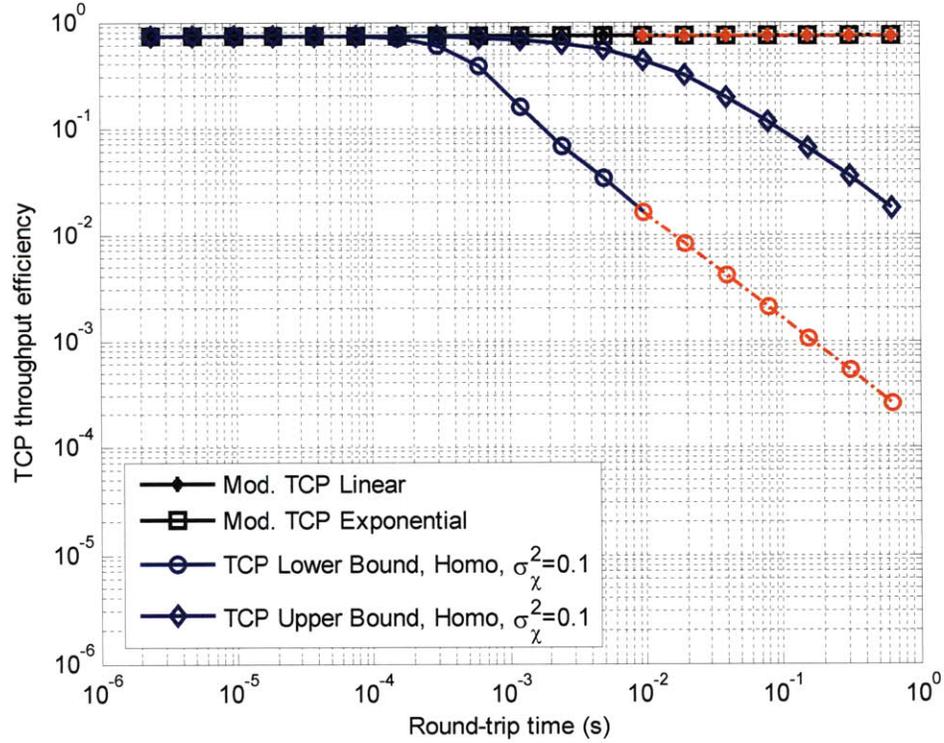
Figure 6.26: Throughput efficiency for congestion loss being modeled as a step function and $\sigma_\chi^2=0.1$
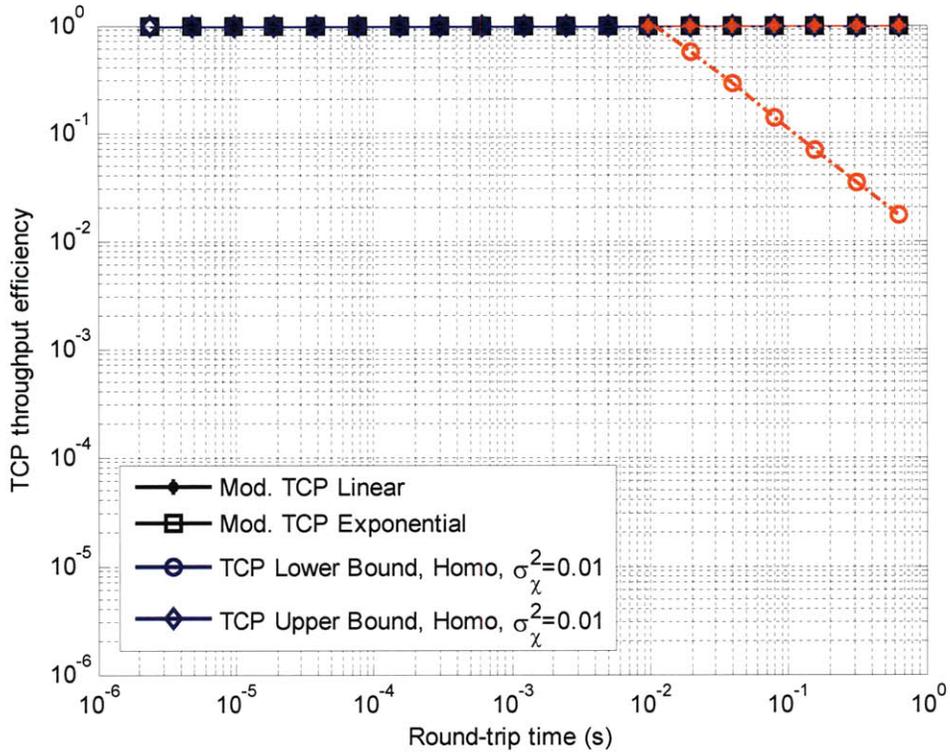
Figure 6.27: Throughput efficiency for congestion loss per packet of 0 and $\sigma_\chi^2=0.01$
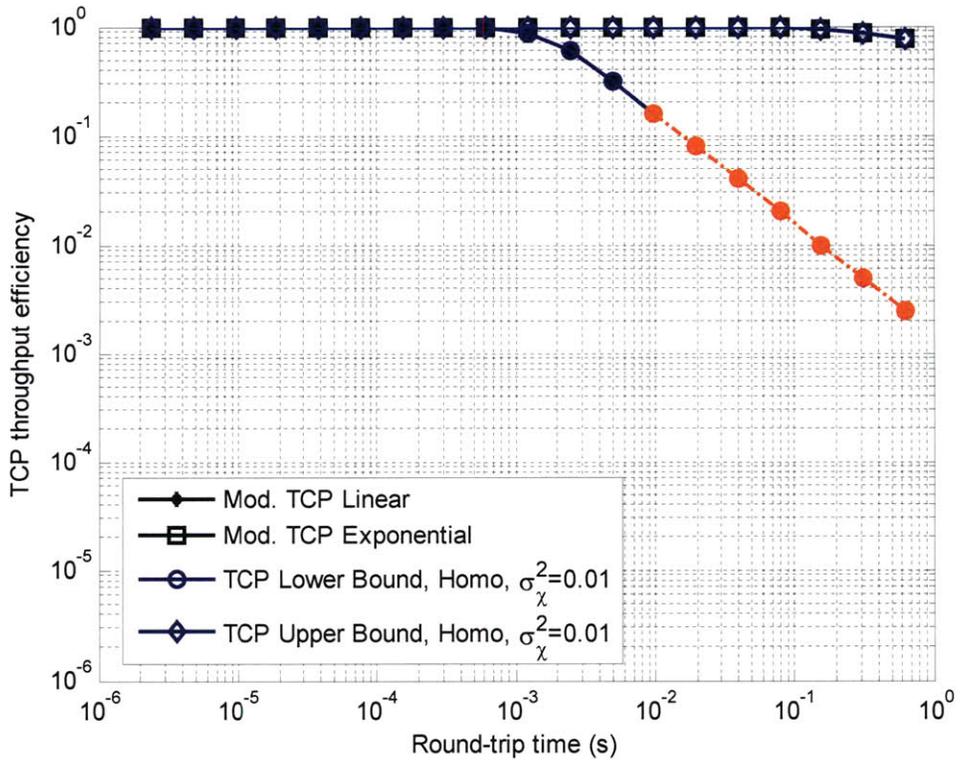
Figure 6.28: Throughput efficiency for congestion loss per packet of 10^{-6} and $\sigma_\chi^2=0.01$
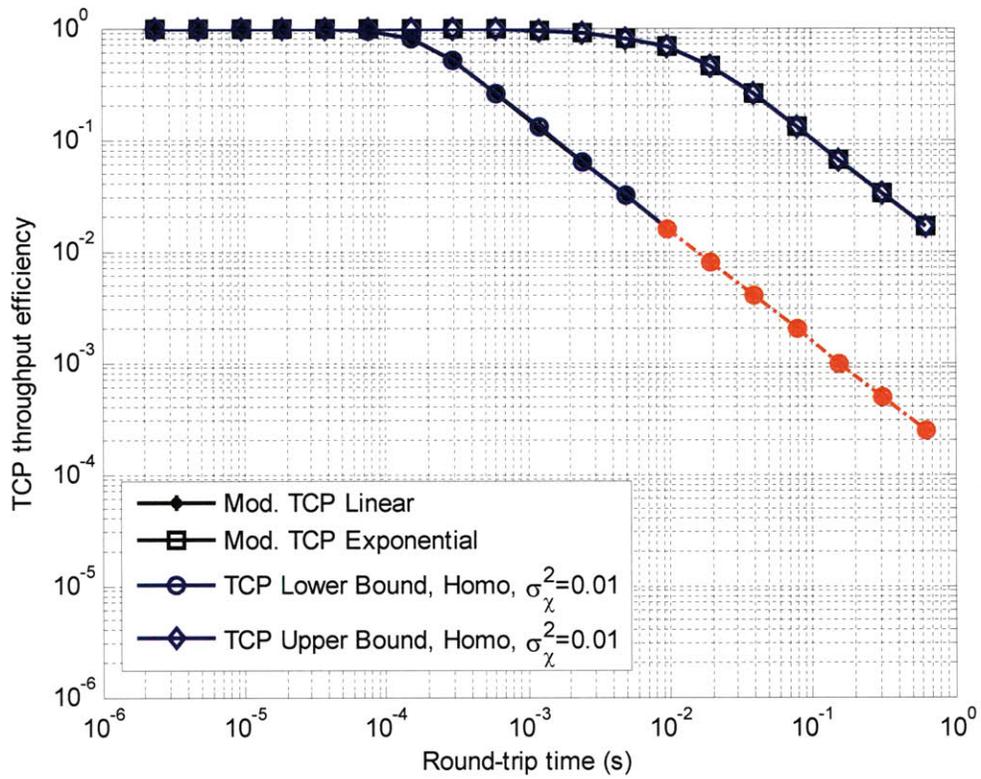
Figure 6.29: Throughput efficiency for congestion loss per packet of 10^{-4} and $\sigma_\chi^2 = 0.01$
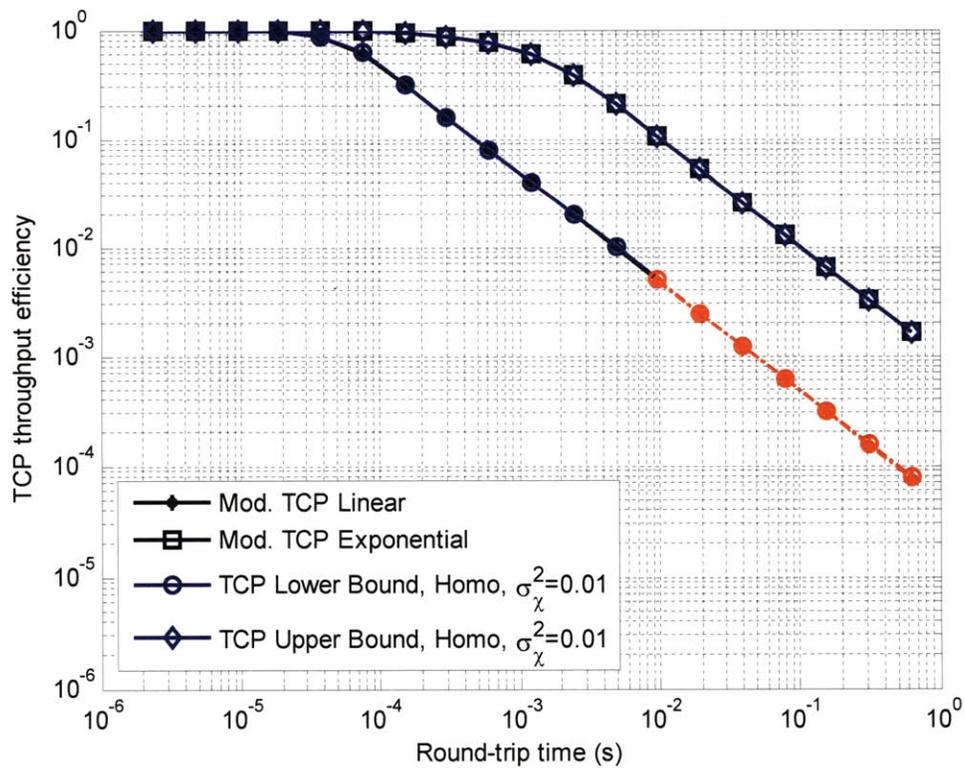
Figure 6.30: Throughput efficiency for congestion loss per packet of 10^{-3} and $\sigma_\chi^2 = 0.01$
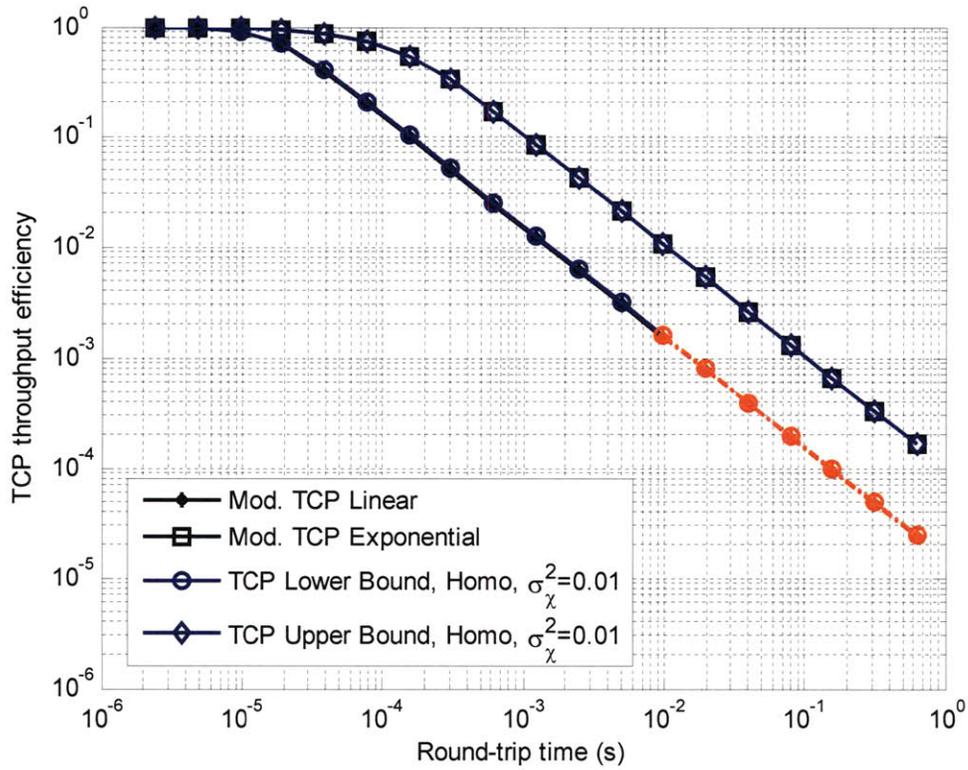
Figure 6.31: Throughput efficiency for congestion loss per packet of 10^{-2} and $\sigma_\chi^2 = 0.01$
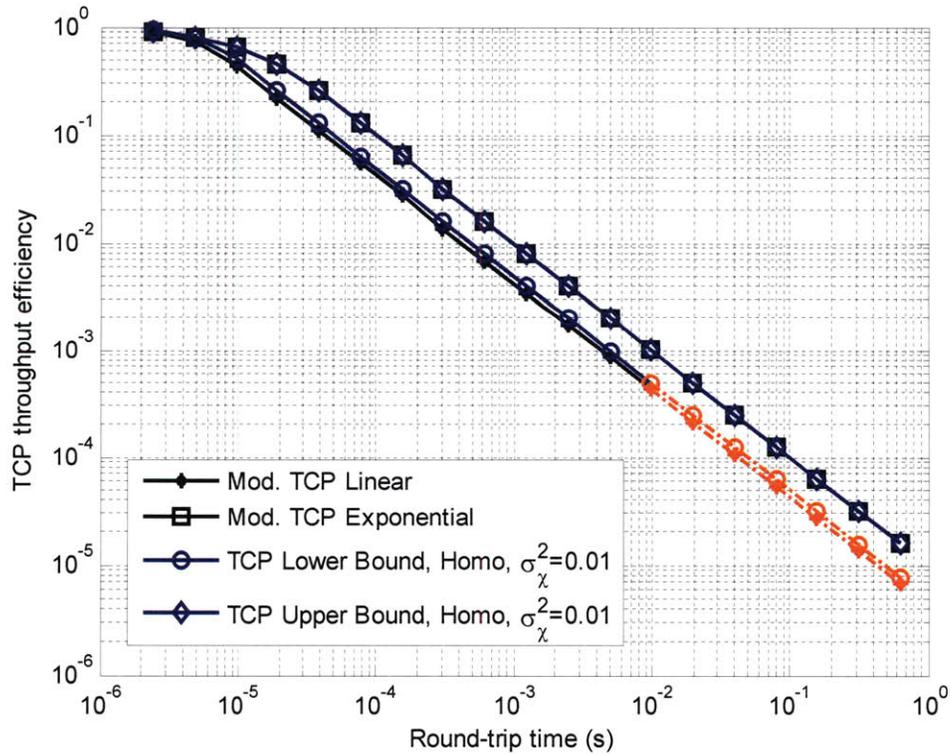
Figure 6.32: Throughput efficiency for congestion loss per packet of 10^{-1} and $\sigma_\chi^2 = 0.01$
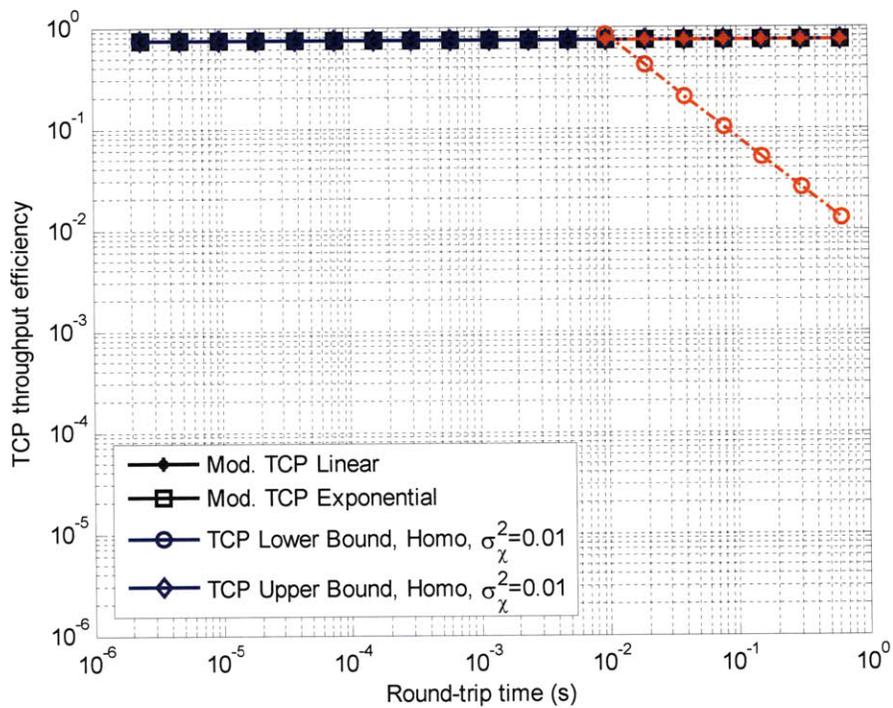
Figure 6.33: Throughput efficiency for congestion loss being modeled as a step function and $\sigma_\chi^2 = 0.01$.
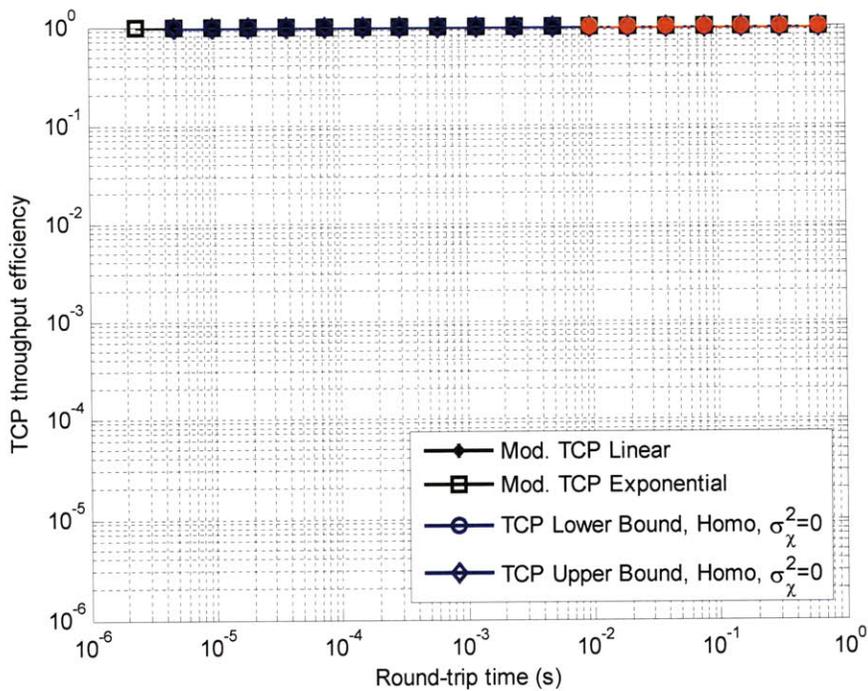
Figure 6.34: Throughput efficiency for congestion loss per packet of 0 and $\sigma_\chi^2 = 0$
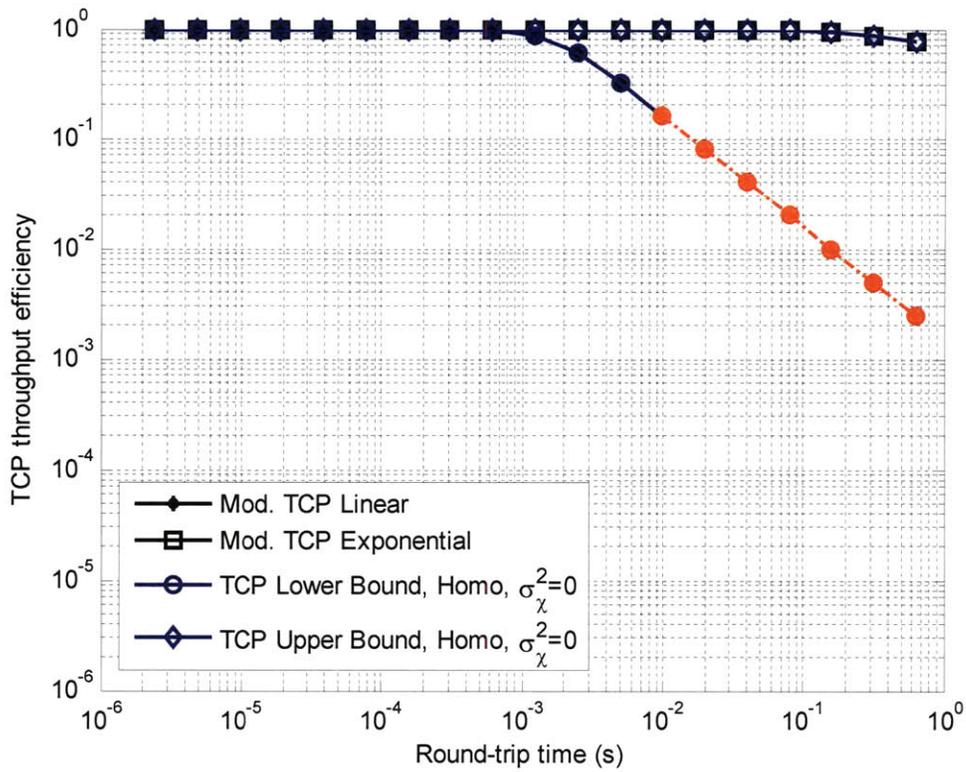
Figure 6.35: Throughput efficiency for congestion loss per packet of 10^{-6} and $\sigma_\chi^2=0$
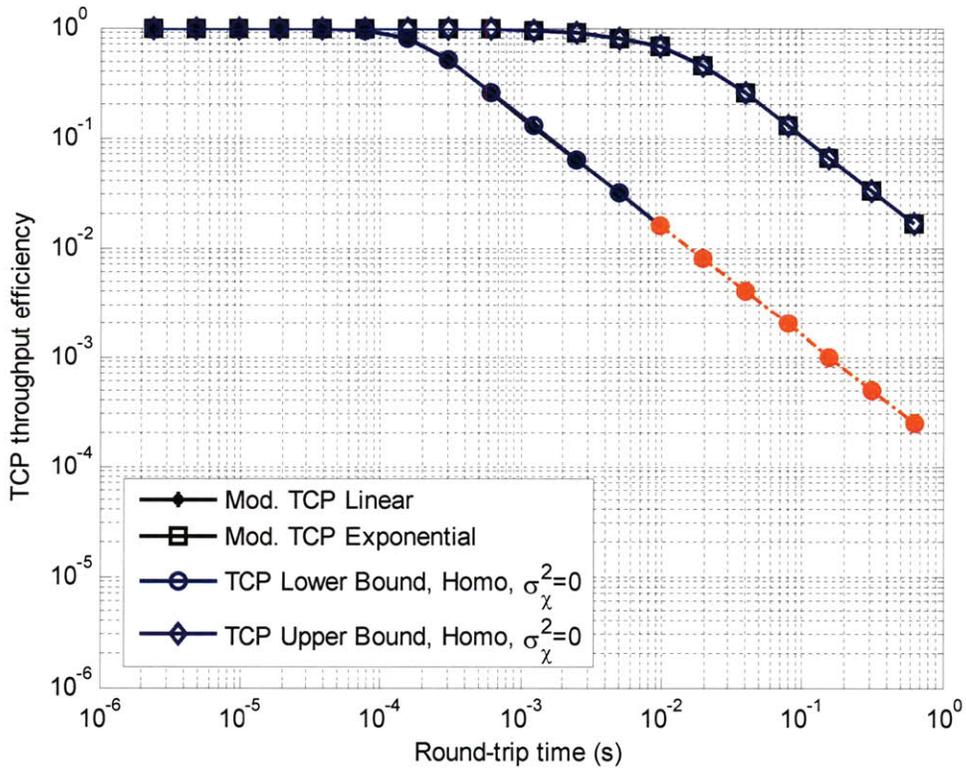
Figure 6.36: Throughput efficiency for congestion loss per packet of 10^{-4} and $\sigma_\chi^2=0$

Figure 6.37: Throughput efficiency for congestion loss per packet of 10^{-3} and $\sigma_\chi^2=0$

Figure 6.38: Throughput efficiency for congestion loss per packet of 10^{-2} and $\sigma_\chi^2=0$

Figure 6.39: Throughput efficiency for congestion loss per packet of 10^{-1} and $\sigma_\chi^2=0$

Figure 6.40: Throughput efficiency for congestion loss being modeled as a step function and $\sigma_\chi^2=0$
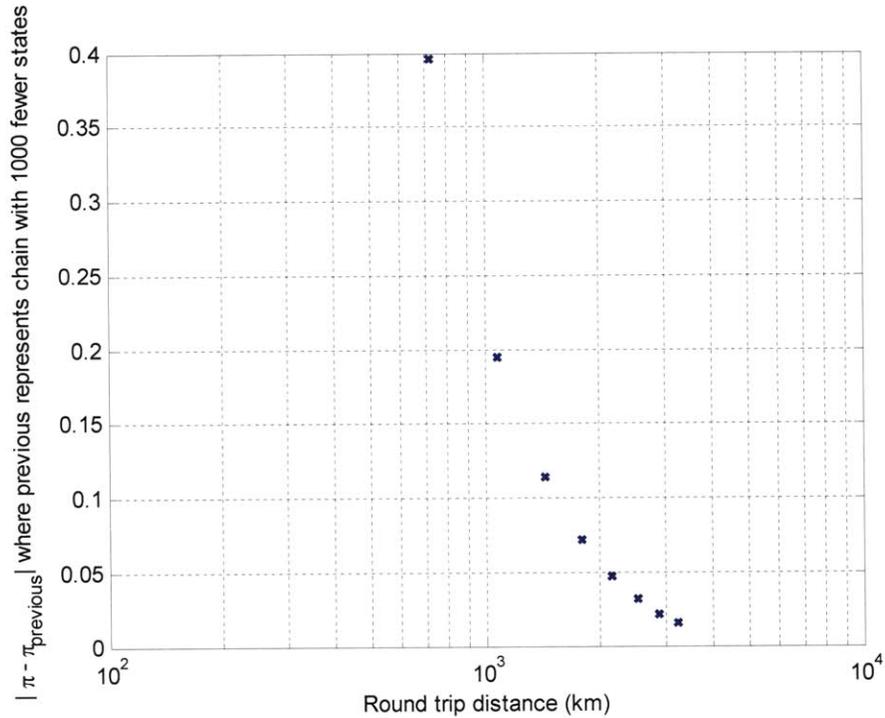
Figure 6.41: L₁ norm of the change in steady state probability distribution, $p_{congperpkt}=0$, $\sigma_\chi^2=0.3$

As we can see in Figures 6.6-6.40, as expected, the steady state TCP throughput goes down when atmospheric turbulence increases. This is because packet losses due to link outages cause the sender to timeout and reduce its window size to one packet, and higher turbulence means more frequent window reduction. Furthermore, for long round-trip time sessions (such as a path from the Earth to a geosynchronous satellite with round-trip time of a quarter of a second), the throughput is very low when turbulence and congestion losses exist. This is because after each window reduction, the window size is increased gradually (every quarter of a second), and it takes the TCP sender a large number of round-trip times to increase its average rate. Even if the window increase is exponential, the rate of increase is still slow due to the long round-trip time. Moreover, before the window builds up to full value, another outage will cut the window again preventing the TCP sender from reaching the maximum possible number of packets in flight. For large round-trip time links

through the atmosphere (and large sender transmission rate), TCP congestion control is far from ideal as a congestion control protocol.

Modified TCP gains back the performance loss due to outages. This is because Modified TCP does not cut its window in response to outage losses and thus also does not have to suffer the slow increase in its window after the outage. Over large round-trip time free-space optical links, Modified TCP performs significantly better than TCP when the turbulence is strong such that the outages dominate the congestion losses. However, as previously mentioned, TCP and Modified TCP have an issue of slow window increase over high bandwidth-delay products, which we do not address in this thesis. Moreover, this slow window increase over high bandwidth-delay product paths results in unfairness to sessions with large round-trip times because the sessions with small round-trip times increase their window more quickly (due to ACKs returning more quickly) and are able to use up more of the available network capacity. See Chapter 8 for other ways that have been suggested to improve the slow window increase, and why we suggest that adding the 'Congestion Loss' feedback to the existing proposed schemes may allow for better performance over FSO networks.

For short round-trip times (less than a 10 microseconds), the throughput of TCP and Modified TCP is nearly optimal. This is because the maximum number of packets in flight is small so when an outage causes the window to be reduced, the window can increase to full value in only a few round-trip times. For example, for a round-trip time of 10 microseconds, packet size of 1.5 kilobytes, and a maximum transmission rate of 10 Gb/s, the maximum number of packets in flight is 9. So even if the window increases linearly starting from a window size of one packet, TCP could increase its window size to the maximum possible packets in flight in only 9 round-trip times (90 microseconds). Since TCP performance is good for short round-trip time paths, the

marginal performance increase of Modified TCP in this regime may not be worth the effort. It is over long round-trip time paths that the Modified TCP provides significant and worthwhile benefit.

We depict in Figure 6.42 the regions of atmospheric turbulence and round-trip times over which it is worthwhile to use Modified TCP rather than TCP (for the parameters given in the figure caption, which correspond to the parameters of Figures 6.6-6.40). These regions are determined by inspection of the throughput plots of Figures 6.6-6.40. The areas labeled "TCP" indicate regions where TCP performance is sufficient without using Modified TCP. "Sufficient" is taken to be when the TCP linear increase throughput efficiency is >10%. Modified TCP is worthwhile to use over TCP when the turbulence is not weak and round-trip times are large, and these regions are labeled "Modified TCP". For large round-trip times, even though Modified TCP offers improvement over TCP, the throughput is still not high due to the limited rate of window increase, as we discussed. In the figure, we labeled regions where Modified TCP throughput efficiency is less than 10% as "Other" to denote that another congestion control protocol that allows for faster window increase would provide improved sender throughput compared to Modified TCP. The points with an 'x' or 'o' are points that are determined by inspection of the steady state throughput plots, and the lines that connect these markers are interpolated. For $\sigma_\chi = 0.1$, TCP and Modified TCP's throughput is roughly the same. This is why we extrapolated the boundary between the TCP and Modified TCP regions to level off above $\sigma_\chi = 0.1$. As the congestion loss probability increases, the TCP and Modified TCP throughputs decrease. Thus, the boundaries between "TCP" and "Modified TCP" and between "Modified TCP" and "Other" get squeezed to the left. The pairs of numbers shown as (number1, number2) are the expected outage length and outage probability corresponding to the 'x' marker points.

192
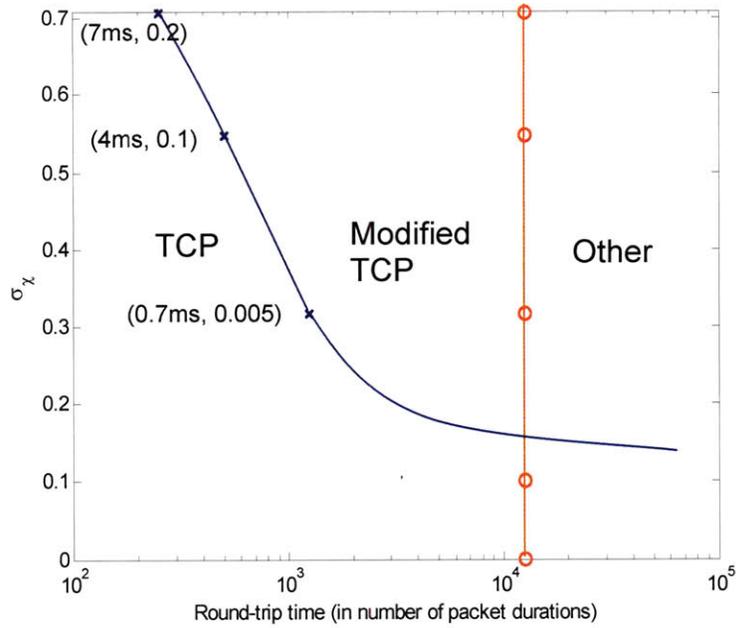
Figure 6.42 (a) $p_{congperpkt} = 10^{-6}$
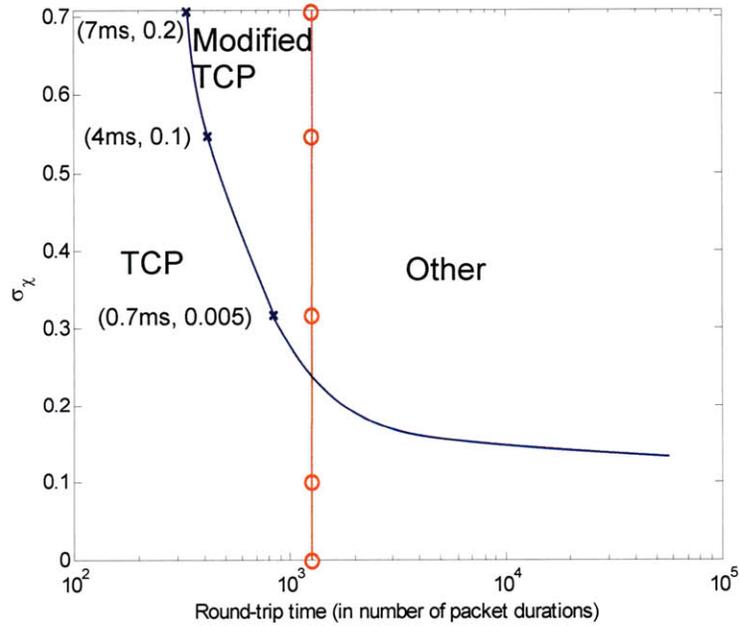
Figure 6.42 (b) $p_{congperpkt} = 10^{-4}$
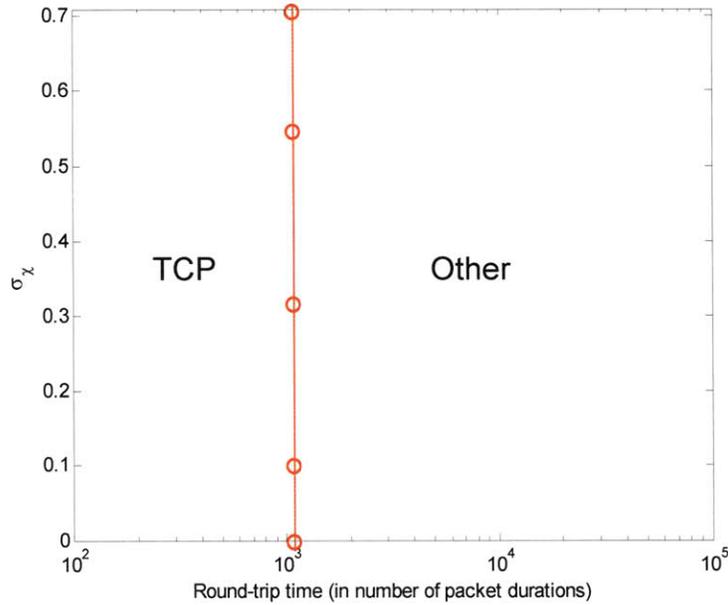
Figure 6.42 (c) $p_{congperpkt}=10^{-2}$

Figure 6.42: Diagram of regions, in steady state, where it is worthwhile to use Modified TCP rather than TCP when probability of congestion loss per packet is (a) 10^{-6} (b) 10^{-4} and (c) 10^{-2} and where $R_{max}=10Gb/s$, $G=1.5kbytes$, $m=8dB$, $P_e^{thresh}=0.1$, $v_t=10km/hr$, $N_n=1$, outages cause timeouts. The boundaries indicate that the algorithm on the left has throughput lower bounds of less than 10% to the right of the boundary. The pairs of numbers shown as (number1, number2) are the expected outage length and outage probability corresponding to the 'x' marker points.

As we saw in Chapter 3, adding diversity in the Physical Layer decreases the expected outage length and outage probability. However, for high data rate, stationary FSO links with reasonable link margins (a few dB) and diversity values (<25), and under typical wind speeds (<40 km/hr), the expected outage length is still multiple orders of magnitude larger than packet sizes; it is typical for outages to cause TCP senders to timeout. For a diversity-N channel where N>1, it is not clear that the channel is well modeled by a 2-state continuous-time Markov process. Thus, we did not calculate the throughputs for N>1 for the range of atmospheric turbulence and congestion parameters as we did for N=1. The appropriate model for a diversity-N channel is an

194

area for future exploration. Assuming that a diversity-16 channel is well modeled by a 2-state continuous-time Markov process, we show in Figure 6.43, the TCP throughput upper bound when a diversity of 16 is used. Even with a diversity of 16, the TCP throughput is poor for long round-trip distances. This is again due to outages causing window reduction to one packet and the slow window increase afterwards. For long round-trip distances, diversity alone is not enough to allow TCP to achieve high throughputs, and Modified TCP or another protocol is needed. The plot assumes strong turbulence ($\sigma_\chi^2 = 0.5$) because we want to show that if we communicate through strong atmospheric turbulence, the throughput is low even with a significant amount of diversity.
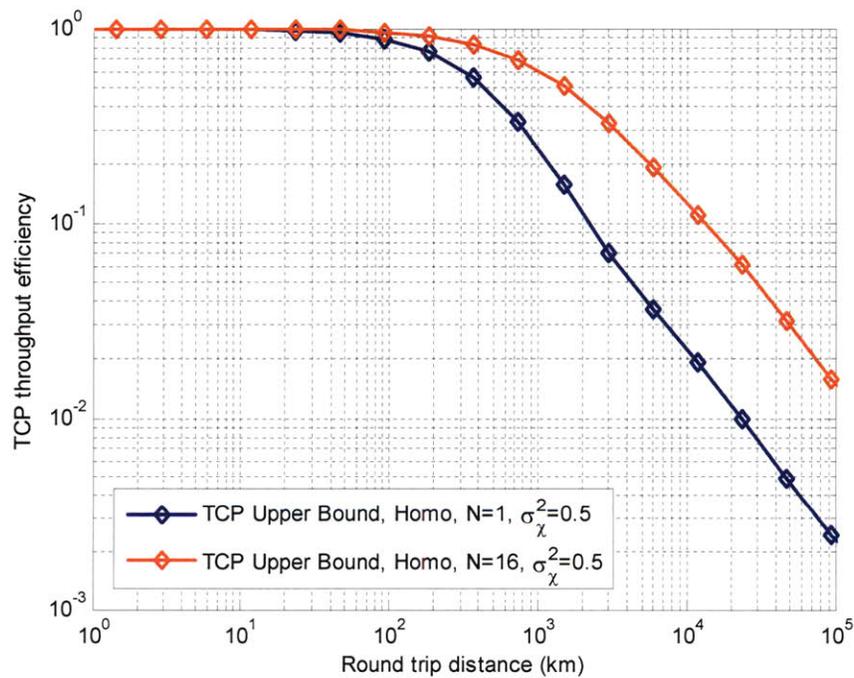
Figure 6.43: TCP throughput efficiency for zero congestion loss probability, R=10 Gb/s, G=1.5 kbytes, m=8 dB, P_e^{thresh}=0.1, v_t=10km/hr, N_n=10^{-6}

6.5.2 Transient Throughput of TCP and Modified TCP

In this section, we show plots of and discuss the expected number of packets sent in the Kth and in the first K round-trip times after session initiation. We also provide expressions of the value to which the expected number of packets sent in the Kth RTT converges and discuss the files sizes for which Modified TCP provides significant benefit over TCP.

See Figures 6.44-6.113 for plots of the amount of data transmitted in the Kth RTT and up to the Kth RTT by TCP and Modified TCP for various congestion loss probabilities and turbulence levels. These plots use the same parameters as used for the steady state throughput plots but where the round-trip time is fixed to 0.3 seconds[16]. The dotted curves (approximate curves) correspond to the expressions given in (6.33)-(6.36). These approximate curves are good for the first few time steps for both Modified TCP and TCP and are particularly good for small congestion probabilities and turbulence. The approximate curves diverge from the actual curves as time increases to higher values because as window sizes increase, it becomes more likely that congestion or outages cause the window to be reduced. We cut off the approximate curves after they diverge from the actual curves since they are no longer useful and are visually distracting. When plotting the linear increase curves, in order to reduce the computer memory and calculations required (especially important for long round-trip distances), we cut off the number of states at the maximum number of time units (in RTTs) for which we plotted[17] i.e. K=2000. This elimination of states has no effect on the final calculated throughput value because the states that are cut

---

[16] A significant round-trip time of 0.3 seconds is chosen to highlight that a) if the window increase is always linear, it takes a large number of round-trip times to reach steady state, and b) for typical congestion loss probabilities and when link fading exists, the expected window size never reaches a value large enough to achieve full link rate

[17] We plotted up to the 2000th RTT because the program we used, MATLAB on a 32-bit Windows machine, runs out of application memory when doing matrix multiplications for K>2000.

off are unoccupied up from the first RTT to the K=2000th RTT; when the window increases linearly, it cannot increase to a value larger than the current time (measured in RTTs) and thus cannot be larger than the maximum time.
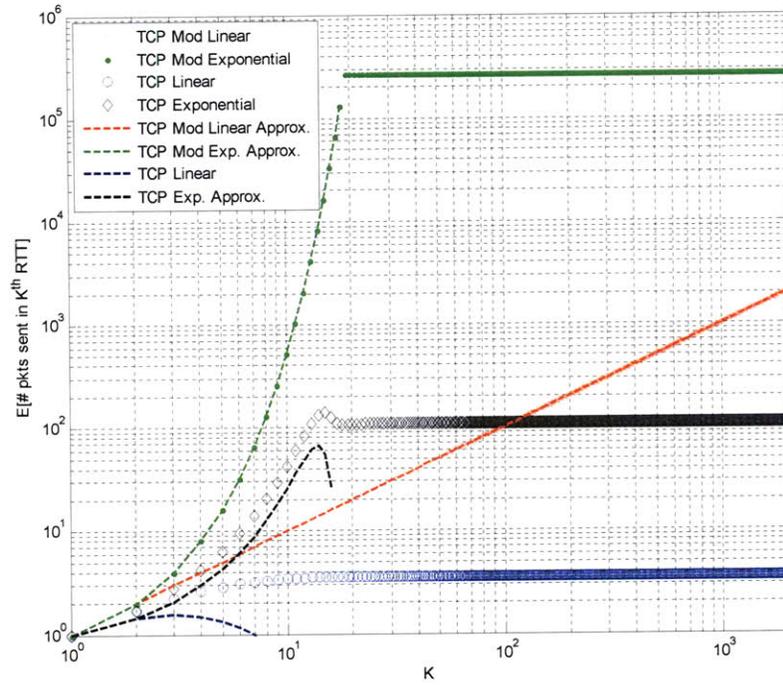
Figure 6.44: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 0 and $\sigma_x^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)
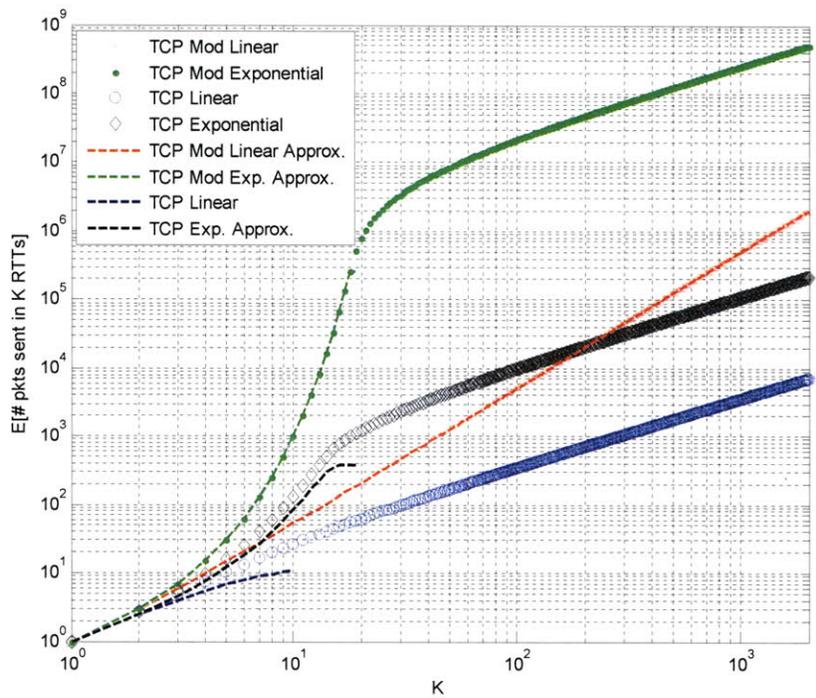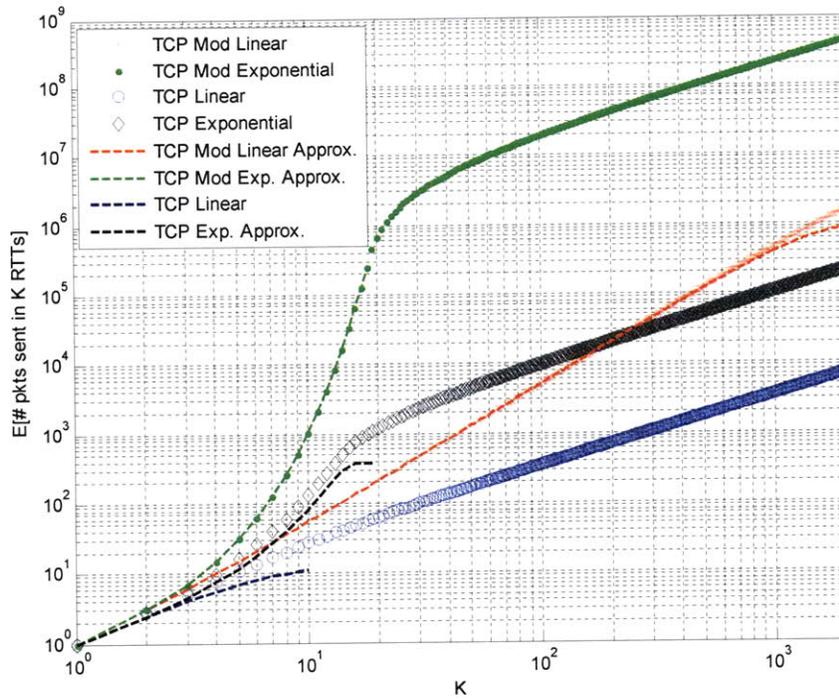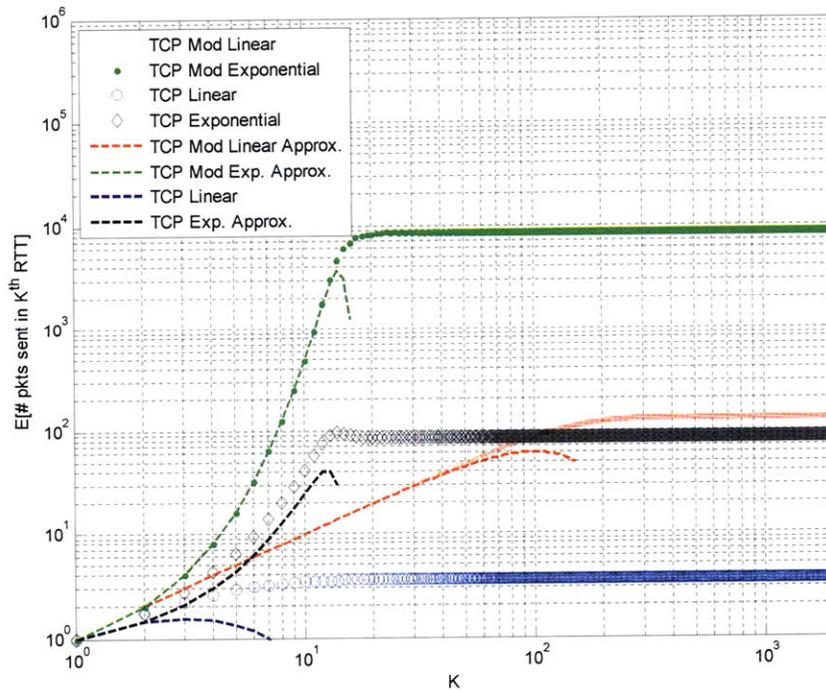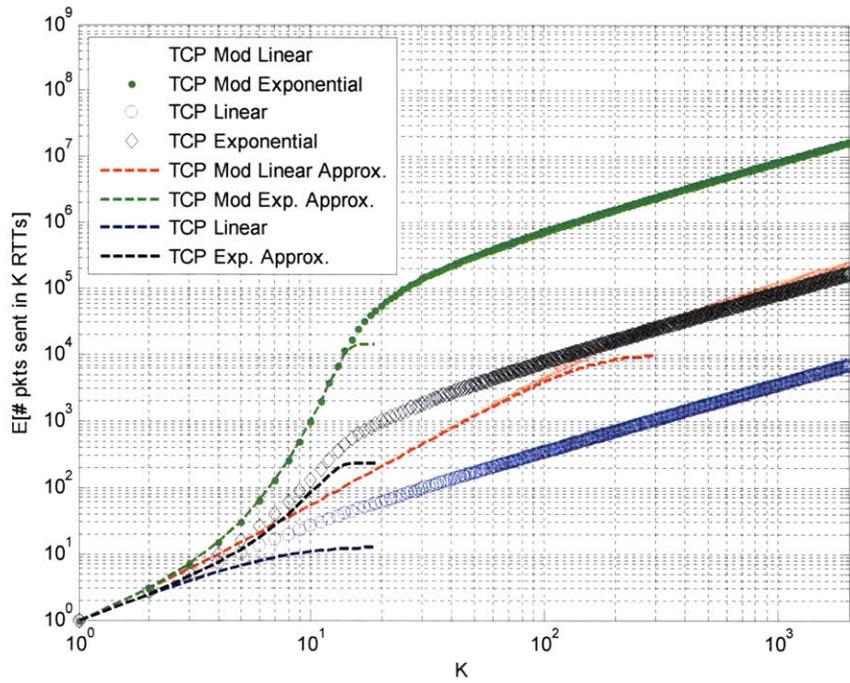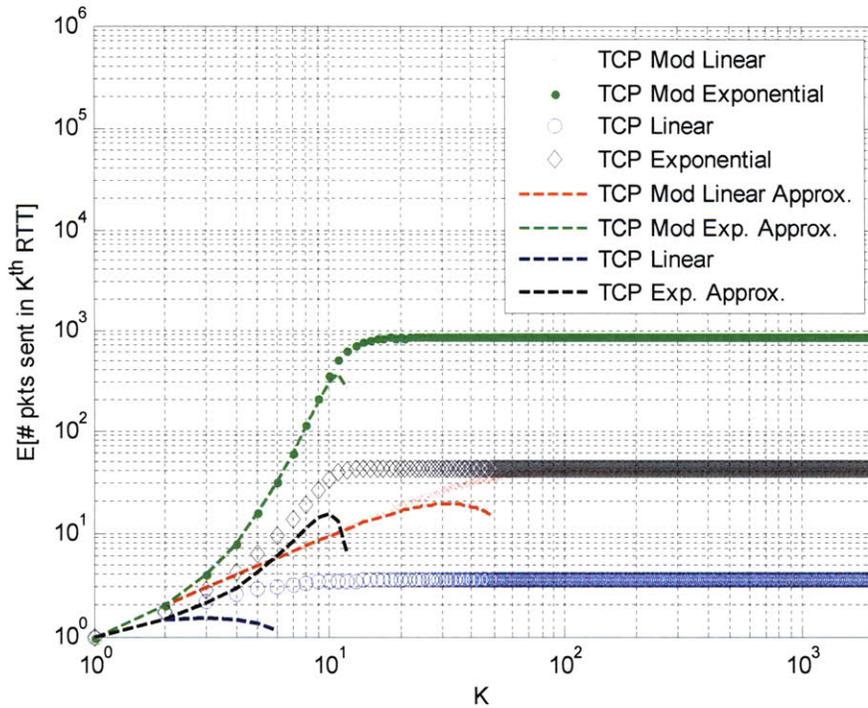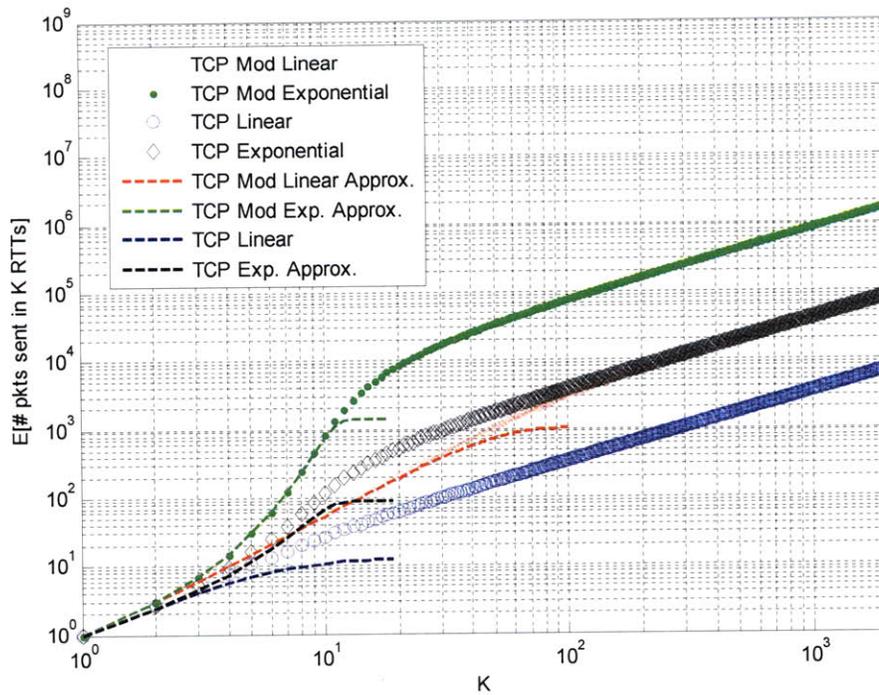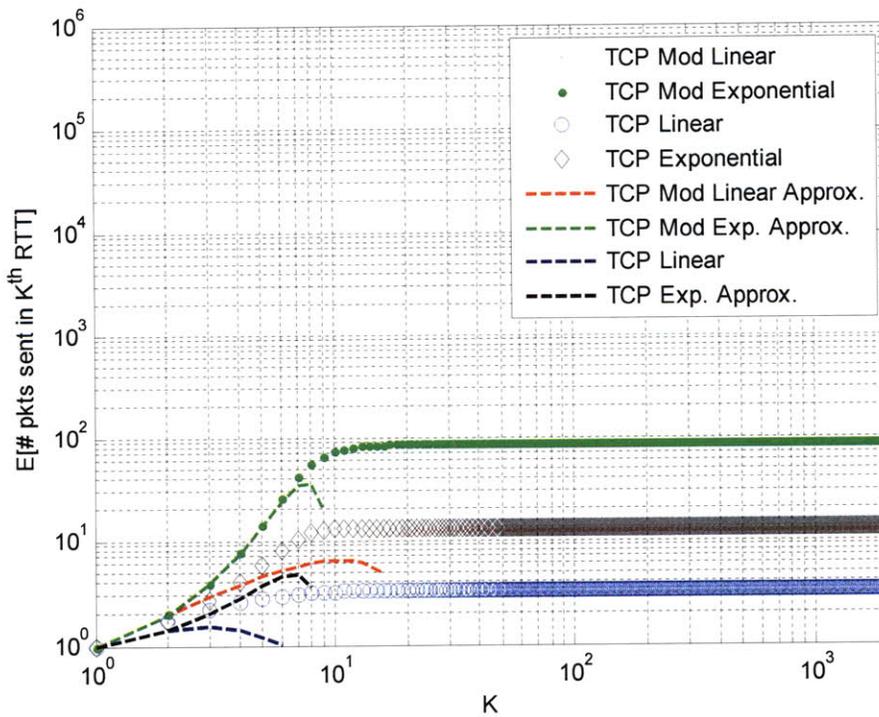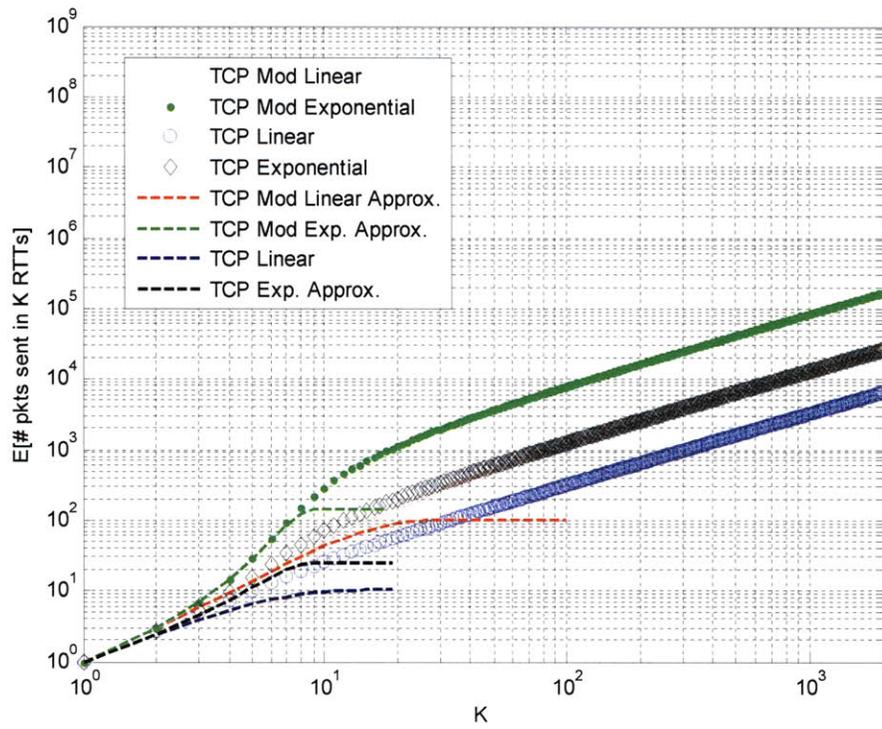
Figure 6.45: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 0 and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)
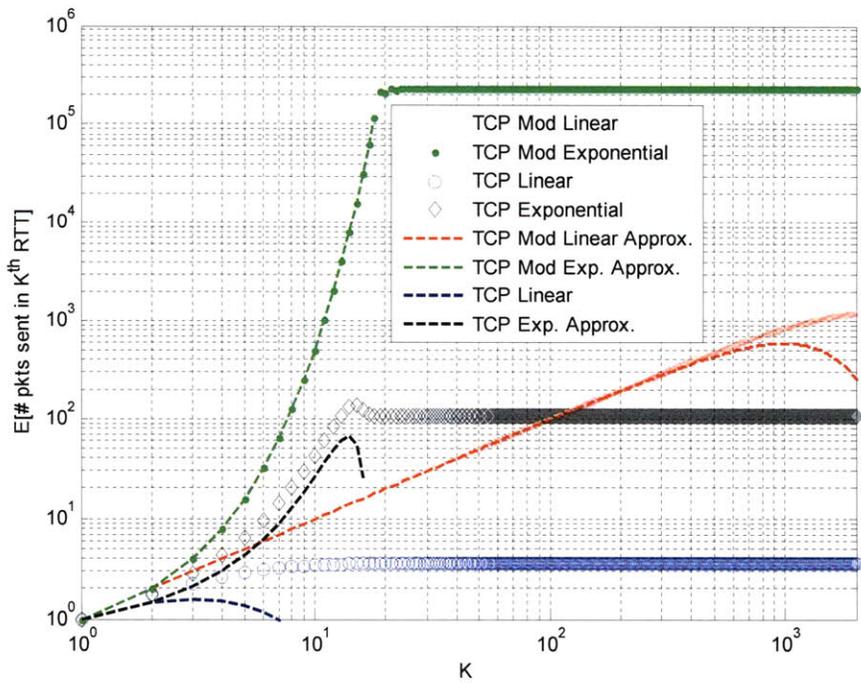
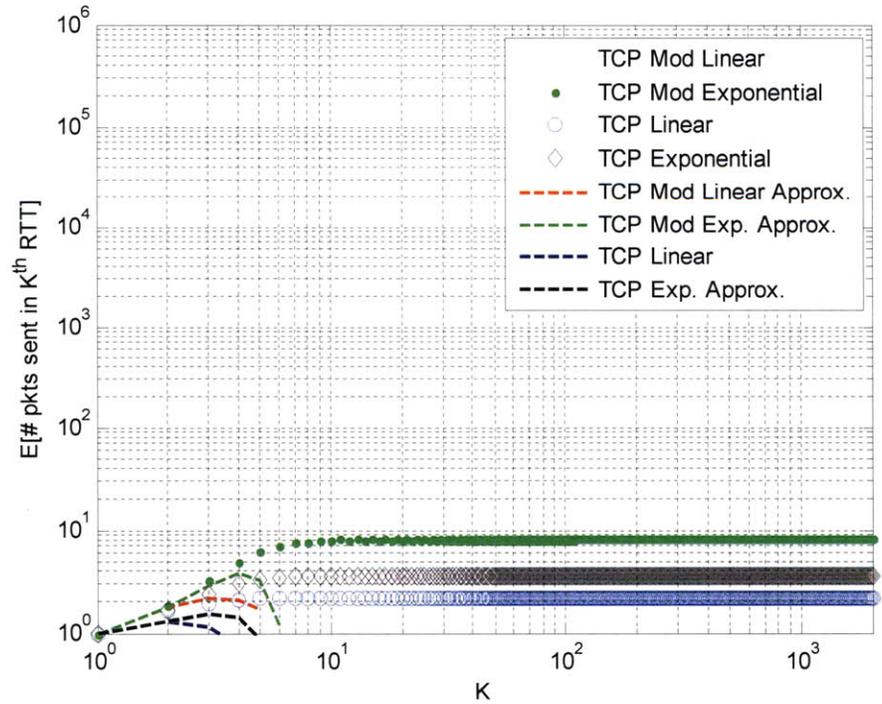Figure 6.46: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-6} and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)
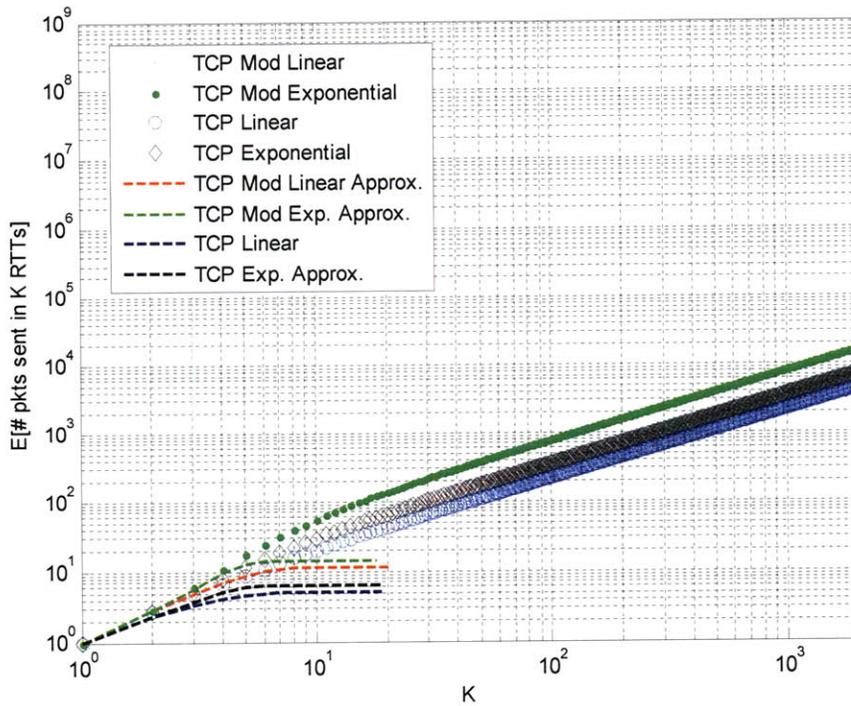
Figure 6.47: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-6} and $\sigma_X^2 = 0.5$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.48: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-4} and $\sigma_X^2 = 0.5$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.49: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-4} and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)
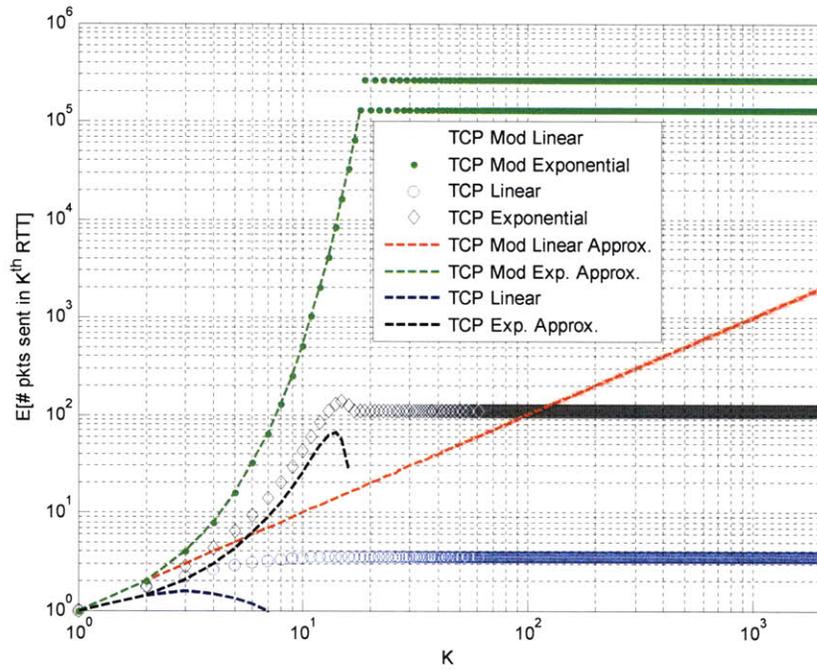

Figure 6.50: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-3} and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.51: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-3} and $\sigma_X^2 = 0.5$ and $M = 2^{18}$ (RTT=0.3 sec)
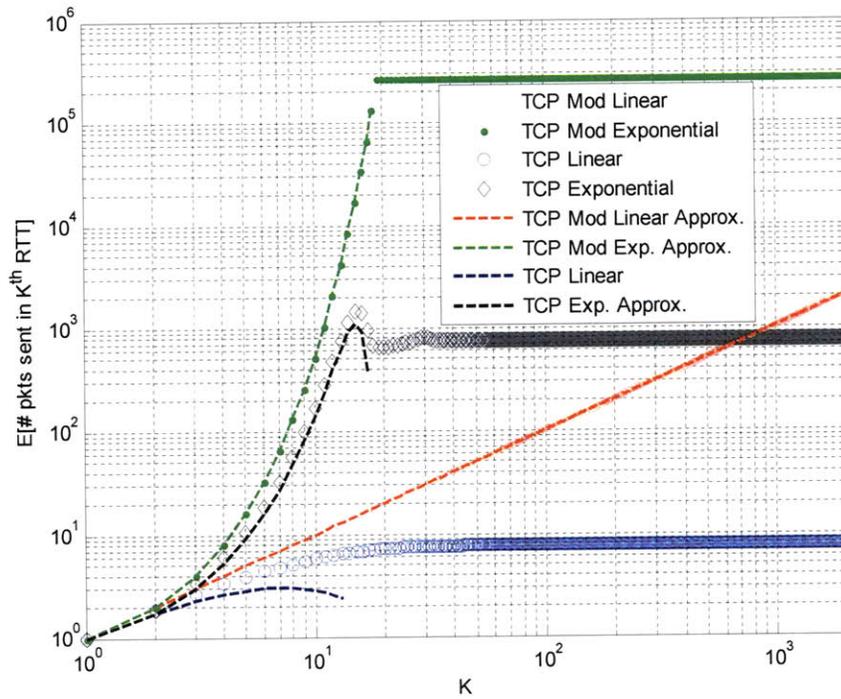

Figure 6.52: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-2} and $\sigma_X^2 = 0.5$ and $M = 2^{18}$ (RTT=0.3 sec)
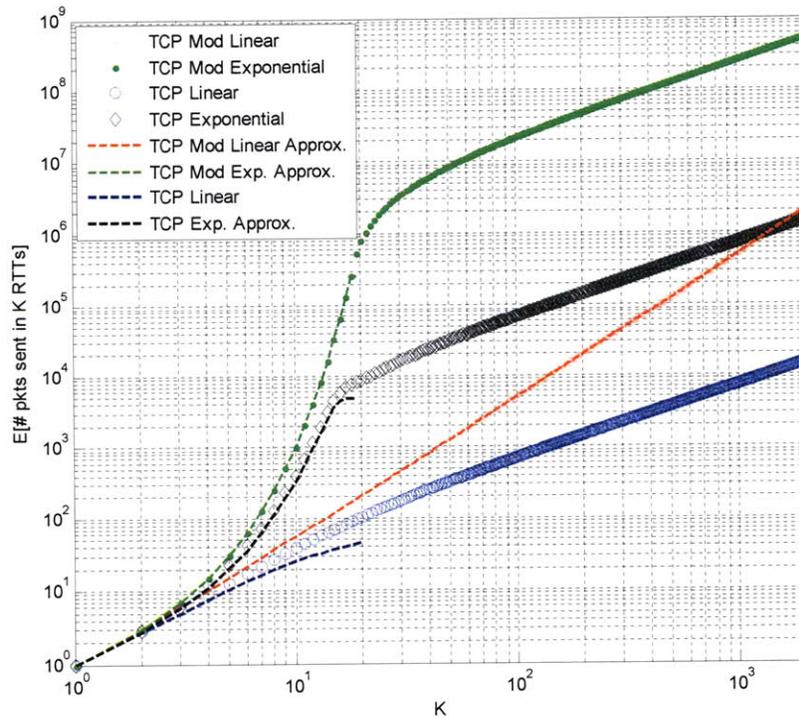
Figure 6.53: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-2} and $\sigma_x^2 = 0.5$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.54 (a) on same scale as other E[# pkts sent in K^{th} RTT] plots
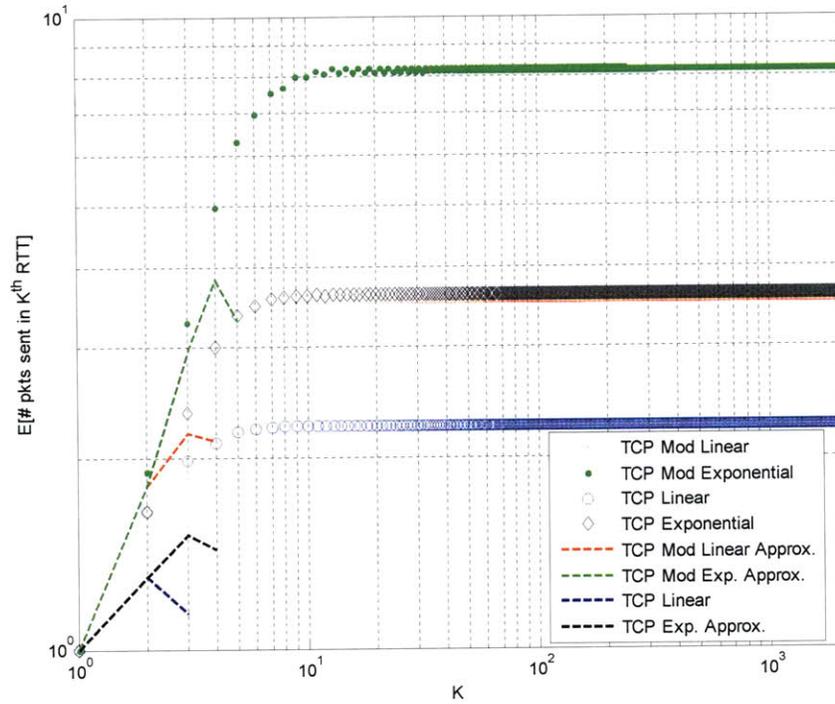
Figure 6.54 (b) zoomed on vertical axis in for better resolution

Figure 6.54: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-1} and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.55: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-1} and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)
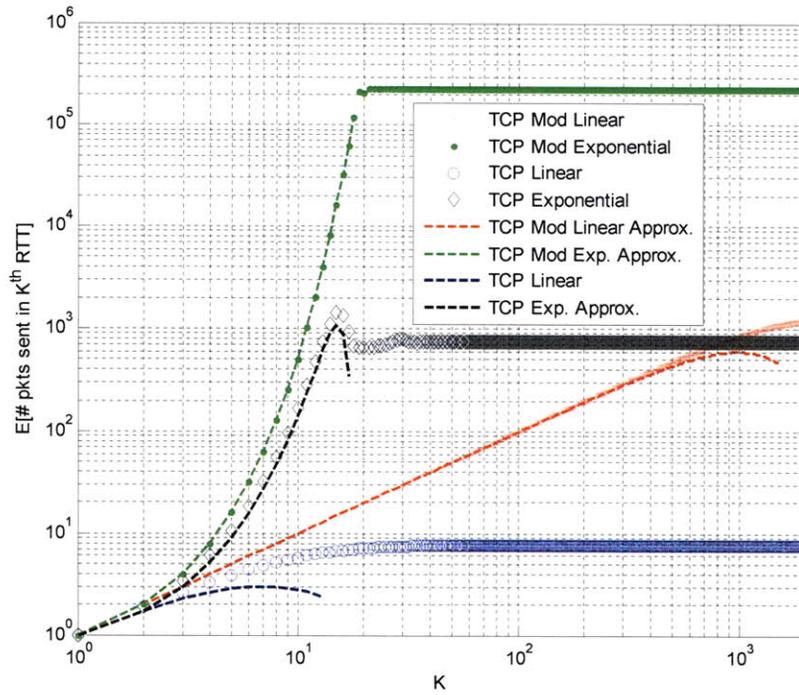
Figure 6.56: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet being modeled as a step function and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)
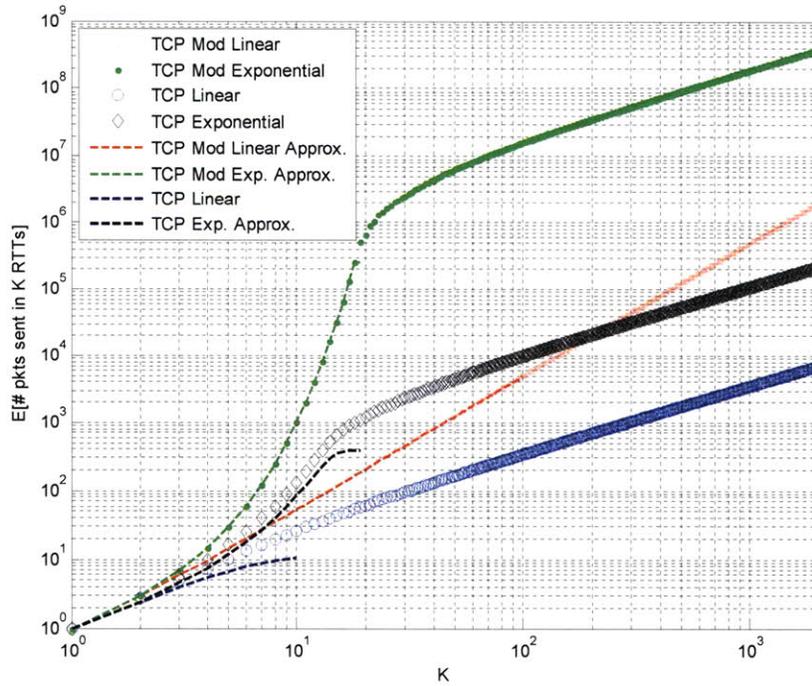
Figure 6.57: Expected number of packets sent in K round-trip intervals for congestion loss per packet being modeled as a step function and $\sigma_X^2=0.5$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.58: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 0 and $\sigma_X^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)
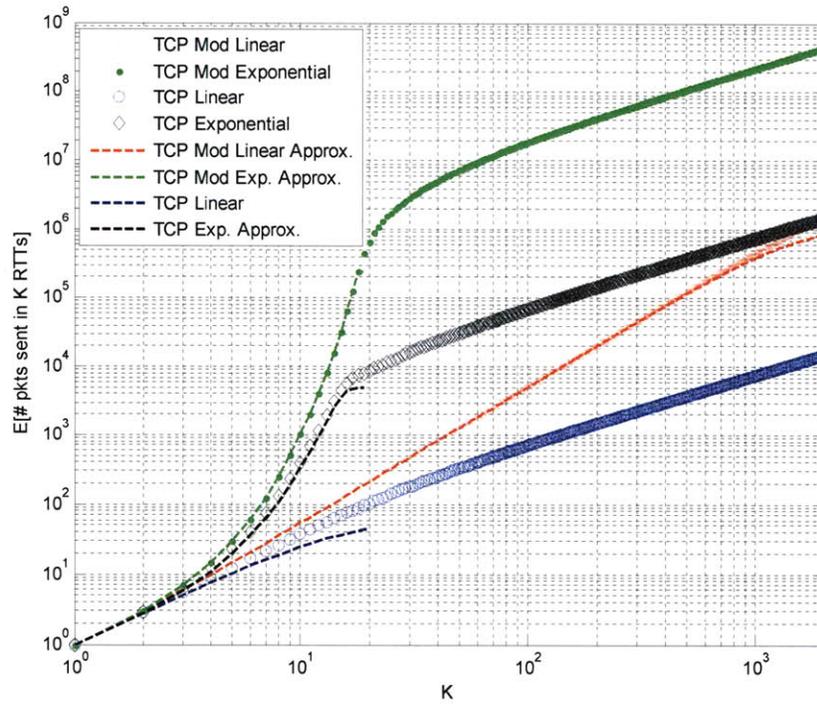

Figure 6.59: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 0 and $\sigma_X^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)
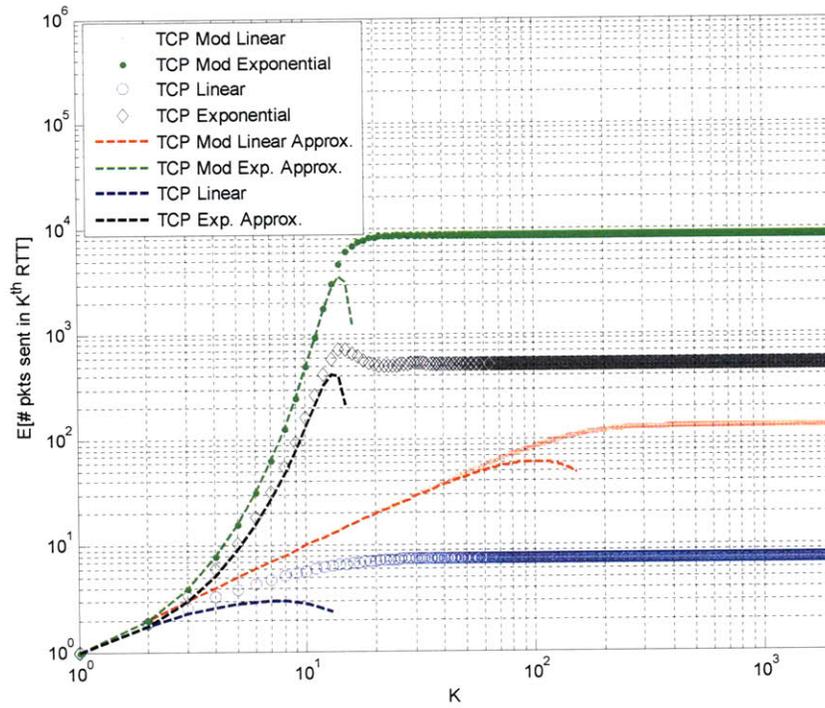
Figure 6.60: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-6} and $\sigma_X^2 = 0.3$ and $M = 2^{18}$ (RTT=0.3 sec)
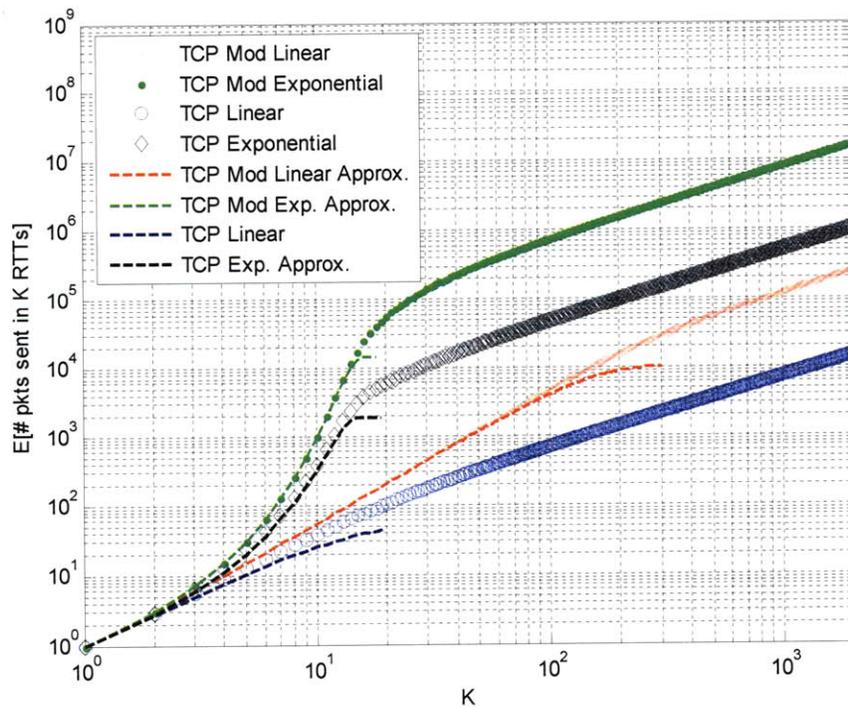

Figure 6.61: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-6} and $\sigma_X^2 = 0.3$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.62: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-4} and $\sigma_X^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.63: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-4} and $\sigma_X^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)
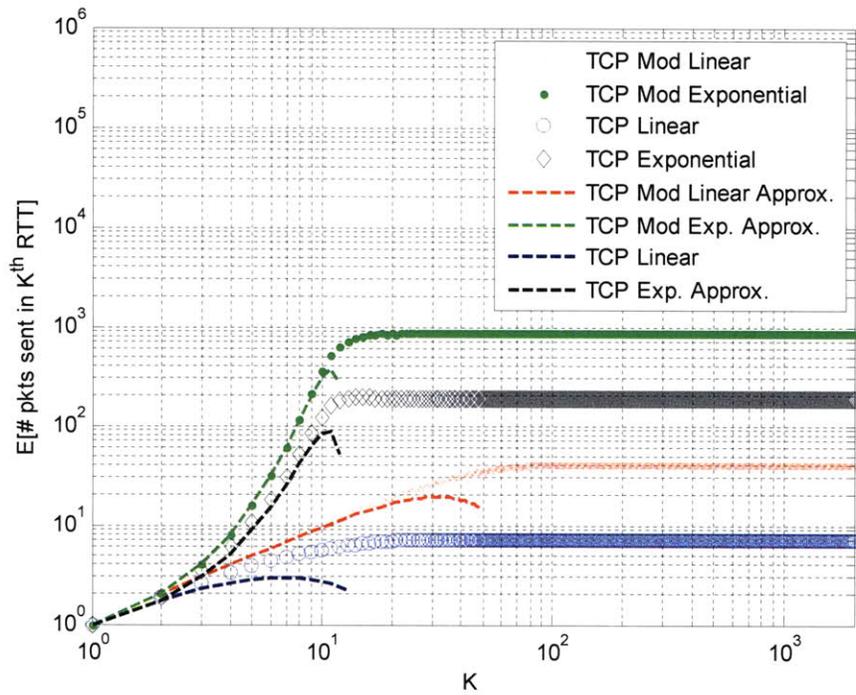
Figure 6.64: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-3} and $\sigma_X^2 = 0.3$ and $M = 2^{18}$ (RTT = 0.3 sec)
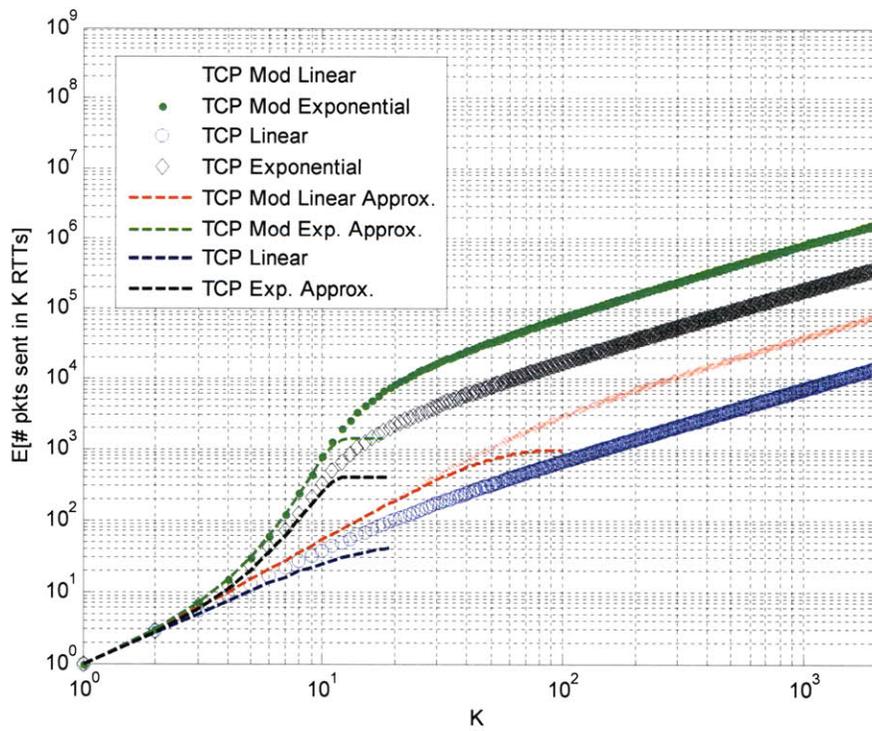

Figure 6.65: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-3} and $\sigma_X^2 = 0.3$ and $M = 2^{18}$ (RTT = 0.3 sec)
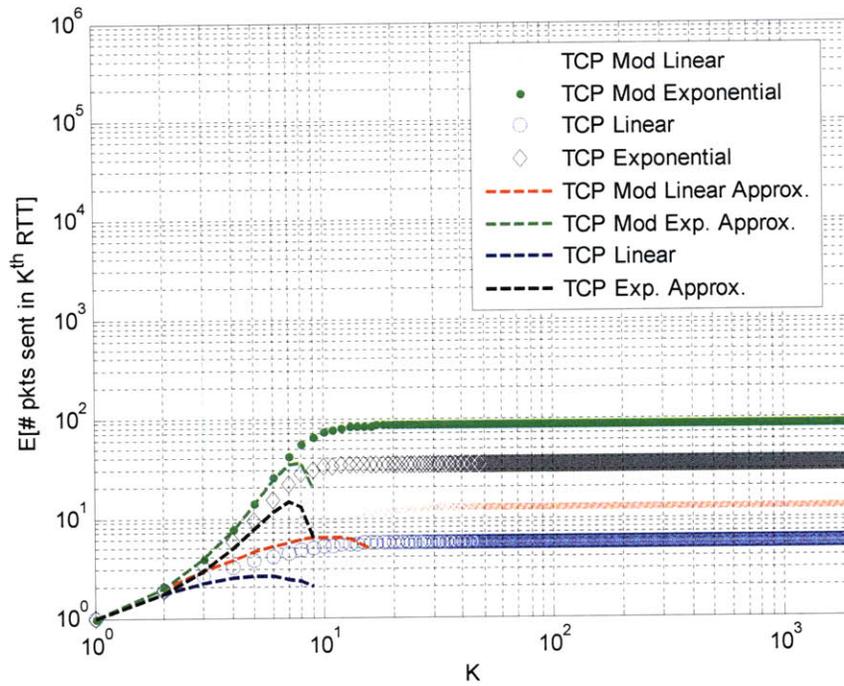
Figure 6.66: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-2} and $\sigma_x^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)
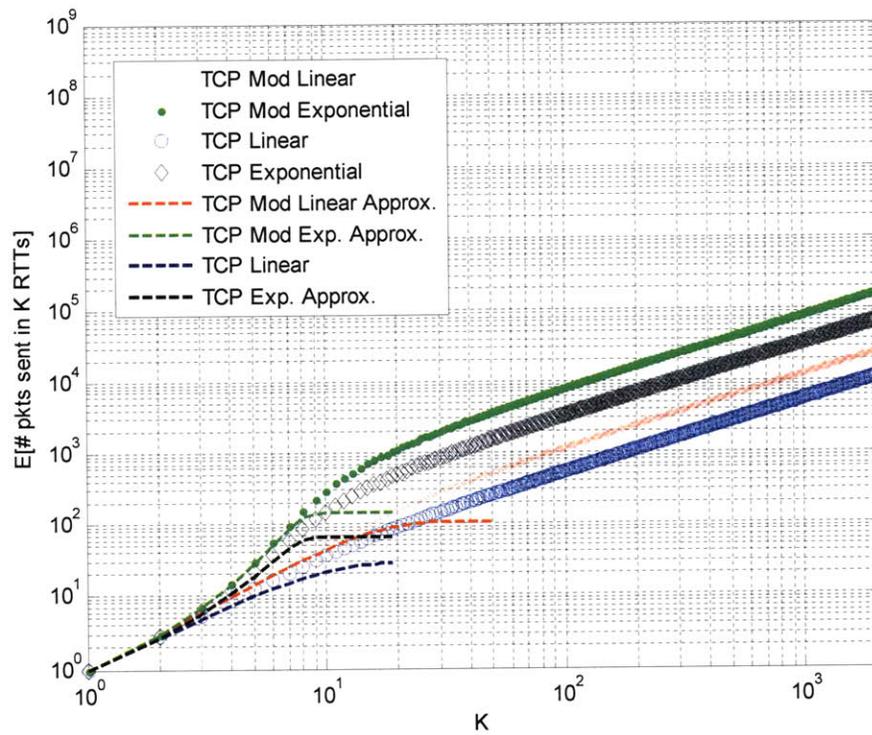

Figure 6.67: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-2} and $\sigma_x^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.68 (a) on same scale as other E[# packets sent in K^{th} RTT] plots

Figure 6.68 (b) zoomed in on vertical axis for better resolution

Figure 6.68: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-1} and $\sigma_X^2 = 0.3$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.69: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-1} and $\sigma_X^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)
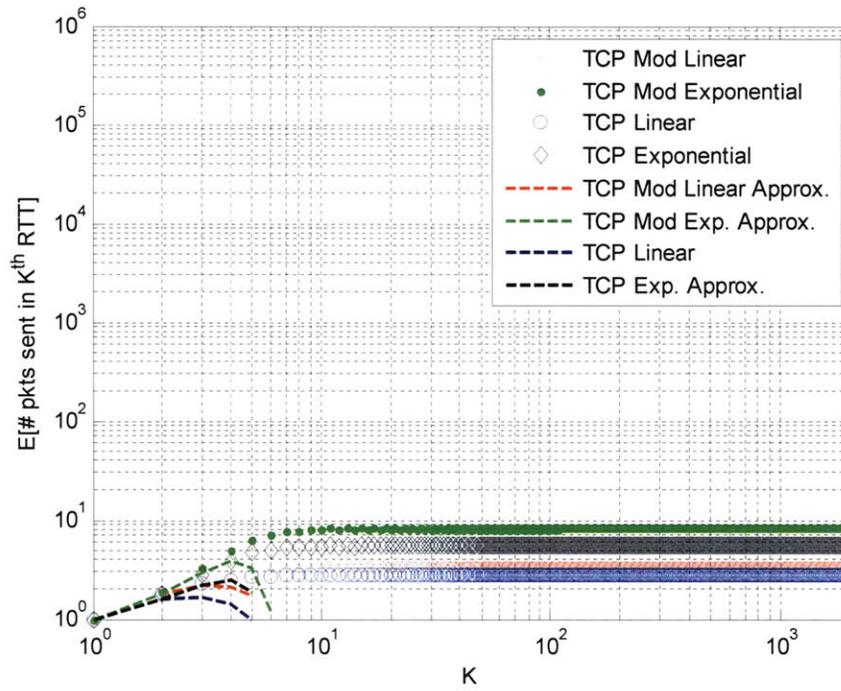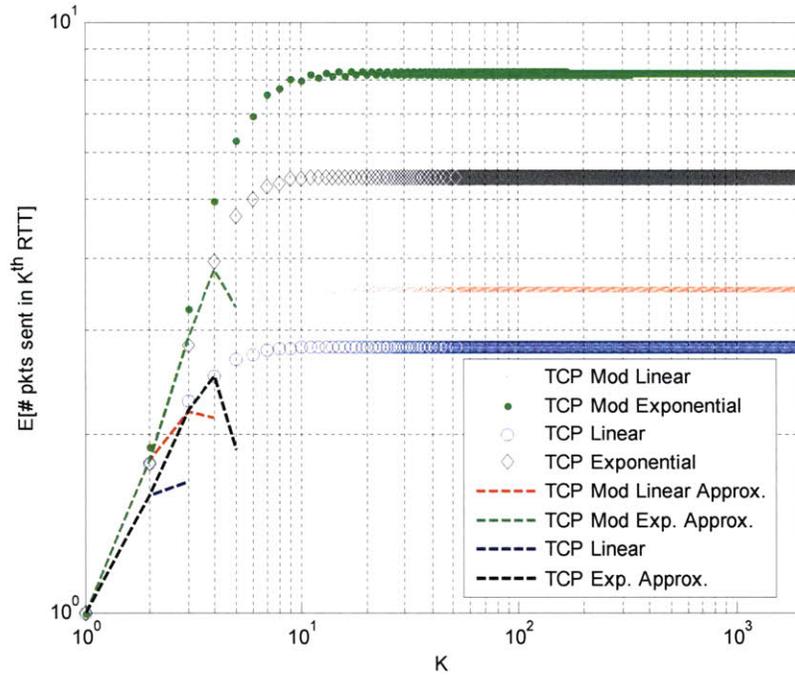

Figure 6.70: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet being modeled as a step function and $\sigma_X^2=0.3$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.71: Expected number of packets sent in K round-trip intervals for congestion loss per packet being modeled as a step function and $\sigma_x^2 = 0.3$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.72: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 0 and $\sigma_x^2 = 0.1$ and $M = 2^{18}$ (RTT=0.3 sec)
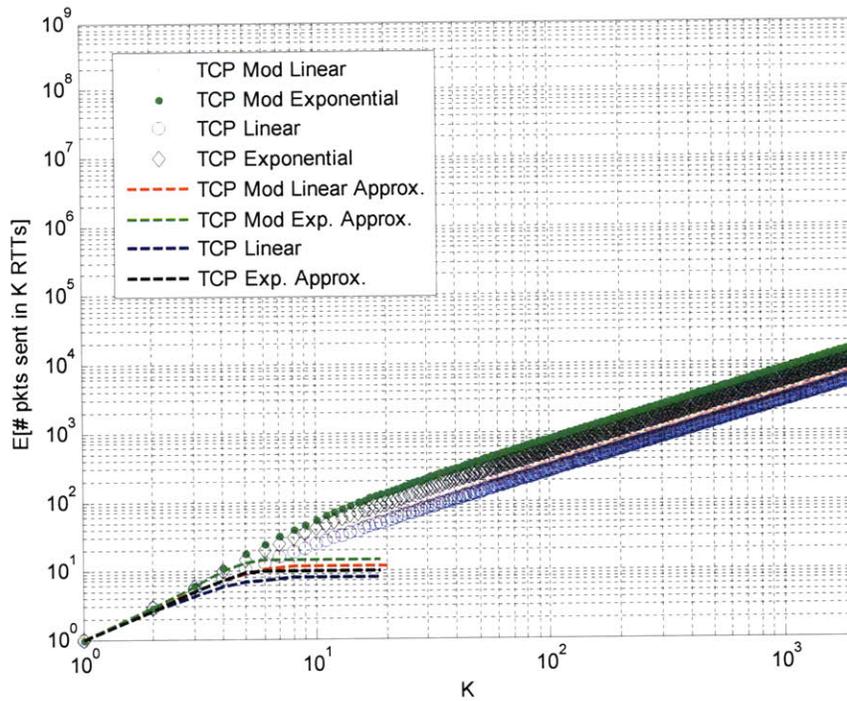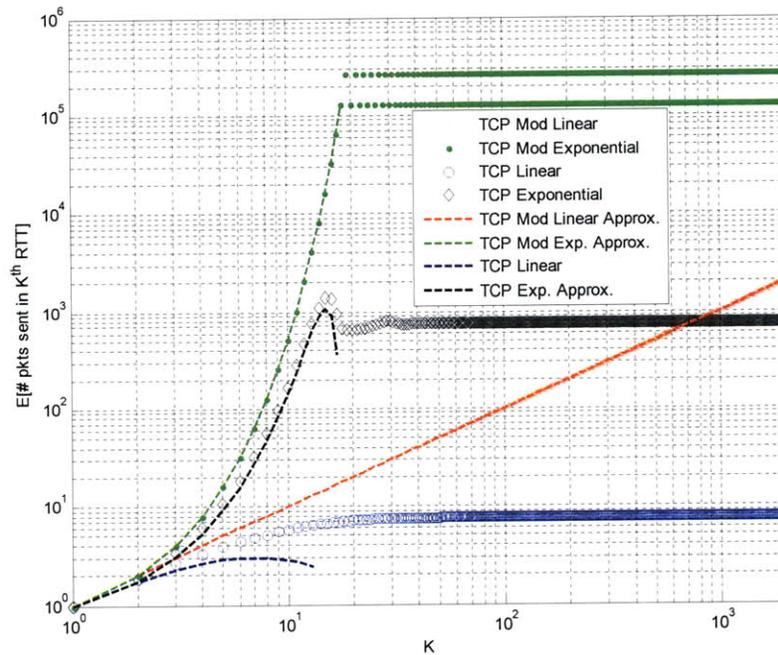
Figure 6.73: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 0 and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.74: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-6} and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.75: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-6} and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.76: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-4} and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)
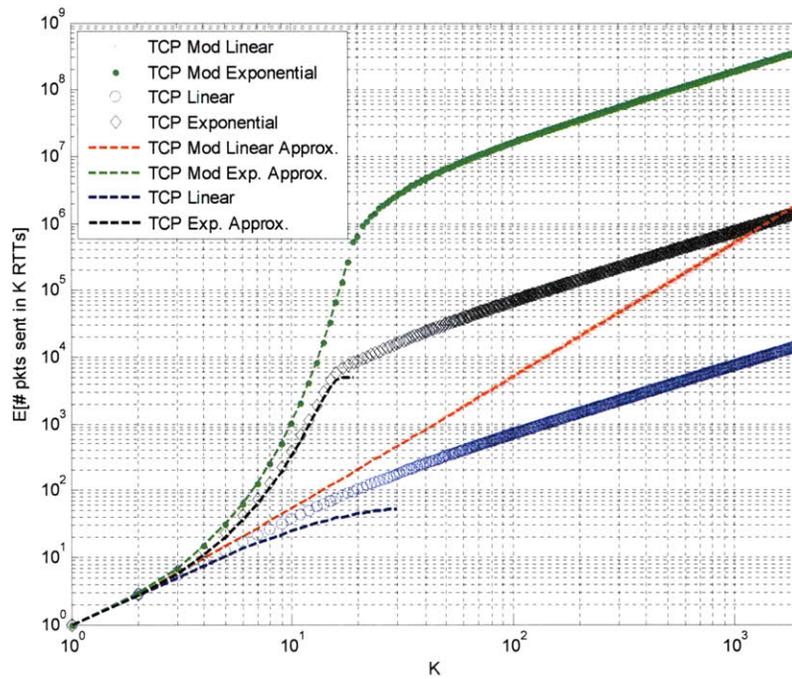
Figure 6.77: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-4} and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)
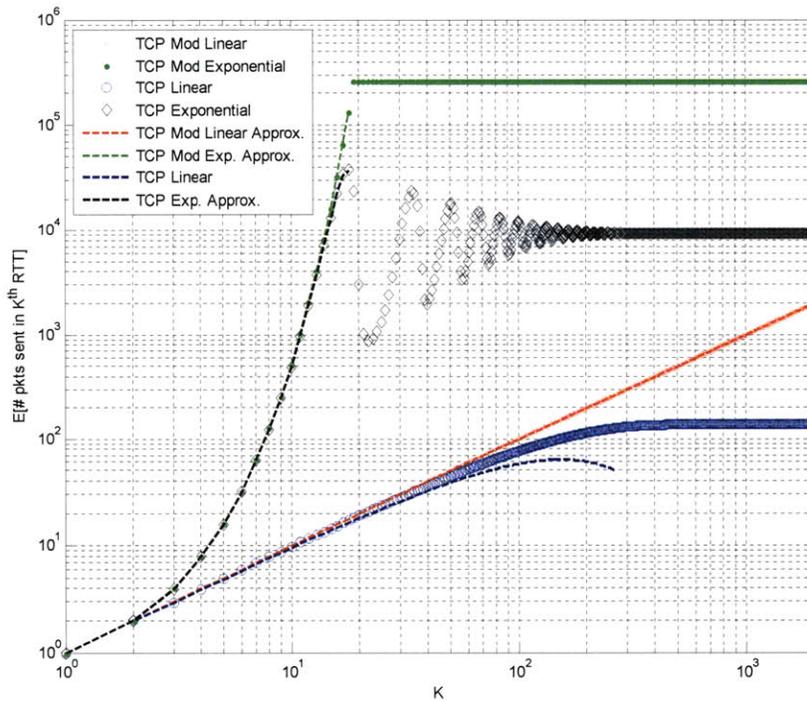

Figure 6.78: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-3} and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)
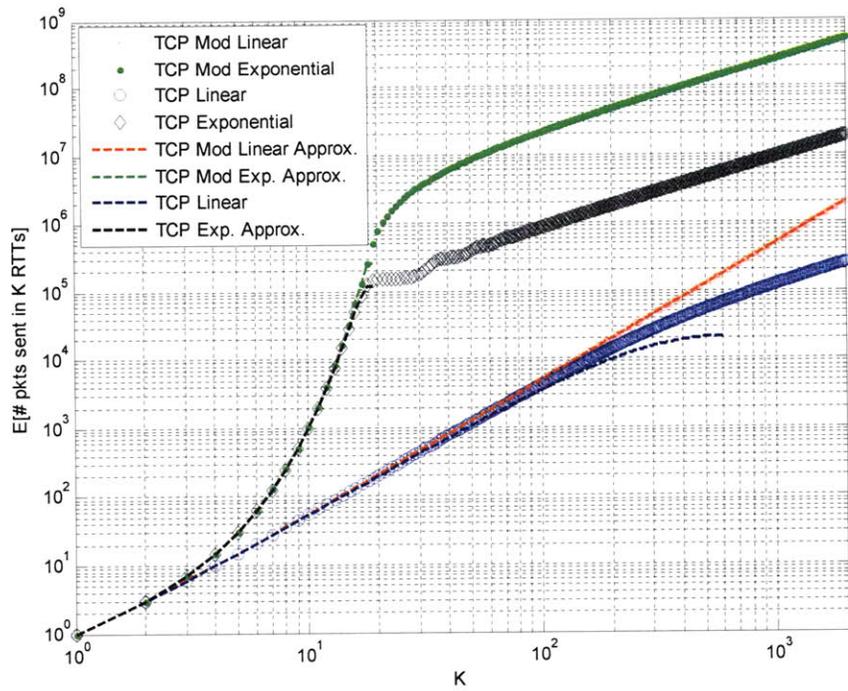
Figure 6.79: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-3} and $\sigma_x^2 = 0.1$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.80: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-2} and $\sigma_x^2 = 0.1$ and $M = 2^{18}$ (RTT=0.3 sec)

Figure 6.81: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-2} and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)
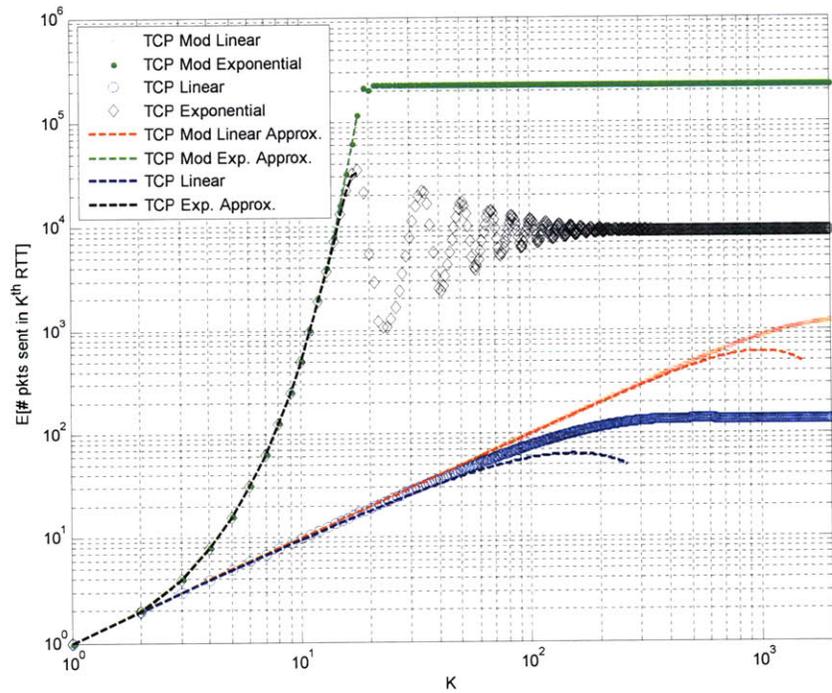

Figure 6.82: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-1} and $\sigma_X^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)
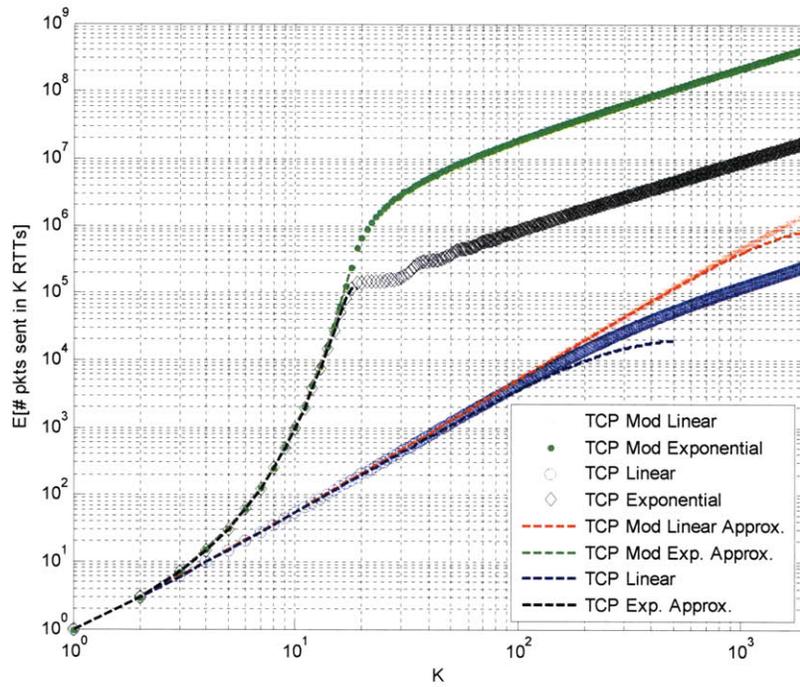
Figure 6.83: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-1} and $\sigma_x^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)
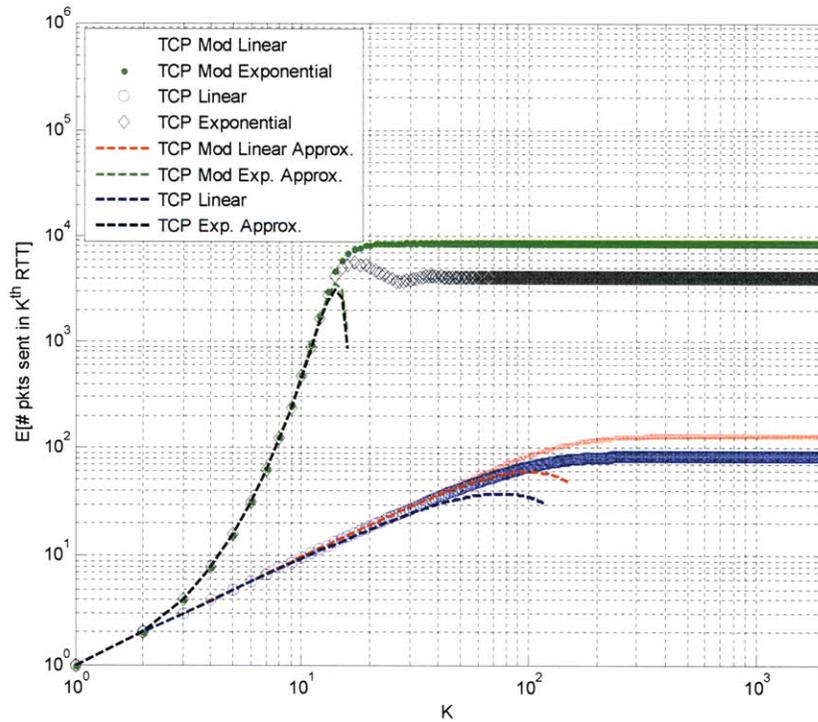

Figure 6.84: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet being modeled as a step function and $\sigma_x^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.85: Expected number of packets sent in K round-trip intervals for congestion loss per packet being modeled as a step function and $\sigma_x^2=0.1$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.86: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 0 and $\sigma_x^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
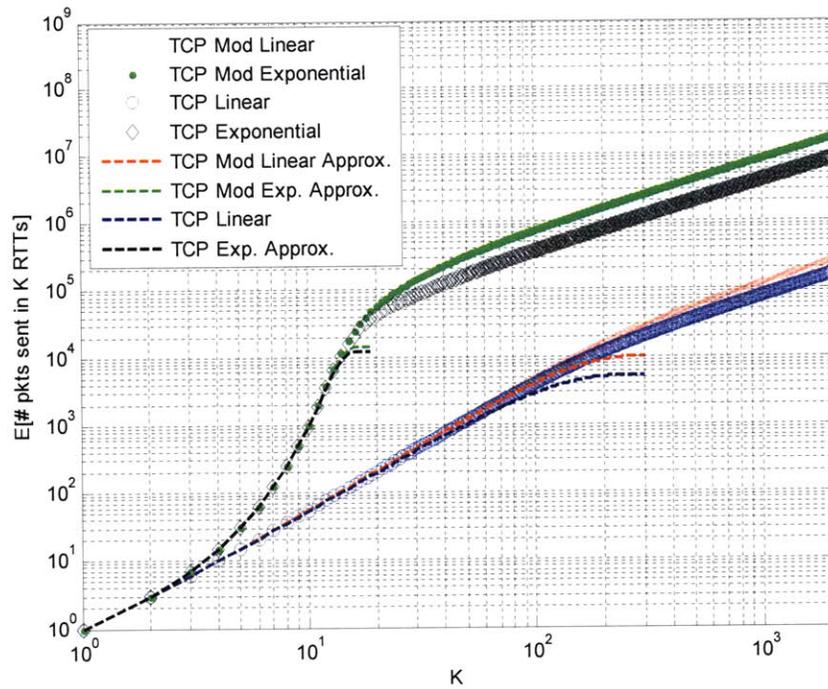
Figure 6.87: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 0 and $\sigma_x^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.88: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-6} and $\sigma_x^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.89: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-6} and $\sigma_x^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
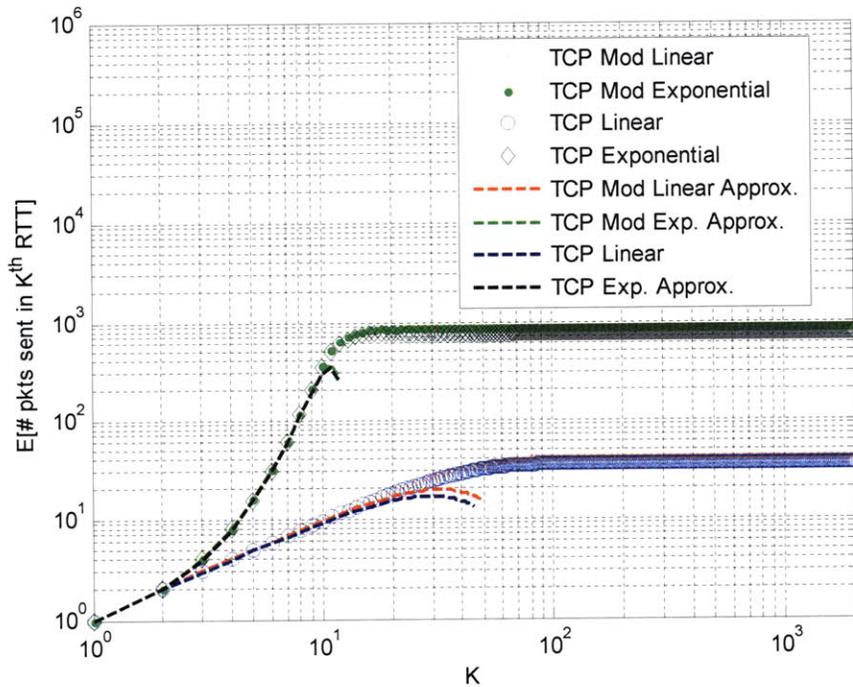

Figure 6.90: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-4} and $\sigma_x^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
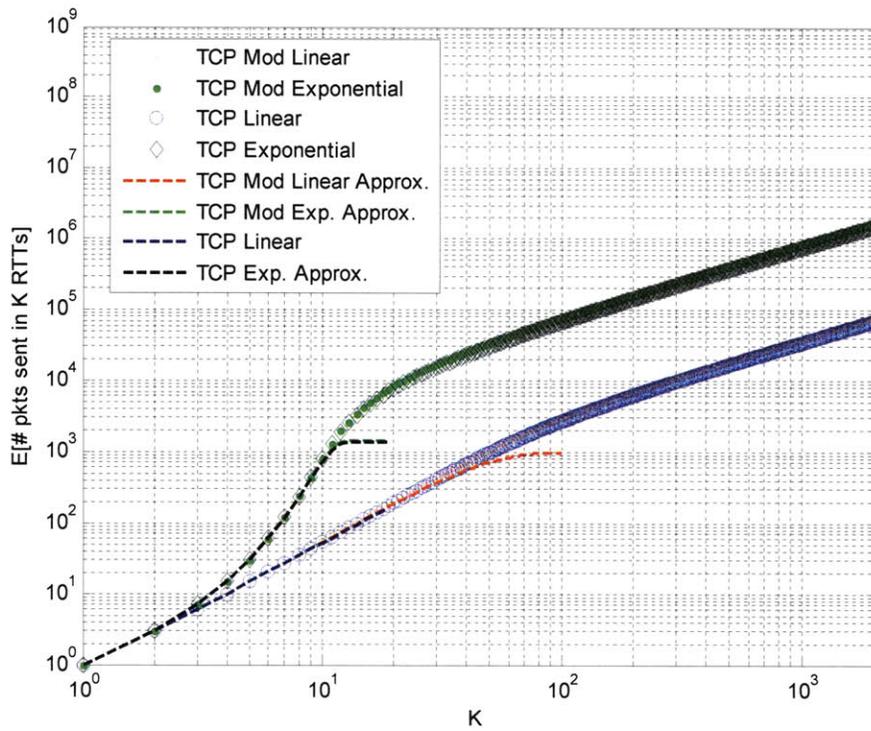
221

Figure 6.91: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-4} and $\sigma_X^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)

Figure 6.92: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-3} and $\sigma_X^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
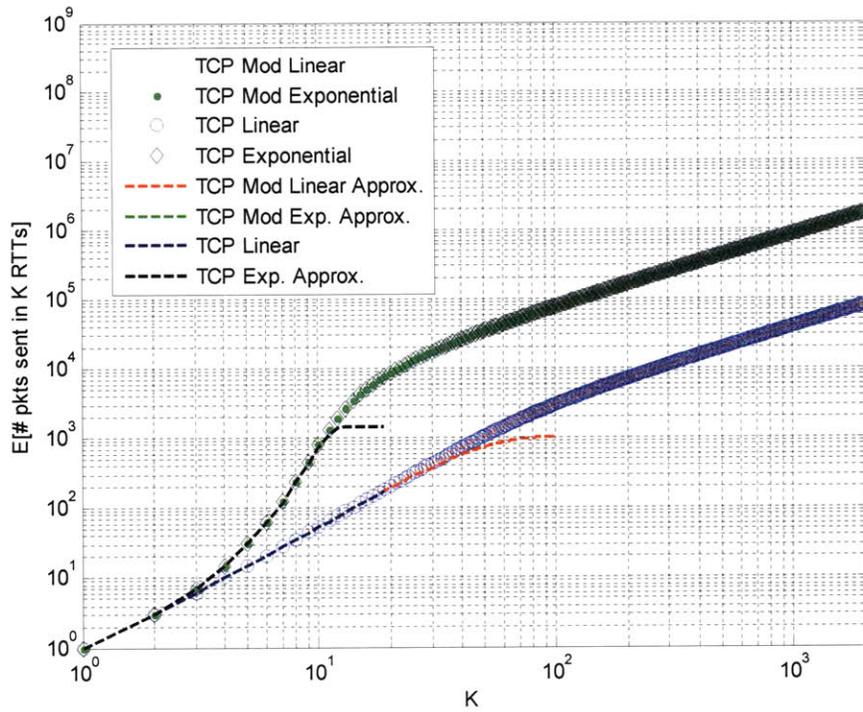
Figure 6.93: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-3} and $\sigma_X^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
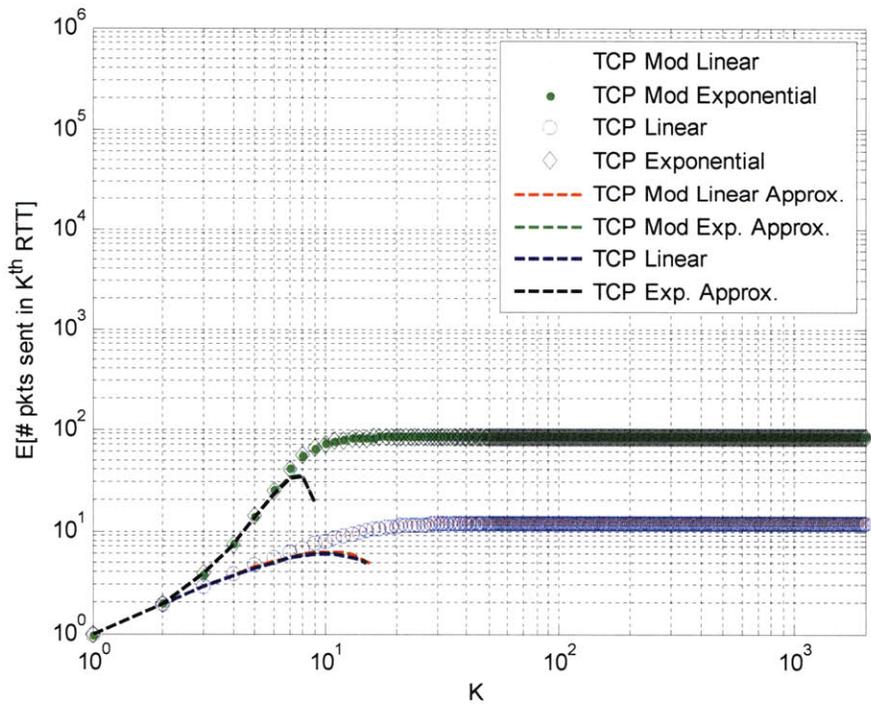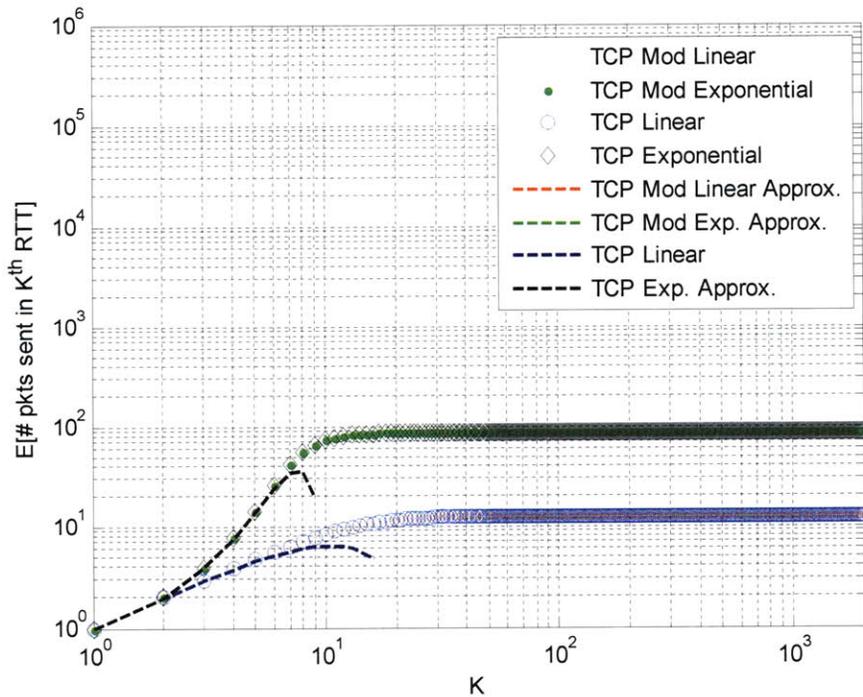

Figure 6.94: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-2} and $\sigma_X^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
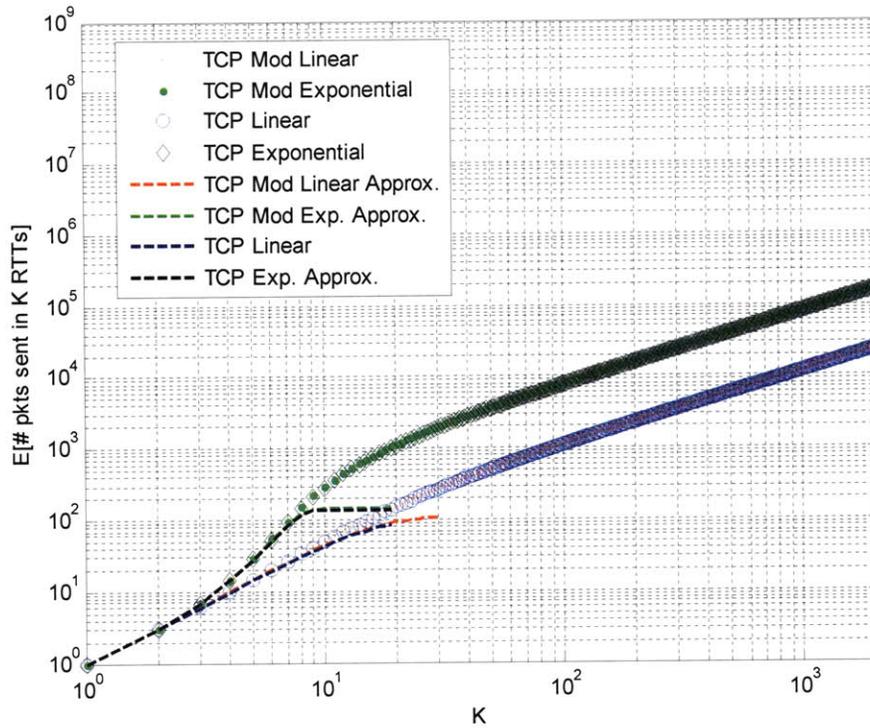
223
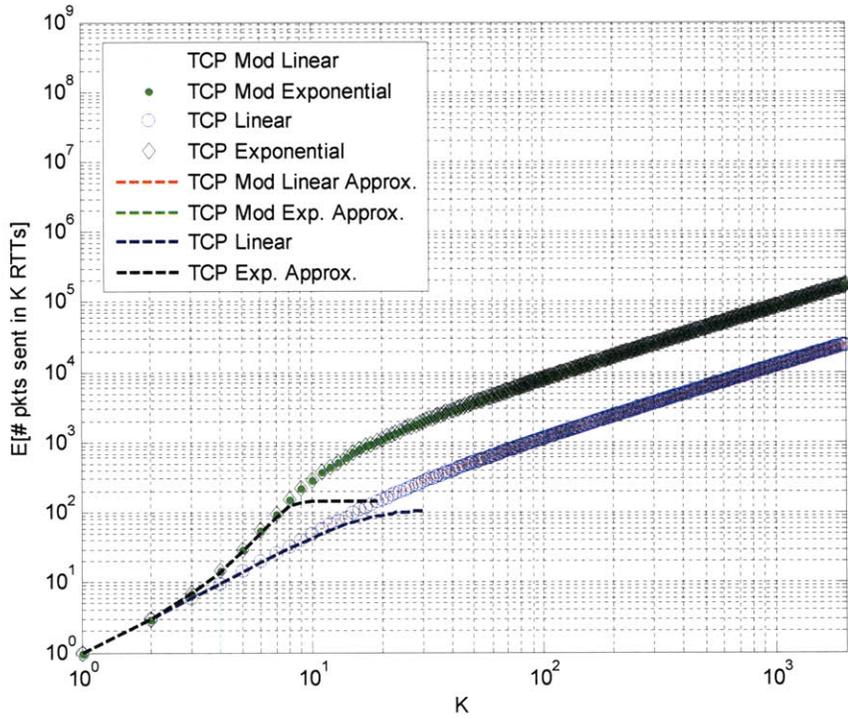
Figure 6.95: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-2} and $\sigma_X^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
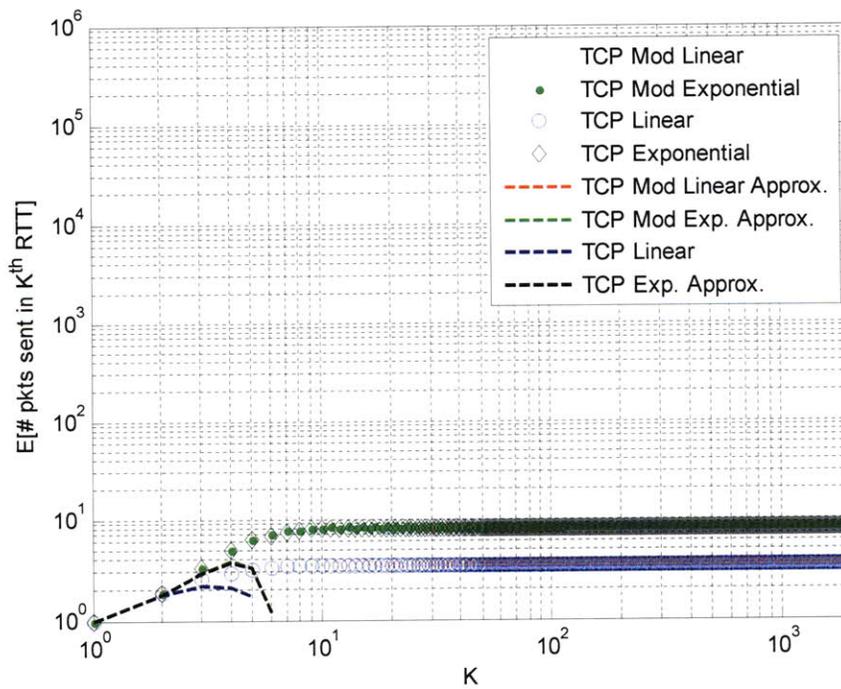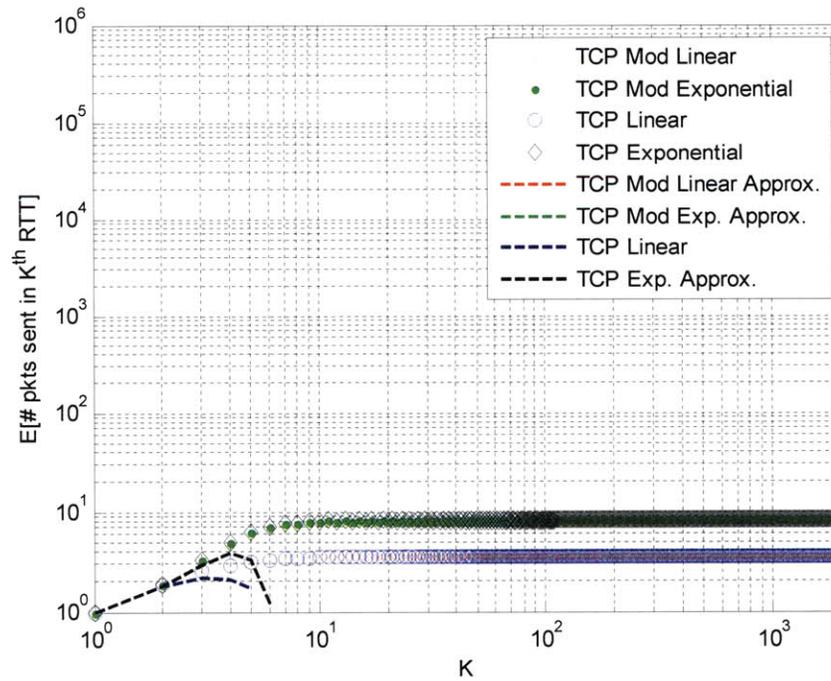

Figure 6.96: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-1} and $\sigma_X^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
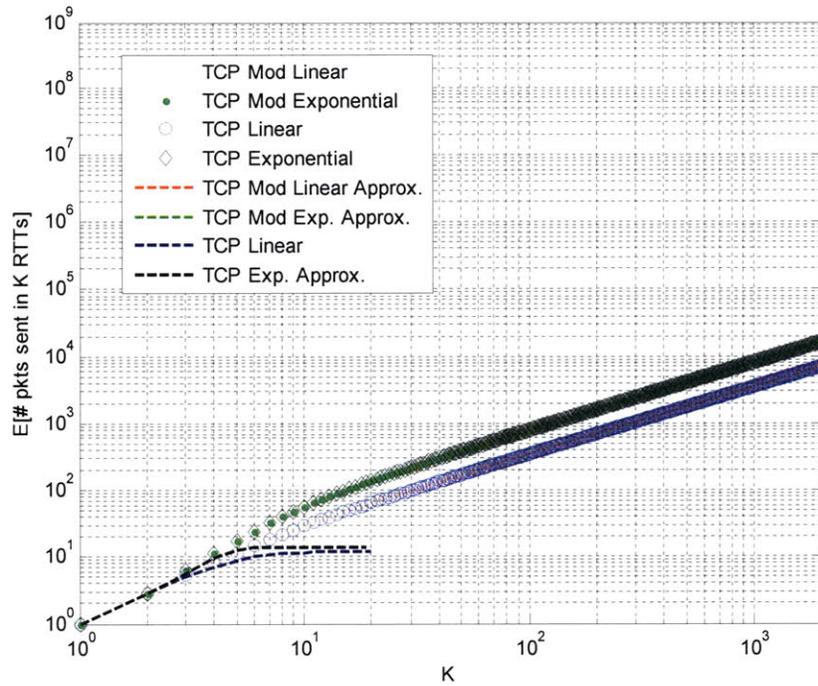
Figure 6.97: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-1} and $\sigma_x^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
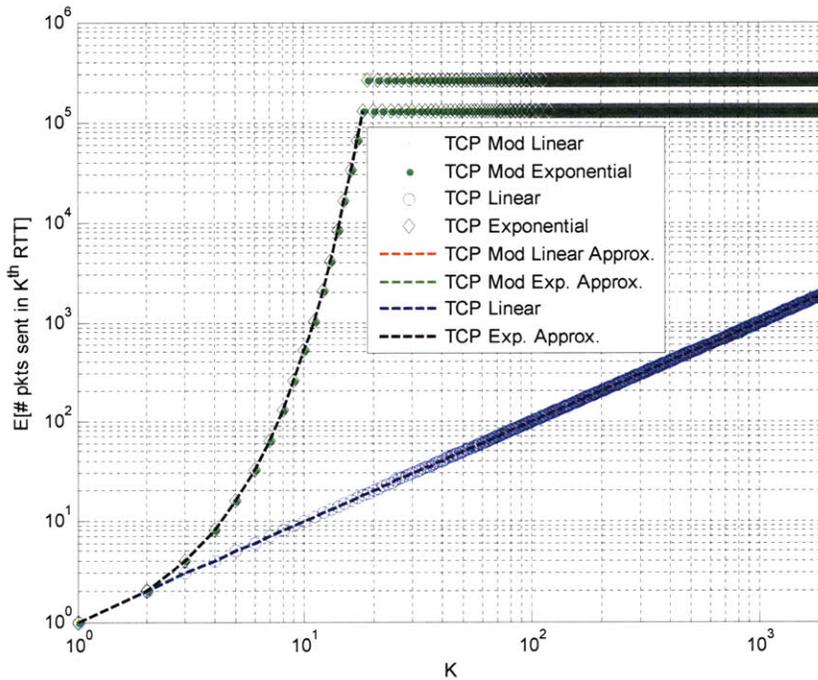
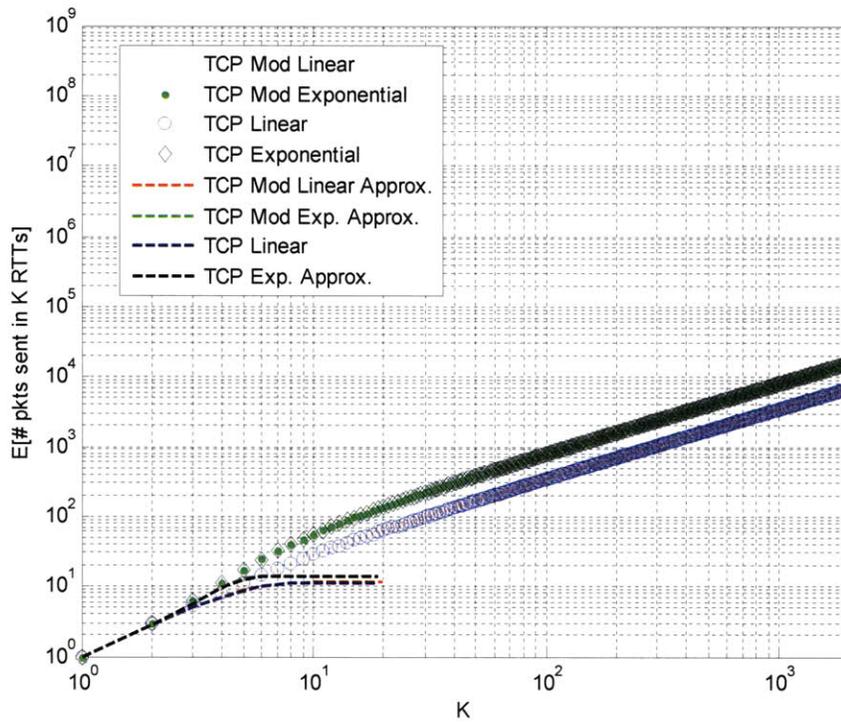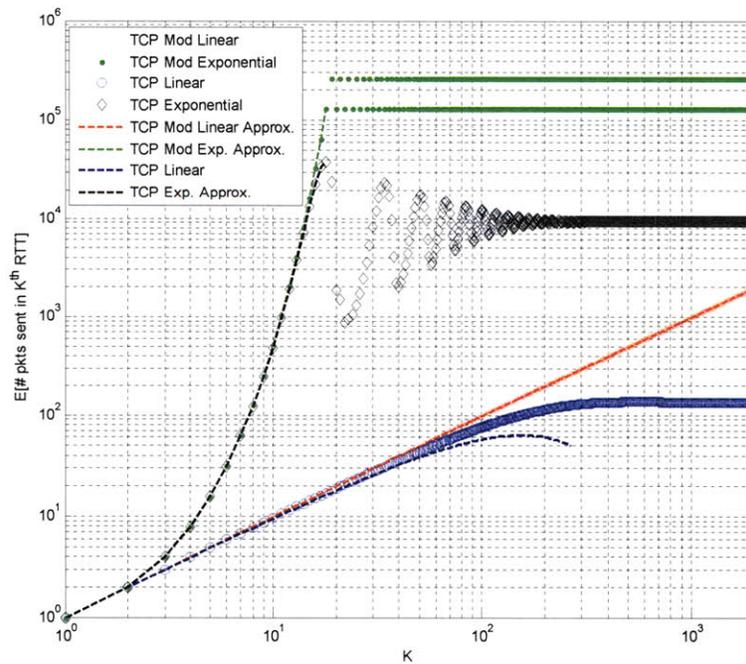Figure 6.98: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet being modeled as a step function and $\sigma_x^2=0.01$ and $M=2^{18}$ (RTT=0.3 sec)
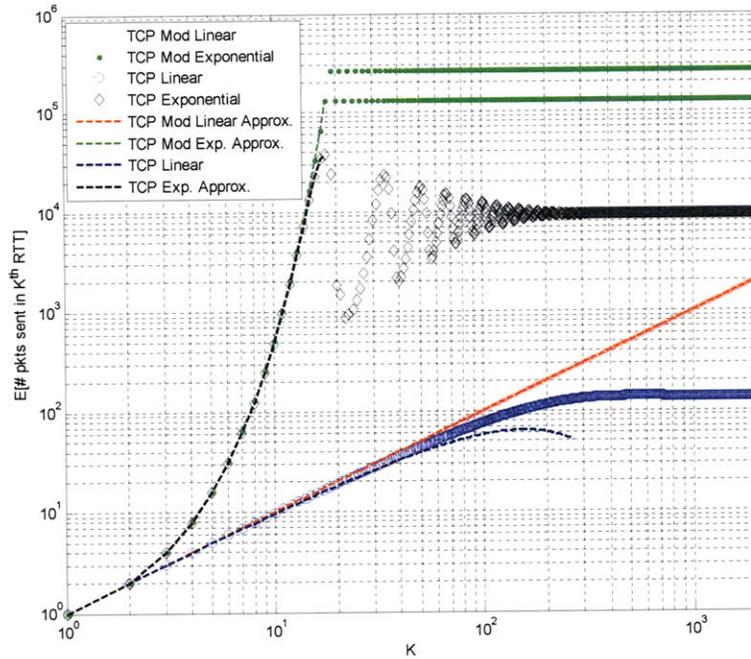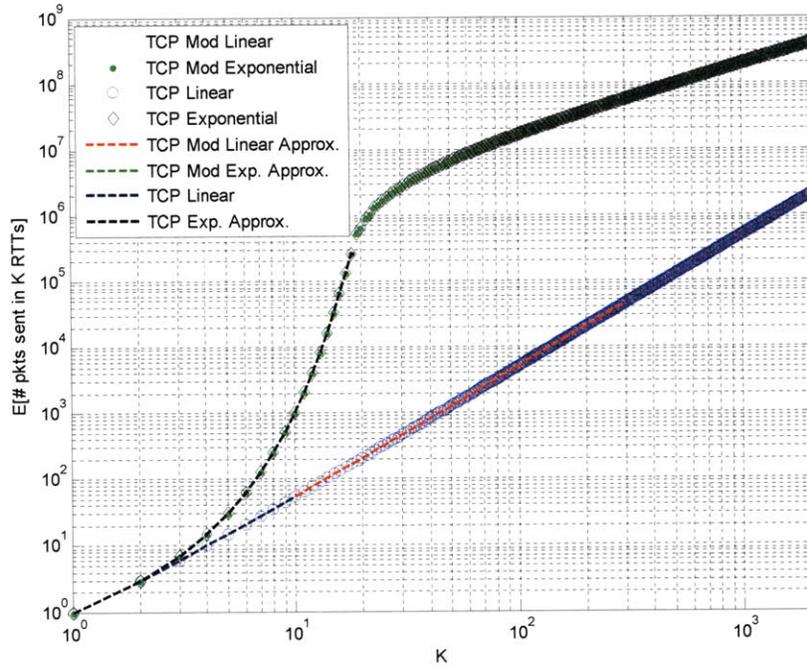
Figure 6.99: Expected number of packets sent in K round-trip intervals for congestion loss per packet being modeled as a step function and $\sigma_X^2 = 0.01$ and $M = 2^{18}$ (RTT=0.3 sec)
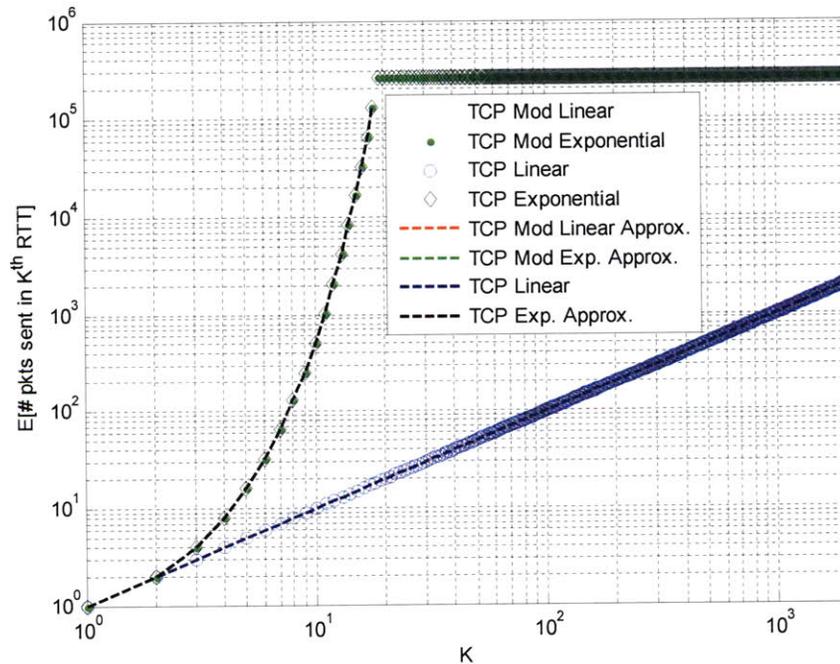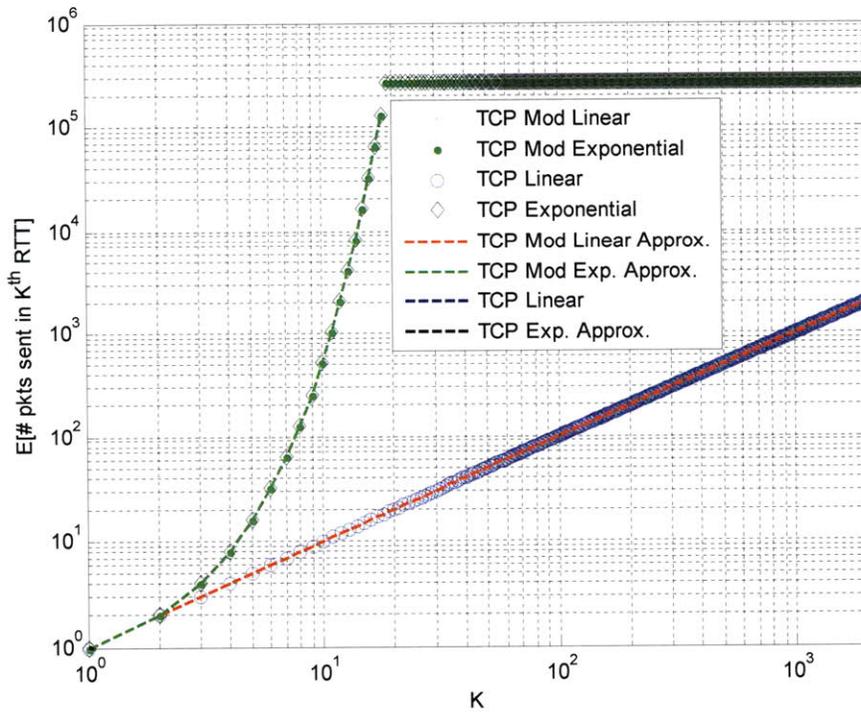

Figure 6.100: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 0 and $\sigma_X^2 = 0$ and $M = 2^{18}$ (RTT=0.3 sec)
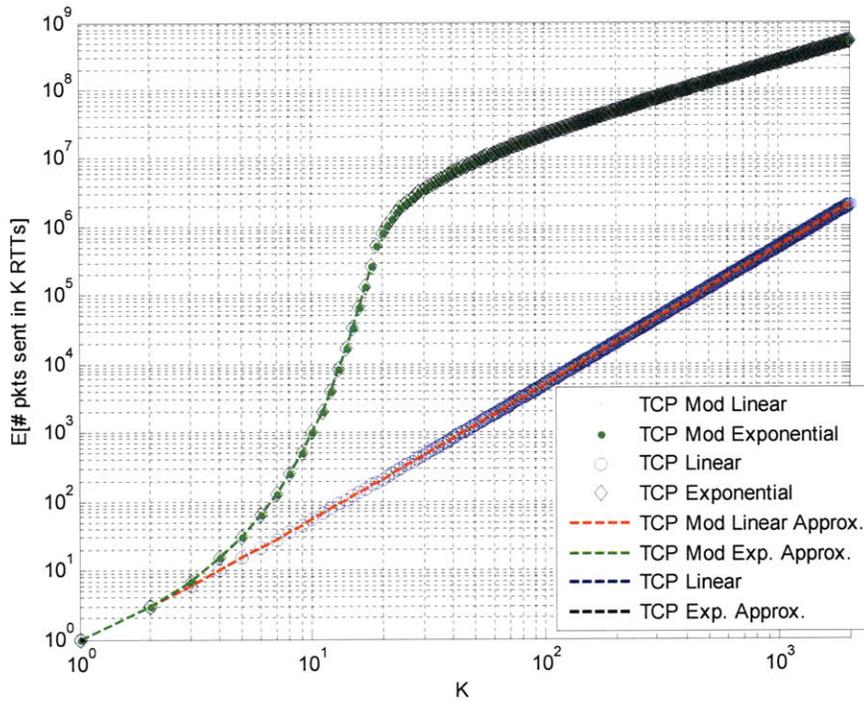
Figure 6.101: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 0 and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
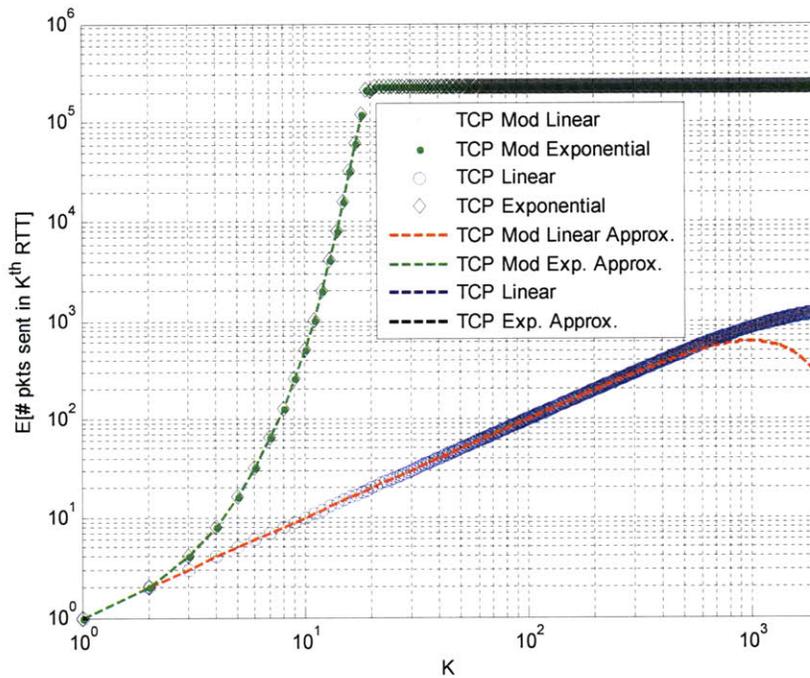
Figure 6.102: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-6} and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
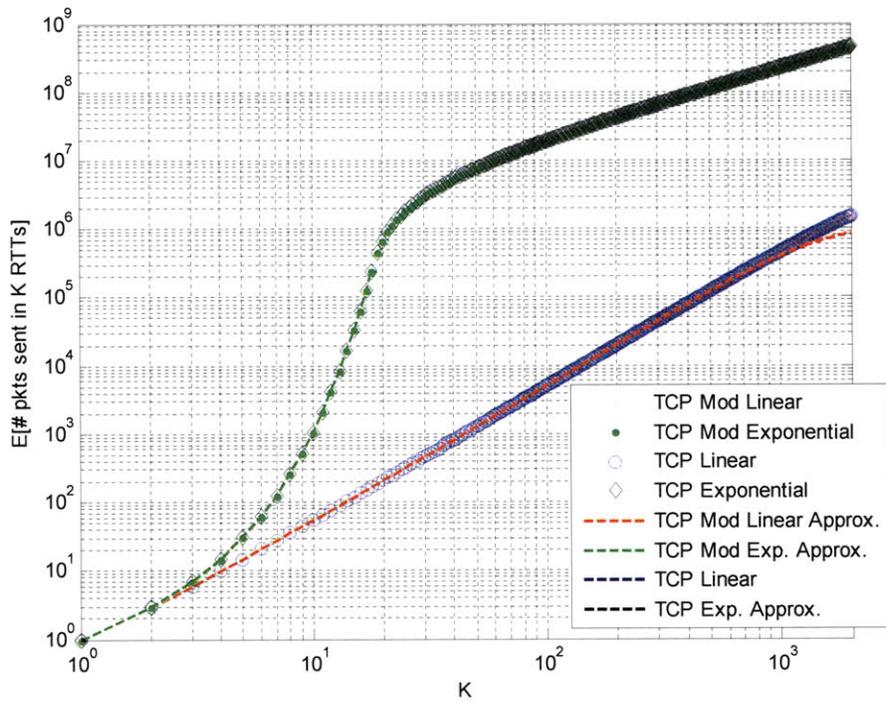
Figure 6.103: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-6} and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
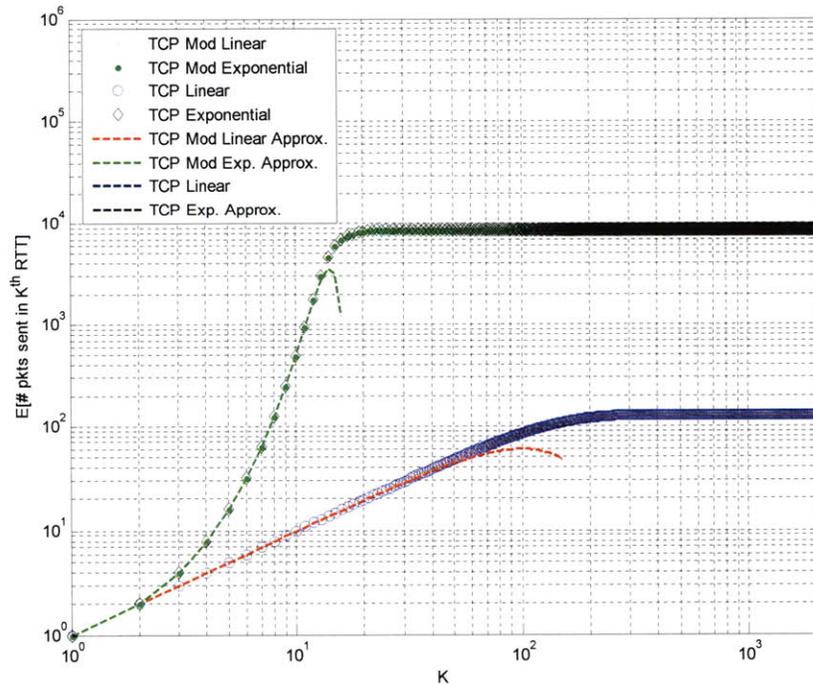
Figure 6.104: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-4} and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
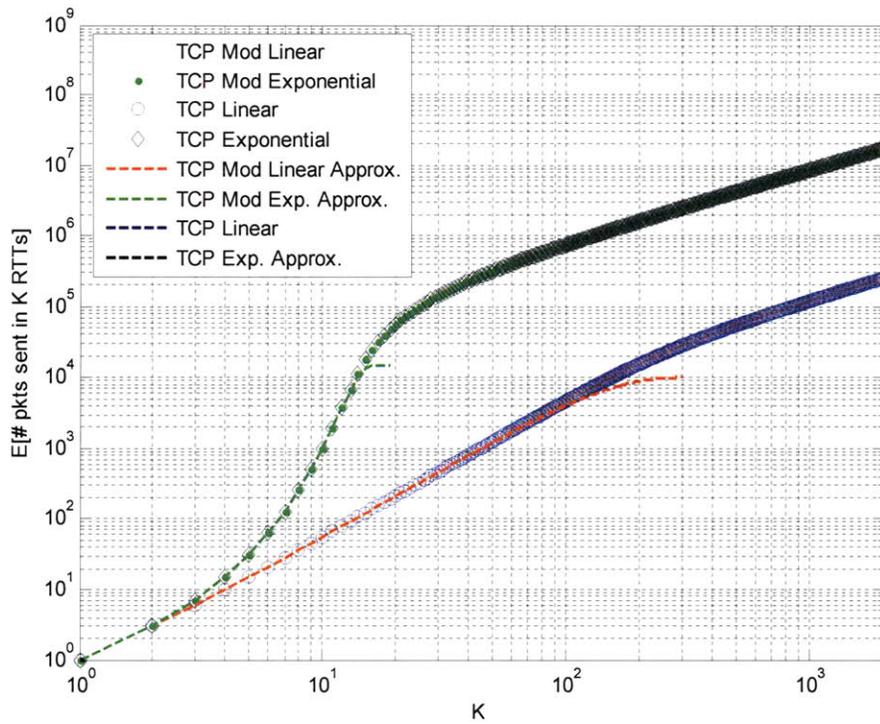
Figure 6.105: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-4} and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
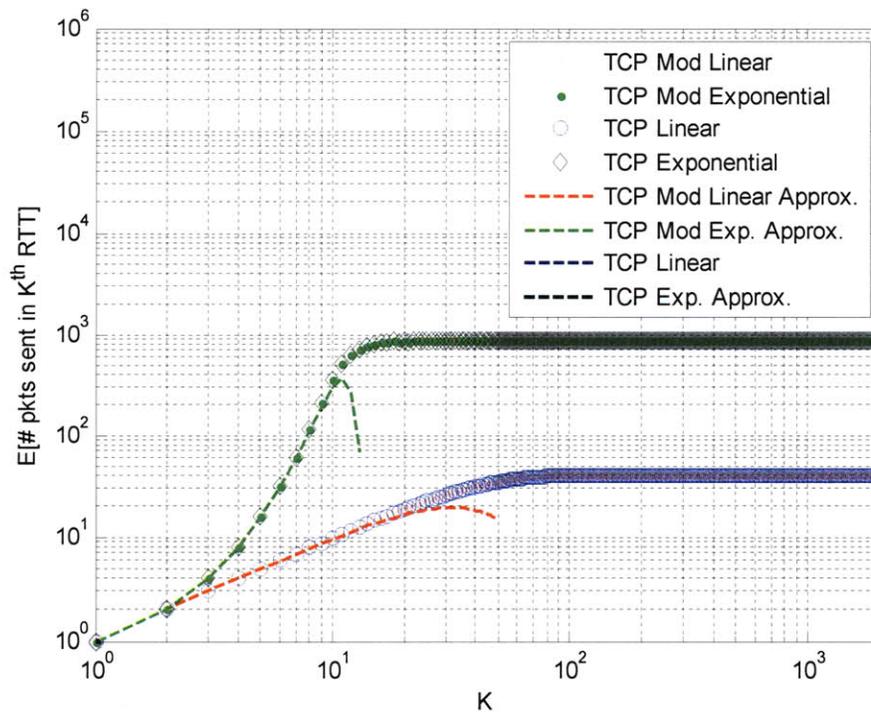
Figure 6.106: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-3} and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
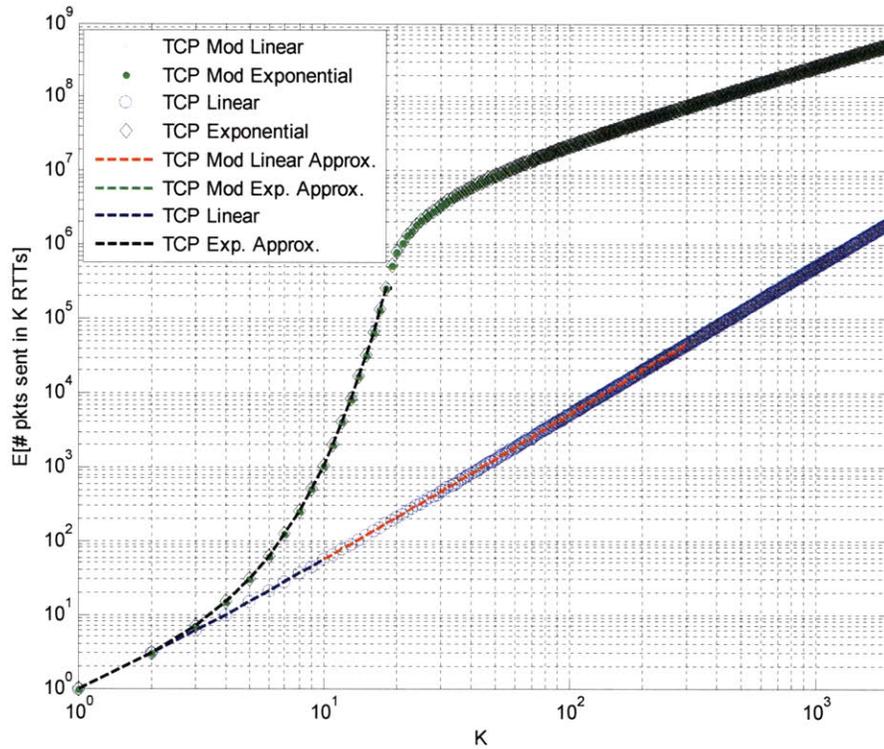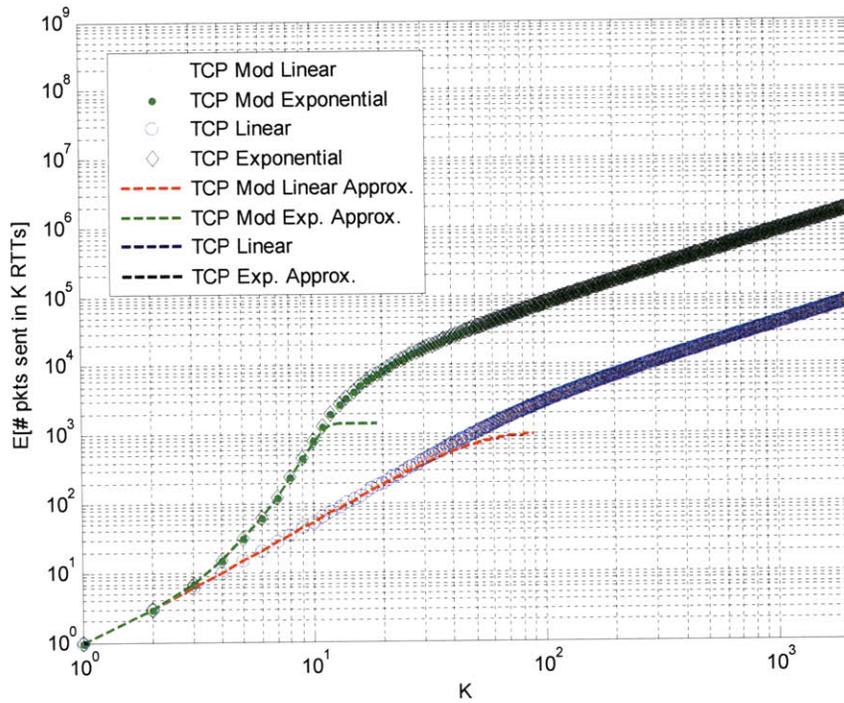
Figure 6.107: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-3} and $\sigma_x^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
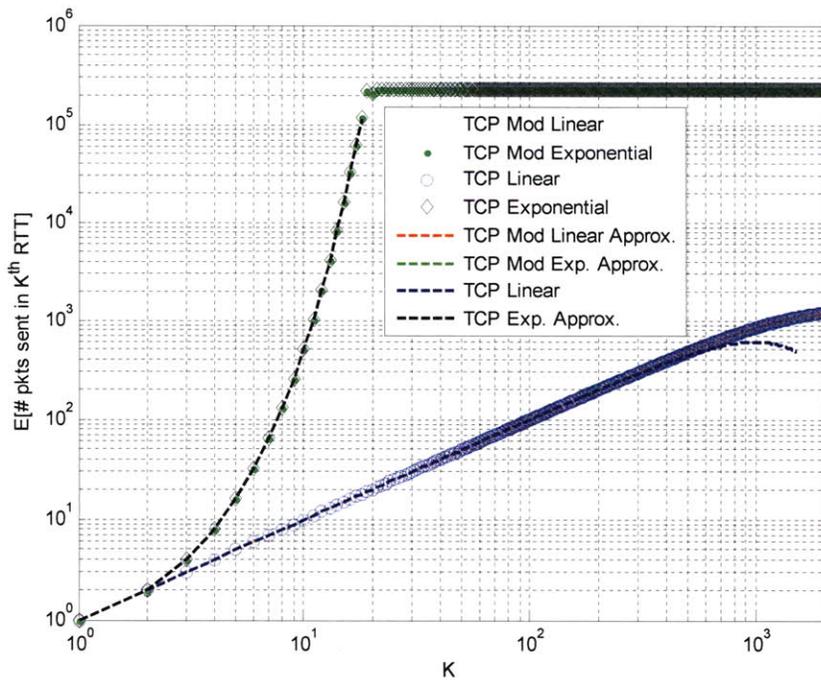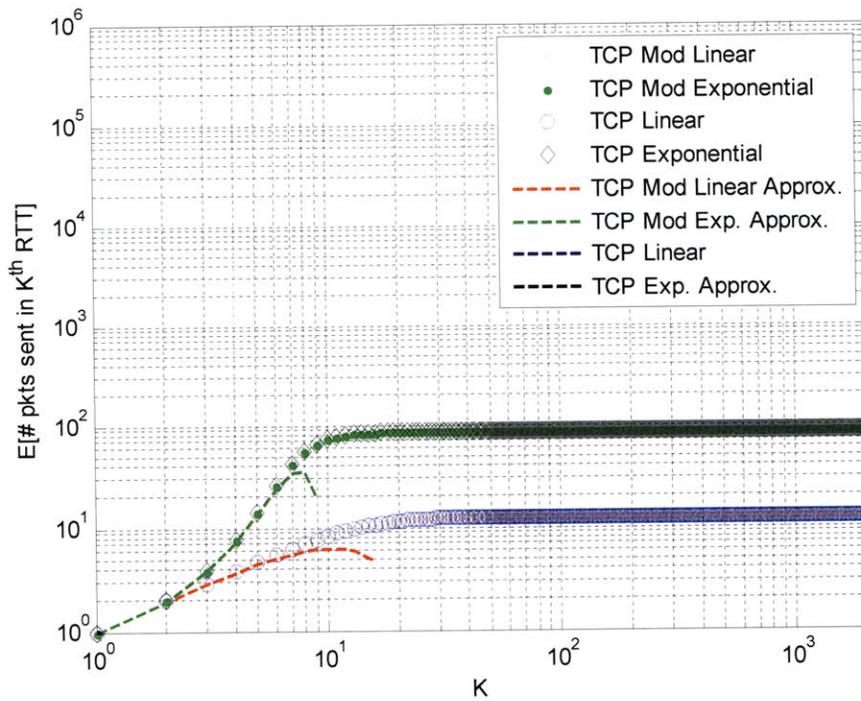

Figure 6.108: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-2} and $\sigma_x^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
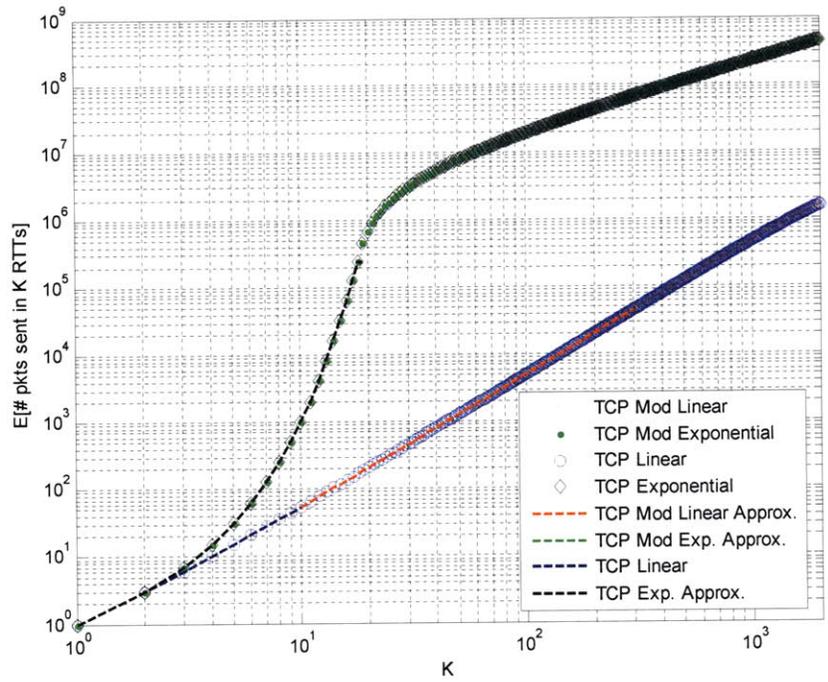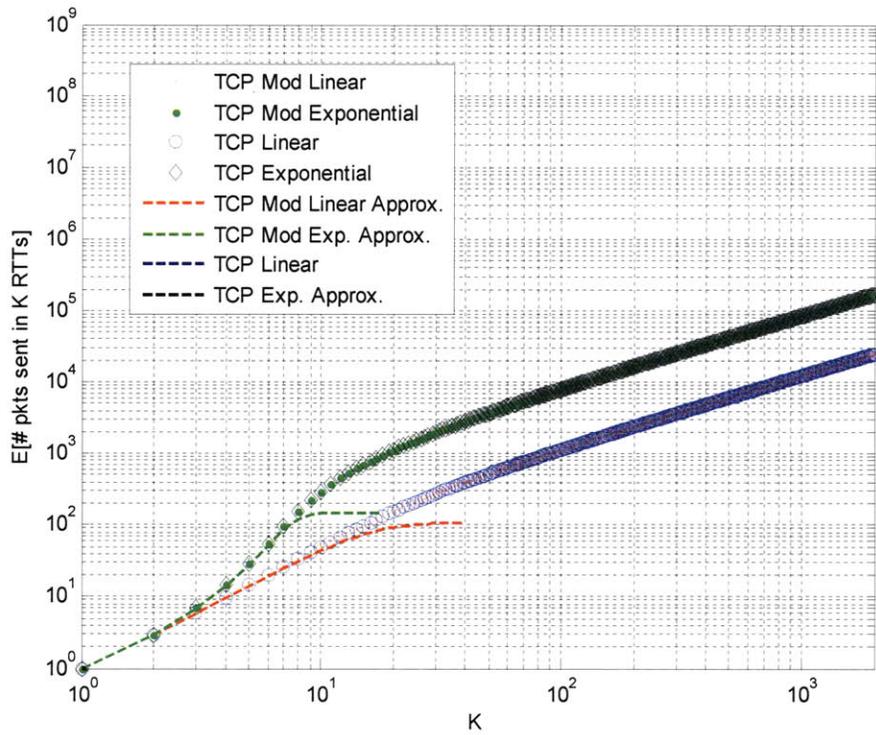
Figure 6.109: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-2} and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
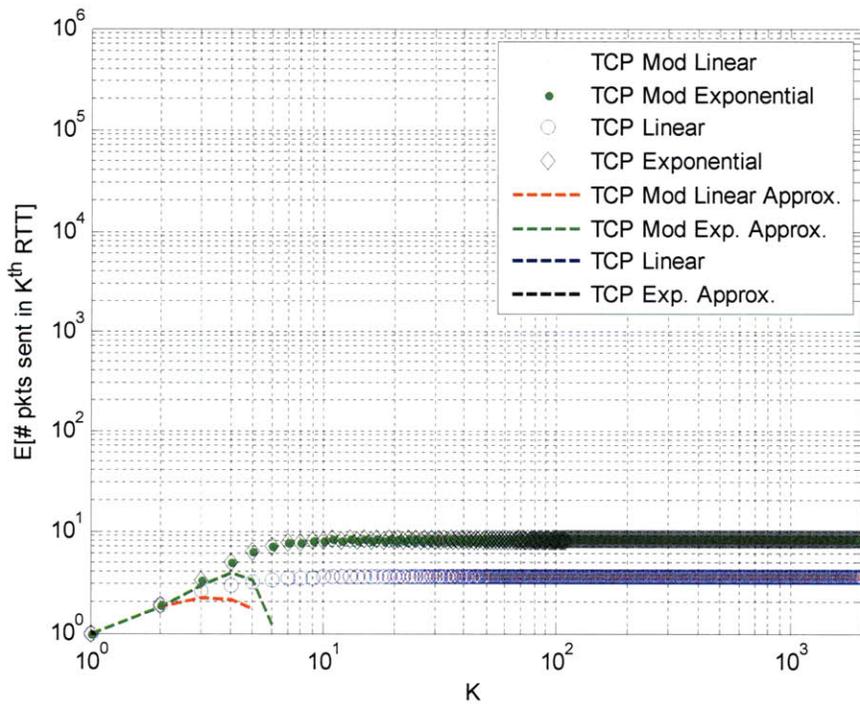
Figure 6.110 (a) on same scale as other E[# packets sent in K^{th} RTT] plots
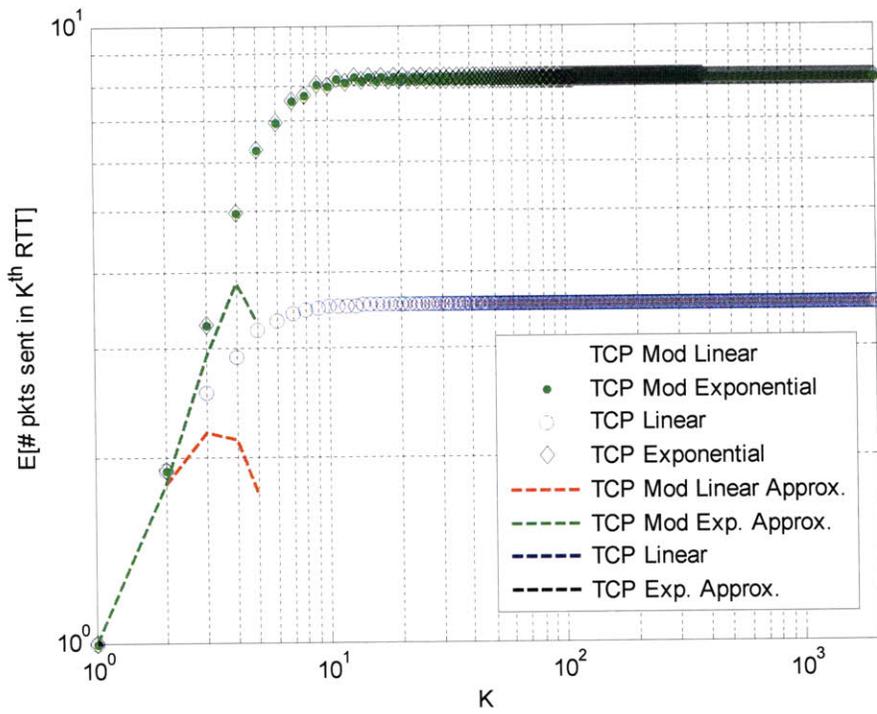
Figure 6.110 (b) zoomed in on vertical axis for better resolution

Figure 6.110: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet of 10^{-1} and $\sigma_x^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
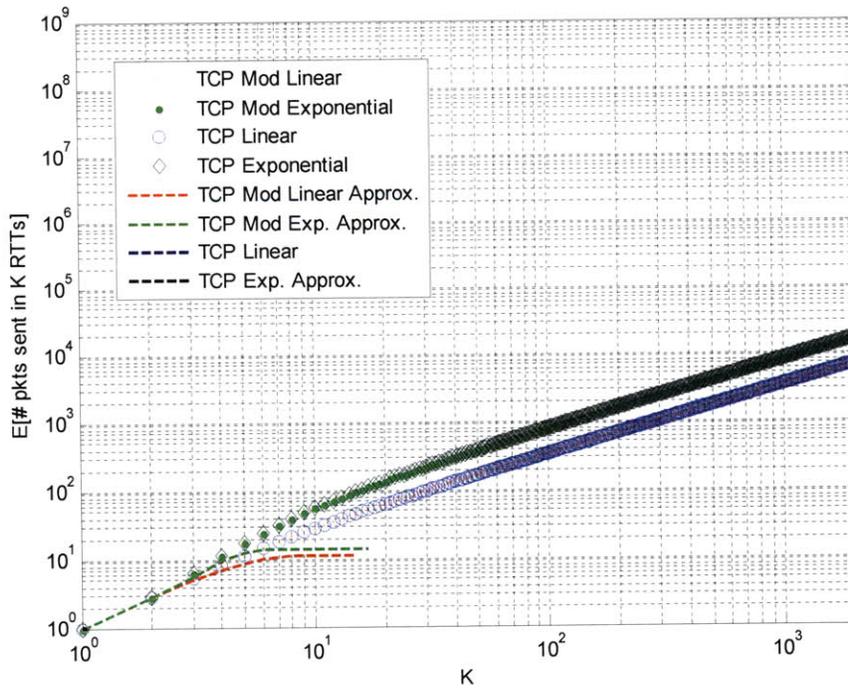
Figure 6.111: Expected number of packets sent in K round-trip intervals for congestion loss per packet of 10^{-1} and $\sigma_x^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
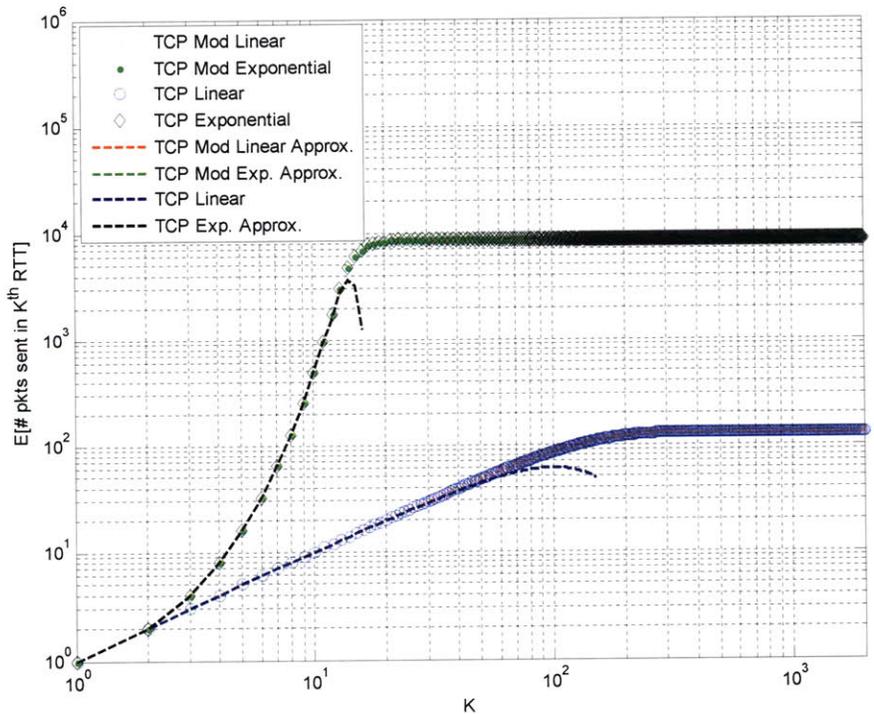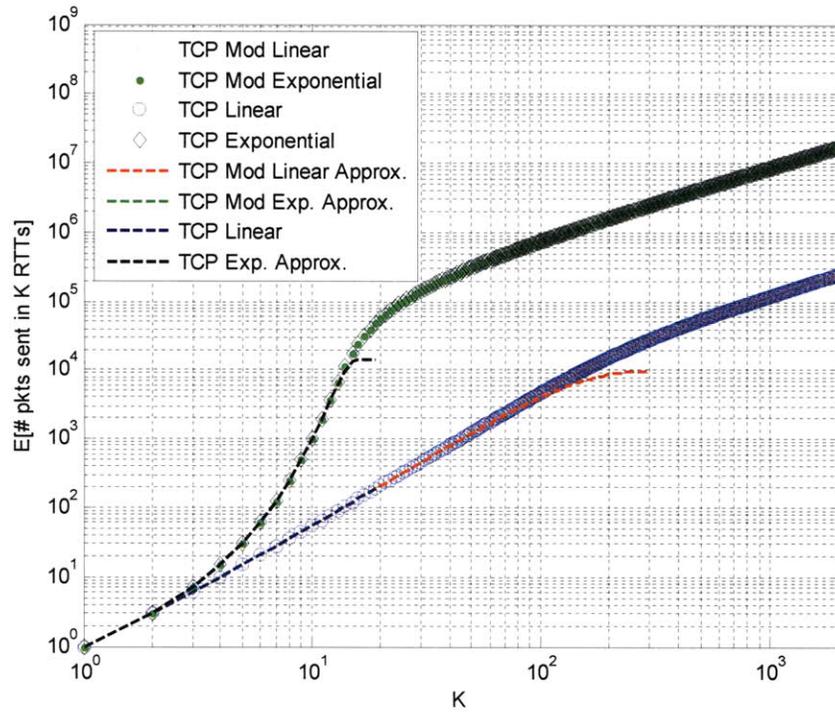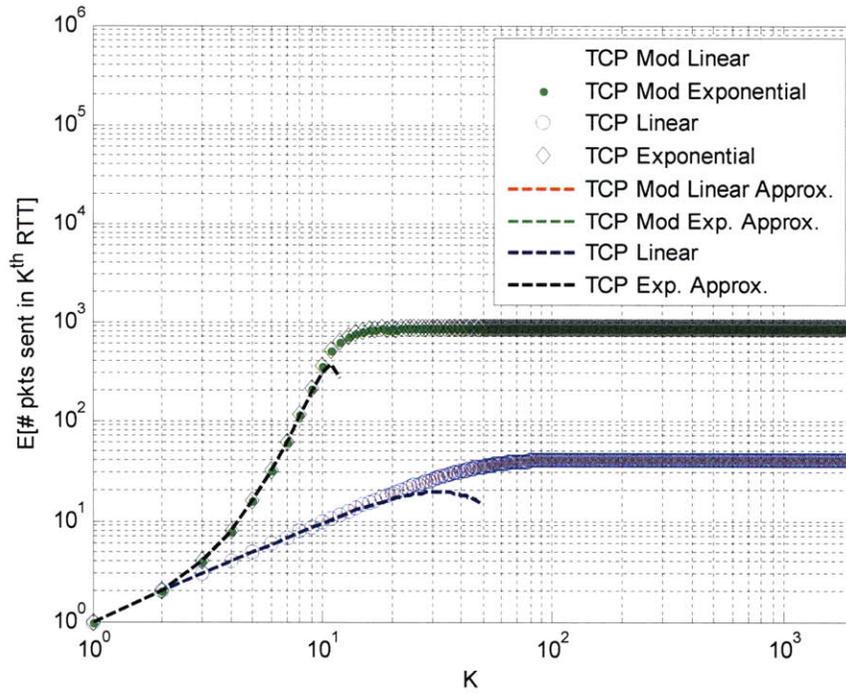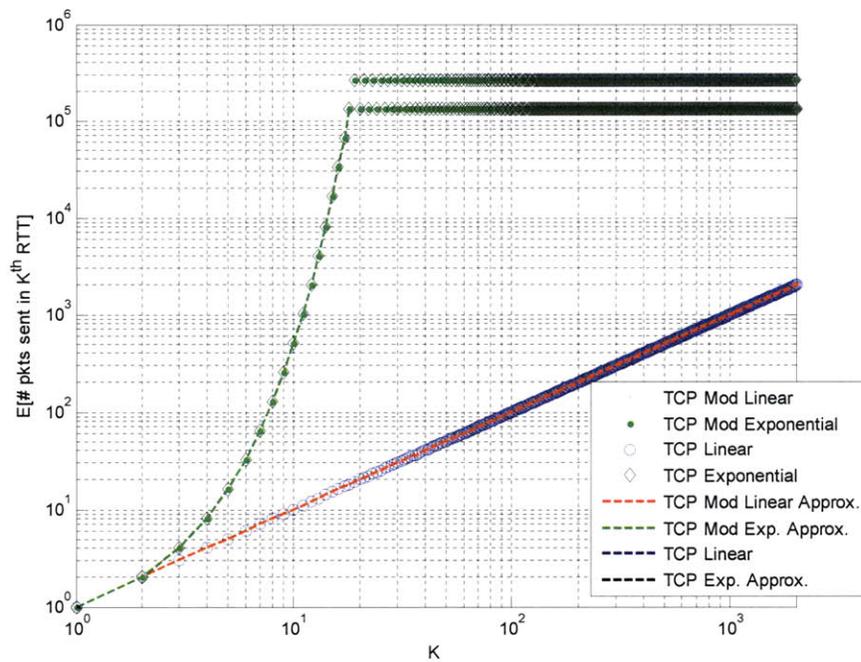
Figure 6.112: Expected number of packets sent in the Kth round-trip interval for congestion loss per packet being modeled as a step function and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)
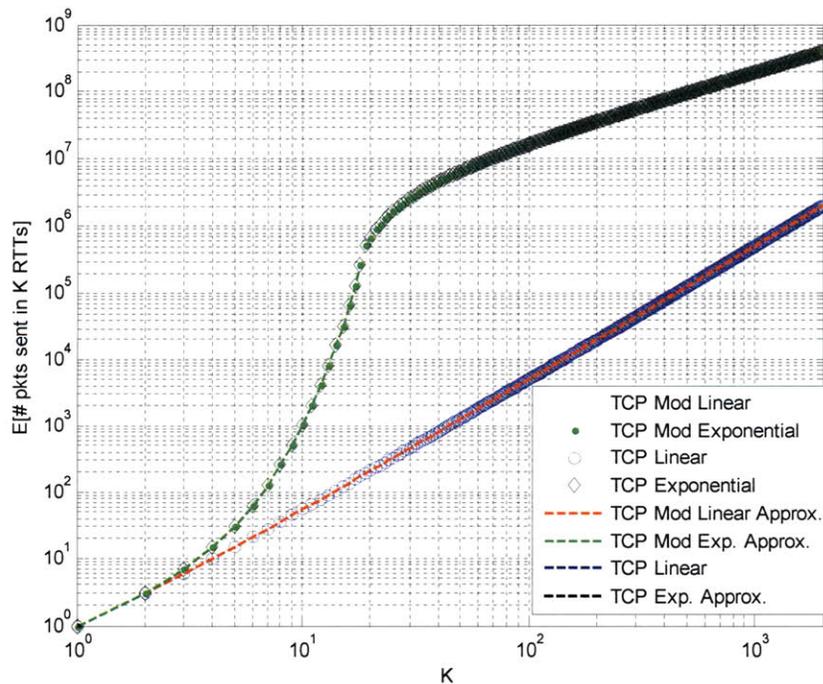
Figure 6.113: Expected number of packets sent in K round-trip intervals for congestion loss per packet being modeled as a step function and $\sigma_X^2=0$ and $M=2^{18}$ (RTT=0.3 sec)

For long round-trip distances, when packet losses due to congestion occur with independent probability $p_{congperpkt}$, the value at which the expected number of packets sent in the Kth RTT levels off is approximately

$$E\left[\text{number of packets sent in K}^{th}\text{ RTT}\right] \rightarrow \min\left(M, \frac{0.75}{p_{congperpkt}}\right)$$

(6.37)

for Modified TCP exponential increase, and approximately

$$E\left[\text{number of packets sent in K}^{th}\text{ RTT}\right] \rightarrow \min\left(M, \sqrt{\frac{2}{p_{congperpkt}}}\right)$$

(6.38)

for Modified TCP linear window increase. M is the maximum possible number of packets in flight for the given RTT. The reasoning of why the expected number of packets sent in the Kth RTT levels off at the above values follows:

The Modified TCP sender only reduces its window upon a packet loss due to congestion. If packet losses due to congestion occur with independent probability $p_{congperpkt}$, then on average, the number of packets that are sent before a congestion loss occurs is $p_{congperpkt}^{-1}$. For example, for $p_{congperpkt} = 10^{-4}$, $p_{congperpkt}^{-1} = 10,000$ packets are sent on average before the congestion loss is experienced. For exponential window increase, when we start at a window size of 1, the window increases to $p_{congperpkt}^{-1}$ by the time $p_{congperpkt}^{-1}$ packets are sent. On average, this window is cut in half, and on average, no loss occurs with a window size of $\frac{1}{2}p_{congperpkt}^{-1}$ and the window increases

234

again to $p_{congperpkt}^{-1}$ in the next RTT. At the new window size, the window is cut in half again and so forth. Thus, the expected window size is thus roughly $0.75 p_{congperpkt}^{-1}$. If the maximum packets in flight M is less than $0.75 p_{congperpkt}^{-1}$, then the expected number of packets in the Kth RTT would not be larger than M. For linear window increase, when starting at a window size of 1, it takes roughly n= $\sqrt{\dfrac{2}{p_{congperpkt}}}$ round-trip times (actually a little less) to send $p_{congperpkt}^{-1}$ packets (since $1+2+3+...+n = \dfrac{n(n-1)}{2}$ for any integer n, n>1). So on average, when the window size reaches n, the window gets cut in half. The window will on average, hover around n. If the window size is less than or equal to n after the window cut, then it will increase on average, to a value between n and $(\sqrt{2}-1)n$ before getting cut again. (See Appendix I for the calculation that the window reaches size $(\sqrt{2}-1)n$ before it is cut if the window starts at size n and increases linearly.) Thus, it makes sense that the average window size is roughly n= $\sqrt{\dfrac{2}{p_{congperpkt}}}$. Again, if the maximum packets in flight M is less than $\sqrt{\dfrac{2}{p_{congperpkt}}}$, then the expected number of packets in the Kth RTT would not be larger than M.

For large M (large RTT), we can see from (6.37) and (6.38) that for a packet congestion loss probability of less than or equal to 10^{-2}, there is at least an order of magnitude difference between the steady state level of the expected number of packets sent in the Kth RTT for exponential increase and that for linear increase. This is confirmed in Figures 6.44-6.113. For example, in Figure 6.62 which corresponds to a packet congestion loss probability of 10^{-4}, the curve for Modified TCP with exponential window increase levels off at roughly 10^4 packets and the curve for Modified TCP with linear window increase levels off at roughly 10^2 packets. Thus, in

this example, there are two orders of magnitude difference between the values at which the two curves level off.

In the plots of expected number of packets sent in the Kth RTT, when the congestion loss probability is low or zero, the TCP exponential curve oscillates prior to reaching steady state. See Figure 6.74 for example. We now explain the reason for the oscillation. For low congestion loss probabilities, the window is rarely reduced due to congestion loss. Outage losses are the dominant cause of window reduction. At the start of the session, as the window starts to build up, the probability of packet loss is low due to the small window size and large round-trip time. As time progresses from time 0, only a small portion of the probability distribution moves down to state 1 and starts to increase again. The remainder and majority of the probability distribution follows the exponential window increase. However, as the majority of the distribution increases to larger window values at the exponential rate, it suddenly becomes very probable (due to packets occupying a large fraction of the RTT) that an outage causes packet loss and the window reduces to one packet. This sudden increase in outage loss probability is due to the window size doubling every RTT. Thus, there is a sudden decrease in the average number of packets sent. The oscillation in the figure does not go down to one because the expected number of packets is averaged with some of the probability that had previously moved down to state 1 and had subsequently increased to higher window sizes. The large amount of probability weight that went to state 1 increases to higher window values, and then the same window cutting happens again. This is why we see the oscillation. Each time the majority portion of the probability increases to larger window sizes, it gets smaller because some of it moves to state 1. Eventually the probability distribution settles on the steady state distribution. However, the individual sample functions of the state of the protocol still oscillate. The TCP linear increase curves do not exhibit

236

this oscillation because as the window increases linearly, the probability of outage loss increases gradually, and the probability distribution across the states shifts up more evenly.

The time at which the expected number of packets sent in the Kth RTT levels off signifies that the system has reached steady state. After this time, the steady state throughput, as given in Section 6.3, applies. In the absence of congestion loss and atmospheric turbulence, the exponential window increase curves take $1+\log_2 M$ round-trip times to reach steady state and the linear window increase curves take M round-trip times. If congestion loss probability is added, it takes all of the curves less time to reach steady state because the congestion losses cause the majority of the probability distribution to occupy a lower number of states. If atmospheric turbulence is added, it again takes the linear increase chain less time to reach steady state because the probability distribution occupies a lower number of states. However, the exponential curves take longer to reach steady state with than without atmospheric turbulence because the oscillation that we discussed in the previous paragraph takes time to settle down.

As we can see from Figures 6.44-6.113, in the transient time frame prior to steady state, when communicating through a large amount of atmospheric turbulence, it takes Modified TCP less time to transmit a moderate sized file than it takes TCP. The reason is the same as for steady state: the Modified TCP sender distinguishes outage loss from congestion loss and does not reduce its window upon outage losses. The benefit of Modified TCP is greatest when the turbulence level is strong and the file size is moderate to large.

We depict in Figure 6.114 the regions of atmospheric turbulence and file sizes in which it may be worthwhile to use Modified TCP rather than TCP (for the parameters given in the figure caption). These regions are determined by inspection of Figures 6.44-6.113. The areas labeled "TCP" indicate regions where Modified TCP offers little benefit over TCP. The threshold of when Modified TCP offers significant benefit over TCP is taken to be when Modified TCP that increases its window linearly takes half the time to send the file that TCP that increases its window linearly takes. At a packet congestion loss probability of 10^{-6} (which corresponds to Figure 6.114a), when $\sigma_\chi = 0.1$, it takes TCP and Modified TCP roughly the same time to transmit a file. This is why, in Figure 6.114a, we extrapolated the boundary between the TCP and Modified TCP regions to level off above $\sigma_\chi = 0.1$. Similarly, at a packet congestion loss probability of 10^{-4} and 10^{-2}, when $\sigma_\chi = \sqrt{0.1} = 0.3$, it takes TCP and Modified TCP roughly the same time to transmit a file. Thus, in Figures 6.114b and 6.114c, we extrapolated the boundary between the TCP and Modified TCP regions to level off above $\sigma_\chi = \sqrt{0.1} = 0.3$. Modified TCP offers the most benefit when the turbulence is large and file sizes are moderate to large, and these regions are labeled "Modified TCP". However, since the round-trip time is large (0.3 seconds), the rate of window increase is not very high for TCP or Modified TCP; all regions in Figure 6.114 could benefit from another congestion control protocol that allows for faster window increase.

238
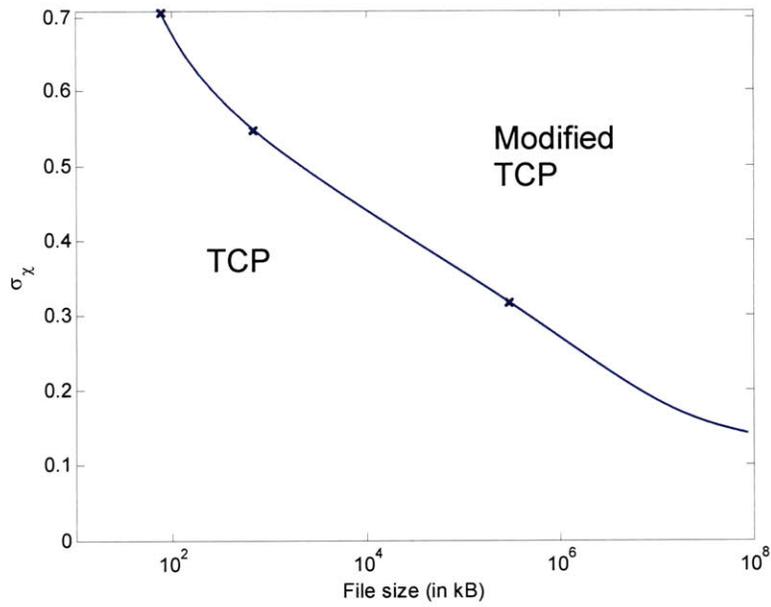
Figure 6.114 (a) $p_{congperpkt}=10^{-6}$
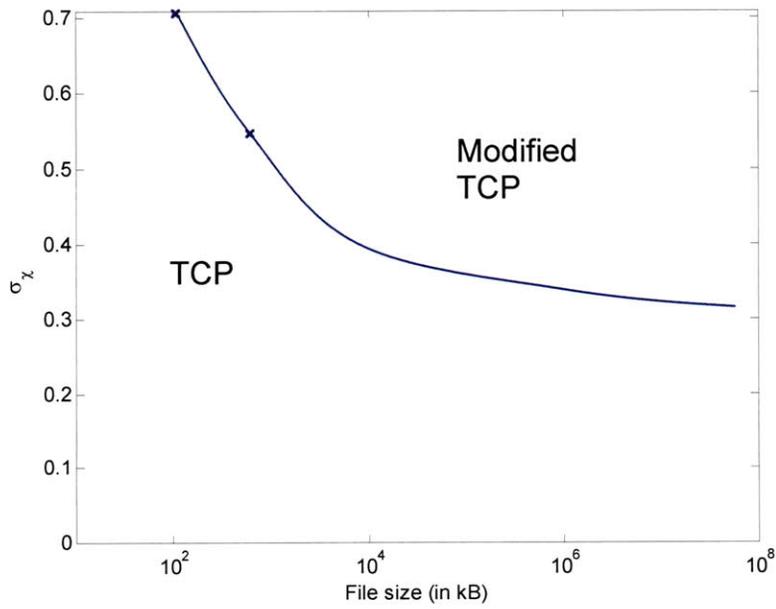
Figure 6.114 (b) $p_{congperpkt}=10^{-4}$
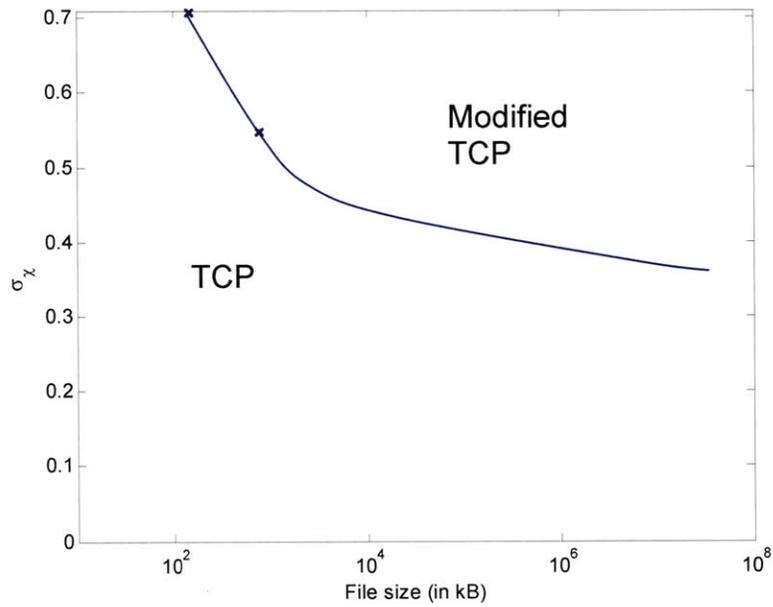
Figure 6.114 (c) $p_{congperpkt}=10^{-2}$

Figure 6.114: Diagram of regions where it is worthwhile to use Modified TCP over TCP when probability of congestion loss per packet is (a) 10^{-6} (b) 10^{-4} and (c) 10^{-2} where RTT=0.3 sec., R_{max}=10Gb/s, G=1.5kbytes, m=8dB, P_e^{thresh}=0.1, v_t=10km/hr, N_n=1, outages cause timeouts.

Chapter 7

Modified TCP Deployment, 'Congestion Loss' Feedback Generation and Processing, and Drawbacks

In the previous chapter, we analyzed the throughput that can be achieved by a class of TCP-based protocols whose senders distinguish whether a packet loss is due to an FSO link outage or due to congestion and react appropriately. We discussed an implementation in which routers provide 'Congestion Loss' feedback packets to the senders to help them distinguish outage loss from congestion loss. In this chapter, we discuss at a high level a possible direction that can be taken to deploy this implementation with 'Congestion Loss' feedback packets but without modifying all existing routers. We also discuss the impact on the network of generating 'Congestion Loss' feedback packets, and of the senders releasing extra packets into the network when the 'Congestion Loss' feedback packets are lost.

7.1 Deployment of Modified TCP in Subset of Routers

One of the concerns about a new Transport Layer protocol, or even a slightly modified Transport Layer protocol, is that it is expensive to change or modify all of the existing routers and senders to incorporate the new protocol. The Modified TCP implementation we considered in Chapter 6 requires routers and senders to be modified: routers have to have the functionality of providing 'Congestion Loss' feedback when they drop packets; senders have to have a different window update algorithm and possibly get feedback from the network about the mean and standard deviation of outage lengths[18].

In this section, we introduce at a high level, a possible direction that can be taken to deploy a Modified TCP protocol in a heterogeneous network without replacing all routers. The network for which this deployment is useful is one in which there are only a few FSO links and the remainder of the links are highly reliable. Instead of adding the functionality of generating 'Congestion Loss' feedback packets to all of the routers, entities which we call special gateways can be deployed around the reliable sub-networks. The special gateways act as routers but with the added functionality of creating 'Congestion Loss' feedback packets when they drop packets, as well as some other functionality and complexity, which we will discuss. There are many challenges associated with this deployment. We identify a few of the key ones as we discuss the deployment details. However, the intricacies of the deployment are really a subject for future work.

Consider the network shown in Figure 7.1 which consists of two sub-networks, A and B, with highly reliable links (such as fiber links) connected to each other via a

---

[18] As we discussed in Section 6.2, the senders, instead of getting feedback about the outage length statistics to set the Backup-Timer value, can set the Backup-Timer to a value that is larger than most outage lengths at all turbulence conditions. However, this leads to an unnecessarily large Backup-Timer duration.

satellite sub-network C. The up and down links to and from the satellite network suffer outages due to atmospheric turbulence. Within the satellite sub-network itself, there are highly reliable links due to communicating through space. Congestion can occur anywhere in the network i.e. in the ground sub-networks or in the satellite network. The source and destination pairs reside outside of the network described and traverse through the network to get from source to destination.
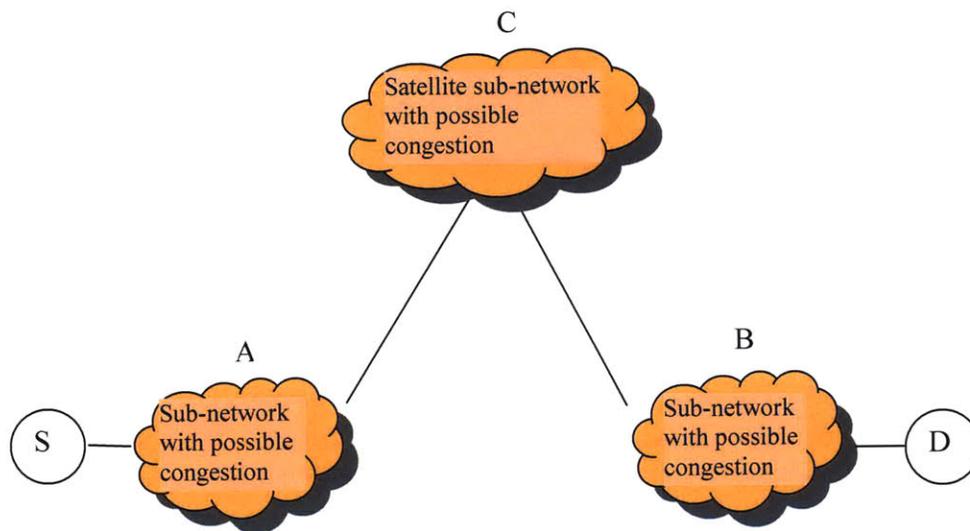
Figure 7.1: Diagram of a network that consists of two sub-networks (A and B) with highly reliable links connected via a satellite sub-network (C). The source and destination nodes are denoted by S and D respectively.

Special gateways can be placed around the edges of the satellite sub-network and the two fiber sub-networks as shown in Figure 7.2. The special gateways are denoted by M1-M6. The job of the special gateways, in addition to routing, is to discern when a congestion loss occurs within any of the sub-networks (or at itself) and to feed back a 'Congestion Loss' message to the sender of the packet. This feedback helps the sender to distinguish whether a packet loss is due to a link outage or congestion. If a congestion loss occurs at a special gateway rather than within a sub-network, the gateway can easily determine whose packet it is dropping by looking at the header of

the packet before dropping it. It would then feed back a 'Congestion Loss' packet to the sender.
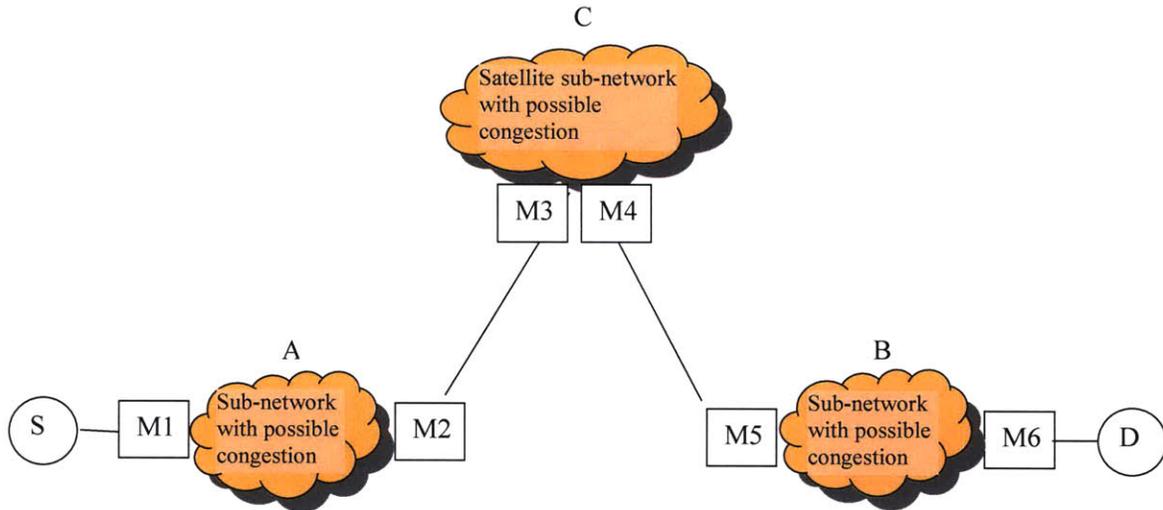
Figure 7.2: Diagram of a network that consists of two sub-networks (A and B) with highly reliable links connected via a satellite sub-network (C) and where special gateways (M1-M6) are placed around the sub-networks. The special gateways provide 'Congestion Loss' feedback to the sender if they detect a congestion loss in a sub-network. The source and destination nodes are denoted by S and D respectively.

The special gateways at the entrance and exit of a sub-network work together to determine when a congestion loss occurs in the sub-network by doing the following: the special gateway at the entrance to the sub-network encapsulates the incoming packet, adds a sequence number (SN), and tunnels the packet to the special gateway at the exit of the sub-network i.e. packets are tunneled between M1-M2, between M3-M4, and between M5-M6 in Figure 7.2. Using one set of SNs between the special gateway at the entrance and exit of the sub-network is simpler than using a separate set of SNs for each TCP flow because using a separate set for each TCP flow would require the special gateways to keep track of each flow's last SN. If a sequence number is missing in the received tunneled packets by a special gateway (by M2, M4

244

or M6 in Figure 7.2), then the receiving special gateway can assume the missing packet is due to congestion and feedback a 'Congestion Loss' message to the sending gateway (M1, M3 or M5 in the figure) with the missing SN. The sending gateway would forward this 'Congestion Loss' message to the source (the sender of the packet). In order for the gateway to determine which source to which it should forward the 'Congestion Loss' message, (i.e. which source's packet was dropped in the sub-network), the sending gateway stores the source addresses and SNs of all packets that it sends. By doing so, if the sending special gateway receives a 'Congestion Loss' feedback message from the receiving special gateway with the SN of the packet that was dropped in the sub-network, it can look up the sender of the missing packet and forward the 'Congestion Loss' feedback back to the sender. The sending special gateway has to store the SNs and addresses of the packets that it sends to the receiving special gateway until the packets have cleared the sub-network. If ACKs are implemented between the special gateways, the sending gateway can discard an SN and address upon receiving the corresponding ACK. If ACKs are not used between the special gateways, the sending gateway can discard the SN and address of the earliest packet it sent when its memory gets full. In either case, the sending gateway memory size should be large enough that it can hold the SNs and addresses of packets that it released into the sub-network and for which a 'Congestion Loss' message may return to the sending gateway. If a packet already has the maximum size allowed by the sub-network before being encapsulated, the special gateway would need to be fragment the packet before encapsulating it. The reassembly could be done either by the receiving gateway or the destination node.

This deployment scheme has added complexity compared to the simple modification that would be done to all routers in order to provide 'Congestion Loss' feedback. However, this complexity is in the special gateways rather than in all routers. This

method of determining the sender of the packet dropped in the sub-network due to congestion is similar to the method that would be needed to provide 'Outage Loss' feedback to senders instead of 'Congestion Loss' feedback. However, if 'Outage Loss' feedback is provided instead of 'Congestion Loss' feedback, as we discussed in Chapter 6, the feedback may not arrive at the sender early enough to avoid a timeout.

If a packet is lost within one of the sub-networks due to link loss rather than congestion, the gateway at the exit of the sub-network assumes the loss is due to congestion, and feeds back a 'Congestion Loss' message to the source. This would cause the TCP sender to unnecessarily reduce its window (just as in regular TCP) thereby reducing its average throughput. However, for links in the sub-network with very low packet loss rate, such as fiber or space, this unnecessary window reduction is rare and the average TCP sender throughput reduction is small.

In Figure 7.2, for simplicity, we showed only one source and destination and two special gateways surrounding each reliable sub-network. In actuality, there can be many senders whose packets are aggregated and enter the sub-networks at different physical locations. In order to provide 'Congestion Loss' feedback to these senders when their packets are lost due to congestion in one of the sub-networks, there should be multiple special gateways placed around the sub-networks, one at each point where senders' packets can enter the sub-networks. There needs to be a separate SN between the gateways that surround each sub-network. In addition, there can be "other" ground sub-networks that are also connected to any of the sub-networks shown in Figure 7.2. To provide senders with the benefit of 'Congestion Loss' feedback if a congestion loss occurs within one of the "other" sub-networks, special gateways should be placed around the "other" sub-networks. These special

gateways would work together to determine if a congestion loss occurs in the "other" sub-network and provide 'Congestion Loss' feedback to the sender.

Let us briefly discuss an incremental deployment of the special gateways. In particular, consider the effect of the "other" sub-networks not having special gateways placed around them. The senders whose packets enter the network of Figure 7.2 (which has special gateways) from the "other" sub-networks can still benefit from the 'Congestion Loss' feedback that the network in Figure 7.2 provides. However, there may be negative effects on network congestion if congestion losses occur in the "other" sub-networks to cause the Modified TCP sender to timeout. This is because the sender, upon timeout, will not cut its window and will release more packets into the congested network until the Backup-Timer expires. The challenges related to an incremental deployment need to be considered carefully and they are a subject that requires further exploration.

Another issue that needs further investigation in this deployment is how to re-route the TCP flows to another less congested gateway if a particular gateway is congested.

7.2 Generation of 'Congestion Loss' Feedback Packets and Loss of 'Congestion Loss' Feedback packets– Impact on Network Congestion

In this section, we re-consider the Modified TCP implementation in which the router (or specialized gateway) generates and feeds back 'Congestion Loss' packets to the sender, and the sender keeps a Backup-Timer. Specifically, we discuss the effect of this implementation on network congestion.

The 'Congestion Loss' feedback messages are separate packets that are generated and fed back to senders. The feedback packets do not need to carry the original payload, but rather, only a header (with a single bit of the header being used to signify 'Congestion Loss'). If the feedback is in-band, one can think of the router congestion drops and creation of feedback messages as dropping some data packets and replacing them with smaller feedback packets. Although replacing a data packet with a 'Congestion Loss' feedback packet reduces the total bytes in the network immediately when the network is congested, the total number of packets (including both data packets and 'Congestion Loss' feedback packets) is not immediately reduced. This puts extra load on routers when they are congested and dropping packets compared to when the feedback packets are not generated. Since routers' processing of incoming packets depends heavily on the number of packets (and a little on the size of the packets), and the number of packets is not reduced upon congestion loss, routers do not achieve congestion relief until the feedback packets get back to the sender. The router will only get relief after at least one RTT to the sender. Under the most extreme case, window closing (possibly to one packet) may occur for every flow through the congested router. Thus, in order to avoid this situation, the router should aggressively drop packets before the buffer gets as full as it would if the 'Congestion Loss' feedback message were not generated. For example, if Random Early Discard (RED) [13] is used, it should be tuned so that drops occur at an earlier buffer size threshold and with higher probability. Alternatively, in order to have the functionality of providing 'Congestion Loss' feedback but also immediately reduce the number of packets in addition to the number of bytes when congestion losses occur, the router (or specialized gateway) can create a 'Congestion Loss' feedback message for every few packets that it drops rather than for every packet that it drops. Further analysis of the effectiveness of this scheme needs to be carried out.

The network congestion is not only affected by the added 'Congestion Loss' feedback packets, but also by extra packets that the sender releases into the network if the 'Congestion Loss' feedback packets continuously get lost. Consider the situation where severe congestion is dropping all of the sender's packets but none of the 'Congestion Loss' feedback packets are received by the sender (due a return link that is down for example). The Backup-Timer will expire and the sender's response time to congestion is longer than in TCP since the Backup-Timer duration is longer than the RTO. During this longer period, the Modified TCP sender will send more packets into the network than a TCP sender. This will affect the length of time for which the congestion persists and may cause additional packets of other sessions to be dropped. However, provided the chance that the sender's packets are all dropped and the return link is down for the entire Backup-Timer duration is small, the Backup-Timer will expire infrequently, even when severe congestion exists[19]. If there is severe congestion, the sender would likely receive at least one of the 'Congestion Loss' feedback packets that are generated and fed back to the sender by the router in the congested region that is closest to the sender.

---

[19] Let us consider in detail on a packet level why the Backup-Timer will not expire often. If the 'Congestion Loss' feedback corresponding to packet number i is lost and the RTO expires, the packet is retransmitted an RTO after the original transmission. If the congestion improves, the sender may receive an ACK corresponding to the retransmitted packet and the corresponding Backup-Timer is turned off and will not expire. If the congestion persists and the retransmitted packet is dropped, there is another chance for the sender to receive the (new) 'Congestion Loss' feedback and respond to the congestion. Moreover, the sender has the chance to receive duplicate ACKs for packet i corresponding to any of packets i+1,i+2,... that also try to get through the network before the retransmission of packet i. If packets i+1, i+2... are dropped in congestion, the sender has the chance to receive the corresponding 'Congestion Loss' feedbacks. It is when no ACKs and no 'Congestion Loss' feedback messages are received for an entire Backup-Timer duration that the backup-timer will expire.

Chapter 8

Future Transport Layer Work

In this chapter, we discuss aspects of the Transport Layer's congestion control for FSO networks that we did not consider in this thesis, but that still need to be addressed.

In this thesis, we addressed the problem of free-space optical links causing sender timeouts that lead to reduced throughput. We derived the performance that can be achieved if a Modified TCP sender is able to correctly distinguish outage loss from congestion loss and not reduce its window in response to an outage loss. Moreover, we compared the performance to that of a TCP sender and found the performance gain of the Modified TCP sender to be significant for paths with strong atmospheric turbulence and large bandwidth-delay product.

In this thesis, we did not address the issue of the TCP sender being slow to grab available rate on high bandwidth-delay product paths. TCP's linear window increase during the Congestion Avoidance phase is not aggressive enough for high bandwidth-delay links. Even exponential increase is not fast enough for high bandwidth-delay

product paths as we showed in throughput plots. It takes the TCP sender a very long time to increase its window to a value high enough to utilize the available rate on the path to the destination. The TCP sender's slow window increase causes its throughput over these high bandwidth-delay product paths to be poor.

As discussed in Chapter 6, this slow rate increase may be improved by congestion control algorithms such as XCP [23,24] or TCP variants which allow for faster window ramp up for longer RTT sessions [6,14,22,25,34]. XCP is a congestion control algorithm in which the routers monitor their queue size and incoming traffic rate and feed back to each sender the amount by which the sender can increase or should decrease its window. XCP requires a new header format which includes the sender's window size, round-trip time and router feedback. The router, for each packet that goes through it, looks at the sender's window size and round-trip time. The router uses the sender's window size and round-trip time to calculate the sender's feedback, and enters the feedback in the header of the packet. Note that because of the explicit feedback of allowed window increase to senders, XCP allows senders to increase their window faster than TCP variants whose senders gradually increase their window over many round-trip times. The XCP sender's faster window increase is at the cost of routers having to do additional calculations every time they process a packet, and a new header format being needed.

All of these congestion control algorithms (TCP variants and XCP) however, include an RTO timer as a way to detect and respond to long gaps in missing packet acknowledgements (meaning severe congestion). For the TCP variants, the sender's response to a timeout is to cut its window to one packet. For XCP, the appropriate response to a timeout needs to be investigated since a drastic window cut may be unnecessarily extreme and a less extreme response may be sufficient [24]. An XCP

sender's default method to detect and react to congestion loss is to reduce its window to one packet if the RTO expires. Although the XCP sender has an advantage over the TCP variant sender in being able to increase its window to a larger value more quickly (within two round-trip times) after window reduction, it still suffers from the timeout.

Since the 'Congestion Loss' feedback discussed in this thesis helps the problem of outages causing timeouts, using such feedback together with a TCP variant or XCP may allow for fast window ramp up over FSO networks that have high bandwidth-delay product together with senders not cutting their window upon outage loss. It would be worthwhile to investigate the performance of the 'Congestion Loss' feedback with TCP variants and with XCP. In the case of XCP, whether the timeout response is changed to a less drastic cut or left as cutting to one packet, the 'Congestion Loss' feedback would still improve throughput by preventing timeouts due to outages. However, there would be less to be gained by adding the 'Congestion Loss' feedback if the timeout response is less drastic than cutting the window to one packet.

Another issue to explore is allowing the congestion control algorithm to take into account users with different demands rather than assuming all senders are equal and that they should be given equal rate. The rate allocation aspect of Asynchronous Transfer Mode (ATM) [12] was designed to take into account senders with different throughput demands. However, ATM's rate allocation scheme tracks per flow information and is very burdensome on the routers, especially for a large number of flows and as data rates increase. TCP is good in that low demand senders do not have the unused rate in the network allocated to them; they simply do not send at a rate higher than the packet rate their application has to send. TCP self regulates the

different demand users. However, TCP and the TCP variants do not allow senders to grab available rate as quickly as XCP. Although XCP allows senders to increase their rate quickly, it tries to give equal rate to all users thereby leaving unused capacity on the network when there are senders with widely different demands. Thus, it would be worthwhile to explore an alternative and simpler congestion control algorithm that that takes into account different sender demands. An algorithm similar to XCP but one that includes demand in the packet headers and where routers allocate to senders a rate proportional to the fraction of total demand (as in [10]) may be a useful avenue to pursue. This would require routers to keep track of the total incoming rate and demand. Also, there would have to be a way to prevent senders from artificially inflating their true demand.

Chapter 9

Conclusions

When communicating optically through the atmospheric channel, in addition to fading due to atmospheric turbulence, there may be interference whose signal deteriorates communication performance. We evaluated the mitigation of fading (due to clear atmospheric turbulence) and off-axis interference by use of diversity direct detection and diversity coherent detection. Using a log-normal fading model for the clear atmospheric channel, we derived the error probability and outage probability of diversity direct detection and diversity coherent systems in the presence and absence of turbulence and interference. We also derived the worst case interference duty cycle. The worst case duty cycle, error probability, and outage probability are proportional to the ratio of interference to signal photons. In direct detection, increasing diversity beyond an optimal value actually begins to degrade performance because, as diversity increases, the amount of interference and background noise detected increases. This optimal diversity degree for direct detection occurs when the improvement in fading statistics from diversity is counterbalanced by the added background noise and /or interference signal. We derived this optimal diversity degree both in the absence of and presence of

interference. In coherent detection, increasing diversity always improves the performance.

In order to compare diversity coherent detection with diversity direct detection from different perspectives, we analyzed the power gain of diversity coherent detection over diversity direct detection in the presence of and absence of interference, and where direct detection's diversity is fixed to its optimal value and where it varies. See Figures 5.1 and 5.2 for plots of the power gains. If direct detection's diversity is the same as coherent detection's diversity, then for large diversity values N, the power gain increases proportionally to $\sqrt{N}$ in the absence of interference, and proportionally to N in the presence of interference that uses the worst case duty cycle. If direct detection's diversity is fixed to its optimal value, the power gain approaches a constant. This is due to the limit in average fading improvement that diversity provides. The use of diversity coherent detection receivers is a valuable way to mitigate fading in an atmospheric optical communication system with interference. Diversity coherent detection provides significantly better performance over diversity direct detection receivers due to its ability to limit the amount of unwanted background noise and interference detected. Moreover, most of the performance gain of coherent detection can be achieved with a moderate amount of diversity. However, if the received interference signal is strong and becomes comparable to the strength of the received communication signal, power link margin may be required to mitigate the interference.

Although diversity systems can help with the outages due to atmospheric turbulence and interference, they cannot completely eliminate outages. While we have obtained the expected outage lengths of diversity receivers, the full statistics of the diversity channel is still an open question. Outages on the link affect upper layers such as the

Transport Layer and cause significantly worse performance in them if they are not designed for free-space optical networks. In TCP, since FSO link outages are many orders of magnitude longer than packet transmission times, they can unnecessarily cause the sender to drastically reduce its window to one packet. We derived the maximum performance gain that can be achieved by a class of TCP-based protocols in which the sender distinguishes outage loss from congestion loss and does not reduce its window upon an outage loss. This class of protocols can gain back the performance loss to TCP that link outages cause. Moreover, for high bandwidth-delay product FSO networks with atmospheric turbulence, the class of protocols has significant sender throughput improvement compared to TCP. We discussed a possible way to implement a TCP-based protocol that allows the sender to distinguish outage loss versus congestion loss. Specifically, the router provides an explicit 'Congestion Loss' feedback packet to the sender for each packet the router drops in order to help the sender determine if a packet loss was due to an outage or congestion. The sender also has a Backup-Timer as a way to detect congestion if the 'Congestion Loss' feedback packets and ACKs are continuously dropped. The implementation we discussed is not a full protocol proposal; not all of the details and effects on the network were analyzed. The discussion of this implementation is a means to show the feasibility of providing feedback to the sender to help it distinguish outage loss from congestion loss. Although the class of TCP-based protocols that we discussed has improved throughput over TCP, the sender throughput can be further improved if a congestion control algorithm with improved rate of window increase is used.

Appendix A

Derivation of Probability that Channel is in Outage (or Non-Outage) in t Time Units

In this appendix, we use the 2-state Markov channel model described in Chapter 2 to derive the probability that the process is in an outage (or non-outage) state at time t, given it is in a non-outage state (or outage) at time 0. This appendix is referenced in Sections 2.2 and 6.3.1.1. The 2-state Markov process is allowed to transition between the two states any number of times or no times *between* time interval (0,t).

The *Chapman-Kolmogorov equations* of an n-state Markov Process are given by

$$P_{ij}(t) = \sum_{k} P_{ik}(s) P_{kj}(t-s) \tag{A.1}$$

where $P_{ij}(t)$ is the probability that the Markov process is in state j at time t given that it was in state i at time 0, and the summation is over all n states [17]. The *Kolmogorov backward differential equations*, which are derived by letting s→0, are given by

$$\frac{dP_{ij}(t)}{dt} = \sum_{k \neq i} \left[q_{ik} P_{kj}(t) \right] - v_i P_{ij}(t) \tag{A.2}$$

where v_i is the rate of exiting state i and q_{ik} is the rate of transition to state k given that the process is in state i [17]. These equations can be compacted written in matrix notation as

$$\frac{d}{dt}[P(t)] = [Q] \ [P(t)], \qquad\qquad i > 0 \tag{A.3}$$

where the i,jth element of [P(t)] is $P_{ij}(t)$, and the i,jth element of [Q] is q_{ij} for i≠j and −v_i for i=j. The initial condition at time t=0 is the [P(t)]=identity matrix.

For our channel model, the 2-state Markov chain given in Chapter 2, the corresponding Kolmogorov backward differential equations are given by

$$\frac{d}{dt}\begin{bmatrix} P_{11}(t) & P_{12}(t) \\ P_{21}(t) & P_{22}(t) \end{bmatrix} = \begin{bmatrix} -v_{12} & v_{12} \\ v_{21} & -v_{21} \end{bmatrix} \begin{bmatrix} P_{11}(t) & P_{12}(t) \\ P_{21}(t) & P_{22}(t) \end{bmatrix} \tag{A.4}$$

In expanded form, the equations are

$$\frac{d}{dt} P_{11}(t) = v_{12} P_{21}(t) - v_{12} P_{11}(t) \tag{A.5}$$

$$\frac{d}{dt} P_{21}(t) = -v_{21} P_{21}(t) + v_{21} P_{11}(t) \tag{A.6}$$

$$\frac{d}{dt} P_{12}(t) = -v_{12} P_{12}(t) + v_{12} P_{22}(t) \tag{A.7}$$

$$\frac{d}{dt} P_{22}(t) = v_{21} P_{12}(t) - v_{21} P_{22}(t) \tag{A.8}$$

with initial condition $P_{11}(0)=P_{22}(0)=1$, $P_{12}(0)=P_{21}(0)=0$. The first two differential equations can be solved simultaneously, and likewise for the next two differential equations. Let us proceed in solving the first two differential equations.

v_{21} times (A.5) plus v_{12} times (A.6) yields

$$v_{21}\frac{d}{dt}P_{11}(t)+v_{12}\frac{d}{dt}P_{21}(t)=0 \tag{A.9}$$

Integrating, we get

$$v_{21}P_{11}(t)+v_{12}P_{21}(t)=c \tag{A.10}$$

Using the initial conditions $P_{11}(0)=1$ and $P_{21}(0)=0$, this reduces to

$$v_{21}=c \tag{A.11}$$

so (A.10) becomes

$$v_{21}P_{11}(t)+v_{12}P_{21}(t)=v_{21} \tag{A.12}$$

Rewriting (A.6) using (A.12) yields

$$\frac{d}{dt}P_{21}(t)=-v_{21}P_{21}(t)+v_{21}-v_{12}P_{21}(t)$$
$$=v_{21}-(v_{12}+v_{21})P_{21}(t)$$
$$\frac{d}{dt}P_{21}(t)+(v_{12}+v_{21})P_{21}(t)=v_{21} \tag{A.13}$$

We solve this first order differential equation (A.13) by multiplying both sides by $\exp(v_{12}+v_{21})t$ first.

$$\frac{d}{dt}\left[P_{21}(t)e^{(v_{12}+v_{21})t}\right]=v_2e^{(v_{12}+v_{21})t}$$

(A.14)

Then we integrate and isolate for P₂₁(t) as follows

$$P_{21}(t)e^{(v_{12}+v_{21})t} = v_{21}\int e^{(v_{12}+v_{21})t}dt$$

$$= \frac{v_{21}}{v_{12}+v_{21}}e^{(v_{12}+v_{21})t}+c'$$

$$P_{21}(t) = \frac{v_{21}}{v_{12}+v_{21}}+c'e^{-(v_{12}+v_{21})t}.$$

(B.15)

Using the initial condition P₂₁(0)=0 gives us

$$c' = -\frac{v_{21}}{v_{12}+v_{21}}$$

(A.16)

Therefore,

$$\boxed{P_{21}(t) = \frac{v_{21}}{v_{12}+v_{21}} - \frac{v_{21}}{v_{12}+v_{21}}e^{-(v_{12}+v_{21})t}}$$

(A.17)

We can also solve for P₁₁(t) by substituting the above equation into (A.12).

$$v_{21}P_{11}(t)+v_{12}\left(\frac{v_{21}}{v_{12}+v_{21}} - \frac{v_{21}}{v_{12}+v_{21}}e^{-(v_{12}+v_{21})t}\right)=v_{21}$$

$$P_{11}(t) = -\frac{v_{12}}{v_{12}+v_{21}} + \frac{v_{12}}{v_{12}+v_{21}}e^{-(v_{12}+v_{21})t}+1$$

$$\boxed{P_{11}(t) = \frac{v_{21}}{v_{12}+v_{21}} + \frac{v_{12}}{v_{12}+v_{21}}e^{-(v_{12}+v_{21})t}}$$

(A.18)

Similarly, we can solve for the other two differential equations (A.7) and (A.8) and yield

$$P_{12}(t) = \frac{v_{12}}{v_{12} + v_{21}} - \frac{v_{12}}{v_{12} + v_{21}} e^{-(v_{12} + v_{21})t} \quad \text{and} \tag{A.19}$$

$$P_{22}(t) = \frac{v_{12}}{v_{12} + v_{21}} + \frac{v_{21}}{v_{12} + v_{21}} e^{-(v_{12} + v_{21})t} \tag{A.20}$$

Note that these solutions used the Kolmogorov backward differential equations and did not assume that there are no transitions between time (0,t). i.e. they only stipulate the states at times 0 and t.

Substituting (2.11) and (2.12) into (A.17)-(A.20) gives

$$P_{21}(t) = (1 - P_{outage}) - (1 - P_{outage}) e^{-(v_{12} + v_{21})t} \tag{A.21}$$

$$P_{11}(t) = (1 - P_{outage}) + P_{outage} e^{-(v_{12} + v_{21})t} \tag{A.22}$$

$$P_{12}(t) = P_{outage} - P_{outage} e^{-(v_{12} + v_{21})t} \tag{A.23}$$

$$P_{22}(t) = P_{outage} + (1 - P_{outage}) e^{-(v_{12} + v_{21})t} \tag{A.24}$$

263

Appendix B

Accuracy of the Poisson Detection Model

In this appendix, we show when the Poisson detection model is a good approximate model for single mode detection. This appendix is referenced in Section 3.2.1.

For single mode detection in the presence of only background noise, the photon count is typically modeled by the Bose-Einstein probability distribution [16], namely

$$\Pr(count = k) = \frac{N_n^k}{(N_n + 1)^{k+1}} \tag{B.1}$$

where N_n is the average number of received noise photons. We can approximate the photon counting in the presence of a single mode of background noise with the Poisson distribution

$$\Pr(count = k) = \frac{N_n^k e^{-N_n}}{k!}. \tag{B.2}$$

265

Figure B.1 plots the Bose-Einstein and Poisson distributions for various amounts average background noise photon counts. The Poisson model is a good approximation when the count is low (0, 1 or 2), and up to an average received background noise photons of 1.
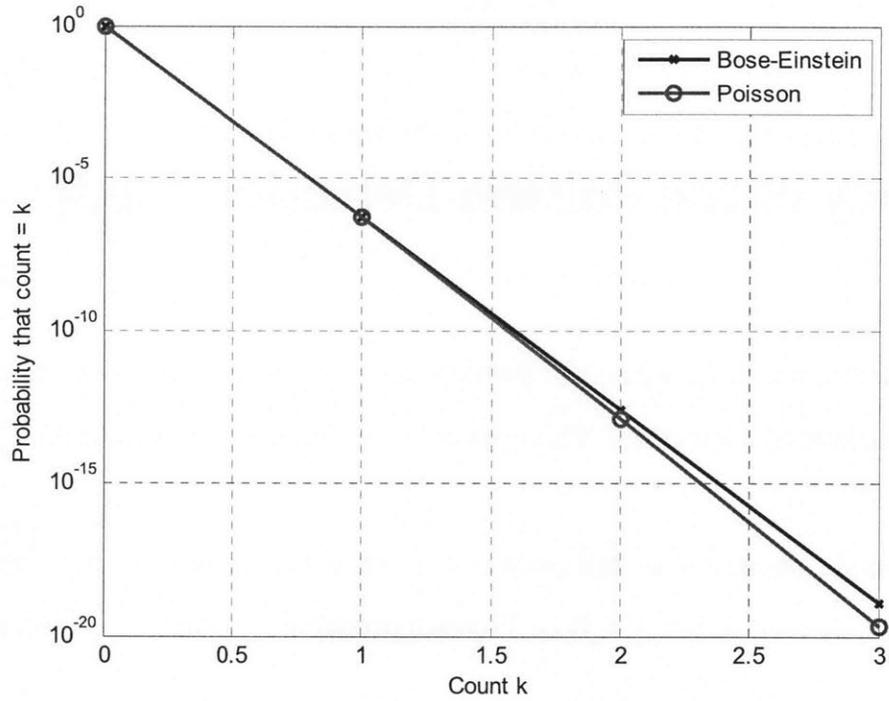
Figure B.1 (a) 0.5×10^{-6} average received photons in the half bit time
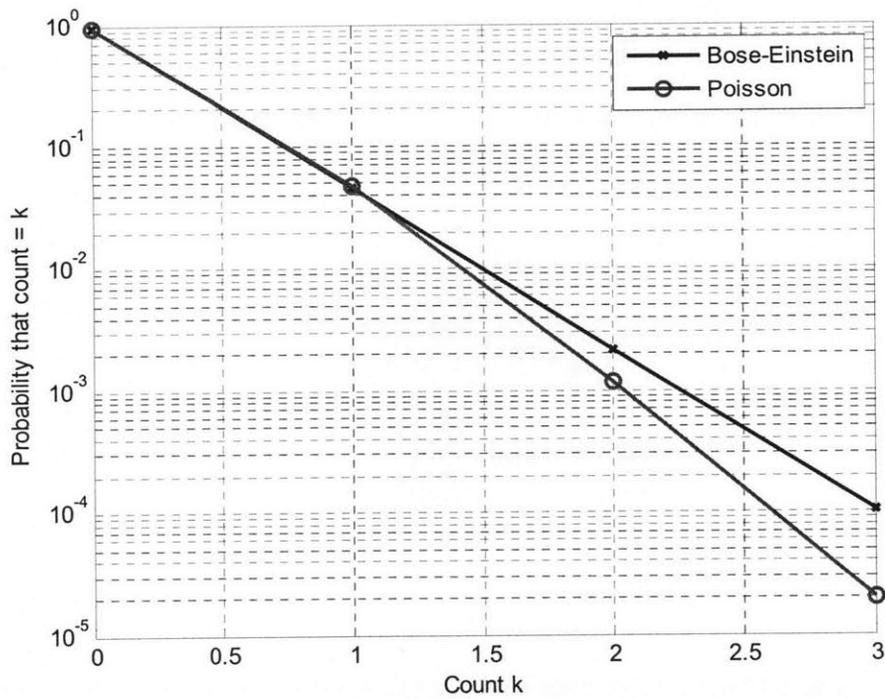
Figure B.1 (b) 0.05 average received photons in the half symbol time
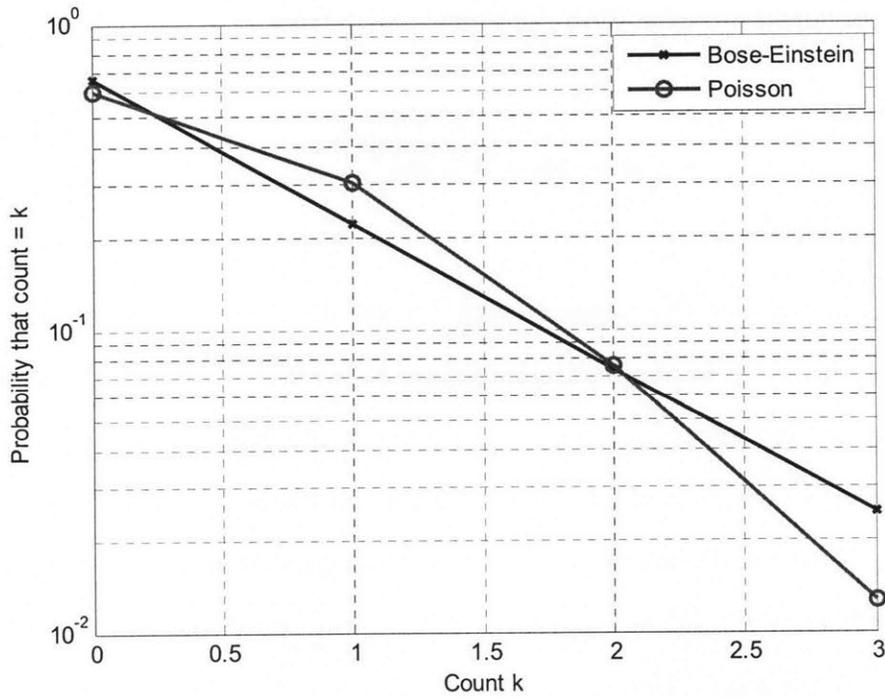
Figure B.1 (c) 0.5 average received photons in the half symbol time

Figure B.1: Photon count probability in a half symbol time using Bose-Einstein and Poisson distributions when the average received photons in the half symbol time is (a) 0.5×10^{-6} (b) 0.05 (c) 0.5

For single mode detection in the presence of background noise and a constant signal, the photon count per symbol is modeled by the Laguerre probability distribution [16], namely

$$\Pr(count = k) = \frac{N_n^k}{(N_n + 1)^{k+1}} \exp\left(\frac{-N_s}{N_n + 1}\right) L_k\left(\frac{-N_s}{N_n(N_n + 1)}\right) \tag{B.3}$$

$$\text{where } L_k(x) = \sum_{j=0}^{k} \binom{k}{k-j} \frac{(-x)}{j!}$$

and N_s and N_n are the average received signal and background noise photons per symbol respectively. We can approximate the photon counting in the presence of a single mode of constant signal with the Poisson distribution

$$\Pr(count = k) = \frac{(N_s + N_n)^k e^{-(N_s + N_n)}}{k!}. \tag{B.4}$$

Figures B.2 and B.3 plot the Laguerre and Poisson distributions for various amounts average background noise photon counts where N_s is fixed to be 10 and 5 respectively. The distributions are very close to each other when the average received background noise photons per symbol is less than 1. Thus, when the signal photons per symbol is on the order of 10, the Poisson counting model is a good approximation for detection of single mode background noise and constant signal when the average background noise photon count is less than 1.
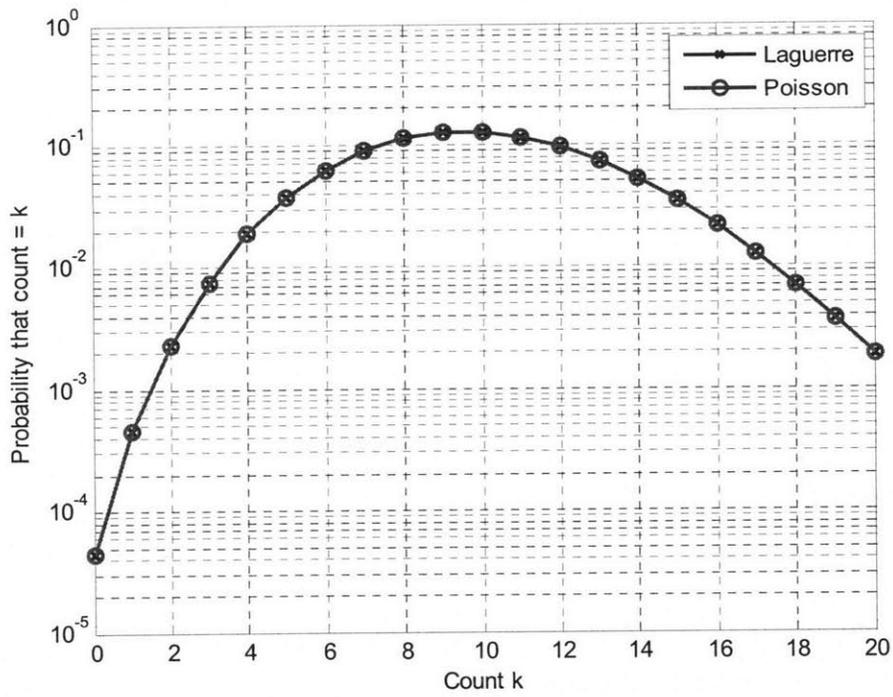
Figure B.2 (a) 10 average received signal photons and 0.5x10⁻⁶ average received background noise photons in the half symbol time

Figure B.2 (b) 10 average received signal photons and 0.05 average received background noise photons in the half symbol time
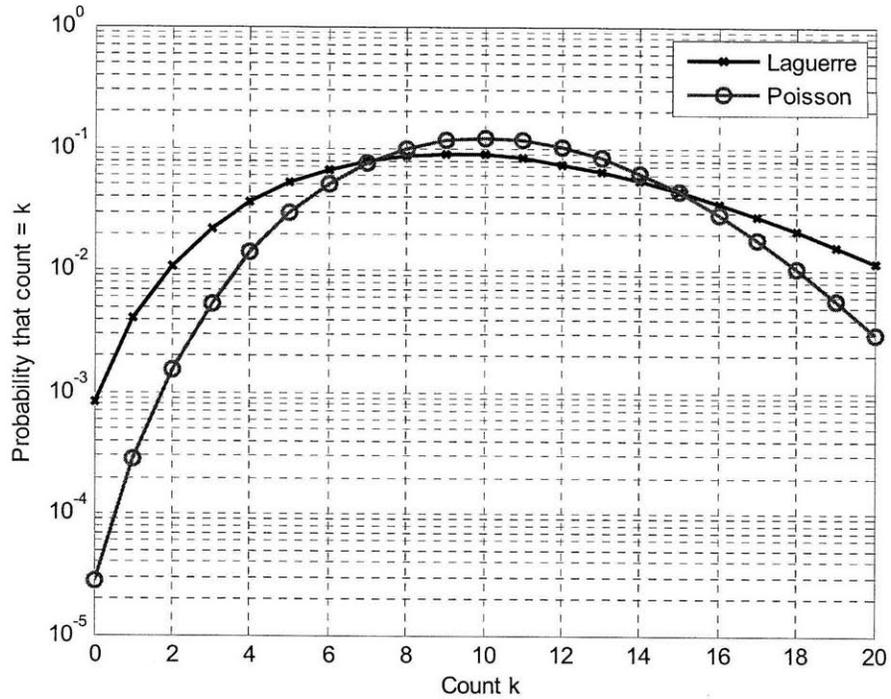
Figure B.2 (c) 10 average received signal photons and 0.5 average received background noise photons in the half symbol time

Figure B.2: Photon count probability in a half symbol time using Laguerre and Poisson distributions when the average received signal photons in the half symbol time is 10 and the average received background noise photons in the half symbol time is (a) 0.5×10^{-6} (b) 0.05 (c) 0.5
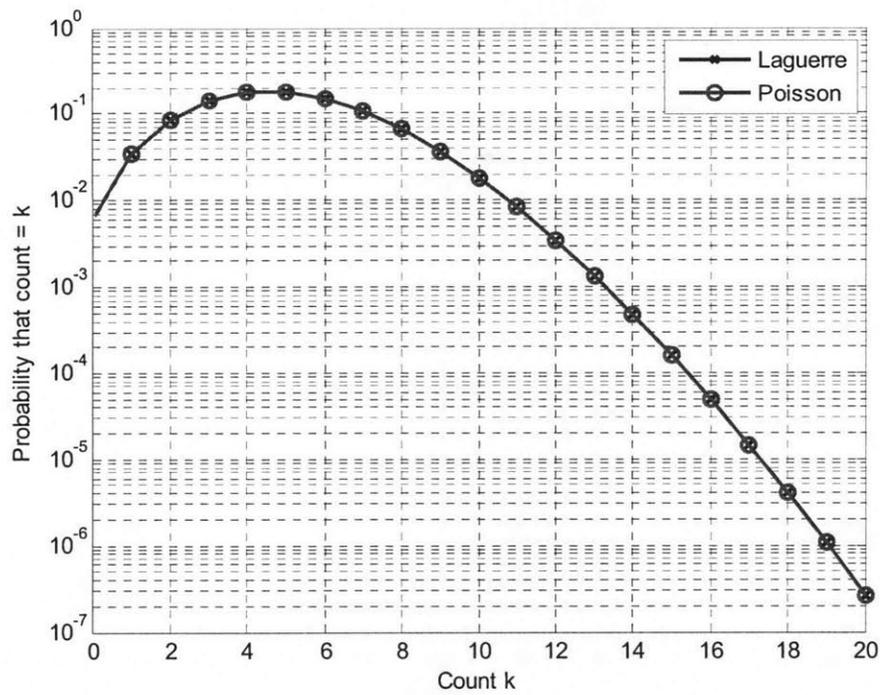
Figure B.3 (a) **5** average received signal photons and 0.5×10^{-6} average received background noise photons in the half symbol time
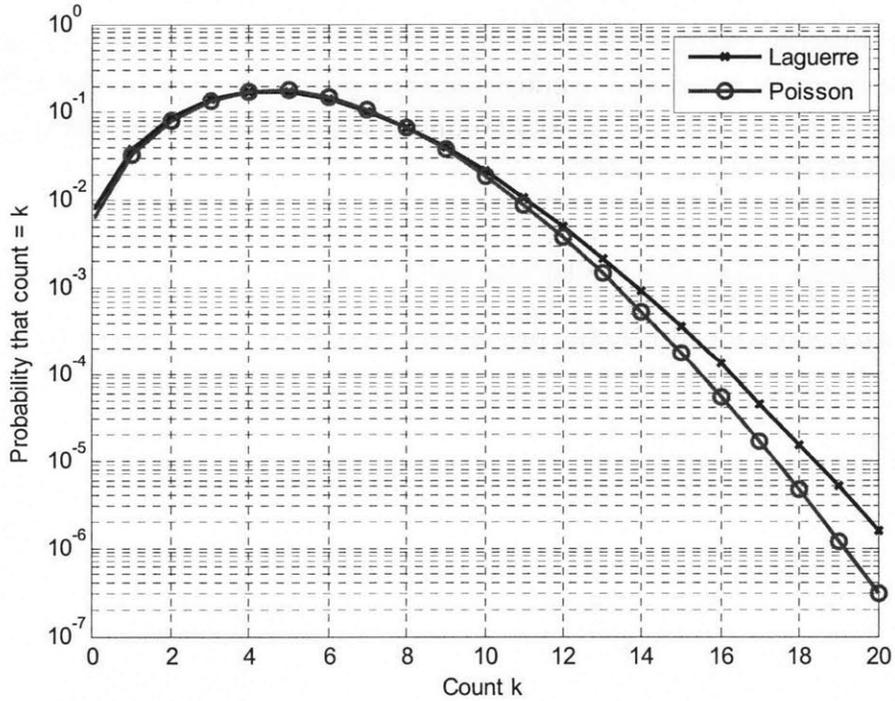
Figure B.3 (b) **5** average received signal photons and 0.05 average received background noise photons in the half symbol time
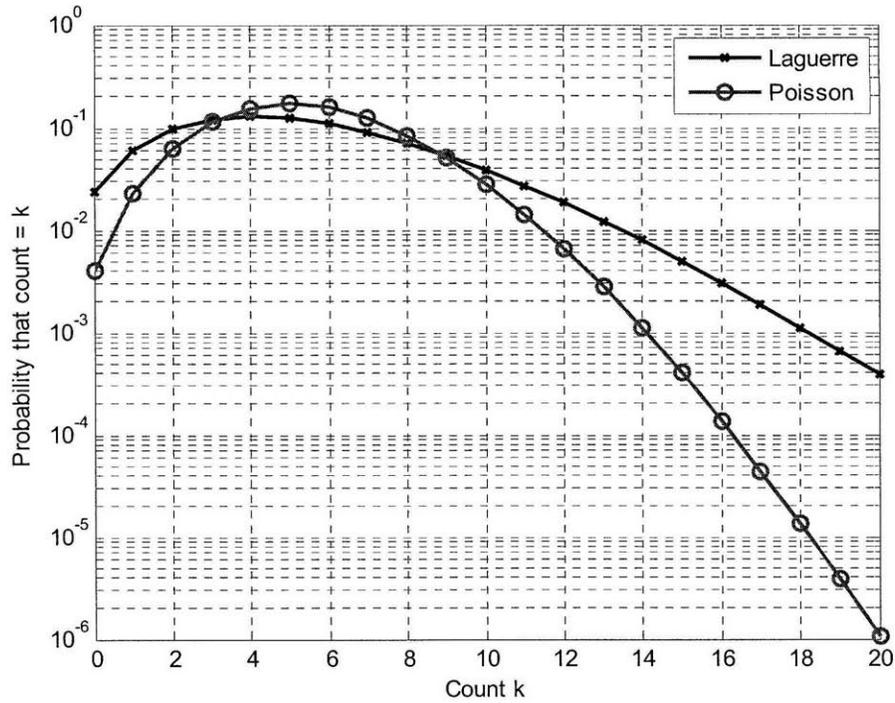
Figure B.3 (c) 5 average received signal photons and 0.5 average received background noise photons in the half symbol time

Figure B.3: Photon count probability in a half symbol time using Laguerre and Poisson distributions when the average received signal photons in the half symbol time is 5 and the average received background noise photons in the half symbol time is (a) 0.5×10^{-6} (b) $N_n = 0.05$ (c) $N_n = 0.5$

In using the Poisson detection model, the value of importance is the error probability. In Direct Detection with BPPM, the probability of symbol error is given by

$$
\begin{aligned}
P(e) &= P(e \mid H_0) \\
&= \sum_{i=0}^{\infty} \Pr(N_1 \le N_0 \mid N_1 = i)\Pr(N_1 = i) \\
&= \sum_{i=0}^{\infty} \Pr(N_1 \le i)\Pr(N_1 = i)
\end{aligned}
\tag{B.5}
$$

where H_0 denotes a '0' being sent by the sender corresponding to signal transmission in the first half symbol, and N_0 and N_1 are the counts in the first and second half symbol intervals. Under hypothesis '0', N_1 is modeled by Laguerre probability and N_0 is modeled by Bose-Einstein probability. Table B.1 tabulates this symbol error probability when N_1 and N_0 are modeled by Laguerre and Bose-Einstein probabilities, and when they are modeled by Poisson probabilities. When $N_s=10$, the Poisson approximation is good when N_n is less than one, and when $N_s=5$, the Poisson approximation is fairly good when N_n is less than or equal to one.

N_n	N_s	Symbol error probability P(e) when using Laguerre and Bose-Einstein model	Symbol error probability P(e) when using Poisson model
10^{-6}	10	4.5×10^{-5}	4.5×10^{-5}
0.1	10	1.1×10^{-4}	0.67×10^{-4}
1	10	5.1×10^{-3}	0.38×10^{-3}
10^{-6}	5	6.7×10^{-3}	6.7×10^{-3}
0.1	5	1.0×10^{-2}	0.8×10^{-2}
1	5	6.2×10^{-2}	2×10^{-2}

Table B.1 Error probability values when using the Laguerre and Bose-Einstein model and when using the Poisson model

Appendix C

Derivation of Optimal Local Oscillator Weights in Diversity Coherent Detection

In this appendix, we derive local oscillator weights γ_i of the diversity coherent detection system that minimize the error probability. This appendix is referenced in Section 3.3.2. The conditional error probability of the diversity heterodyne system is given by (3.28), namely

$$P(e \mid \alpha_1, \alpha_2, \ldots, \alpha_N) = \frac{1}{2} \exp \left(-\frac{\left(\sum_{i=1}^{N} \gamma_i \sqrt{\alpha_i} \right)^2}{N} \left(\frac{N_S}{1 + N_n} \right) \right)$$

The weights that minimize the error probability are the weights that maximize the exponent. In order to incorporate the fact that $\sum_{i=1}^{N} \gamma_i^2 = 1$, we re-write the weights as

$$\gamma_i = \frac{k_i}{\sqrt{\sum_{j=1}^{N} k_j^2}} \tag{C.1}$$

where $k_i \geq 0$. Substituting (C.1) into the above error probability, we get

$$P(e \mid \alpha_1, \alpha_2, \ldots, \alpha_N) = \frac{1}{2} \exp\left(-\left(\sum_{i=1}^{N} \frac{k_i \sqrt{\alpha_i}}{\sqrt{\sum_{j=1}^{N} k_j^2}}\right)^2 \left(\frac{N_S}{N(1+N_n)}\right)\right) \tag{C.2}$$

We now take the derivative of the exponent with respect to k_l and set it equal to 0.

$$
\begin{aligned}
0 &= \frac{\partial}{\partial k_l}\left[\sum_{i=1}^{N} \frac{k_i \sqrt{\alpha_i}}{\sqrt{\sum_{j=1}^{N} k_j^2}}\right] \\
&= \left(\sum_{i \neq l} \frac{-k_l k_i \sqrt{\alpha_i}}{\left(\sum_{j=1}^{N} k_j^2\right)^{3/2}}\right) + \left(\frac{\sqrt{\alpha_l \sum_{j=1}^{N} k_j^2} - k_l \sqrt{a_l} \left(\sum_{j=1}^{N} k_j^2\right)^{-1/2} k_l}{\sum_{j=1}^{N} k_j^2}\right) \tag{C.3} \\
&= -k_l \sum_{i \neq l} k_i \sqrt{\alpha_i} + \sqrt{\alpha_l}\left[\sum_{j=1}^{N} k_j^2 - k_l^2\right] \\
&= -k_l \sum_{i \neq l} k_i \sqrt{\alpha_i} + \sqrt{\alpha_l} \sum_{i \neq l} k_i^2
\end{aligned}
$$

Isolating for k_l we get

$$k_l = \frac{\sqrt{\alpha_l} \sum_{i \neq l} k_i^2}{\sum_{i \neq l} k_i \sqrt{\alpha_i}} \qquad (C.4)$$

To find the values of k_i, for all i, that satisfy the above equation, let us start with the simple case where N=2. For N=2, (C.4) is

$$k_1 = \frac{\sqrt{\alpha_1} k_2^2}{k_2 \sqrt{\alpha_2}} = \frac{\sqrt{\alpha_1}}{\sqrt{\alpha_2}} k_2 \qquad (C.4)$$

of which the solution is $k_i = \sqrt{\alpha_i}$ by inspection. For general values of N, let us postulate that the solution is also $k_i = \sqrt{\alpha_i}$. Substituting this solution into (C.4), we indeed find that the equation (C.4) is satisfied. Thus, $k_i = \sqrt{\alpha_i}$ minimizes the error probability. Substituting this solution $k_i = \sqrt{\alpha_i}$ into (C.1), the optimal weights γ_i are given by

$$\gamma_i = \frac{\sqrt{\alpha_i}}{\sqrt{\sum_{j=1}^{N} \alpha_i}}$$

277

Appendix D

Derivation of Direct Detection Optimal Diversity in Absence of Interference

In this appendix, we derive the optimal diversity value (an integer ≥ 1) in direct detection that minimizes the amount of signal power required to achieve a given outage probability $P_{outage,DD}$ in the absence of interference. This appendix is referenced in Section 3.4. Isolating for N_s in the outage probability expression for direct detection (given in the last line of (3.10)) gives

$$N_{S,DD} = \left(\theta_{thresh} + 2\sqrt{NN_n\theta_{thresh}}\right)\sqrt{1 + \frac{e^{4\sigma_\chi^2} - 1}{N}} \cdot \exp\left\{-\sqrt{-2\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)\ln\left(2P_{outage,DD}\right)}\right\} \quad (D.1)$$

Rearranging, we get

$$N_{S,DD} = \theta_{thresh} \sqrt{1 + \frac{e^{4\sigma_\chi^2} - 1}{N}} \cdot \exp\left\{ -\sqrt{-2\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)} \ln\left(2P_{outage,DD}\right) \right\}$$

$$+ 2\sqrt{N_n \theta_{thresh}} \sqrt{N + e^{4\sigma_\chi^2} - 1} \exp\left\{ -\sqrt{-2\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)} \ln\left(2P_{outage,DD}\right) \right\} \qquad \text{(D.2)}$$

Taking the derivative with respect to N and setting equal to zero, we get

$$0 = \exp\left\{ -\sqrt{-2\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)} \ln\left(2P_{outage,DD}\right) \right\} \left\{ \frac{-\theta_{thresh} a N^{-2} \sqrt{-\ln\left(2P_{outage,DD}\right)}}{\sqrt{1 + \frac{a}{N}} \sqrt{2\ln\left(1 + \frac{a}{N}\right)}} - \frac{\theta_{thresh} a}{2N^2 \sqrt{1 + \frac{a}{N}}} - \frac{2\sqrt{N_n \theta_{thresh} N} \sqrt{-\ln\left(2P_{outage,DD}\right)} a N^{-2}}{\sqrt{2\left(1 + \frac{a}{N}\right) \ln\left(1 + \frac{a}{N}\right)}} + \sqrt{\frac{N_n \theta_{thresh}}{N + a}} \right\} \qquad \text{(D.3)}$$

where $a = e^{4\sigma_\chi^2} - 1$. Multiplying both sides by

$$\frac{N^2 \sqrt{2\ln\left(1 + \frac{a}{N}\right)} \sqrt{1 + \frac{a}{N}}}{\exp\left\{ -\sqrt{-2\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)} \ln\left(2P_{outage,DD}\right) \right\}}$$

results in

$$0 = \left\{ -\theta_{thresh} a \sqrt{-\ln\left(2P_{outage,DD}\right)} - \frac{\theta_{thresh} a}{2} \sqrt{2\ln\left(1 + \frac{a}{N}\right)} \atop -2a\sqrt{-N_n \theta_{thresh} N \ln\left(2P_{outage,DD}\right)} + N^{3/2} \sqrt{2N_n \theta_{thresh} \ln\left(1 + \frac{a}{N}\right)} \right\} \qquad \text{(D.4)}$$

Approximating $\dfrac{a}{N} = \dfrac{e^{4\sigma_\chi^2} - 1}{N} \ll 1$ (reasonable for mild turbulence) allows us use the

approximation $\ln\left(1 + \dfrac{a}{N}\right) \cong \dfrac{a}{N}$. Thus,

$$0 = \left\{ \begin{array}{l} -\theta_{thresh}a\sqrt{-\ln\left(2P_{outage,DD}\right)} - \dfrac{\theta_{thresh}a}{2}\sqrt{2\dfrac{a}{N}} \\[3mm] -2a\sqrt{-N_n\theta_{thresh}N\ln\left(2P_{outage,DD}\right)} + N^{3/2}\sqrt{2N_n\theta_{thresh}\dfrac{a}{N}} \end{array} \right\} \qquad (D.5)$$

Multiplying both sides by $\sqrt{\dfrac{N}{\theta_{thresh}a}}$ gives

$$0 = -\sqrt{-N\theta_{thresh}a\ln\left(2P_{outage,DD}\right)} - a\sqrt{\dfrac{\theta_{thresh}}{2}} - 2N\sqrt{-aN_n\ln\left(2P_{outage,DD}\right)} + N^{3/2}\sqrt{2N_n} \qquad (D.6)$$

Let us denote the optimal direct detection diversity value in the absence of as N_{opt}.

Then re-writing (D.6), N_{opt} is the root (that is ≥ 1)[20] of

$$0 = a_3 N_{opt}^{3/2} + a_2 N_{opt} + a_1 N_{opt}^{1/2} + a_0$$

where

$$a_3 = \sqrt{2N_n}$$

$$a_2 = -2\sqrt{-N_n\left(e^{4\sigma_\chi^2} - 1\right)\ln\left(2P_{outage}\right)}$$

$$a_1 = -\sqrt{-\theta^{thresh}\left(e^{4\sigma_\chi^2} - 1\right)\ln\left(2P_{outage}\right)}$$

$$a_0 = -\sqrt{\dfrac{\theta^{thresh}}{2}\left(e^{4\sigma_\chi^2} - 1\right)}$$

---

[20] When the roots of (D.6) were found for a wide range of outage probabilities, background noise and turbulence, only one of the roots was a real number.

and where $\dfrac{e^{4\sigma_\chi^2}-1}{N} \ll 1$. In cases where N_{opt} is not an integer, it is sufficient to check

if rounding up or down yields the best diversity value. This is because the second

derivative with respect to N of (D.2) is positive, as we now summarize. Taking the

second derivative with respect to N of (D.2) is straightforward but tedious, and results

in

$$
\frac{\partial^2}{\partial N^2} N_s = \exp\left\{-\sqrt{-2\ln\left(1+\frac{a}{N}\right)\ln\left(2P_{outage,DD}\right)}\right\}
\left[
\begin{array}{l}
\dfrac{-\theta_{thresh}\,a\sqrt{-\ln\left(2P_{outage,DD}\right)}}{2N^4\sqrt{2\left(1+\frac{a}{N}\right)\ln\left(1+\frac{a}{N}\right)}}\left[\dfrac{a}{\left(1+\frac{a}{N}\right)\ln\left(1+\frac{a}{N}\right)}+\dfrac{a}{\left(1+\frac{a}{N}\right)}-4N^7\right] \\[2em]
-\dfrac{\theta_{thresh}a}{N^3\sqrt{1+\frac{a}{N}}}\left(\dfrac{a}{2N\left(1+\frac{a}{N}\right)}-2\right) \\[2em]
-\dfrac{a\sqrt{-N_n\theta_{thresh}\,\ln\left(2P_{outage,DD}\right)}}{N^{7/2}\sqrt{2\left(1+\frac{a}{N}\right)\ln\left(1+\frac{a}{N}\right)}}\left[\dfrac{a}{\left(1+\frac{a}{N}\right)\ln\left(1+\frac{a}{N}\right)}+\dfrac{a}{1+\frac{a}{N}}-3N\right] \\[2em]
+\dfrac{\sqrt{N_n\theta_{thresh}}}{2N^{3/2}\sqrt{1+\frac{a}{N}}}\left(\dfrac{a}{N\left(1+\frac{a}{N}\right)}+1\right) \\[2em]
-\dfrac{a\sqrt{-\theta_{thresh}\,\ln\left(2P_{outage,DD}\right)}}{N^4\left(1+\frac{a}{N}\right)\sqrt{2\ln\left(1+\frac{a}{N}\right)}}\left[\begin{array}{l}\dfrac{-a\sqrt{-\ln\left(2P_{outage,DD}\right)}\left(\sqrt{\theta_{thresh}}+2\sqrt{N_nN}\right)}{\sqrt{2\ln\left(1+\frac{a}{N}\right)}} \\[1.2em] -a\sqrt{\theta_{thresh}}+\sqrt{N_n}N^{3/2}\end{array}\right]
\end{array}
\right]
$$

(D.7)

where $a = e^{4\sigma_\chi^2}-1$. By careful and tedious inspection, for $\sigma_\chi^2 \in [0,0.5]$, the second

derivative can be seen to be positive.

282

Appendix E

Asymptotic Outage Probability of Homodyne Detection for Large Diversity

In this appendix, we derive the asymptotic outage probability of diversity homodyne detection as a function of N. This appendix is referenced in Section 3.4. From (3.35), the outage probability of homodyne detection is

$$P_{outage,Homo,noInterference} \cong \frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right]^2\right\}$$

Substituting m_U and σ_U^2 (as given in (3.9)) into the above outage probability,

$$P_{outage, Homo, noInterference} \cong \frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right]^2\right\}$$

$$= \frac{1}{2}\exp\left\{-\frac{1}{2\ln\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right)}\left[-\frac{1}{2}\ln\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right) - \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right]^2\right\} \quad (E.1)$$

For large N, $\ln\left(1+\frac{e^{4\sigma_\chi^2}-1}{N}\right) \cong \frac{e^{4\sigma_\chi^2}-1}{N}$. Thus,

$$P_{outage, Homo, noInterference} \cong \frac{1}{2}\exp\left\{-\frac{1}{2\left(\frac{e^{4\sigma_\chi^2}-1}{N}\right)}\left[-\frac{1}{2}\left(\frac{e^{4\sigma_\chi^2}-1}{N}\right) - \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right]^2\right\}$$

$$= \frac{1}{2}\exp\left\{-\frac{1}{8}\left(\frac{e^{4\sigma_\chi^2}-1}{N}\right) - \frac{1}{2} - \frac{1}{2\left(e^{4\sigma_\chi^2}-1\right)}\left[\ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right]^2 N\right\} \quad (E.2)$$

$$= \frac{1}{2}\exp\left\{-\frac{1}{8}\left(\frac{e^{4\sigma_\chi^2}-1}{N}\right) - \frac{1}{2} - cN\right\}$$

where in the last line we let $c = \dfrac{\left[\ln\left(\dfrac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_S}\right)\right]^2}{2\left(e^{4\sigma_\chi^2}-1\right)}$.

Thus, for large N, the outage probability of homodyne detection goes as

$$P_{outage, Homo, noInterference} \sim \frac{1}{2}\exp\{-cN\}$$

$$\text{where} \quad c = \frac{\left[\ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_S} \right) \right]^2}{2\left(e^{4\sigma_\chi^2} - 1 \right)}$$

Appendix F

Derivation of Expected Outage Length

This appendix is referenced in Section 3.5. As found in Chapter 3, the outage probability of an N receiver diversity direct detection system is given by (3.10), namely

$$P_{outage,DD,noInterference} \cong Q\left(\frac{m_U - \ln\left(\frac{1}{N_S}\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]\right)}{\sigma_U}\right)$$

$$\cong \frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\left[\theta_{thresh} + \sqrt{2\theta_{thresh}NN_n}\right]}{N_S}\right)\right]^2\right\}$$

(F.1)

and the outage probability of an N receiver diversity homodyne detection system is given by (3.35), namely

$$P_{outage,Homo,nolnterference} \cong Q\left(\frac{m_U - \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)}{\sigma_U}\right)$$

$$\cong \frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{(1+N_n)(-\ln 2 + \theta_{thresh})}{2N_s}\right)\right]^2\right\} \qquad (F.2)$$

Using level crossing theory [11], a Gaussian process W, has an expected outage length of

$$E[outage\ length] = \frac{P_{outage}}{E[C_\zeta(0,1)]} \qquad (F.3)$$

where P_{outage} is probability of the Gaussian process W being below level ζ (and is given by (F.1) for direct detection and (F.2) for homodyne detection), and $E[C_\zeta(0,1)]$ is the mean number of ζ-level crossings of the Gaussian process W per second and is given by

$$E[C_\zeta(0,1)] = \frac{1}{\pi}\sqrt{\frac{\lambda_2^W}{\lambda_0^W}}\exp\left\{-\frac{(\zeta - E[W])^2}{2\lambda_0^W}\right\} \qquad (F.4)$$

$$\text{where } \lambda_0^W = \frac{1}{2\pi}\int S_W(\omega)d\omega = \int S_W(f)df = K(0) \qquad (F.5)$$

$$\text{and } \lambda_2^W = \frac{1}{2\pi}\int \omega^2 S_W(\omega)d\omega = (2\pi)^2 \int f^2 S_W(f)df = -K_W^{''}(0) \qquad (F.6)$$

$S_W(f)$ and $K_W(\tau)$ denote the power spectral density (psd) and covariance function of the Gaussian process W.

The power fading factor $e^{2\chi}$ of a single transmitter, single receiver system is modeled as a log-normal (as described in Chapter 2) where χ is Gaussian with mean $E[\chi]=-\sigma_\chi^2$.

288

For an N receiver system, the fading factor is $\left(e^{2\chi_1} + e^{2\chi_2} + ... e^{2\chi_N}\right)/N$ and we approximate this sum of log-normal random variables to be log-normal [35,43]. We further define this factor to be

$$e^{2W} = \frac{1}{N}\left(e^{2\chi_1} + e^{2\chi_2} + ... e^{2\chi_N}\right) \qquad \text{(F.7)}$$

where W is a Gaussian random. Finding the expected outage length of the log-normal process e^{2W} is equivalent to the expected outage length of W. Since we know P_{outage} is given by (F.1) for direct detection and (F.2) for homodyne detection, if we can find the mean E[W], 0th spectral moment λ_0 and second spectral moment λ_2 of W, then given level ζ, we can use (F.3)-(F.6) to determine the expected outage length of the averaged diversity branches.

In Chapter 3, we defined $e^U = \frac{1}{N}\sum_{i=1}^{N}\alpha_i = \frac{1}{N}\left(e^{2\chi_1} + e^{2\chi_2} + ... e^{2\chi_N}\right)$ and quoted the mean of U as being $m_U = -\frac{1}{2}\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)$ (derived in [43]). Then $W = U/2$, and the mean of W is half the mean of U, namely

$$E[W] = -\frac{1}{4}\ln\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right) \qquad \text{(F.8)}$$

Now derive the 0th and second spectral moment of χ (λ_0^χ and λ_2^χ) and then express the 0th and second moment of the spectral density of W (λ_0^W and λ_2^W) in terms of λ_0^χ and λ_2^χ. The psd of χ for a plane wave propagating through a medium with refractive index spectrum given in [27] to be

289

$$S_\chi(f) = 0.15\langle \chi^2 \rangle \frac{1}{f_0} \left[1 + 0.48 \left(\frac{f}{f_0} \right)^{4/3} \right] \qquad for \ \ f \ll f_0$$

$$S_\chi(f) = 1.14 \langle \chi^2 \rangle \frac{f^{-8/3}}{f_0^{-5/3}} \qquad for \ \ f \gg f_0$$

(F.9)

where $f_0 = v_\perp / \sqrt{2\pi\lambda L}$ at wavelength λ, path distance L and transverse wind speed $v_\perp$. We approximate this psd by

$$S_\chi(f) = \begin{cases} 0.15 \langle \chi^2 \rangle \dfrac{1}{f_0} \left[1 + 0.48 \left(\dfrac{f}{f_0} \right)^{4/3} \right] & \text{for } f < f_0 \\[4mm] 1.14 \langle \chi^2 \rangle \dfrac{f^{-8/3}}{f_0^{-5/3}} & \text{for } f_0 \leq f \leq f_{cutoff} \\[4mm] 0 & \text{for } f > f_{cutoff} \end{cases}$$

(F.10)

The value $1/f_0$ is the time it takes for a turbule of width equal to the first Fresnel zone to cross the path of the optical wave at speed $v_\perp$. The psd is cutoff above frequency f_{cutoff} to ensure that $\int f^2 S(f) df$ is not infinite. f_{cutoff} is taken to be much larger than f_0. Substituting (F.10) into the definition of the 0th and 2nd spectral moment (F.5) and (F.6) (where we replace W by χ) and performing the simple integration, λ_0^χ and λ_2^χ are given by

$$\lambda_0^\chi = 2(0.15)\langle \chi^2 \rangle \left[1 + \frac{3(0.48)}{7} \right] - \frac{6}{5}(1.14)\langle \chi^2 \rangle \left[\left(\frac{f_{cutoff}}{f_0} \right)^{-5/3} - 1 \right]$$

(F.11)

and

$$\lambda_2^\chi = 2(2\pi)^2 (0.15)\langle \chi^2 \rangle \left[\frac{1}{3} + \frac{3(0.48)}{13} \right] f_0^2 + 6(2\pi)^2 (1.14)\langle \chi^2 \rangle \left[f_{cutoff}^{1/3} f_0^{5/3} - f_0^2 \right]$$

(F.12)

The correlation function of $e^{2\chi}$ is

$$R_{e^{2\chi}e^{2\chi}}(\tau) = E\left[e^{2\chi(t)}e^{2\chi(t+\tau)}\right]$$

$$= \iint \frac{e^{2y_1+2y_2}}{2\pi\sigma_\chi^2\sqrt{1-\rho^2}}\exp\left\{-\frac{(y_1-m_\chi)^2\sigma_\chi^2 - 2(y_1-m_\chi)(y_2-m_\chi)\rho\sigma_\chi^2 + (y_2-m_\chi)^2\sigma_\chi^2}{2\sigma_\chi^2(1-\rho^2)}\right\}dy_1dy_2 \quad \text{(F.13)}$$

By using the fact that $\chi(t)$ and $\chi(t+\tau)$ are jointly Gaussian with correlation coefficient

$$\rho = \frac{K_{\chi\chi}(\tau)}{\sigma_\chi^2} \quad \text{(F.14)}$$

and that

$$m_\chi = -\sigma_\chi^2, \quad \text{(F.15)}$$

we can simplify (F.13) by expanding all terms in the integrand and substituting (F.14) and (F.15). After some simple and tedious algebra, we get

$$R_{e^{2\chi}e^{2\chi}}(\tau) = E\left[e^{2\chi(t)}e^{2\chi(t+\tau)}\right] = \exp\left[4K_{\chi\chi}(\tau)\right] \quad \text{(F.16)}$$

The covariance function of $e^{2\chi}$ is then

$$K_{e^{2\chi}e^{2\chi}}(\tau) = R_{e^{2\chi}e^{2\chi}}(\tau) - m_{e^{2\chi}}^2 = \exp\left(4K_{\chi\chi}(\tau)\right) - 1 \quad \text{(F.15)}$$

Assuming that the fading factor seen by each receiver is independent, the covariance function of e^{2W} is

$$K_{e^{2W}e^{2W}}(\tau) = \frac{1}{N} K_{e^{2\chi}e^{2\chi}}(\tau)$$

$$= \frac{1}{N} \left[\exp\left(4K_{\chi\chi}(\tau)\right) - 1 \right]$$

$$= \exp\left\{ \ln\left(\frac{1}{N} \exp\left[4K_{\chi\chi}(\tau)\right] + \frac{N-1}{N} \right) \right\} - 1 \qquad \text{(F.16)}$$

where the last line is of the form of (F.15) but where the expression inside of the exponent is different. Thus,

$$K_{WW}(\tau) = \frac{1}{4} \ln\left\{ \frac{1}{N} \exp\left[4K_{\chi\chi}(\tau)\right] + \frac{N-1}{N} \right\} \qquad \text{(F.17)}$$

and

$$\lambda_0^W = K_{WW}(0) = \frac{1}{4} \ln\left(\frac{1}{N} \exp\left(4\lambda_0^\chi\right) + \frac{N-1}{N} \right) \qquad \text{(F.18)}$$

$$\lambda_2^W = -K_{WW}''(0) = -\left[\frac{K_{\chi\chi}''(\tau) e^{4K_{\chi\chi}(\tau)}}{e^{4K_{\chi\chi}(\tau)} + N - 1} \right]_{\tau=0} = \frac{\lambda_2^\chi e^{4\lambda_0^\chi}}{e^{4\lambda_0^\chi} + N - 1} \qquad \text{(F.19)}$$

where the second last equality is true because we assumed that process $\chi(t)$ does not have an infinite number of u-level crossings ($K_{\chi\chi}'(0)=0$). Thus, we have found E[W], λ_0^W and λ_2^W in terms if σ_χ^2, λ_0^χ and λ_2^χ.

We now describe the level ζ at which W is in an outage. From (3.6) and (3.34), the average number of detected signal photons per symbol needed by an N-receiver direct detection and homodyne system to achieve P_e^{thresh} in the absence of fading are

$$N_{s,DD,N}^{thresh} = \left(\sqrt{-\ln\left(P_e^{thresh}\right)} + \sqrt{\frac{NN_n}{2}} \right)^2 - \frac{NN_n}{2} \qquad \text{(F.20)}$$

and

$$N_{s,Homo,N}^{thresh} = -\frac{(1+N_n)}{2}\ln(2P_e^{thresh}) = N_s^* \qquad \text{(F.21)}$$

respectively. If the system provides a link margin of m beyond N_S^* (the signal photons needed by a single receiver homodyne system to achieve P_e^{thresh}), the system is in outage if

$$e^{2W} < \frac{N_{s,DD,N}^{thresh}}{mN_s^*} \quad \Rightarrow \quad W < \frac{1}{2}\ln\left(\frac{\left(\sqrt{-\ln(P_e^{thresh})}+\sqrt{\frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{-m\frac{(1+N_n)}{2}\ln(2P_e^{thresh})}\right) \qquad \text{(F.22)}$$

in direct detection and

$$e^{2W} < \frac{N_{s,Homo,N}^{thresh}}{mN_s^*} \quad \Rightarrow \quad W < \frac{1}{2}\ln\left(\frac{1}{m}\right) \qquad \text{(F.23)}$$

in homodyne detection. Thus,

$$\zeta = \frac{1}{2}\ln\left(\frac{\left(\sqrt{-\ln(P_e^{thresh})}+\sqrt{\frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{-m\frac{(1+N_n)}{2}\ln(2P_e^{thresh})}\right) \qquad \text{(F.24)}$$

in direct detection and

$$\zeta = \frac{1}{2}\ln\left(\frac{1}{m}\right) \qquad \text{(F.25)}$$

in homodyne detection.

We can now calculate the expected outage length of W (i.e. the expected outage length when diversity is used) as a function of diversity N, link margin m, transmitter-receiver separation L, wavelength λ, turbulence strength σ_χ^2, and transverse wind speed $v_\perp$ by substituting (F.1) (or (F.2)), (F.8), (F.18), (F.19) and (F.24) (or (F.25)) into (F.3) and (F.4).

Now let us show that the expected outage length is inversely proportional to the transverse wind speed i.e. E[outage length]$\propto 1/v_t$. The expected outage length expression (F.3) consists of P_{outage} which is not a function of v_t and $E\lfloor C_\zeta(0,1)\rfloor$. All variables in the expression for $E\lfloor C_\zeta(0,1)\rfloor$ (in (F.4)) are not a function of v_t except for possibly λ_0^W and λ_2^W. Let us now see the dependency of λ_0^W and λ_2^W on v_t. Let the cutoff frequency be a multiple of f of f₀. Substituting $f_{cutoff} = g f_0$ (where g>>1) into (F.11), we get

$$\lambda_0^\chi = 2(0.15)\langle\chi^2\rangle\left[1+\frac{3(0.48)}{7}\right]-\frac{6}{5}(1.14)\langle\chi^2\rangle\left[(g)^{-5/3}-1\right] \qquad \text{(F.26)}$$

which does not change with v_t. Thus, from (F.18), λ_0^W does not change with v_t. Substituting $f_{cutoff} = g f_0$ and $f_0 = v_\perp/\sqrt{2\pi\lambda L}$ into (F.12), we get

$$
\begin{aligned}
\lambda_2^\chi &= 2(2\pi)^2(0.15)\langle\chi^2\rangle\left[\frac{1}{3}+\frac{3(0.48)}{13}\right]f_0^2 + 6(2\pi)^2(1.14)\langle\chi^2\rangle\left[g^{1/3}f_0^2 - f_0^2\right] \\
&= f_0^2\left\{2(2\pi)^2(0.15)\langle\chi^2\rangle\left[\frac{1}{3}+\frac{3(0.48)}{13}\right]+6(2\pi)^2(1.14)\langle\chi^2\rangle\left[g^{1/3}-1\right]\right\} \qquad \text{(F.27)} \\
&= \frac{v_\perp^2}{2\pi\lambda L}\left\{2(2\pi)^2(0.15)\langle\chi^2\rangle\left[\frac{1}{3}+\frac{3(0.48)}{13}\right]+6(2\pi)^2(1.14)\langle\chi^2\rangle\left[g^{1/3}-1\right]\right\}
\end{aligned}
$$

which is proportional to the square of the transverse wind speed. As we can see from (F.19), since λ_0^{χ} is not a function of $v_\perp$ and λ_2^{χ} is proportional to the square of $v_\perp$, λ_2^{W} is proportional to $v_\perp$. Thus, from (F.4), $E\left[C_\zeta(0,1)\right]$ is proportional to $v_\perp$ and from (F.3), the expected outage length is inversely proportional to the transverse wind speed.

Appendix G

Derivations of Various Expressions in Chapter 4

In this appendix, we provide derivations for various expressions in Chapter 4. We summarize in Table G.1 descriptions of the quantities derived, the section numbers in this appendix in which they are derived, the sections in the thesis to which the derivations apply, and the corresponding equation numbers in the thesis. When deriving the optimal duty cycles, we assume $NN_I < mN_S^*$. This is reasonable for off-axis interference that is scattered into the receivers.

Quantity derived	Section in this appendix	Section in thesis to which derivation applies	Equation derived
The tightest upper bound for error probability of direct detection in the presence of constant interference that is on for the first half symbol	G.1	4.4.1	4.9
The worst case duty cycle for constant interference that is on for the first half of the symbol in direct detection	G.2	4.4.1	4.10, 4.15
The mean and variance of the photodetector output for direct detection in the presence of Gaussian interference that is on for the entire symbol	G.3	4.4.2	4.17, 4.18
The worst case duty cycle for Gaussian interference that is on for the entire symbol in direct detection	G.4	4.4.2	4.20, 4.24
The worst case duty cycle for Gaussian interference that is on for the first half of the symbol in direct detection	G.5	4.4.3	4.29, 4.33
The average photon count for canceling interference that is on for the entire symbol in direct detection	G.6	4.4.4	4.35, 4.36
The tightest upper bound for error probability of direct detection in the presence of canceling interference that is on for the entire symbol	G.7	4.4.4	4.37
The worst case duty cycle for canceling interference that is on for the entire symbol in direct detection	G.8	4.4.4	4.39
The average photon count for canceling interference that is on for the first half of the symbol in direct detection	G.9	4.4.5	4.47, 4.48
The worst case duty cycle for canceling interference that is on for the first half of the symbol in direct detection	G.10	4.4.5	4.53
The worst case duty cycle for Gaussian interference in homodyne detection	G.11	4.5.1	4.60, 4.64
The worst case duty cycle for canceling interference in homodyne detection	G.12	4.5.2	4.71, 4.76
The average photon count difference between the first and second half of the symbol in direct detection with Gaussian interference that is on for first half symbol and with Gaussian interference that is on for entire	G.13	4.7.1	N/A

Table G.1 Quantities derived in Appendix G and corresponding section numbers

298

G.1 Tightest Upper Bound for Error Probability of Direct Detection in Presence of Constant Interference that is on for First Half Symbol

In this section, we find the tightest Chernoff Bound of (4.8), namely of

$$P(e) \le \frac{\beta}{2}\min_{s\ge 0} E\left[e^{s(N_1-N_0)}\middle|\text{interference is on, H}_0\right] + \frac{\beta}{2}\min_{t\ge 0} E\left[e^{t(N_0-N_1)}\middle|\text{| interference is on, H}_1\right]$$

We introduced, in Section 4.4.1, that if the sender sends a '0', the average photon counts in the first half and second half symbol intervals are $mN_s^* + \frac{NN_I}{\beta} + \frac{NN_n}{2}$ and $\frac{NN_n}{2}$ respectively and if the sender sends a '1', the total average photon counts in the first and second half symbol intervals are $\frac{NN_I}{\beta} + \frac{NN_n}{2}$ and $mN_s^* + \frac{NN_n}{2}$ respectively.

Thus,

$$P(e) \le \frac{\beta}{2} \min_{s \ge 0} E\left[E\left[e^{s(N_1 - N_0)} \Big| \text{interference is on}, H_0, N_0\right] \Big| \text{interference is on}, H_0\right]$$

$$+ \frac{\beta}{2} \min_{t \ge 0} E\left[E\left[e^{t(N_0 - N_1)} \Big| \text{interference is on}, H_1, N_0\right] \Big| \text{interference is on}, H_1\right]$$

$$= \frac{\beta}{2} \min_{s \ge 0} E\left[e^{-sN_0} E\left[e^{sN_1} \Big| \text{interference is on}, H_0\right] \Big| \text{interference is on}, H_0\right]$$

$$+ \frac{\beta}{2} \min_{t \ge 0} E\left[e^{tN_0} E\left[e^{-tN_1} \Big| \text{interference is on}, H_1\right] \Big| \text{interference is on}, H_1\right]$$

$$= \frac{\beta}{2} \min_{s \ge 0} E\left[e^{-sN_0} \exp\left\{\frac{NN_n}{2}\left(e^s - 1\right)\right\} \Big| \text{interference is on}, H_0\right]$$

$$+ \frac{\beta}{2} \min_{t \ge 0} E\left[e^{tN_0} \exp\left\{\left(mN_S^* + \frac{NN_n}{2}\right)\left(e^{-t} - 1\right)\right\} \Big| \text{interference is on}, H_1\right]$$

$$= \frac{\beta}{2} \min_{s \ge 0} \exp\left\{\frac{NN_n}{2}\left(e^s - 1\right)\right\} E\left[e^{-sN_0} \Big| \text{interference is on}, H_0\right]$$

$$+ \frac{\beta}{2} \min_{t \ge 0} \exp\left\{\left(mN_S^* + \frac{NN_n}{2}\right)\left(e^{-t} - 1\right)\right\} E\left[e^{tN_0} \Big| \text{interference is on}, H_1\right]$$

$$= \frac{\beta}{2} \min_{s \ge 0} \exp\left\{\frac{NN_n}{2}\left(e^s - 1\right)\right\} \exp\left\{\left(mN_S^* + \frac{NN_I}{2} + \frac{NN_n}{2}\right)\left(e^{-s} - 1\right)\right\} \tag{G.1}$$

$$+ \frac{\beta}{2} \min_{t \ge 0} \exp\left\{\left(mN_S^* + \frac{NN_n}{2}\right)\left(e^{-t} - 1\right)\right\} \exp\left\{\left(\frac{NN_I}{2} + \frac{NN_n}{2}\right)\left(e^t - 1\right)\right\}$$

where in the second and fourth equality, we used the fact that for a Poisson random variable with rate parameter λ, $E\left[e^{sX}\right] = \exp\left\{\lambda\left(e^s - 1\right)\right\}$. Optimizing over s (by taking the first derivative of $\exp\left\{\left(mN_S^* + \frac{NN_I}{2} + \frac{NN_n}{2}\right)\left(e^{-s} - 1\right) + \frac{NN_n}{2}\left(e^s - 1\right)\right\}$ and setting it equal to zero), gives us

$$e^s = \sqrt{\frac{mN_S^* + \dfrac{NN_I}{2} + \dfrac{NN_n}{2}}{\dfrac{NN_n}{2}}}. \tag{G.2}$$

Optimizing over t (by taking the first derivative of

$$\exp\left\{\left(mN_S^* + \frac{NN_n}{2}\right)\left(e^{-t} - 1\right) + \left(\frac{NN_I}{2} + \frac{NN_n}{2}\right)\left(e^t - 1\right)\right\}$$ and setting it equal to zero), gives us

$$e^t = \sqrt{\frac{mN_S^* + \dfrac{NN_n}{2}}{\dfrac{NN_I}{2} + \dfrac{NN_n}{2}}} \cdot \qquad \text{(G.3)}$$

Substituting (G.2) and (G.3) into the last line of (G.1), the tightest upper bound is

$$P(e) \le \frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_I}{\beta} + \frac{NN_n}{2}} - \sqrt{\frac{NN_n}{2}}\right)^2\right\} + \frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\} \quad \text{(G.4)}$$

G.2 Worst Case Duty Cycle for Constant Interference that is on for First Half Symbol in Direct Detection

<u>Worst case duty cycle in absence of fading</u>

From (4.9), the error probability of direct detection in the presence of interference that is on for the first half of the symbol as a constant signal and has duty cycle β is

$$P(e) \cong \frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_I}{\beta} + \frac{NN_n}{2}} - \sqrt{\frac{NN_n}{2}}\right)^2\right\} + \frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\}$$

Assuming that errors which occur when the communication and interference signals are in the same half symbol are negligible i.e. the first term in the above equation is negligible, the error probability is approximately

301

$$P(e) \cong \frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\} \qquad (G.5)$$

To find the duty cycle that maximizes error probability, we take the first derivative of the error probability with respect to β and set it equal to 0,

$$0 = \frac{\partial}{\partial\beta}\left[\frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\}\right]$$

$$= \frac{\beta}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\}(-2)\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)\left(\frac{NN_I}{2\beta^2\sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}}\right)$$

$$+ \frac{1}{2}\exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2\right\}$$

$$= -\left(\sqrt{mN_S^* + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)\left(\frac{NN_I}{\beta\sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}}\right) + 1$$

$$= -\left(\sqrt{\frac{mN_S^* + \frac{NN_n}{2}}{\frac{NN_I}{\beta} + \frac{NN_n}{2}}} - 1\right)\left(\frac{NN_I}{\beta}\right) + 1$$

$$(G.6)$$

Assuming $N_n \ll \frac{N_I}{\beta}$,

$$0 = -\left(\sqrt{\frac{mN_S^* + \frac{NN_n}{2}}{\frac{NN_I}{\beta}}} - 1\right)\left(\frac{NN_I}{\beta}\right) + 1$$

$$= -\sqrt{\frac{NN_I}{\beta}\left(mN_S^* + \frac{NN_n}{2}\right)} + \frac{NN_I}{\beta} + 1 \qquad (G.7)$$

$$= -\sqrt{\beta NN_I\left(mN_S^* + \frac{NN_n}{2}\right)} + NN_I + \beta$$

302

Thus,

$$\sqrt{\beta} = \frac{\sqrt{NN_I\left(mN_S^* + \frac{NN_n}{2}\right)} \pm \sqrt{NN_I\left(mN_S^* + \frac{NN_n}{2}\right) - 4NN_I}}{2} \qquad (G.8)$$

For large mN_S^* and N_I, the only possible root that is in $(0,1]$ is the one that corresponds to using the negative sign rather than the positive sign. The other root is >1. The root that is in $(0,1]$ is

$$\sqrt{\beta} = \frac{\sqrt{NN_I\left(mN_S^* + \frac{NN_n}{2}\right)} - \sqrt{NN_I\left(mN_S^* + \frac{NN_n}{2}\right) - 4NN_I}}{2} \qquad (G.9)$$

$$\beta = \frac{1}{4}\left(\sqrt{NN_I\left(mN_S^* + \frac{NN_n}{2}\right)} - \sqrt{NN_I\left(mN_S^* + \frac{NN_n}{2}\right) - 4NN_I}\right)^2$$

$$= \frac{NN_I mN_S^*}{4}\left(\sqrt{1 + \frac{NN_n}{mN_S^*}} - \sqrt{1 + \frac{NN_n}{mN_S^*} - \frac{4}{mN_S^*}}\right)^2$$

Assuming $mN_S^* \gg NN_n - 4$, we can use the approximation that $\sqrt{1 - \delta} \approx 1 - \frac{\delta}{2}$ for small δ,

$$\beta = \frac{NN_I mN_S^*}{4}\left(1 + \frac{NN_n}{2mN_S^*} - \left(1 + \frac{NN_n - 4}{2mN_S^*}\right)\right)^2$$

$$= \frac{NN_I mN_S^*}{4}\left(\frac{2}{mN_S^*}\right)^2 \qquad (G.10)$$

$$= \frac{NN_I}{mN_S^*}$$

If $NN_I < mN_S^*$, for $\frac{NN_I}{mN_S^*} < \beta \leq 1$, the first derivative of (G.5) is negative and for

$0 < \beta < \dfrac{NN_I}{mN_S^*}$, it is positive. Thus, the worst case duty cycle is

$$\beta_{wc} = \frac{NN_I}{mN_S^*} \tag{G.11}$$

Worst case duty cycle in presence of fading

From (4.13), the outage probability of direct detection the presence of interference that is on for the first half symbol as a constant signal and has duty cycle β is

$$P_{outage} \cong \frac{\beta}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}} \right)^2 - \frac{NN_n}{2}}{mN_S^*} \right) \right]^2 \right\}$$

To find the duty cycle that maximizes the outage probability, we take the first derivative of the outage probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial \beta}\left[\frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{mN_S^*}\right)\right]^2\right\}\right]$$

$$= \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{mN_S^*}\right)\right]^2\right\}$$

$$\cdot \left(\frac{-1}{\sigma_U^2}\right)\left[m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{mN_S^*}\right)\right]$$

$$\cdot \left(\frac{-mN_S^*}{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}\right)\left(\frac{2}{mN_S^*}\right)\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)$$

$$\cdot \left(\frac{1}{2}\right)\frac{1}{\sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}}\left(-\frac{NN_I}{\beta}\right)$$

$$+ \frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{mN_S^*}\right)\right]^2\right\}$$

$$= -\frac{NN_I}{\beta\sigma_U^2}\left[m_U - \ln\left(\frac{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2 - \frac{NN_n}{2}}{mN_S^*}\right)\right]$$

$$\cdot \left(\frac{\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}}{\left(\sqrt{-\ln 2 + \theta_{thresh}} + \sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}\right)^2 - NN_n}\right)\frac{1}{\sqrt{\frac{NN_I}{\beta} + \frac{NN_n}{2}}} \tag{G.12}$$

$$+ 1$$

Assuming $\dfrac{N_I}{\beta} \gg N_n$ and $\dfrac{NN_I}{\beta} \gg -\ln 2 + \theta_{thresh}$ (the latter is reasonable since

$-\ln 2 + \theta_{thresh} = 1.6$ for an error probability threshold of 0.1 and $\dfrac{N_I}{\beta}$ is much larger than 1

in scenarios where the interference is much larger than worst case background noise

of $N_n=1$), the above equation reduces to

$$0 = \frac{NN_I}{\beta\sigma_U^2}\left[m_U - \ln\left(\frac{NN_I}{\beta m N_S^*}\right)\right]\frac{\beta}{NN_I} + 1$$

$$m_U - \ln\left(\frac{NN_I}{\beta m N_S^*}\right) = \sigma_U^2$$

$$\beta = \frac{NN_I}{m N_S^*}\exp\left(-m_U + \sigma_U^2\right) \qquad (G.13)$$

$$= \frac{NN_I}{m N_S^*}\exp\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^{3/2}$$

where in the last line, we substituted in for m_U and σ_U^2 given in (3.9). Thus, the

value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{NN_I}{m N_S^*}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^{3/2} \qquad (G.14)$$

G.3 Mean and Variance of Photodetector Output for Direct Detection in Presence of Gaussian Interference that is on for Entire Symbol

As discussed in Section 4.4.2, the output of the photodetector in the presence of a

random Gaussian field is a doubly stochastic Poisson process where the rate parameter

varies with the impinging field.

Let the random variable K denote the photodetector's photon count over the duration of half a symbol time $T/2$. For a given rate parameter of the doubly stochastic Poisson process, the distribution of K is Poisson. Let us denote this Poisson random variable's mean by λ. Since K is Poisson distributed, it has mean λ and variance λ. The expected count K is given by

$$
\begin{aligned}
E[K] &= \int_0^\infty \left(\sum_{k=0}^\infty k \Pr(K = k \mid \lambda) \right) p(\lambda) d\lambda \\
&= \int_0^\infty E[K \mid \lambda] p(\lambda) d\lambda \\
&= \int_0^\infty \lambda p(\lambda) d\lambda \\
&= E[\lambda]
\end{aligned}
\tag{G.15}
$$

In Chapter 4, when the interference is on, we denote the average received interference photons per symbol by N_I/β. Thus, the expected count over the duration of half a symbol is $N_I/(2\beta)$ (for symbols during which the interference is on). Thus,

$$
E[K] = E[\lambda] = \frac{N_I}{2\beta}
\tag{G.16}
$$

The second moment of the count K is given by

$$
\begin{aligned}
E[K^2] &= \int_0^\infty \left(\sum_{k=0}^\infty k^2 \Pr(K = k \mid \lambda) \right) p(\lambda) d\lambda \\
&= \int_0^\infty E[K^2 \mid \lambda] p(\lambda) d\lambda \\
&= \int_0^\infty (\lambda + \lambda^2) p(\lambda) d\lambda \\
&= E[\lambda] + E[\lambda^2]
\end{aligned}
\tag{G.17}
$$

where the second line is true because $E[K^2] = \sigma_K^2 + (E[K])^2 = \lambda + \lambda^2$ (for Poisson random variable K). Let us define x to be a Gaussian variable representing a scaled version of the envelope of the received interference field. Specifically, let $x = f_I \sqrt{\dfrac{\eta}{h\nu} \dfrac{T}{2}}$ where f_I is the envelope of the received interference field. Since the interference is Gaussian with 0-mean, x has 0-mean. The rate parameter λ is then given by

$$\lambda = x^2 \tag{G.18}$$

Substituting (G.18) into (G.17),

$$E[K^2] = E[\lambda] + E[x^4] \tag{G.19}$$

For a Gaussian random variable x with 0-mean, $E[x^4] = 3(E[x^2])^2$. Thus,

$$
\begin{aligned}
E[K^2] &= E[\lambda] + 3(E[x^2])^2 \\
&= E[\lambda] + 3(E[\lambda])^2 \\
&= E[K] + 3(E[K])^2
\end{aligned}
\tag{G.20}
$$

The variance of K is given by

$$
\begin{aligned}
\sigma_K^2 &= E[K^2] - (E[K])^2 \\
&= E[K] + 3(E[K])^2 - (E[K])^2 \\
&= E[K] + 2(E[K])^2 \\
&= E[\lambda] + 2(E[\lambda])^2 \\
&= \frac{N_I}{2\beta} + 2\left(\frac{N_I}{2\beta}\right)^2
\end{aligned}
\tag{G.21}
$$

308

Modeling the output of the photodetector as Gaussian, this Gaussian random variable has mean $\frac{N_I}{2\beta}$ and variance $\frac{N_I}{2\beta} + 2\left(\frac{N_I}{2\beta}\right)^2$ (from (G.16) and (G.21)).

For an N receiver system where the rate parameters vary together,

$$E[\lambda] = \frac{NN_I}{2\beta} \text{ and} \tag{G.22}$$

$$\begin{aligned} \sigma_K^2 &= E[\lambda] + 2(E[\lambda])^2 \\ &= \frac{NN_I}{2\beta} + 2\left(\frac{NN_I}{2\beta}\right)^2 \end{aligned} \tag{G.23}$$

G.4 Worst Case Duty Cycle for Gaussian Interference that is on for Entire Symbol in Direct Detection

<u>Worst case duty cycle in absence of fading</u>

From (4.19), the error probability of direct detection the presence of Gaussian interference that is on for the entire symbol and has duty cycle β is

$$P(e) \cong \frac{\beta}{2} \exp\left\{ -\frac{\left(mN_S^*\right)^2}{2\left(\frac{NN_I}{\beta} + \left(\frac{NN_I}{\beta}\right)^2\right)} \right\}$$

To find the duty cycle that maximizes the error probability, we take the first derivative of the error probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial \beta}\left[\frac{\beta}{2}\exp\left\{-\frac{\left(mN_S^*\right)^2}{2\left(\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right)}\right\}\right]$$

$$= \frac{\beta}{2}\exp\left\{-\frac{\left(mN_S^*\right)^2}{2\left(\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right)}\right\}\frac{\left(mN_S^*\right)^2\left(-\frac{NN_I}{\beta^2}-\frac{2\left(NN_I\right)^2}{\beta^3}\right)}{2\left(\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right)^2}$$

$$+\frac{1}{2}\exp\left\{-\frac{\left(mN_S^*\right)^2}{2\left(\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right)}\right\}$$

$$= -\frac{\left(mN_S^*\right)^2}{2}\frac{\frac{NN_I}{\beta}\left(1+\frac{2NN_I}{\beta}\right)}{\left(\frac{NN_I}{\beta}+\left(\frac{NN_I}{\beta}\right)^2\right)^2}+1 \tag{G.24}$$

$$= -\frac{\left(mN_S^*\right)^2}{2}\frac{\left(1+\frac{2NN_I}{\beta}\right)}{\frac{NN_I}{\beta}\left(1+\frac{NN_I}{\beta}\right)^2}+1$$

Assuming $\frac{NN_I}{\beta} \gg 1$,

$$0 = -\frac{\left(mN_S^*\right)^2}{2}\frac{\left(\frac{2NN_I}{\beta}\right)}{\frac{NN_I}{\beta}\left(\frac{NN_I}{\beta}\right)^2}+1$$

$$= -\frac{\left(mN_S^*\right)^2}{\left(\frac{NN_I}{\beta}\right)^2}+1 \tag{G.25}$$

$$\beta = \frac{NN_I}{mN_S^*}$$

Thus, the value of β that maximizes the error probability is

$$\beta_{wc} = \frac{NN_I}{mN_S^*} \qquad\qquad\qquad\qquad (G.26)$$

Worst case duty cycle in presence of fading

From (4.23), the outage probability of direct detection the presence of Gaussian interference that is on for the entire symbol and has duty cycle β is

$$P_{outage} \cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\sqrt{2\left(\frac{NN_I}{\beta} + \left(\frac{NN_I}{\beta}\right)^2\right)(-\ln 2 + \theta_{thresh})}}{mN_S^*}\right)\right]^2\right\}$$

Assuming $\dfrac{NN_I}{\beta} << \left(\dfrac{NN_I}{\beta}\right)^2$,

$$P_{outage} \cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\left(\frac{NN_I}{\beta mN_S^*}\right)\sqrt{2(-\ln 2 + \theta_{thresh})}\right)\right]^2\right\} \qquad (G.27)$$

To find the duty cycle that maximizes the outage probability, we take the first derivative of the outage probability with respect to β and set it equal to 0.

311

$$0 = \frac{\partial}{\partial \beta} \left[\frac{\beta}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\left(\frac{NN_I}{\beta m N_S^*} \right) \sqrt{2(-\ln 2 + \theta_{thresh})} \right) \right]^2 \right\} \right]$$

$$= \frac{\beta}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\left(\frac{NN_I}{\beta m N_S^*} \right) \sqrt{2(-\ln 2 + \theta_{thresh})} \right) \right]^2 \right\} \left\{ \left(\frac{-1}{\sigma_U^2} \right) \left[m_U - \ln\left(\left(\frac{NN_I}{\beta m N_S^*} \right) \sqrt{2(-\ln 2 + \theta_{thresh})} \right) \right] \right.$$

$$\cdot \left(\frac{\beta m N_S^*}{NN_I \sqrt{2(-\ln 2 + \theta_{thresh})}} \right) \left(\frac{NN_I \sqrt{2(-\ln 2 + \theta_{thresh})}}{\beta^2 m N_S^*} \right)$$

$$+ \frac{1}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\left(\frac{NN_I}{\beta m N_S^*} \right) \sqrt{2(-\ln 2 + \theta_{thresh})} \right) \right]^2 \right\}$$

$$= \frac{-1}{\sigma_U^2} \left[m_U - \ln\left(\left(\frac{NN_I}{\beta m N_S^*} \right) \sqrt{2(-\ln 2 + \theta_{thresh})} \right) \right] + 1$$

(G.28)

Rearranging, we get

$$\ln\left(\left(\frac{NN_I}{\beta m N_S^*} \right) \sqrt{2(-\ln 2 + \theta_{thresh})} \right) = m_U - \sigma_U^2$$

(G.29)

$$\beta = \frac{NN_I}{m N_S^*} \sqrt{2(-\ln 2 + \theta_{thresh})} \left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N} \right)^{3/2}$$

Thus, the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{NN_I}{m N_S^*} \sqrt{2(-\ln 2 + \theta_{thresh})} \left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N} \right)^{3/2}$$

(G.30)

G.5 Worst Case Duty Cycle for Gaussian Interference that is on for First Half Symbol in Direct Detection

<u>Worst case duty cycle in absence of fading</u>

From (4.28), the error probability of direct detection the presence of Gaussian interference that is on for the first half symbol and has duty cycle β is

$$P(e) \cong \frac{\beta}{2} \exp\left\{-\frac{\left(mN_s^* - \frac{NN_I}{\beta}\right)^2}{2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)}\right\}$$

To find the duty cycle that maximizes the error probability, we take the first derivative of the error probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial \beta}\left[\frac{\beta}{2}\exp\left\{-\frac{\left(mN_S^* - \frac{NN_I}{\beta}\right)^2}{2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)}\right\}\right]$$

$$= \frac{\beta}{2}\exp\left\{-\frac{\left(mN_S^* - \frac{NN_I}{\beta}\right)^2}{2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)}\right\}$$

$$\cdot\left[\frac{-2\left(mN_S^* - \frac{NN_I}{\beta}\right)^2\left[\frac{NN_I}{\beta^2} + 4\frac{(NN_I)^2}{\beta^{3/2}}\right] + 4\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)\left(mN_S^* - \frac{NN_I}{\beta}\right)\frac{NN_I}{\beta^2}}{4\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)^2}\right]$$

$$+ \frac{1}{2}\exp\left\{-\frac{\left(mN_S^* - \frac{NN_I}{\beta}\right)^2}{2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)}\right\}$$

$$= 2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)^2 - \left(mN_S^* - \frac{NN_I}{\beta}\right)^2\left(\frac{NN_I}{\beta} + 4\left(\frac{NN_I}{\beta}\right)^2\right) - 2\left(\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2\right)\left(mN_S^* - \frac{NN_I}{\beta}\right)\frac{NN_I}{\beta}$$

$$(G.31)$$

Assuming $\frac{NN_I}{\beta} + 2\left(\frac{NN_I}{\beta}\right)^2$,

$$0 = 8\left(\frac{NN_I}{\beta}\right)^4 - 4\left(\frac{NN_I}{\beta}\right)^2\left(mN_S^* - \frac{NN_I}{\beta}\right)^2 - 4\left(\frac{NN_I}{\beta}\right)^3\left(mN_S^* - \frac{NN_I}{\beta}\right)$$

$$= 2\left(\frac{NN_I}{\beta}\right)^2 - \left(mN_S^*\right)^2 + mN_S^*\frac{NN_I}{\beta} \qquad (G.32)$$

$$= \beta^2 - \frac{NN_I}{mN_S^*}\beta - 2\left(\frac{NN_I}{\beta}\right)^2$$

Thus,

$$\beta = \frac{\frac{NN_I}{mN_S^*} \pm \sqrt{\left(\frac{NN_I}{mN_S^*}\right)^2 + 8\left(\frac{NN_I}{mN_S^*}\right)^2}}{2}$$

$$= \frac{1}{2}\left(\frac{NN_I}{mN_S^*} \pm 3\frac{NN_I}{mN_S^*}\right) \qquad (G.33)$$

The root that corresponds to using the negative sign rather than the positive sign is negative. The other root is

$$\beta = \frac{2NN_I}{mN_S^*} \qquad (G.34)$$

If $2NN_I < mN_S^*$, for $\frac{2NN_I}{mN_S^*} < \beta \leq 1$, the first derivative of (G.33) is negative and for $0 < \beta < \frac{2NN_I}{mN_S^*}$, it is positive. Thus, the value of β that maximizes the error probability is

$$\beta_{wc} = \frac{2NN_I}{mN_S^*} \qquad (4.35)$$

Worst case duty cycle in presence of fading

From (4.32), the outage probability of direct detection the presence of Gaussian interference that is on for the first half symbol and has duty cycle β is

$$P_{outage} \cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\frac{\sqrt{2\left(\frac{NN_I}{\beta} + 2\left[\frac{NN_I}{\beta}\right]^2\right)\left(-\ln 2 + \theta_{thresh}\right)} + \frac{NN_I}{\beta}}{mN_S^*}\right)\right]^2\right\}$$

315

Assuming $\dfrac{NN_I}{\beta} << \left(\dfrac{NN_I}{\beta}\right)^2$

$$P_{outage} \cong \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\left[\frac{NN_I}{\beta m N_S^*}\right]\left(2\sqrt{(-\ln 2 + \theta_{thresh})} + 1\right)\right)\right]^2\right\} \tag{G.36}$$

To find the duty cycle that maximizes the outage probability, we take the first derivative of the outage probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial \beta}\left[\frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left(\left[\frac{NN_I}{\beta m N_S^*}\right]\left(2\sqrt{(-\ln 2 + \theta_{thresh})} + 1\right)\right)\right]^2\right\}\right] \tag{G.37}$$

After straightforward algebra, this reduces to

$$\beta = \frac{NN_I}{m N_S^*}\left(2\sqrt{-\ln(2) + \theta_{thresh}} + 1\right)\exp\left(\sigma_U^2 - m_U\right) \tag{G.38}$$

Substituting in for m_U and σ_U^2 given in (3.9), this becomes

$$\beta = \frac{NN_I}{m N_S^*}\left(2\sqrt{-\ln(2) + \theta_{thresh}} + 1\right)\left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right)^{3/2} \tag{G.39}$$

Thus, the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{NN_I}{m N_S^*}\left(2\sqrt{-\ln 2 + \theta_{thresh}} + 1\right)\left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right)^{3/2} \tag{G.40}$$

316

G.6 Average Photon Count for Canceling Interference that is on for Entire Symbol in Direct Detection

If the sender sends a '0' in the presence of canceling interference, then in the first half symbol, the average number of received photons by receiver i, $1 \le i \le N$ is

$$
\begin{aligned}
N_{0,i} &= \frac{\eta}{h\upsilon} \frac{A}{N} |U_S - U_I|^2 + \frac{N_n}{2} \\
&= \frac{\eta}{h\upsilon} \frac{A}{N} \left\{ |U_S|^2 + |U_I|^2 - 2\,\mathrm{Re}[U_S U_I] \right\} + \frac{N_n}{2} \\
&= \frac{\eta}{h\upsilon} \frac{A}{N} \left\{ U_S^2 + U_I^2 - 2 U_S U_I \right\} + \frac{N_n}{2} \\
&= \frac{mN_S^*}{N} + \frac{N_I}{2\beta} - 2\sqrt{\frac{mN_S^* N_I}{2\beta N}} + \frac{N_n}{2}
\end{aligned}
\tag{G.41}
$$

where U_S and U_I are the envelopes of the received signal and interference field, and where in the second last line U_S and U_I are taken to be real. The total average photon count is the sum of the average counts from each receiver. Thus,

$$
\begin{aligned}
N_0 &= N \left(\frac{mN_S^*}{N} + \frac{N_I}{2\beta} - 2\sqrt{\frac{mN_S^* N_I}{2\beta N}} + \frac{N_n}{2} \right) \\
&= mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2NN_I mN_S^*}{\beta}} + \frac{N_n}{2}
\end{aligned}
\tag{G.42}
$$

The average received photon count in the second half symbol by receiver i is

$$
N_{1,i} = \frac{N_I}{2\beta} + \frac{N_n}{2}
\tag{G.43}
$$

Thus the total average received photon count in the second half symbol is

$$N_1 = \frac{NN_I}{2\beta} + \frac{NN_n}{2} \qquad\qquad\text{(G.44)}$$

G.7 Tightest Upper Bound for Error Probability of Direct Detection in Presence of Canceling Interference that is on for Entire Symbol

In this section, we find the tightest Chernoff Bound of (4.8), namely of

$$P(e) \le \min_{s \ge 0} \beta E\left[e^{s(N_1 - N_0)} \mid \text{interferer is on}, H_0\right]$$

We introduced, in Section 4.4.4, that if the sender sends a '0', the average photon counts in the first half and second half symbol intervals are

$$N_0 = \left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{2\beta}}\right)^2 + \frac{NN_n}{2} = mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2mN_S^* NN_I}{\beta}} + \frac{NN_n}{2} \quad \text{and} \quad N_1 = \frac{NN_I}{2\beta} + \frac{NN_n}{2}$$

respectively and if the sender sends a '1', the total average photon counts in the first and second half symbol intervals is reversed.

$$
\begin{aligned}
P(e) &\le \beta \min_{s \ge 0} E\left[E\left[e^{s(N_1 - N_0)} \middle| \text{interference is on}, H_0, N_0\right] \middle| \text{interference is on}, H_0\right] \\
&= \beta \min_{s \ge 0} E\left[e^{-sN_0} E\left[e^{sN_1} \middle| \text{interference is on}, H_0\right] \middle| \text{interference is on}, H_0\right] \\
&= \beta \min_{s \ge 0} E\left[e^{-sN_0} \exp\left\{\left(\frac{NN_I}{2\beta} + \frac{NN_n}{2}\right)(e^s - 1)\right\} \middle| \text{interference is on}, H_0\right] \qquad\text{(G.45)}\\
&= \beta \min_{s \ge 0} \exp\left\{\left(\frac{NN_I}{2\beta} + \frac{NN_n}{2}\right)(e^s - 1)\right\} E\left[e^{-sN_0} \middle| \text{interference is on}, H_0\right] \\
&= \beta \min_{s \ge 0} \exp\left\{\left(\frac{NN_I}{2\beta} + \frac{NN_n}{2}\right)(e^s - 1)\right\} \exp\left\{\left(mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2mN_S^* NN_I}{\beta}} + \frac{NN_n}{2}\right)(e^{-s} - 1)\right\}
\end{aligned}
$$

where in the third and fifth equality, we used the fact that for a Poisson random variable with rate parameter λ, $E[e^{sX}] = \exp\{\lambda(e^s - 1)\}$. Optimizing over s (by taking the first derivative of

$$\exp\left\{\left(\frac{NN_I}{2\beta} + \frac{NN_n}{2}\right)(e^s - 1)\right\} \exp\left\{\left(mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2mN_S^* NN_I}{\beta}} + \frac{NN_n}{2}\right)(e^{-s} - 1)\right\}$$

and setting it equal to zero), gives us

$$e^s = \sqrt{\frac{mN_S^* + \dfrac{NN_I}{2\beta} - \sqrt{\dfrac{2mN_S^* NN_I}{\beta}} + \dfrac{NN_n}{2}}{\dfrac{NN_I}{2\beta} + \dfrac{NN_n}{2}}}. \tag{G.46}$$

Substituting (G.46) into the last line of (G.45), the tightest upper bound is

$$P(e) \le \beta \exp\left\{-\left(\sqrt{mN_S^* + \frac{NN_I}{2\beta} - \sqrt{\frac{2NN_I mN_S^*}{\beta}} + \frac{NN_n}{2}} - \sqrt{\frac{NN_I}{2\beta} + \frac{NN_n}{2}}\right)^2\right\}$$

G.8 Worst Case Duty Cycle for Canceling Interference that is on for Entire Symbol in Direct Detection

Worst Case Duty Cycle in Absence of Fading

From (4.38), the error probability of direct detection the presence of canceling interference that is on for the entire symbol and has duty cycle β is

$$P(e) \cong \beta \exp\left\{-\left(\sqrt{mN_S^*} - \sqrt{\frac{2NN_I}{\beta}}\right)^2\right\}$$

Taking the derivative with respect to β and setting equal to 0, we get

$$
\begin{aligned}
0 &= \beta \exp\left\{-\left(\sqrt{mN_S^*}-\sqrt{\frac{2NN_I}{\beta}}\right)^2\right\}(-2)\left(\sqrt{mN_S^*}-\sqrt{\frac{2NN_I}{\beta}}\right)\left(\frac{\sqrt{2NN_I}}{2\beta^{3/2}}\right)+\exp\left\{-\left(\sqrt{mN_S^*}-\sqrt{\frac{2NN_I}{\beta}}\right)^2\right\} \\
&= -\left(\sqrt{mN_S^*}-\sqrt{\frac{2NN_I}{\beta}}\right)\sqrt{\frac{2NN_I}{\beta}}+1 \\
1 &= \sqrt{\frac{2NN_I}{\beta}}\left(\sqrt{mN_S^*}-\sqrt{\frac{2NN_I}{\beta}}\right) \\
\beta &= \sqrt{2NN_I}\left(\sqrt{\beta mN_S^*}-\sqrt{2NN_I}\right) \\
0 &= \beta - \sqrt{2NN_I mN_S^*\beta}+2NN_I
\end{aligned}
$$

(G.47)

Thus,

$$
\begin{aligned}
\sqrt{\beta} &= \frac{\sqrt{2NN_I mN_S^*}\pm\sqrt{2NN_I mN_S^*-8NN_I}}{2} \\
\beta &= \frac{1}{4}\left(\sqrt{2NN_I mN_S^*}\pm\sqrt{2NN_I mN_S^*-8NN_I}\right)^2 \\
&= \frac{1}{4}\left(\sqrt{2NN_I mN_S^*}\pm\sqrt{2NN_I mN_S^*}\sqrt{1-\frac{4}{mN_S^*}}\right)^2 \\
&= \frac{2NN_I mN_S^*}{4}\left(1\pm\sqrt{1-\frac{4}{mN_S^*}}\right)^2 \\
&\cong \frac{NN_I mN_S^*}{2}\left(1\pm\left(1-\frac{2}{mN_S^*}\right)\right)^2
\end{aligned}
$$

(G.48)

where the last line is true if $mN_S^* \gg 4$. For $mN_S^* \gg 1$, the only possible root that is in $(0,1]$ is the one that corresponds to using the negative sign rather than the positive sign. The other root is >1. The root that is in $(0,1]$ is

$$
\beta = \frac{NN_I}{mN_S^*}
$$

(G.49)

320

If $NN_I < mN_S^*$, for $\dfrac{NN_I}{mN_S^*} < \beta \le 1$, the first derivative of the error probability (4.38) (and re-stated at the start of this section) is negative and for $0 < \beta < \dfrac{NN_I}{mN_S^*}$, it is positive.

Thus, the worst case interference duty cycle is approximately

$$\beta_{wc} = \frac{NN_I}{mN_S^*} \tag{G.50}$$

G.9 Average Photon Count for Canceling Interference that is on for First Half Symbol in Direct Detection

Average Photon Counts in First and Second Half Symbols

If the sender sends a '0', the average photon counts in the first half and second half of the symbol is the same as in the presence of the presence of the canceling interferer that is on for the entire symbol except that $NN_I /(2\beta)$ is replaced with NN_I / β.

If the sender sends a '0', the communication signal is in the first half symbol and the interference is in the second half symbol. Thus, the average received photon count per receiver in the first and second half symbol intervals is

$$N_{0,i} = \frac{mN_S^*}{N} + \frac{N_n}{2} \text{ and}$$

$$N_{1,i} = \frac{N_I}{\beta} + \frac{N_n}{2} \tag{G.51}$$

and the total average received photon count in the first and half symbol intervals is

321

$$N_0 = NN_{0,i} = N\left(\frac{mN_S^*}{N} + \frac{N_n}{2}\right) = mN_S^* + \frac{NN_n}{2}$$

$$N_1 = N\left(\frac{N_I}{\beta} + \frac{N_n}{2}\right) = \frac{NN_I}{\beta} + \frac{NN_n}{2} \tag{G.52}$$

G.10 Worst Case Duty Cycle for Canceling Interference that is on for First Half Symbol in Direct Detection

Worst Case Duty Cycle in Absence of Fading

From (4.52), the error probability of direct detection the presence of canceling interference that is on for the first half symbol and has duty cycle β is

$$P(e) \cong \beta \exp\left\{-\left(\sqrt{mN_S^*} - \sqrt{\frac{NN_I}{\beta}}\right)^2\right\}$$

This is the same as the error probability of diversity detection in the presence of canceling interference that is on for the entire symbol except $2NN_I/\beta$ is replaced with NN_I/β. The worst case duty cycle is calculated with the same equations except where $2NN_I/\beta$ is replaced with NN_I/β. Thus, the worst case duty cycle is approximately

$$\beta_{wc} = \frac{NN_I}{mN_S^*} \tag{G.53}$$

322

G.11 Worst Case Duty Cycle for Gaussian Interference in Homodyne Detection

Worst Case Duty Cycle in Absence of Fading

From (4.59), the error probability of homodyne detection in the presence of Gaussian interference that has duty cycle β is

$$P(e) \cong \frac{\beta}{2} \exp\left\{ -\frac{2mN_s^*}{1+\frac{N_I}{\beta}} \right\}$$

To find the duty cycle that maximizes the error probability, we take the first derivative of the error probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial \beta}\left[\frac{\beta}{2}\exp\left\{ -\frac{2mN_s^*}{1+\frac{N_I}{\beta}} \right\} \right]$$

$$= \frac{\beta}{2}\exp\left\{ -\frac{2mN_s^*}{1+\frac{N_I}{\beta}} \right\} \frac{2mN_s^*}{\left(1+\frac{N_I}{\beta}\right)^2}\left(-\frac{N_I}{\beta^2} \right) + \frac{1}{2}\exp\left\{ -\frac{2mN_s^*}{1+\frac{N_I}{\beta}} \right\} \qquad \text{(G.54)}$$

$$= \frac{-2mN_s^* N_I}{\beta\left(1+\frac{N_I}{\beta}\right)^2} + 1$$

Assuming $\frac{N_I}{\beta} \gg 1$, this reduces to

$$\beta = \frac{N_I}{2mN_s^*} \qquad \text{(G.55)}$$

Thus, the value of β that maximizes the error probability is

$$\beta_{wc} = \frac{N_I}{2mN_S^*} \tag{G.56}$$

Worst Case Duty Cycle in Presence of Fading

From (4.63), the outage probability of homodyne detection in the presence of Gaussian interference that has duty cycle β is

$$P_{outage} \cong \frac{\beta}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\frac{\left(1 + \frac{N_I}{\beta}\right)(-\ln 2 + \theta_{thresh})}{2mN_S^*} \right) \right]^2 \right\}$$

To find the duty cycle that maximizes the outage probability, we take the first derivative of the outage probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial \beta} \left[\frac{\beta}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\frac{\left(1+\frac{N_I}{\beta}\right)(-\ln 2 + \theta_{thresh})}{2mN_S^*} \right) \right]^2 \right\} \right]$$

$$= \frac{\beta}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\frac{\left(1+\frac{N_I}{\beta}\right)(-\ln 2 + \theta_{thresh})}{2mN_S^*} \right) \right]^2 \right\}$$

$$\cdot \left(\frac{1}{\sigma_U^2} \right) \left[m_U - \ln\left(\frac{\left(1+\frac{N_I}{\beta}\right)(-\ln 2 + \theta_{thresh})}{2mN_S^*} \right) \right] \frac{2mN_S^*}{\left(1+\frac{N_I}{\beta}\right)(-\ln 2 + \theta_{thresh})} \frac{(-N_I)(-\ln 2 + \theta_{thresh})}{2mN_S^* \beta^2}$$

$$+ \frac{1}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left(\frac{\left(1+\frac{N_I}{\beta}\right)(-\ln 2 + \theta_{thresh})}{2mN_S^*} \right) \right]^2 \right\}$$

$$= -\frac{N_I}{\beta \sigma_U^2 \left(1+\frac{N_I}{\beta}\right)} \left[m_U - \ln\left(\frac{\left(1+\frac{N_I}{\beta}\right)(-\ln 2 + \theta_{thresh})}{2mN_S^*} \right) \right] + 1$$

(G.57)

Assuming $\frac{N_I}{\beta} \gg 1$, this reduces to

$$0 = -\frac{1}{\sigma_U^2} \left[m_U - \ln\left(\frac{N_I(-\ln 2 + \theta_{thresh})}{2mN_S^* \beta} \right) \right] + 1$$

(G.58)

Substituting in for m_U and σ_U^2 given in (3.9), this reduces to

$$\beta = \frac{N_I(-\ln 2 + \theta_{thresh})}{2mN_S^*} \left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N} \right)^{3/2}$$

(G.59)

Thus, the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{N_I \left(-\ln 2 + \theta_{thresh}\right)}{2mN_S^*} \left(1 + \frac{e^{4o_\chi^2} - 1}{N}\right)^{3/2} \tag{G.60}$$

G.12 Worst Case Duty Cycle for Canceling Interference in Homodyne Detection

<u>Worst Case Duty Cycle in Absence of Fading</u>

From (4.70), the error probability of homodyne detection in the presence of canceling interference that has duty cycle β is

$$P(e) \cong \frac{\beta}{4}\left[\exp\left\{\frac{-2\left(\sqrt{mN_S^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1 + N_n}\right\} + \exp\left\{\frac{-2\left(\sqrt{mN_S^*} + \sqrt{\frac{N_I}{\beta}}\right)^2}{1 + N_n}\right\}\right]$$

Assuming that errors when the communication and interference signals add are negligible,

$$P(e) \cong \frac{\beta}{4}\exp\left\{\frac{-2\left(\sqrt{mN_S^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1 + N_n}\right\} \tag{G.61}$$

To find the duty cycle that maximizes the error probability, we take the first derivative of the error probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial \beta}\left[\frac{\beta}{4}\exp\left\{\frac{-2\left(\sqrt{mN_s^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1 + N_n}\right\}\right]$$

$$= \frac{\beta}{4}\exp\left\{\frac{-2\left(\sqrt{mN_s^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1 + N_n}\right\}\left(\frac{-4}{1 + N_n}\right)\left(\sqrt{mN_s^*} - \sqrt{\frac{N_I}{\beta}}\right)\frac{\sqrt{N_I}}{2\beta^{3/2}} + \frac{1}{4}\exp\left\{\frac{-2\left(\sqrt{mN_s^*} - \sqrt{\frac{N_I}{\beta}}\right)^2}{1 + N_n}\right\} \quad \text{(G.62)}$$

$$= -2\left(\sqrt{mN_s^*} - \sqrt{\frac{N_I}{\beta}}\right)\sqrt{\frac{N_I}{\beta}} + 1$$

$$= (1 + N_n)\beta - 2\sqrt{mN_s^* N_I \beta} + 2N_I$$

From the above,

$$\sqrt{\beta} = \frac{2\sqrt{mN_s^* N_I} \pm \sqrt{4mN_s^* N_I - 4(1 + N_n)2N_I}}{2(1 + N_n)}$$

$$\beta = \frac{\left(\sqrt{mN_s^* N_I} \pm \sqrt{mN_s^* N_I}\sqrt{1 - \frac{2(1 + N_n)}{mN_s^*}}\right)^2}{(1 + N_n)^2} \quad \text{(G.63)}$$

$$= \frac{\left(\sqrt{mN_s^* N_I} \pm \sqrt{mN_s^* N_I}\left(1 - \frac{(1 + N_n)}{mN_s^*}\right)\right)^2}{(1 + N_n)^2}$$

where in the last line we assumed $mN_s^* \gg (1 + N_n)$. For large mN_s^* and N_I, the only possible root in $(0,1]$ is the one that corresponds to using the negative sign rather than the positive sign. This root is

$$\beta = \frac{\left(\sqrt{mN_s^* N_I}\frac{(1 + N_n)}{mN_s^*}\right)^2}{(1 + N_n)^2} = \frac{N_I}{mN_s^*} \quad \text{(G.64)}$$

If $N_I < mN_S^*$, for $\dfrac{N_I}{mN_S^*} < \beta \le 1$, the first derivative of the error probability in (4.70) (which is re-stated at the start of this section), is negative and for $0 < \beta < \dfrac{N_I}{mN_S^*}$, it is positive. Thus, the value of β that maximizes the error probability is

$$\beta_{wc} = \frac{N_I}{mN_S^*} \tag{G.65}$$

Worst Case Duty Cycle in Presence of Fading

From (4.74), the outage probability of homodyne detection in the presence of canceling interference that has duty cycle β is

$$P_{outage} \cong \frac{\beta}{2} \exp\left\{ -\frac{1}{2\sigma_U^2} \left[m_U - \ln\left\{ \frac{1}{mN_S^*} \left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}} \right)^2 \right\} \right]^2 \right\}$$

To find the duty cycle that maximizes the outage probability, we take the first derivative of the outage probability with respect to β and set it equal to 0.

$$0 = \frac{\partial}{\partial\beta}\left[\frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left\{\frac{1}{mN_S^*}\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^2\right\}\right]^2\right\}\right]$$

$$= \frac{\beta}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left\{\frac{1}{mN_S^*}\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^2\right\}\right]^2\right\}$$

$$\cdot\left(\frac{-1}{\sigma_U^2}\right)\left[m_U - \ln\left\{\frac{1}{mN_S^*}\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^2\right\}\right]$$

$$\cdot\left(\frac{-\sqrt{N_I}}{2\beta^{3/2}}\right)\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^{-1}$$

$$+ \frac{1}{2}\exp\left\{-\frac{1}{2\sigma_U^2}\left[m_U - \ln\left\{\frac{1}{mN_S^*}\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^2\right\}\right]^2\right\}$$

$$= -\frac{1}{\sigma_U^2}\sqrt{\frac{N_I}{\beta}}\left[m_U - \ln\left\{\frac{1}{mN_S^*}\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^2\right\}\right]\left(\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} + \sqrt{\frac{N_I}{\beta}}\right)^{-1} + 1$$

(G.66)

Assuming $\sqrt{\frac{(1+N_n)(-\ln 4 + \theta_{thresh})}{2}} \ll \sqrt{\frac{N_I}{\beta}}$, this reduces to

$$0 = -\frac{1}{\sigma_U^2}\left[m_U - \ln\left(\frac{N_I}{mN_S^*\beta}\right)\right] + 1$$

$$\beta = \frac{N_I}{mN_S^*}\exp\left(-m_U + \sigma_U^2\right) \qquad\qquad \text{(G.67)}$$

$$= \frac{N_I}{mN_S^*}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^{3/2}$$

Thus, the value of β that maximizes the outage probability is

$$\beta_{wc} = \frac{N_I}{mN_S^*}\left(1 + \frac{e^{4\sigma_\chi^2} - 1}{N}\right)^{3/2} \qquad\qquad \text{(G.68)}$$

G.13 Average Photon Count Difference Between First and Second Half of Symbol in Direct Detection with Gaussian Interference that is on for First Half Symbol and with Gaussian Interference that is on for Entire

First, let us assume background noise is negligible compared to the interference.

When the Gaussian interference that is on for the first half symbol uses the worst case duty cycle, and the sender sends a '1', the average count in the first half symbol during which the interference is on is

$$\frac{NN_I}{\beta_{wc}} = \frac{mN_S^*}{2} \tag{G.69}$$

where β_{wc} represents this interference type's worst case duty cycle (4.29). The average photon count in the second half symbol is mN_S^*. Thus, the difference in the average count of the two half symbols intervals is $mN_S^*/2$.

For the Gaussian interference that is on for the entire symbol and uses the worst case duty cycle, the average count due to interference in any half symbol during which the interference is on is

$$\frac{NN_I}{2\beta_{wc}} = \frac{mN_S^*}{2} \tag{G.70}$$

where β_{wc} represents this interference type's worst case duty cycle (4.20). Thus, the average photon count in the half symbol without the communication signal is $mN_S^*/2$

and in the other half symbol is $\dfrac{mN_S^*}{2} + mN_S^* = \dfrac{3mN_S^*}{2}$. Thus, the difference in average

count of the two half symbol intervals is mN_S^*.

If background noise is not negligible compared to the interference, then for each half

symbol intervals, the average photon count would increase by $\dfrac{NN_n}{2}$ and the average

photon count difference between the two half symbols would not change.

Appendix H

Solution of Steady State Probability Distribution for Modified TCP Exponential Window Increase Markov Chain

In this appendix, we derive the steady state probability distribution of the Modified TCP exponential increase Markov chain of Figure 6.5. The detailed balance equations of the Markov Chain of Figure 6.5 (which models the window progression of the Modified TCP assuming exponential window increase) are given by (6.26), namely

$$\pi_{i-1}(1 - p_{c,i-1}) = \pi_i p_{c,i} \quad \text{for i=2,3,...,n}_{max}$$

Thus,

$$
\begin{aligned}
\pi_i &= \pi_{i-1}\left(\frac{1 - p_{c,i-1}}{p_{c,i}}\right) \\
&= \pi_{i-2}\left(\frac{1 - p_{c,i-2}}{p_{c,i-1}}\right)\left(\frac{1 - p_{c,i-1}}{p_{c,i}}\right) \\
&\quad \cdots \\
&= \pi_1 \prod_{j=1}^{i-1}\left(\frac{1 - p_{c,j}}{p_{c,j+1}}\right)
\end{aligned}
\tag{H.1}
$$

Since the sum of the steady state probabilities equals 1,

$$
\begin{aligned}
1 &= \sum_{i=1}^{n_{\max}} \pi_i \\
&= \pi_1 \left(1 + \sum_{i=1}^{n_{\max}} \prod_{j=1}^{i-1} \left(\frac{1 - p_{c,j}}{p_{c,j+1}} \right) \right) \\
\pi_1 &= \left(1 + \sum_{i=1}^{n_{\max}} \prod_{j=1}^{i-1} \left(\frac{1 - p_{c,j}}{p_{c,j+1}} \right) \right)^{-1}
\end{aligned}
\tag{H.2}
$$

Thus, the closed form solution for the steady state probability distribution of the Markov chain of Figure 6.5 is given by

$$
\pi_1 = \left(1 + \sum_{i=1}^{n_{\max}} \prod_{j=1}^{i-1} \left(\frac{1 - p_{c,j}}{p_{c,j+1}} \right) \right)^{-1} \quad \text{and}
\tag{H.3}
$$

$$
\pi_i = \pi_1 \prod_{j=1}^{i-1} \left(\frac{1 - p_{c,j}}{p_{c,j+1}} \right)
$$

Appendix I

Derivation of Number of Round-trip Times to Send ($1/p_{congperpkt}$) packets if Window Increase is Linear and Starts at Value n

In this appendix, we show that if a TCP sender's window starts at size n packets where $\frac{n^2}{2} = \frac{1}{p_{congperpkt}}$, and the window increases linearly, then the window will increase to a size of approximately $(\sqrt{2}-1)n$ by the time $1/p_{congperpkt}$ packets have been sent. This appendix is referenced in Section 6.5.2.

Consider a window size that starts at n packets where $\frac{n^2}{2} = \frac{1}{p_{congperpkt}}$ and let the window size increase by one packet every round-trip time. Let m denote the number of round-trip times until $1/p_{congperpkt}$ packets have been sent. Then

335

$$\frac{1}{P_{congperpkt}} = (n+1) + (n+2) + \ldots (n+m)$$

$$= 1 + 2 + \ldots + m + nm$$

$$= \frac{m(m-1)}{2} + mn \qquad\qquad (I.3)$$

$$\frac{2}{P_{congperpkt}} = m(m-1) + 2mn$$

$$0 = m^2 + m(2n-1) - \frac{2}{P_{congperpkt}}$$

Thus,

$$m = \frac{-(2n-1) \pm \sqrt{(2n-1)^2 + \dfrac{8}{P_{congperpkt}}}}{2}$$

$$= \frac{-(2n-1) \pm \sqrt{(2n-1)^2 + 4n^2}}{2}$$

$$= -n + \frac{1}{2} \pm \frac{\sqrt{4n^2 - 4n + 1 + 4n^2}}{2}$$

$$= -n + \frac{1}{2} \pm \frac{\sqrt{8n^2 - 4n + 1}}{2} \qquad\qquad (I.3)$$

$$\cong -n + \frac{1}{2} \pm \sqrt{2}n$$

$$\cong n(\sqrt{2} - 1)$$

where in the second last line, we used the fact than for very large n, $8n^2 - 4n + 1 \cong 8n^2$, and in the last line we assumed that the term of 1/2 in the second last line is negligible compared to $n(\sqrt{2} - 1)$.

336

Bibliography

[1]   M. Allman, "TCP Congestion Control", *Internet Engineering Task Force*, RFC 2581, April 1999. [Online]. Available: http://www.ietf.org/rfc/rfc2581.txt.

[2]   L. C. Andrews and R. L. Phillips, *Laser Beam Propagation through Random Media*. Washington: SPIE Optical Engineering Press, 1998.

[3]   H. Balakrishnan et al., "A Comparison of Mechanisms for Improving TCP Performance over Wireless Links," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 756-769, December 1997.

[4]   D. Bertsekas and R. Gallager, *Data Networks*. New Jersey: Prentice Hall Inc., 1992.

[5]   S. Biaz and X. Wang, "RED for Improving TCP over Wireless Networks," *International Conference on Wireless Networks*, Las Vegas, NV, pp. 628-633, June 2004.

[6]   C. Caini and R. Firrincieli, "TCP Hybla: A TCP enhancement for heterogeneous networks", *International Journal of Satellite Communications and Networking*, vol. 22, no. 5, pp. 547-566, August 2004.

[7]   V. W. S. Chan, "Optical Space Communications,", *IEEE Journal on Selected Topics in Quantum Electronics*, vol. 6, no. 6, pp. 959-975, November/December, 2000.

[8]   V. W. S. Chan, "Free-Space Optical Communications," *Journal of Lightwave Technology (Invited Paper)*, vol. 24, no. 12, pp. 4750-4762, Dec. 2006.

[9]   J. P. Choi, "Channel Prediction and Adaptation over Satellite Channels with Weather-Induced Impairments," Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2000.

[10] J. P. Choi, "Resource allocation and scheduling for communication satellites with advanced transmission antennas," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2006.

[11] H. Cramer and M. R. Leadbetter, *Stationary and Related Stochastic Processes.* New York: John Wiley & Sons, Inc., 1967.

[12] R. H. Davis, *ATM for Public Networks.* New York: McGraw-Hill Inc., 1999.

[13] S. Floyd, "Random Early Discard (RED) gateways for Congestion Avoidance," *IEEE/ACM Transactions on Networking,* vol. 1 no. 4, pp. 397-413, August 1993.

[14] S. Floyd, "HighSpeed TCP for Large Congestion Windows," *Internet Engineering Task Force,* RFC 3649, December 2003. [Online]. Available: http://www.ietf.org/rfc/rfc3649.txt.

[15] C. P. Fu and S. C. Liew, "TCP Veno: TCP Enhancement for Transmission Over Wireless Access Networks," *IEEE Journal on Selected Areas in Communications,* vol. 21, no. 2, pp. 216-228, February 2003.

[16] R. M. Gagliardi and S. Karp, *Optical Communications.* New York: Wiley & Sons, Inc., 1995.

[17] R. Gallager. *Discrete Stochastic Processes.* Boston: Kluwer Academic Publishers, 1999.

[18] J. W. Goodman, *Introduction to Fourier Optics.* New York: McGraw-Hill, 1968.

[19] E. V. Hoversten, R. O. Harger and S. J. Halme, "Communication Theory for the Turbulent Atmosphere," *Proceedings of the IEEE,* vol. 58, no. 10, pp. 1626-1650, October 1970.

[20] V. Jacobson, "Congestion Avoidance and Control," *Computer Communication Review,* vol. 18, no. 4, pp. 314-329, 1988.

[21] V. Jacobson, R. Braden, and D. Borman, "TCP Extensions for High Performance," *Internet Engineering Task Force*, RFC 1323, May 1992. [Online]. http://www.ietf.org/rfc/rfc1323.txt.

[22] C. Jin, D. X. Wei, and S. H. Low, "FAST TCP: motivation, architecture, algorithms, performance," *IEEE INFOCOM*, Hong Kong, China, pp. 2490-2501, March 2004.

[23] D. Katabi, M. Handley, and C. Rohrs, "Congestion Control for High Bandwidth-Delay Product Networks," *Computer Communication Review*, vol. 32, no. 4, pp. 89-102, October 2002.

[24] D. Katabi, "Decoupling Congestion Control and Bandwidth Allocation Policy With Application to High Bandwidth-Delay Product Networks", Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2003.

[25] T. Kelly, "Scalable TCP: Improving Performance in Highspeed Wide Area Networks," *Computer Communication Review*, vol. 33, no. 2, pp. 83-91, April 2003.

[26] I. Kim et al., "Measurement of scintillation and link margin for the Teralink communication system", *Proceedings of SPIE*, vol. 3232, pp. 100-108, 1998.

[27] R. S. Lawrence and J. W. Strohbehn, "A Survey of Clear-Air Propagation Effects Relevant to Optical Communications," *Proceedings of the IEEE*, vol. 58, no. 10, pp. 1523-1545, October 1970.

[28] E. J. Lee and V. W. S. Chan, "Part 1: Optical Communication Over the Clear Atmospheric Channel Using Diversity," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 9, pp. 1896-1906, November 2004.

[29] E. L. Lee and V. W. S. Chan, "Performance of the Transport Layer Protocol for Diversity Communication over the Clear Turbulent Atmospheric Optical Channel," *IEEE International Conference on Communications*, Seoul, S. Korea, pp. 333-339, May 2005.

[30] E. J. Lee and V. W. S. Chan, "Diversity Coherent Receivers for Optical Communication over the Clear Turbulent Atmosphere," *IEEE International Conference on Communications*, Glasgow, Scotland, pp. 2485-2492, June 2007.

[31] E. J. Lee and V. W. S. Chan, "Performance of diversity coherent and diversity incoherent receivers for optical communication over the clear turbulent atmosphere in the presence of an interferer," *Proceedings of SPIE*, vol. 6709, article 670900, September 2007.

[32] E. J. Lee and V. W. S. Chan, "Power Gain of Homodyne Detection over Direct Detection Receivers for Free Space Optical Communication in the Presence of Interference," *Conference on Laser and Electro-Optics*, San Jose, CA, pp. 1-2, May 2008.

[33] E. J. Lee and V. W. S. Chan, "Diversity Coherent and Incoherent Receivers for Free Space Optical Communication in the Presence of and Absence of an Interferer". *IEEE Journal of Optical Communications and Networking*, vol. 1, no. 5, pp. 463-483, October 2009.

[34] S. Mascolo et al., "TCP Westwood: Bandwidth Estimation for Enhanced Transport over Wireless Links," *International Conference on Mobile Computing and Networking*, Rome, Italy, pp. 287-297, July 2001.

[35] R. L. Mitchell, "Permanence of the Log-Normal Distribution," *Journal of the Optical Society of America*, vol. 58, no. 9, pp. 1267-1272, September 1968.

[36] R. Padhye, V. Firoiu, D. Townsley and J. Kurose, "Modeling TCP Throughput: A Simple Model and its Empirical Validation," *Computer Communication Review*, vol. 28, no. 4, pp. 303-314, 1998.

[37] D. C. Palter, *Satellites and the Internet Challenges and Solutions.* California: Satnews Publishers, 2003.

[38] J. G. Proakis, *Digital Communications.* Boston: McGraw-Hill, 2001.

[39] K. Ramakrishnan, "The Addition of Explicit Congestion Notification (ECN) to IP," *Internet Engineering Task Force*, RFC 3168, September 2001. [Online]. http://www.ietf.org/rfc/rfc3168.txt.

[40] R. Ramaswami, *Optical Networks, A Practical Perspective*. Boston: Boston Morgan Kaufmann, 2010.

[41] J. H. Shapiro, "Propagation-medium limitations on phase-compensated atmospheric imaging," *Journal of the Optical Society of America*, vol. 66, no. 5, pp. 460-469, May 1976.

[42] J. H. Shapiro and R. C. Harney, "Simple algorithms for calculating optical communication performance through turbulence," *SPIE Control and Communication Technology in Laser Systems*, vol. 295, pp. 41-54, 1981.

[43] E. J. Shin, "Diversity Optical Communication over the Turbulent Atmospheric Channel," Master's thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2002.

[44] J. W. Strohbehn, ed., *Laser Beam Propagation in the Atmosphere*. New York: Springer-Verlag, 1978.

[45] R. K. Tyson, *Principles of Adaptive Optics*. Boston: Academic Press, 1998.