# Communication for Wideband Fading Channels: on Theory and Practice

by

Cheng Luo

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

Doctor of Philosophy in Computer Science and Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2006





Author .....
Department of Electrical Engineering and Computer Science
er 30, 2005

Certified by .....
Muriel Médard
Harold E. and Esther Edgerton Associate Professor
Thesis Supervisor

Certified by .....
Lizhong Zheng
Assistant Professor
rvisor

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

# Communication for Wideband Fading Channels: on Theory and Practice

by

Cheng Luo

Submitted to the Department of Electrical Engineering and Computer Science
on November 30, 2005, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Computer Science and Engineering

## Abstract

This dissertation investigates some information theoretic aspects of communication over wideband fading channels and their applicability to design of signaling schemes approaching the wideband capacity limit. This work thus leads to enhanced understanding of wideband fading channel communication, and to the proposal of novel efficient signaling schemes, which perform very close to the optimal limit. The potential and limitations of such signaling schemes are studied.

First, the structure of the optimal input signals is investigated for two commonly used channel models: the discrete-time memoryless Rician fading channel and the Rayleigh block fading channel. When the input is subject to an average power constraint, it is shown that the capacity-achieving input amplitude distribution for a Rician channel is discrete with a finite number of mass points in the low SNR regime. A similar discrete structure for the optimal amplitude is proven to hold over the entire SNR range for the average power limited Rayleigh block fading channel. Channels with a peak power constraint are also analyzed. When the input is constrained to have limited peak power, we show that if its Kuhn-Tucker condition satisfies a sufficient condition, the optimal input amplitude is discrete with a finite number of values. In the low SNR regime, the discrete structure becomes binary.

Next, we consider signaling over general fading models. Multi-tone FSK, a signaling scheme which uses low duty cycle frequency-shift keying signals (essentially orthogonal binary signals), is proposed and shown to be capacity achieving in the wideband limit. Transmission of information over wideband fading channels using Multi-tone FSK is considered by using both theoretic analysis and numerical simulation. With a finite bandwidth and noncoherent detection, the achievable data rate of the Multi-tone FSK scheme is close to the wideband capacity limit. Furthermore, a feedback scheme is proposed for Multi-tone FSK to improve the codeword error performance. It is shown that if the receiver can feedback received signal quality to the transmitter, a significant improvement in codeword error probability can be achieved. Experimental results are also obtained to demonstrate features and practicality of Multi-tone FSK.

Thesis Supervisor: Muriel Medard
Title: Harold E. and Esther Edgerton Associate Professor

Thesis Supervisor: Lizhong Zheng
Title: Assistant Professor

# Acknowledgments

First and foremost, I would like to express my deepest gratitude to my advisors Professor Muriel Médard and Professor Lizhong Zheng for their guidance and unconditional support throughout my years at MIT. I thank them for the countless hours they have spent with me, discussing everything from research to career choices, reading my papers, and critiquing my talks. I have always been excited to have the opportunity to work with them. Words cannot describe my gratitude for their numerous interesting research ideas. I realize that sometimes a single discussion with them has led to ideas for a full research paper, and I feel fortunate to have seen a glimpse of their wisdom. I thank them for pushing me to achieve higher standards. With their encouragement, inspiration, constant guidance and dedication, I feel I have improved myself to become a better researcher and writer, and I consider myself privileged to have experienced a rewarding graduate study.

I am also thankful to Professor Vincent W. S. Chan, Professor Dennis L. Goeckel, and Professor Gregory W. Wornell for agreeing to supervise my work and tolerating the last-minute schedule. I have benefited a lot from the discussions with them. Their example as excellent scholars has set high standards I could strive for. Special thanks are also extended to Professor Alan S. Willsky for his guidance in my first year as my academic advisor, and Professor Alex Megretski for his guidance in my remaining years as my academic advisor at MIT.

I owe a huge debt of gratitude to my parents, Guangxian Luo and Xiaohua Jin, for their endless support and love throughout not only the course of my Ph.D. studies, but my entire life.

Finally, I would like to thank my wife and best friend, Xiaona. Her support and love is unfailing, and her patience and understanding have been a great comfort. She keeps my life filled with laughter and love. This dissertation, which has greatly benefited from her proof reading, is dedicated to my wife.

This work was supported in part by the National Science Foundation, "CAREER", under Grant CCR-00993349; in part by Hewlett-Packard, "Advanced Concepts in

Wireless Networking", under Grant 008542-008; and in part by "Medards Child for UW Wireless," under Grant ANI-0335256.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Motivation

Recent years have witnessed intensive study of information theoretic aspects of wireless channels, including not only capacity but also efficient signaling schemes which can approach the capacity. This rising interest in the information theoretic analysis of wireless channels is motivated by the successful commercial deployment of wireless communications, the rapid technology advances in wireless technology and the need to use scarce resources, such as bandwidth and power, as efficiently as possible under wireless communication conditions.

Information theoretic measures such as capacity and error exponents provide the ultimate performance limit for communication systems. We use them as benchmarks to which the performance of practical communication systems can be compared. Furthermore, with the recent discovery and development of codes that operate very close to the Shannon limit, such as LDPC codes and Turbo codes, information theoretic limits have gained practical relevance.

Although the capacity and other information theoretic aspects of fading channels were investigated as early as the 1960's ( [1]), it is only recently that information theoretic considerations of wideband fading channels under various practically related input and channel constraints have received much attention. One reason behind this interest is that wideband communication techniques are fundamentally differ-

ent from the narrowband communication techniques because they employ extremely large bandwidth to communicate between transmitters and receivers. Another factor fuelling this intensive research interest is the imminent market demand on wideband communication techniques. In February 2002, Federal Communications Commission (FCC) approved the use of Ultra-Wideband modulation (UWB) for communications. It allows the signal to extend over a large bandwidth (several gigahertzs). Figure 1-1 shows the bandwidth allocation for the UWB and 802.11a (Orthogonal frequency-division multiplexing, OFDM).

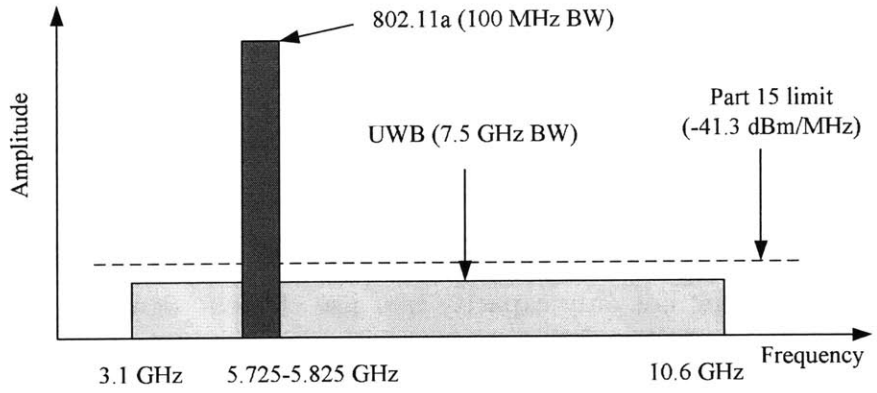


Figure 1-1: The bandwidth allocation of ultra-wideband applications

This introduction provides a brief overview of wideband communications, starting with Shannon wideband capacity limit. Next the discussion turns to the possible signal structures achieving the wideband capacity limit, including the literature background. The discussion in this chapter eliminates the common misconception about wideband communications. It examines the strengths and weaknesses of wideband communication techniques. Furthermore, the main thread of this thesis is explained. The chapter ends with a concise overview of the chapters in this thesis.

## 1.2 Wideband Capacity Limit

As shown by Shannon ( [2], [3]), the capacity of an additive white Gaussian noise (AWGN) channel is a function of the average power $P$ and the bandwidth $W$ given by

$$C_{AWGN} = W \log\left(1 + \frac{P}{N_0 W}\right)$$

where $N_0$ is the spectral noise density. When the bandwidth grows to infinity, the capacity mathematically approaches a limit decided by the average power:

$$\lim_{W\to\infty} W \ln\left(1 + \frac{P}{N_0 W}\right) = \frac{P}{N_0}.$$

We call this limit as the *Shannon wideband capacity limit* or the *AWGN wideband capacity limit.*

It is well known that the channel capacity of multi-path fading channels is upper bounded by the AWGN capacity. Kennedy [1], Gallager [4], and Telatar and Tse [5] further proved that the capacity of a multi-path fading channel approaches the AWGN wideband capacity limit when the bandwidth grows indefinitely. That is

$$\lim_{W\to\infty} C_{fading} = \frac{P}{N_0}. \tag{1.1}$$

The limit on the right hand side (RHS) of (1.1) is called as the *wideband capacity limit* throughout this thesis without specifying the channel type. This result is true even if the receiver does not have any channel state information (CSI).

Given a fading channel's bandwidth is large enough, the capacity of the fading channel approaches the wideband capacity limit. We should be able to find a signaling scheme whose achievable data rate is very close to the wideband capacity limit. Indeed, in this thesis, we propose a family of signaling schemes whose performance approaches the limit.

If the bandwidth of a channel is $W$ and the transmission duration is $T$, the degree of freedom of transmission is $2WT$, i.e., we can transmit $2WT$ symbols during the time period $T$ in the specified bandwidth according to the Nyquest Sampling Theorem. If we index all symbols as indicated in Figure 1-2, we can obtain a symbol sequence as that in Figure 1-3.

The average power constraint $P$ restricts the expected energy consumption during

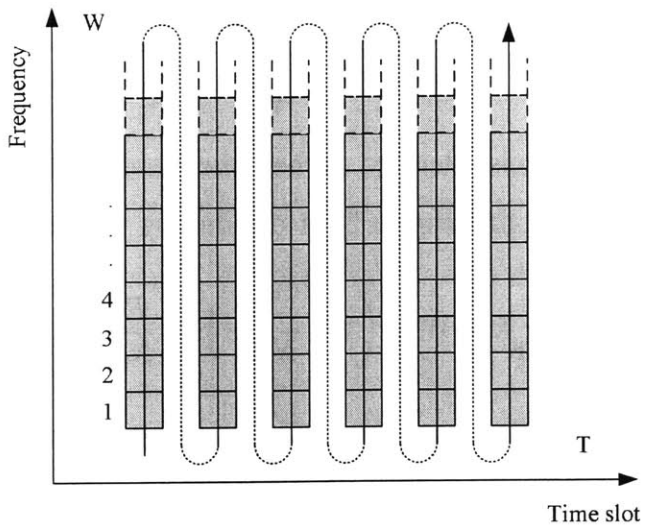


Figure 1-2: Index symbols in both frequency and time

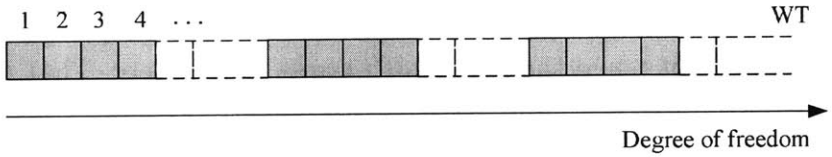


Figure 1-3: The indexed symbol sequence

time $T$ to be $PT$. While bandwidth $W$ increases as required by the wideband assumption, the energy per symbol $PT/(2WT) = P/(2W)$ decreases inversely. When the bandwidth is very large, the signal to noise ratio (SNR) $\frac{P}{N_0 W}$ is very small such that the achievable data rate of the scheme is mainly limited by the power. Results in this thesis are generally concerned with the power-limited case, i.e., communications in the power-limited region are considered.

We also assume that channel state information is not initially known at either the transmitter or the receiver. In this case, as the average signal energy in each coherence bandwidth decreases, the cost of estimating the channel fading coefficients becomes more and more significant. While the capacity of non-coherent channels is the same as that of coherent channels in the wideband limit, the lack of CSI severely affects the performance of many commonly used signaling schemes in the wideband regime.

## 1.3 Signaling Schemes

To answer which kind of signaling schemes can achieve the wideband capacity limit in fading channels, we summarize the main signaling schemes considered in the literature by categorizing them into four classes: no peakiness in either time or frequency, peakiness in frequency but not time, peakiness in time but not frequency, and peakiness in both time and frequency. Although any set of orthogonal signals achieves the capacity limit for the unfaded Gaussian channel, orthogonal signals that are peaky in time or frequency are needed in the presence of fading [4, Sec. 8.6].

For wideband fading channels, the frequency selective effect caused by the channel delay spread should be considered because the frequency coherence of the channel, denoted here as $\mathcal{F}_c$, is usually small compared to the system bandwidth. Also, we assume the time length of the transmission is larger than the time coherence of the channel such that the channel fading coefficient is changing over time. The time coherence is denoted as $\mathcal{T}_c$. The frequency-selective time-varying block fading channel model is used.

In Figure 1-4, 1-5, 1-6, and 1-7, we plot the fading channel in a frequency-time plane. The width and length of the squares are the frequency coherence and the time coherence of the channel. The different colors of the squares indicate different fading coefficients. The transparent shadow is the distribution of the signal's energy density over the frequency-time plane.

### Without Peakiness in either Time or Frequency

The first category of signaling schemes spreads energy over time and frequency as shown in Figure 1-4. This kind of schemes employs symbols occupying the whole system bandwidth, and transmits symbols continuously without duty cycle.

One example is the *bandwidth-scaled spread spectrum signaling* defined by Médard and Gallager in [6]. When the bandwidth $W$ increases, the energy and fourth moment of the signal in each fixed frequency band scale with $1/W$ and $1/W^2$ respectively. To put it simply, the signaling spreads the signal energy evenly over the available

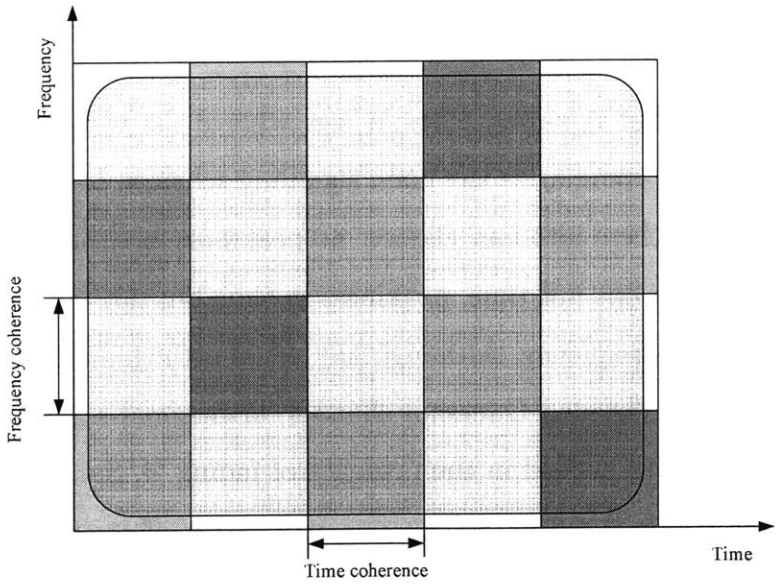


Figure 1-4: Signaling schemes without peakiness in either time and frequency

bandwidth and time without peakiness in either. Such signals encompass Gaussian signals, which are optimal in the coherent case (perfect receiver CSI). However, although bandwidth-scaled signaling has a good performance in the coherent case, it performs poorly in non-coherent scenarios. It has been shown by Médard and Gallager [6], Subramanian and Hajek [7] that, as the bandwidth increases, the achievable data rate using bandwidth-scaled signaling in fact decreases to zero. Therefore such signals, including most common DS-CDMA schemes, are not suitable for wideband transmission in frequency-selective fading channels. Intuitively, this is due to the fact that when we transmit over a large bandwidth, we are essentially transmitting over a large number of independent channels. If signal energy is spread evenly over the available bandwidth, then the capacity of each channel decreases very fast as the bandwidth increases. Decreasing in the average power per channel causes the lack of channel state information which severely affects the performance of the signaling schemes. And, despite the diversity gain, there is a severe performance degradation in the limit.

### Peakiness in Frequency but not Time

In the second kind of signaling schemes (Figure 1-5), the signal energy is peaky in frequency but not time. The signal's energy is distributed over a small portion of the whole bandwidth.

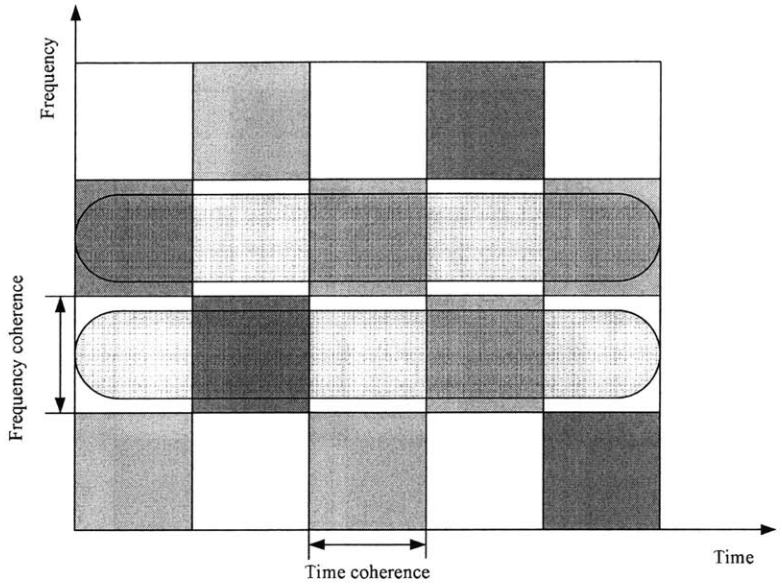


Figure 1-5: Signaling schemes with peakiness in frequency but not time

Many narrowband schemes have the pattern shown in Figure 1-5. They cannot achieve the wideband capacity limit, because these schemes are designed for the bandwidth limited region. Some signaling schemes using several parallel narrowband channels were proposed for improving data rates in wideband channels. For example, the OFDM signaling scheme uses a bundle of subchannels. However, if we assume the bandwidth for each subchannel is $W_s$ and the number of total subchannels is $n$, the capacity of the scheme is bounded by the Shannon capacity with bandwidth $nW_s$, not achieving the wideband capacity limit.

$$C < nW_s \log\left(1 + \frac{P}{N_0 nW_s}\right)$$

If the number of subchannels $n$ increases, we go back to the previous case without peakiness in frequency, the performance of these schemes deteriorates significantly.

### Peakiness in Time but not Frequency

The third class of signaling schemes spreads energy over frequency, but is peaky in time (Figure 1-6), for example, some low-duty cycle pulse-position modulation (PPM) signals.

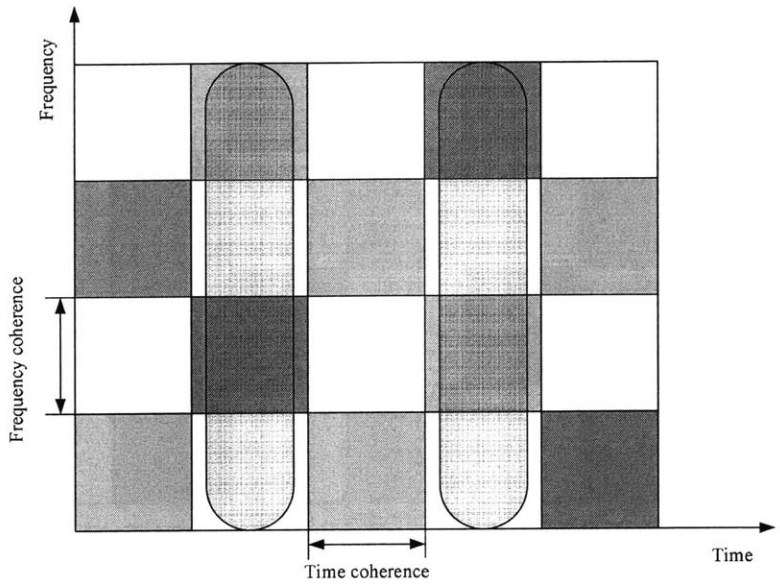


Figure 1-6: Signaling schemes with peakiness in time but not frequency

According to Porrat et. al. [8], this type of signaling schemes can achieve the wideband capacity limit. However, the achievable data rates are limited by the channel uncertainty. They considered direct sequence spread spectrum and pulse position modulation systems with duty cycle, and proved that these two schemes achieve the channel capacity if the increase of the number of channel paths with the bandwidth is not too rapid. They showed that PPM systems have zero throughput if $\frac{L}{\ln W} \to \infty$, where $L$ is the number of independently fading channel paths and $W$ is the bandwidth.

The property that the achievable data rates are limited by the channel uncertainty imposes a big constraint on these schemes' application in wideband fading channels. In [9], Rusch et. al. presented residential channel characterizations from 2 to 8 GHz and showed that the increase of the number of channel paths appears to be sub-linear in bandwidth.

### Peakiness in both Time and Frequency

Another set of interesting signaling schemes is peaky both in time and frequency (Figure 1-7). The frequency shift keying (FSK) scheme used in the literature [4, Sec. 8.6] [5] to achieve the infinite bandwidth capacity of multipath fading channels belongs to this category. Since the scheme transmits with a low duty cycle and its signals burst like flash light, we call it *flash FSK*. In [5], Telatar and Tse used the flash FSK scheme with a threshold decoding rule to achieve the wideband capacity limit in general multipath fading channels.

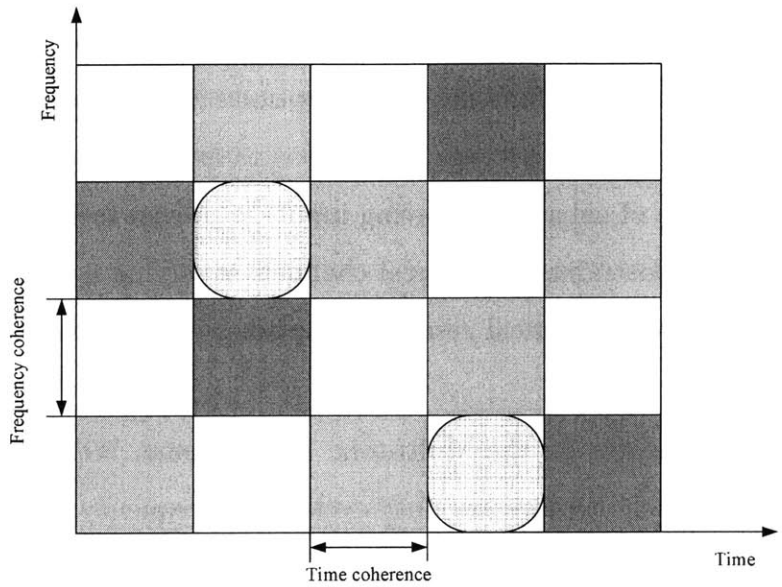


Figure 1-7: Signaling schemes with peakiness in both time and frequency

The Multi-tone FSK scheme which is introduced later in this thesis extends the family of peaky signaling schemes. It will be referred to as "capacity achieving," since it achieves capacity in the wideband limit.

## 1.4 Dissertation Outline

In this thesis, we first consider the capacity-achieving input distribution of fading channels. We prove that in the low SNR region (as we pointed out that wideband channels will be in), the binary amplitude distribution is desirable for a set of channels

and this generalizes what is known for other channels. We also consider a Multi-tone FSK scheme in which each tone has binary amplitude distribution and all tones are coded over frequency and time. We prove that the Multi-tone FSK scheme is capacity achieving in the wideband limit. Furthermore, we study how feedback can help to improve Multi-tone FSK's error performance and define a feedback scheme which reduces the error probability significantly. Finally, we implement the Multi-tone FSK scheme on an ultra-wideband test bed and show the practicality of the scheme.

The organization of the dissertation is as follows:

In Chapter 2, we study the structures of capacity-achieving input distributions of noncoherent memoryless Rician fading channels and Rayleigh block fading channels. We show the discreteness and finiteness of the optimal input amplitude distributions when the input is imposed an average power constraint or a peak power constraint. The discrete structure of capacity-achieving inputs is proven to hold in the presence of only peak power constraints for general channels satisfying a sufficient condition. In addition, we provide numerical results for optimal input distributions in Rician fading channels.

Chapter 3 concentrates on the Multi-tone FSK scheme. We propose the Multi-tone FSK scheme, which employs low duty cycle multi-frequency FSK symbols and a noncoherent receiver. We investigate the performance of Multi-tone FSK and derive its capacity limit and error exponent. We prove that the Multi-tone FSK scheme achieves the wideband capacity limit. A similar analysis is also conducted for the Multi-tone FSK scheme with receiver diversity, and its capacity limit and error exponent are derived. Moreover, we present numerical results to show the performance of Multi-tone FSK at finite bandwidths.

Chapter 4 is devoted to the investigation of the Multi-tone FSK schemes with feedback. We introduce a Feedback Multi-tone FSK scheme in which a feedback link is setup to monitor the transmission quality. We prove that, using a small amount of feedback, the scheme has much better error performance than the Multi-tone FSK scheme.

Chapter 5 includes conclusions and directions for future research.

Finally, in Appendix D, we present some of the experimental results we obtained by testing the Multi-tone FSK scheme on an ultra-wideband platform. By implementing the Multi-tone FSK scheme, we verify the idea of Multi-tone FSK and the validity of model assumptions.

# Chapter 2

# Structure of Optimal Input Distributions

## 2.1 Introduction

The effort to seek information theoretic guidelines for designing efficient signaling schemes dates back to the birth of information theory. Proved by Shannon [2], the capacity-achieving input distribution for an average power limited AWGN channel is a Gaussian distribution.

Smith [10], Shamai and Bar-David [11] considered the peak power limited AWGN channel and proved that the capacity-achieving input amplitude distribution is discrete with a finite number of mass points. The change in the power constraint markedly effects the structure of the optimal input.

Recently, Katz and Shamai [12] considered the noncoherent AWGN channels where the input is distorted by phase noise and have shown that the optimal input amplitude distribution is again discrete but with an infinite number of mass points. This result has been extended to the block-independent case by Nuriyev and Anastasopoulos [13].

Among channel models, it is of particular interest to consider the case when the channel experiences fading, which is a realistic assumption especially in mobile communications. Recent work on noncoherent fading channels, where neither the receiver

nor the transmitter knows the fading, has led to many interesting results that are significantly different from those of unfaded Gaussian channels.

In [14], Richters conjectured that the capacity-achieving distribution for discrete-time Rayleigh memoryless channels is discrete under an average power constraint. The conjecture has not been rigorously proved until 1997. Abou-Fayçal et al. have proven in [15] that, for this basic channel, the capacity-achieving amplitude distribution is discrete with a finite number of mass points.

Gursoy et al. investigated communications over a noncoherent discrete-time memoryless Rician fading channel in [16], and showed that, under an average power constraint, the optimal input amplitude distribution is bounded. They also studied Rician channels with peak power constraints and fourth moment constraints for amplitude, and proved that the amplitude distribution is discrete with a finite number of mass points.

Moreover, Marzetta and Hochwald [17] characterized the optimal input distribution for the multi-antenna Rayleigh block fading channel as a product of two statistically independent matrices: a random matrix which is diagonal, real and nonnegative times an isotropically distributed unitary matrix. For the single antenna case, Palanki [18] proved that the optimal input amplitude distribution for Rayleigh block fading channels is discrete.

In this chapter, we consider the noncoherent discrete-time Rician fading channel and the Rayleigh block fading channel, and study the optimal input structure.

As mentioned before, the optimal input amplitude distribution for the average power limited Rician fading channel was considered in [16], and proven to be bounded. Our result in this chapter is much stronger. We characterize the optimal input amplitude distribution as discrete with a finite number of mass points under a low SNR condition. We show that, there exists a threshold, for a Rician channel subject to an average power constraint less than the threshold, the support of an optimal amplitude distribution is not only bounded, but also discrete with a finite number of mass points. For peak power limited channels, we find a general principle for the optimal input amplitude distribution. The principle provides a sufficient condition

for the support to be discrete with a finite number of mass points. According to this rule, for most common channels subject to a peak power constraint, including Rician channels, capacity-achieving input amplitude distributions have discrete amplitudes with a finite number of mass points.

In [18], the optimal input amplitude distribution for the average power limited Rayleigh block fading channel was considered, and proven to be discrete. We prove, in this chapter, that the optimal input amplitude distribution is not only discrete but also with a finite number of mass points.

The organization of the chapter is as follows. In Section 2.2, we introduce the Kuhn-Tucker conditions with different power constraints, which are used to eliminate impossible distributions for the optimal inputs. In Section 2.3, we introduce the memoryless Rician fading channel model, and characterize the structure of the capacity achieving input distribution when the channel input is subject to average power and peak power constraints. A similar analysis is conducted for the Rayleigh block fading channel case in Section 2.4. Section 2.5 contains our conclusions.

## 2.2 Kuhn-Tucker Condition

To describe a channel statistically, we often use the conditional probability density $p_{y|x}(y|x)$, where $x$ is the input of the channel and $y$ the output of the channel. From an information theoretic point of view, the goal of efficient signaling scheme design is to maximize the achievable data rate for reliable communications, subject to some input signal constraints, such as the average power constraint, $E(|x|^2) \leq P_{av}$, or the peak power constraint, $|x| \leq P_{peak}$. The signal constellation and its associated probability density, i.e., $p_x(x)$, is the function which needs to be optimized upon. For example, with an average power constraint, the optimization problem can be formulated as

$$C = \sup_{\substack{p_x(x) \\ E(|x|^2) \leq P_{av}}} \int \int p_{y|x}(y|x) \ln \frac{p_{y|x}(y|x)}{p_y(y; p_x)} dy p_x(x) dx,$$

where $p_y(y; p_x)$ is the probability density of the output $y$ which is determined by the channel input density $p_x(x)$ and $p_{y|x}(y|x)$. The double integral in the objective function is to calculate the mutual information between $x$ and $y$, which equals to the maximum achievable data rate for reliable communications.

To find analytically the optimal input constellation and probability density, $p_x^*(x)$, for most cases, is an impossible task, owing to the intractability of the optimization problem which has an infinite number of degrees of freedom to work with. In this chapter, we manage to rule out continuous distributions and narrow the scope to a searchable size.

For example, we show that, in the low SNR regime, the optimal input amplitude distribution for the average power limited discrete-time memoryless Rician fading channel is discrete with a finite number of mass points. In [19], Gallager showed that binary inputs are optimal when the energy per degree of freedom goes to zero if only discrete inputs are considered. Thus, the optimal input amplitude distribution in the low SNR limit is binary, which can be found numerically. Many practical algorithms for this purpose exist in the literature.

To prove the optimal input amplitude distribution is discrete with a finite number of mass points, we show that all other cases are not possible, because otherwise the Kuhn-Tucker condition, which should be satisfied by the optimal distribution, will be violated. In the rest of this section, we will establish the Kuhn-Tucker conditions for our particular problem.

For average power limited channels, the Kuhn-Tucker condition has already been derived by Abou-Fayçal et al. in [15]. We cite the result here after rewording:

**Proposition 2.1** *(Abou-Fayçal et al. [15]) For a channel with an average power constraint, $E(|x|^2) \leq P_{av}$ ($P_{av} > 0$), an input random variable $X^*$ with density function $p^*$ achieves the capacity $C$ if and only if there exists a $\gamma \geq 0$ such that*

$$k_a(x) = \gamma(|x|^2 - P_{av}) + C - \int p(y \mid x) \ln \frac{p(y \mid x)}{p(y; p^*)} dy \geq 0 \tag{2.1}$$

*for all $x$, with equality if $x$ is in the support of $X^*$. $p(y; p^*)$ is the probability density*

*of $y$ with the density function of $X^*$ is $p^*$.*

The proof of this proposition is in [15]. Note that $\gamma$ is the slope of the capacity versus average power curve at the given average power $P_{av}$.

For peak power limited channels, we prove the following proposition:

**Proposition 2.2** *For a channel with a peak power constraint, $|x|^2 \leq P_{peak}$ ($P_{peak} > 0$), an input random variable $X^*$ with density function $p^*$ achieves the capacity $C$ if and only if*

$$k_p(x) = C - \int p(y \mid x) \ln \frac{p(y \mid x)}{p(y; p^*)} dy \geq 0 \tag{2.2}$$

*for all $x$ located in $|x|^2 \leq P_{peak}$, with equality if $x$ is in the support of $X^*$. Note that $p^*$ is only non-zero on the range $|x|^2 \leq P_{peak}$.*

This Kuhn-Tucker condition is customized for peak power limited channels. Please see Appendix B for the complete proof of this proposition.

**Proposition 2.3** *For a channel with an average power constraint, $E(|x|^2) \leq P_{av}$ ($P_{av} > 0$) and a peak power constraint, $|x|^2 \leq P_{peak}$ ($P_{peak} > 0$), an input random variable $X^*$ with density function $p^*$ achieves the capacity $C$ if and only if there exists a $\gamma \geq 0$ such that*

$$k_{ap}(x) = \gamma(|x|^2 - P_{av}) + C - \int p(y \mid x) \ln \frac{p(y \mid x)}{p(y; p^*)} dy \geq 0 \tag{2.3}$$

*for all $|x|^2 \leq P_{peak}$, with equality if $x$ is in the support of $X^*$. $p^*$ is only non-zero in the regime $|x|^2 \leq P_{peak}$.*

The proof directly follows that of Proposition 2.2 by limiting the support of $p^*$ to $|x|^2 \leq P_{peak}$.

Remark:

1) $x$ can be a scalar input, a complex input, or a vector input. For complex inputs or vector inputs, $|x|^2$ stands for the squared norm. An example of the vector case is a multiple-input multiple-output (MIMO) channel;

2) $|x|^2 \leq P_{peak}$ is interpreted as a peak power (amplitude) constraint in the scalar and complex cases, while in the vector case, it imposes a constraint on the total power of the input vector. For the MIMO example, it is the total power constraint for all transmitting antennas;

3) The equalities in (2.1), (2.2), and (2.3) are only necessary conditions for $x$ being in the support of $X^*$, not sufficient.

In (2.1), (2.2), and (2.3), for referential convenience, we use $k_a(x)$, $k_p(x)$, and $k_{ap}(x)$ to denote the functions on the left hand side of the inequalities. The subscripts stand for the average power constraint, the peak power constraint, and both, respectively. In cases applicable to all three power constraints, we omit the subscript. When the optimal input phase distribution is uniform on $[0, 2\pi)$ and independent of the amplitude distribution, the value of $k(x)$ will be determined by the amplitude of $x$. Denote the amplitude as $r$. We explicitly write the function as $k(r)$ instead of $k(x)$.

In the following sections, we will derive $k(r)$ for Rician channels and Rayleigh block fading channels with different power constraints. By studying the Kuhn-Tucker condition, we rule out impossible distributions for the optimal input.

## 2.3 Rician Fading Channels

Wireless channels often exhibit fading and dispersion. The fading effect is usually modelled as Rician distribution when line-of-sight (LOS) input propagation is present alongside a Rayleigh component. Additive White Gaussian Noise channels and Rayleigh fading channels are two extreme types of Rician fading channels. In the first, only the LOS component is present. In the second, it is altogether absent.

In this section, we consider transmission over a discrete-time memoryless Rician fading channel for capacity-achieving distributions. We prove that, for an optimal input, the phase distribution is uniform and independent of the amplitude. Based on this result, we obtain the Kuhn-Tucker condition for Rician channels subject to different power constraints. And by using the Kuhn-Tucker conditions, we show that,

for an average power limited Rician channel, there exists a threshold such that, when the average power is less than the threshold, the support of the optimal amplitude distribution is discrete with a finite number of mass points. Moreover, for peak power constrained channels, we provide a sufficient condition for the supports to be discrete and finite.

### 2.3.1 Channel Model

The output of the discrete-time memoryless Rician channel we consider is

$$y = (m + r)x + w$$

where $x$ is the input, $m + r$ is the channel fading coefficient, and $w$ is the additive noise with zero-mean complex Gaussian distribution $CN(0, \sigma_w^2)$. $m$ represents the effect of the line-of-sight propagation, and is invariant. $r$ stands for the effect of the large number of independent scatterers, and is a zero-mean complex Gaussian random variable, $CN(0, \sigma_r^2)$, which changes independently from symbol to symbol due to the nature of rapid fading. We assume that $m$ is observed by the receiver and the transmitter, whereas $r$ can be observed by neither the receiver nor the transmitter. The probability of $y$ conditional on $x$ is

$$p_{y|x}(y|x) = \frac{1}{\pi\left(\sigma_r^2 |x|^2 + \sigma_w^2\right)} \exp\left(-\frac{|y - mx|^2}{\sigma_r^2 |x|^2 + \sigma_w^2}\right), \tag{2.4}$$

where $x$ and $y$ are complex.

### 2.3.2 Phase Distribution

Using the conditional probability (2.4), we show the following proposition for the phase distribution.

**Proposition 2.4** *In a Rician fading channel, the phase distribution of the capacity-achieving input is uniformly distributed and independent of the amplitude.*

**Proof:** Assume that $\Phi$ is a unitary complex number. Then $y' = \Phi y$ and $x' = \Phi x$ are variables obtained by rotating $y$ and $x$ on the complex plane by $\Phi$.

We obtain

$$p_{y|x}(y|x) = p_{y|x}(\Phi y|\Phi x),$$

by substituting $\Phi x$ and $\Phi y$ into (2.4). That is, the conditional probability is invariant when $x$ and $y$ have an identical phase shift.

By the definition of mutual information, we have

$$I(y;\Phi x) = \iint p_{y|x}(y|x) \ln \frac{p_{y|x}(y|x)}{p_y(y;p_x^*(\Phi x))} dy p_x^*(\Phi x) dx.$$

The notation $p_y(y;p_x^*)$ denotes the probability density of the output $y$ when the input density is $p_x^*(x)$. Similarly, $p_y(y;p_x^*(\Phi x))$ stands for the output density when the input is $\Phi x$.

Change variables by letting $x = \Phi^{-1}x'$, and notice that the Jacobian $|\Phi^{-1}|$ is 1.

$$I(y;\Phi x) = \iint p_{y|x}(y|\Phi^{-1}x') \ln \frac{p_{y|x}(y|\Phi^{-1}x')}{p_y(y;p_x^*(\Phi x))} dy p_x^*(x') dx'.$$

Noticing that $p_{y|x}(y|x) = p_{y|x}(\Phi y|\Phi x)$, we have similarly $p_{y|x}(y|\Phi^{-1}x') = p_{y|x}(\Phi y|x')$. In addition, we have

$$\begin{aligned} p_y(y;p_x^*(\Phi x)) &= \int p_{y|x}(y|x) p_x^*(\Phi x) dx \\ &= \int p_{y|x}(y|\Phi^{-1}x) p_x^*(x) dx \\ &= \int p_{y|x}(\Phi y|x) p_x^*(x) dx \\ &= p_y(\Phi y;p_x^*). \end{aligned}$$

Thus,

$$I(y;\Phi x) = \iint p_{y|x}(\Phi y|x') \ln \frac{p_{y|x}(\Phi y|x')}{p_y(\Phi y;p_x^*)} dy p_x^*(x') dx'.$$

Substituting $y'$ for $\Phi y$, and using the fact that the Jacobian $|\Phi^{-1}|$ is 1, we obtain

$$I(y;\Phi x) = I(y;x).$$

Hence, the mutual information $I(y;x)$ is invariant to a rotation of $x$. For any probability density of $x$, $p_x(x)$, there exists

$$I(y;x) = \oint \frac{1}{2\pi} I(y;e^{j\omega}x)d\omega.$$

The concavity of mutual information and Jensen's inequality implies

$$\oint \frac{1}{2\pi} I(y;e^{j\omega}x)d\omega \leq I(y;x^*) \tag{2.5}$$

where $x^*$ has a probability density

$$p_x^*(x) = \frac{1}{2\pi} \oint p_x\left(e^{j\omega}x\right) d\omega.$$

The equality in (2.5) holds if and only if $p_x(e^{j\omega}x) = p_x^*(x)$ for any $\omega$. It is clear that $I(y;x^*)$ is the capacity of the channel, because it is an upper bound for all possible mutual information, and can be achieved by letting $p_x(x) = p_x^*(x)$. $p_x^*(x)$ has a uniformly distributed phase independent of the amplitude, and the proposition follows. ■

The independence between the phase and the amplitude greatly simplifies our discussion of the capacity-achieving input distribution. Notice that we did not assume any specific power constraint in the proof.

**Corollary 2.1** *In a Rician fading channel, the capacity-achieving input generates an output with a uniformly distributed phase independent of the amplitude.*

**Proof:** The output probability density is

$$p_y(y;p_x^*) = \int p_{y|x}(y|x)p_x^*(x)\,dx.$$

Notice that $p_{y|x}(y|x) = p_{y|x}(\Phi y|\Phi x)$ and the phase of $p_x^*$ is uniformly distributed, $p_x^*(x) = p_x^*(\Phi x)$,

$$p_y(y; p_x^*) = \int p_{y|x}(\Phi y|\Phi x) p_x^*(\Phi x)\, dx.$$

Performing the change of variable $\Phi x = x'$ yields

$$\begin{aligned} p_y(y; p_x^*) &= \int p_{y|x}(\Phi y|x') p_x^*(x')\, dx' \\ &= p_y(\Phi y; p_x^*). \end{aligned}$$

Hence, $p_y(y; p_x^*)$ is invariant to rotations of $y$ and

$$p_y(y; p_x^*) = \frac{1}{2\pi} \oint p_y(e^{j\omega} y; p_x^*) d\omega,$$

which implies that the output has a uniformly distributed phase independent of the amplitude. ■

### 2.3.3 Amplitude Distribution

By definition, if the input $x$ has a probability density $p_x^*(x)$, the mutual information between output and input is given by

$$M = \iint p_{y|x}(y|x) \ln \frac{p_{y|x}(y|x)}{p_y(y; p_x^*)} dy p_x^*(x) dx.$$

By substituting (2.4) into the above integral, we obtain

$$\begin{aligned} M &= -\int \ln \left[\pi \left(\sigma_r^2 |x|^2 + \sigma_w^2\right)\right] p_x^*(x) dx - 1 \\ &\quad - \iint p_{y|x}(y|x) \ln p_y(y; p_x^*) dy p_x^*(x) dx. \end{aligned}$$

Denote the modulus of the input, $|x|$, as $\nu$ and the absolute square of the output, $|y|^2$, as $\rho$. By using Change of Variables Theorem, we change coordinates from rectangular to polar, $(\nu, \theta_x)$ and $(\rho, \theta_y)$, where $\theta_x$ and $\theta_y$ are the phases of the input and the output. The corresponding Jacobians are $\nu$ and $\frac{1}{2}$, respectively.

Notice two equations:

1) Since the input has a uniformly distributed phase which is independent of the amplitude, we have equation

$$\nu p_x^*(x(\nu,\theta_x)) = \frac{1}{2\pi} p_\nu^*(\nu)$$

where $p_\nu^*(\nu)$ is the input's amplitude pdf;

2) Since the output has a uniformly distributed phase which is independent of the amplitude, we have

$$\frac{1}{2} p_y(y\,(\rho,\theta_y)\,;p_x^*) = \frac{1}{2\pi} p_\rho(\rho;p_x^*),$$

where $p_y(y;p_x^*)$ and $p_\rho(\rho;p_x^*)$ are the probability densities of $y$ and $\rho$ given $p_x^*$.

The mutual information is then

$$\begin{aligned} M = &- \int_0^{2\pi}\int_0^{\infty} \ln\left[\pi\left(\sigma_r^2\nu^2+\sigma_w^2\right)\right]\frac{p_\nu^*(\nu)}{2\pi}d\nu d\theta_x - 1 \\ &- \int_0^{2\pi}\int_0^{\infty}\int_0^{\infty} g(\rho|\nu)\ln\frac{p_\rho(\rho;p_x^*)}{\pi}d\rho\frac{p_\nu^*(\nu)}{2\pi}d\nu d\theta_x \end{aligned} \tag{2.6}$$

where

$$g(\rho|\nu) = \frac{\exp\left[-\frac{\rho+m^2\nu^2}{\sigma_r^2\nu^2+\sigma_w^2}\right] I_0\left(\frac{2\sqrt{\rho}m\nu}{\sigma_r^2\nu^2+\sigma_w^2}\right)}{\sigma_r^2\nu^2+\sigma_w^2} \tag{2.7}$$

is the Rician distribution. $I_0(\cdot)$ is the modified Bessel function of the first kind, i.e.,

$$I_0(x) = \int_0^{\pi}\frac{1}{\pi}\exp\left(x\cos\theta\right)d\theta.$$

The expectation of a Rician random variable is

$$\begin{aligned} E\left(\rho|\nu\right) &= \int_0^{\infty}\rho g(\rho|\nu)d\rho \\ &= m^2\nu^2+\sigma_r^2\nu^2+\sigma_w^2. \end{aligned}$$

By integrating (2.6) with respect to $\theta_x$, we obtain

$$M = \int_0^\infty \left[ -\ln\left(\sigma_r^2\nu^2 + \sigma_w^2\right) - 1 \right.$$
$$\left. - \int_0^\infty g(\rho|\nu) \ln p_\rho(\rho; p_x^*) d\rho \right] p_\nu^*(\nu) d\nu. \tag{2.8}$$

**Average Power Constraint**

The KTC for an average power constrained channel is given by Proposition 2.1. For a Rician channel with an average power constraint, the KTC is

$$k_a(\nu) \triangleq \underbrace{\int_0^\infty g(\rho|\nu) \log p_\rho(\rho; p^*) d\rho}_{(a)} + 1 + C$$
$$+ \ln\left(\sigma_r^2\nu^2 + \sigma_w^2\right) + \gamma(\nu^2 - P_{av}) \geq 0, \tag{2.9}$$

where $C$ is the channel capacity, $\gamma$ the Lagrange multiplier, $P_{av}$ the average power constraint, and

$$p_\rho(\rho; p^*) = \int_0^\infty g(\rho|\nu) p_\nu^*(\nu) d\nu.$$

The condition $k_a(\nu) \geq 0$ is satisfied for $\nu \geq 0$ and the equality holds for all $\nu$ in the support $\mathfrak{R}$ of an optimal input distribution.

Our derivation will use the following lemma, which was proven in [16].

**Lemma 2.1** *With an average power constraint, $E(\nu^2) \leq P_{av}$, a capacity-achieving amplitude distribution for a discrete-time memoryless Rician fading channel has a bounded support.*

Based on this lemma, we assume that the support $\mathfrak{R}$ of an optimal input is in $[0, u]$, where $u$ is the supremum of $\mathfrak{R}$ such that $\int_0^u p_\nu^*(\nu) d\nu = 1$, $\int_u^\infty p_\nu^*(\nu) d\nu = 0$, and $\int_0^{u-\varepsilon} p_\nu^*(\nu) d\nu < 1$ for any $\varepsilon > 0$.

In the KTC, $\gamma$ is a Lagrange multiplier. As such, it represents the slope of the optimum of the object function as a function of the constraint value. In this case, $\gamma$ is the slope of the capacity versus average power curve $C(P)$. The range of $\gamma$ is given by the following lemma.

**Lemma 2.2** *For a discrete-time memoryless Rician fading channel with an average power constraint* $E(\nu^2) \leq P_{av}$, $\gamma$ *is located in* $\left[\frac{m^2+\sigma_r^2}{\sigma_r^2 u^2+\sigma_w^2}, \frac{m^2+\sigma_r^2}{\sigma_w^2}\right]$, *where* $u$ *is the supremum of* $\Re$.

**Proof:** We first prove $\gamma \leq \frac{m^2+\sigma_r^2}{\sigma_w^2}$.

The capacity of an AWGN channel with a channel gain $m^2+\sigma_r^2$ is $\ln\left[1+\frac{(m^2+\sigma_r^2)P}{\sigma_w^2}\right]$, which is evidently an upper bound to the capacity of the Rician channel.

From the convexity and monotonicity of channel capacity in terms of $P$, the largest slope of the capacity curve happens at $P=0$. The slope of $\ln\left[1+\frac{(m^2+\sigma_r^2)P}{\sigma_w^2}\right]$ at $P=0$ is $\frac{m^2+\sigma_r^2}{\sigma_w^2}$. Because the capacity is a continuous function of power $P$, if $\gamma > \frac{m^2+\sigma_r^2}{\sigma_w^2}$ at $P=0$, we could find a small enough $P$ such that the capacity of the fading channel would exceed the capacity of the AWGN channel. As this is not possible, we conclude that $\gamma \leq \frac{m^2+\sigma_r^2}{\sigma_w^2}$ at $P=0$. Hence, $\gamma \leq \frac{m^2+\sigma_r^2}{\sigma_w^2}$ for all $P \geq 0$.

Second, let's prove $\gamma \geq \frac{m^2+\sigma_r^2}{\sigma_r^2 u^2+\sigma_w^2}$.

We can obtain an upper bound for $(a)$ in (2.9). Given

$$D_1 = \int_0^\infty \ln\left[\int_0^u \frac{e^{-\frac{m^2\upsilon^2}{\sigma_r^2\upsilon^2+\sigma_w^2}} I_0\left(\frac{2\sqrt{\rho}m\upsilon}{\sigma_r^2\upsilon^2+\sigma_w^2}\right)}{\sigma_r^2\upsilon^2+\sigma_w^2} p_\nu^*(\upsilon)d\upsilon\right] g(\rho|\nu)d\rho,$$

we have

$$(a) \leq D_1 - \frac{m^2\nu^2+\sigma_r^2\nu^2+\sigma_w^2}{\sigma_r^2 u^2+\sigma_w^2}.$$

Since $I_0(x) \geq 1$ for $x \geq 0$,

$$D_1 > \ln\left[\int_0^u \frac{e^{-\frac{m^2\upsilon^2}{\sigma_r^2\upsilon^2+\sigma_w^2}}}{\sigma_r^2\upsilon^2+\sigma_w^2} p_\nu^*(\upsilon)d\upsilon\right]. \tag{2.10}$$

The right hand side in (2.10) is a constant.

On the other hand, because $I_0(x) < \exp(x)$ for $x \geq 0$,

$$\begin{aligned} D_1 &< \int_0^\infty g(\rho|\nu)\frac{2\sqrt{\rho}mu}{\sigma_w^2}d\rho \\ &\quad \times \ln\left[\int_0^u \frac{e^{-\frac{m^2\upsilon^2}{\sigma_r^2\upsilon^2+\sigma_w^2}}}{\sigma_r^2\upsilon^2+\sigma_w^2}p_\nu^*(\upsilon)d\upsilon\right] \\ &= o(\nu^2). \end{aligned} \tag{2.11}$$

From (2.10) and (2.11), we conclude $D_1 = o(\nu^2)$. Therefore,

$$\begin{aligned} k_a(\nu) &\leq \ln\left(\sigma_r^2\nu^2+\sigma_w^2\right) + \left(\gamma - \frac{m^2+\sigma_r^2}{\sigma_r^2u^2+\sigma_w^2}\right)\nu^2 \\ &\quad + C + \frac{\sigma_r^2u^2}{\sigma_r^2u^2+\sigma_w^2} + o(\nu^2) - \gamma P_{av}. \end{aligned} \tag{2.12}$$

If $\gamma < \frac{m^2+\sigma_r^2}{\sigma_r^2u^2+\sigma_w^2}$, the limit of (2.12) as $\nu \to \infty$ is $-\infty$, which conflicts the KTC. By contradiction, $\gamma \geq \frac{m^2+\sigma_r^2}{\sigma_r^2u^2+\sigma_w^2}$. ■

From the lower bound of the range,

$$\gamma \geq \frac{m^2+\sigma_r^2}{\sigma_r^2u^2+\sigma_w^2},$$

we obtain

$$u^2 \geq \frac{1}{\sigma_r^2}\left(\frac{m^2+\sigma_r^2}{\gamma} - \sigma_w^2\right). \tag{2.13}$$

From the upper bound,

$$\gamma < \frac{m^2+\sigma_r^2}{\sigma_w^2},$$

we have

$$\frac{m^2+\sigma_r^2}{\gamma} - \sigma_w^2 > 0,$$

which lets the RHS of (2.13) be always positive. Thus, we conclude that small $\sigma_r^2$ requires large $u^2$, i.e., the peak power of a capacity-achieving scheme need to be extremely large when the Rayleigh component is very small. This is consistent with the case of an AWGN channel. An AWGN channel has a zero Rayleigh component, and its $u^2$ is infinite.

Having determined the possible range of $\gamma$, we study the capacity-achieving distribution for different $\gamma$.

First, we show a lemma.

**Lemma 2.3** *When $k(\nu)$ is analytic in the region $\nu \geq 0$, a sufficient condition for the support $\mathfrak{R}$ of the capacity-achieving distribution, $p^*(\nu)$, to be discrete with a finite number of mass points is that, there exists an $\nu_0$ such that for any $\nu > r_0$, $k(\nu)$ is not equal to zero.*

**Proof:** We prove the lemma by ruling out all other cases outside the lemma:

1) Assume the support $\mathfrak{R}$ of $p^*(\nu)$ has an infinite number of points and these points increase on a bounded interval. Continuous distributions satisfy this assumption.

Since $k(\nu)$ is analytic in the region $\nu \geq 0$, by the Identity Theorem [20], $k(\nu) = 0$ in the region. Hence, the Kuhn-Tucker condition is satisfied with equality for all $\nu \geq 0$. This conflicts with the given assumption that $k(\nu)$ is not equal to zero as $\nu \to \infty$.

2) Assume the optimal distribution has an infinite number of mass points but only finitely many of them on any bounded interval.

Then $k(\nu)$ should be equal to zero infinitely often as $\nu \to \infty$ which also conflicts with the assumption that $k(\nu)$ is not equal to zero as $\nu > \nu_0$. ■

Next, we show the following proposition.

**Proposition 2.5** *There exists a $P_0 > 0$, such that, for all $P_{av} \in (0, P_0]$, the support of an optimal amplitude distribution of a discrete-time memoryless Rician fading channel with an average power constraint $E(\nu^2) \leq P_{av}$ is discrete with a finite number of mass points.*

**Proof:** According to Lemma 2.2, $\gamma$ is located in the range $\left[\frac{m^2+\sigma_r^2}{\sigma_r^2 u^2+\sigma_w^2}, \frac{m^2+\sigma_r^2}{\sigma_w^2}\right]$.

First, we assume

$$\frac{m^2+\sigma_r^2}{\sigma_r^2 u^2+\sigma_w^2} < \gamma \leq \frac{m^2+\sigma_r^2}{\sigma_w^2}. \tag{2.14}$$

We prove that the capacity-achieving distribution is discrete and finite in this case.

Given (2.14), since $\frac{m^2+\sigma_r^2}{\sigma_r^2 x^2+\sigma_w^2}$ is a continuous function of $x$, we can find a $u_0$ such that $\gamma = \frac{m^2+\sigma_r^2}{\sigma_r^2 u_0^2+\sigma_w^2}$.

Noticing that $I_0(x) > 1$ and the logarithm is monotonically increasing on $(0, \infty)$, we bound $(a)$ in (2.9) by

$$\begin{aligned}(a) &\geq \int_0^\infty g(\rho|\nu) \ln \left[ \int_{u_0}^u g(\rho|\upsilon) p_\nu^*(\upsilon) d\upsilon \right] d\rho \\ &\geq D_2 - \frac{m^2\nu^2 + \sigma_r^2\nu^2 + \sigma_w^2}{\sigma_r^2 u_0^2 + \sigma_w^2}, \end{aligned} \tag{2.15}$$

where

$$D_2 = \ln \left[ \int_{u_0}^u \frac{\exp\left(-\frac{m^2\upsilon^2}{\sigma_r^2\upsilon^2+\sigma_w^2}\right)}{\sigma_r^2\upsilon^2 + \sigma_w^2} p_\nu^*(\upsilon) d\upsilon \right].$$

$D_2$ is a finite real number. By substituting (2.15) into $k_a(\nu)$, we obtain a lower bound of $k_a(\nu)$ given by

$$\begin{aligned} k_a(\nu) &\geq \ln\left(\sigma_r^2\nu^2 + \sigma_w^2\right) + \frac{\sigma_r^2 u_0^2}{\sigma_r^2 u_0^2 + \sigma_w^2} \\ &\quad + D_2 + C - \gamma P_{av}. \end{aligned}$$

The lower bound diverges to infinity as $\nu \to \infty$. Because $k_a(\nu)$ is an analytic function over $[0, \infty)$, according to Lemma 2.3, the support is discrete and finite when $\gamma$ satisfies (2.14).

Therefore, a sufficient condition for $\mathfrak{R}$ to be discrete and finite is $\gamma \neq \frac{m^2+\sigma_r^2}{\sigma_r^2 u^2+\sigma_w^2}$.

As we know, the asymptotic capacity of a Rician channel as the average power $P$ goes to zero is $\frac{m^2+\sigma_r^2}{\sigma_w^2} P + o(P)$. To achieve the asymptotic capacity as $P$ goes to zero, we need the kurtosis of the norm of the input to increase unboundedly [21].

Suppose that $\gamma = \frac{m^2+\sigma_r^2}{\sigma_r^2 u^2+\sigma_w^2}$.

Recall that $\gamma$ is the slope of the capacity versus average power curve. Then approaching the asymptotic capacity, i.e., $\gamma \to \frac{m^2+\sigma_r^2}{\sigma_w^2}$, requires $u^2 \to 0$. However, $u^2 \to 0$ prevents the kurtosis from growing unboundedly. Thus, there must exist a $P_0 > 0$, such that, for all $P_{av} \leq P_0$, the assumption $\gamma = \frac{m^2+\sigma_r^2}{\sigma_r^2 u^2+\sigma_w^2}$ can not be true, and thus $\mathfrak{R}$ is discrete and finite. The proposition follows. ■

### Peak Power Constraint

In [16]. Gursoy et al. have shown that, for a Rician channel with a peak power constraint, the support of an optimal amplitude distribution is discrete with a finite number of mass points.

In this subsection, we show an even more general result for channels with a peak power constraint.

Using Proposition 2.2, we obtain the KTC of a Rician channel with the peak power constraint $|x| \leq \sqrt{P_{pk}}$:

$$\begin{aligned} k_p(\nu) &= \int_0^\infty g(\rho|\nu) \log p_\rho(\rho; p^*) d\rho + 1 + C \\ &+ \ln\left(\sigma_r^2 \nu^2 + \sigma_w^2\right) \geq 0 \quad \forall \nu \in \left[0, \sqrt{P_{pk}}\right]. \end{aligned}$$

Notice that $k_p(\nu)$ is an analytic function.

**Proposition 2.6** *For a channel with an active peak power constraint, $\nu \leq \sqrt{P_{pk}}$, a sufficient condition for $\mathfrak{R}$ to be discrete with a finite number of mass points is that, $k_p(\nu)$ in the KTC is an analytic function over $[0, \infty)$. When the sufficient condition holds, $\mathfrak{R}$ has one mass point at $\sqrt{P_{pk}}$.*

**Proof:** The peak power constraint is active means that increasing or decreasing $P_{pk}$ will strictly increase or decrease the capacity.

Suppose $\mathfrak{R}$ contains an infinite number of mass points (including continuous cases).

Because there are an infinite number of points in the bounded interval, $\left[0, \sqrt{P_{pk}}\right]$, we can find an infinite sequence with a limit. According to the KTC, $k_p(\nu)$ equals to zero on the sequence. Since $k_p(\nu)$ is an analytic function over $[0, \infty)$ and equals to zero on an infinite sequence with a limit, it equals to 0 on $[0, \infty)$ according to the Identity Theory [20].

However, we can prove that there exists some $\nu > \sqrt{P_{pk}}$ such that $k_p(\nu) < 0$.

Suppose that all $\nu$ greater than $\sqrt{P_{pk}}$ have $k_p(\nu) \geq 0$. Then, we can find a $P'_{pk}$ larger than $P_{pk}$ such that in the interval $\left[\sqrt{P_{pk}}, \sqrt{P'_{pk}}\right]$, we have $k_p(\nu) \geq 0$ for all $\nu$. Let a new peak power constraint be $P'_{pk}$. Then, under the new peak power constraint,

the original input still satisfies the KTC, i.e., $k_p(\nu) \geq 0$ for all $\nu \in \left[0, \sqrt{P'_{pk}}\right]$. This means that, even with a larger peak power constraint, we still have the same optimal input, and thus the same capacity, which conflicts the fact that the peak power constraint is active. Thus, there must exist some $\nu > \sqrt{P_{pk}}$ such that $k_p(\nu) < 0$.

Since $k_p(\nu)$ does not equal to 0 on $[0, \infty)$, $\mathfrak{R}$ cannot contain an infinite number of mass points in $\left[0, \sqrt{P_{pk}}\right]$. The only possibility for $\mathfrak{R}$ is to be discrete and finite.

Now, we will prove that $\mathfrak{R}$ has one mass point at $\sqrt{P_{pk}}$.

Because $\mathfrak{R}$ is discrete with a finite number of mass points, we can index all the points,

$$0 \leq r_0 < r_1 < \cdots < r_N \leq \sqrt{P_{pk}}.$$

If $r_N < \sqrt{P_{pk}}$, i.e., $\mathfrak{R}$ has no mass point at $\sqrt{P_{pk}}$, then we can pick a new peak power constraint $P'_{pk}$ such that $r_N < \sqrt{P'_{pk}} < \sqrt{P_{pk}}$. With the smaller peak power constraint, we still have the same optimal input and thus the same capacity. As this contradicts the fact that the peak power constraint is active, the assumption $r_N < \sqrt{P_{pk}}$ cannot be true.

Hence, $\mathfrak{R}$ has one mass point at $\sqrt{P_{pk}}$. ■

Based on Proposition 2.6, we conclude that $\mathfrak{R}$ for a peak power limited Rician channel is discrete and finite, with a mass point at $\sqrt{P_{pk}}$.

In fact, a channel with a peak power constraint is a special case of channels with an average and peak power constraint, $E(\nu^2) \leq P_{av}$ and $\nu \leq \sqrt{P_{pk}}$, by letting $P_{av} = \infty$.

The KTC of a Rician channel with an average and peak power constraint is given by

$$\begin{aligned} k_{ap}(\nu) &= \int_0^\infty g(\rho|\nu) \log p_\rho(\rho; p^*) d\rho + \ln\left(\sigma_r^2 \nu^2 + \sigma_w^2\right) \\ &+ C + 1 + \gamma(\nu^2 - P_{av}) \geq 0 \quad \forall \nu \in \left[0, \sqrt{P_{pk}}\right]. \end{aligned}$$

$k_{ap}(\nu)$ is an analytic function, and according to Proposition 2.6, $\mathfrak{R}$ is discrete and finite, with a point at $\sqrt{P_{pk}}$.

Proposition 2.6 can be applied to any channel with an analytic Kuhn-Tucker

condition $k_p(\nu)$. As long as the peak power constraint is active, the capacity-achieving distribution is discrete with a finite number of mass points and one mass point is at $\sqrt{P_{pk}}$.

### 2.3.4 Numerical Results

In [19], Gallager studied the optimal distribution problem by considering the reliability function. By restricting power-limited channels to discrete inputs, he showed that binary inputs are optimal in the wideband limit (the low SNR limit). We have shown that the support of the optimal input amplitude distribution is discrete with a finite number of mass points. Hence, in the low SNR limit, the binary input distribution is optimal regardless the discrete input restriction. When we use numerical algorithms to search for the optimal input distribution in the low SNR regime, we can first start with binary inputs.

Without loss of generality, $\sigma_r^2 + m^2$ and $\sigma_w^2$ have been normalized to 1 for all numerical results.

In an average power constrained Rician channel, suppose $\sigma_r^2 = 10^{-2}$ and $P_{av} = 10^{-6}$. A numerical optimization indicates that the optimal binary amplitude pair is $(0, 2.371)$ with probability $(0.9999822, 0.0000178)$. Regular vector optimization methods, such as those introduced in [22], can be used to search the optimal support and probability.

In Figure 2-1, we plot $k_a(\nu)$ for the binary signaling. As we can see, $k_a(\nu) \geq 0$ for $\nu \geq 0$ and $k_a(\nu) = 0$ for $\nu = 0$ and 2.371, which means that the binary signaling is really the optimal input. The support $\mathfrak{R} = \{0, 2.371\}$ is discrete and finite.

During the numerical process, we first search for the optimal binary schemes with a zero symbol (one of the two points is at zero). By substituting the obtained optimal points into the Kuhn-Tucker condition, we can check if $k_a(\nu) \geq 0$ for $\nu \geq 0$. If satisfied, the binary scheme is optimal; otherwise, we increase input levels. This is because that the entropy of a binary symbol is limited by 1 bit, binary inputs are no longer optimal as the SNR increases. To achieve capacity, we need to consider schemes with more input levels.

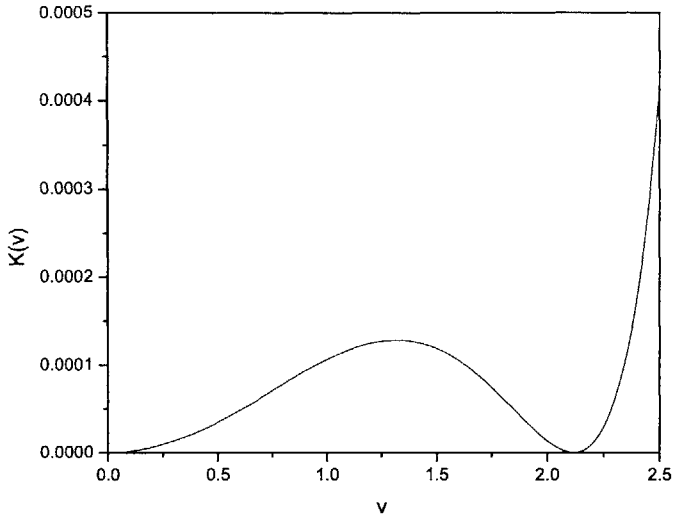


Figure 2-1: The Kuhn-Tucker condition $k(\nu)$ for $\sigma_r^2 = 10^{-2}$, $P_{av} = 10^{-6}$, $\lambda = 0.99015$, and $p(\nu) = 0.9999822\delta(0) + 0.0000178\delta(\nu - 2.371)$.

In Figure 2-2 and Figure 2-3, we plot the obtained location and probability of the nonzero mass point in the binary signaling for a range of average power. Different values of $\sigma_r$ correspond to different Rician channels.

Schemes with no point at zero are also included in our search algorithm by simply setting the probability of the zero symbol to zero. Optimal schemes in Rician channels subject to a peak power constraint can be searched by using a similar method.

However, when we need to search the optimal distributions with a large number of mass points, the computational task involved is formidable. We have to resort to suboptimal methods. In [23], Huang and Meyn introduced a class of algorithms based on the cutting-plane method, which generate discrete distributions that are optimal within a prescribed class.

## 2.4 Rayleigh Block Fading Channels

Discrete-time memoryless fading channel models are often used to model wireless channels that exhibit fast fading. For a channel experiencing slow fading, the fading coefficient changes slowly and remains almost the same for a period of time. Hence,

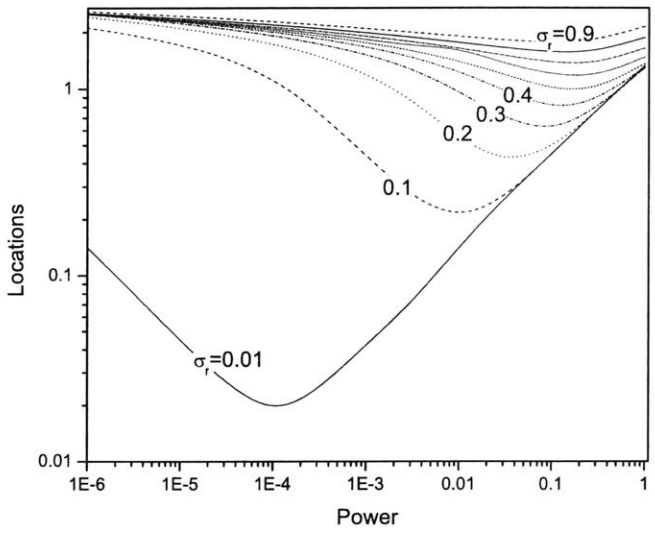


Figure 2-2: Location of the nonzero mass point

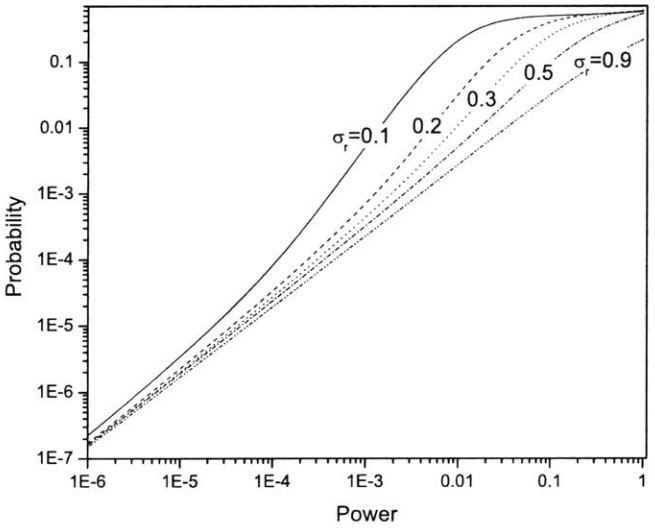


Figure 2-3: Probability of the nonzero mass point

to model the channel's behavior, the channel coherence should be taken into the consideration.

Block fading channel models take the channel coherence into account. In these channel models, the fading coefficient remains constant for a coherence time and changes to another value in the next coherence time. In the following section, we consider the Rayleigh block fading channel, which is a commonly used model for slow fading channels.

In this section, we obtain the Kuhn-Tucker condition for the Rayleigh block fading channel and show that, for the average power limited Rayleigh block fading channel, the support of the optimal amplitude distribution is discrete with a finite number of mass points. Moreover, for the peak power constrained channel, we also show that the support is discrete and finite.

### 2.4.1 Channel Model

We write the Rayleigh block fading channel between the complex input $x_t$ and the complex output $y_t$ as

$$y_t = hx_t + w_t$$

where the channel fading coefficient, $h$, is a complex Gaussian random variable, $CN(0,1)$. The additive white Gaussian noise samples, $\{w_t\}$, are independent, identically distributed (i.i.d.) complex Gaussian random variables, $CN(0,1)$. The subscript, $t$, indexes different time samples. The channel fading coefficient is invariant for a coherence time, $T$, and changes to another value in the next coherence time. Each value is independent, identically distributed. We assume that neither the receiver nor the transmitter knows the value of $h$, i.e., neither has channel state information.

We can consider symbols in the same coherence time altogether since they multiply the same fading coefficient, and write the channel model in a vector form for derivation convenience:

$$Y = hX + W$$

where $Y$, $X$, and $W$ are $T \times 1$ complex vectors. Each entry of $X$ is a symbol in the

coherence time.

With an average power constraint, $E(|X|^2) \leq P_{av}$, the vector form of the channel capacity over $T$ symbols is

$$C = \sup_{\substack{p_X(X) \\ E(|X|^2) \leq P_{av}}} \int \int p_{Y|X}(Y|X) \ln \frac{p_{Y|X}(Y|X)}{p_Y(Y; p_X^*)} dY p_X(X) dX.$$

To obtain the capacity with other constraints, we only need to change the subject function accordingly.

Since the channel exhibits strong correlation between symbols within a coherence time, the transmitter can detect information carried not only by symbols' amplitude, but also by the relative difference between symbols. To express this rigorously, we consider the vector of symbols in the same coherence time. The vector can be decomposed to be a unitary vector times a non-negative real number. The unitary vector indicates the direction of the original vector, and the non-negative real number is the vector's Euclidean norm. We call the unitary vector the vector's "phase" and the Euclidean norm its "amplitude". Clearly, information can be carried not only in the amplitude, but also in the phase.

In the following sections, we study the phase distribution and the amplitude distribution of the optimal input in the Rayleigh block fading channel.

### 2.4.2 Phase Distribution

According to [17], we can decouple the phase distribution and the amplitude distribution when we consider the optimal input in the Rayleigh block fading channel. The following proposition gives the optimal phase distribution.

**Proposition 2.7** *(Marzetta and Hochwald [17]) The optimal distribution for Rayleigh block fading channel should be a magnitude times an independent isotropically distributed phase, i.e. the input is*

$$X = r\Phi,$$

*where $r$ is a non-negative real number, and $\Phi$ is an independent isotropically distributed unitary vector.*

Please see [17] for the proof. Proposition 2.7 is the single antenna case in their result.

The received signal can be then written as

$$Y = hr\Phi + W = \sqrt{R}\Phi_y$$

where $R$ is a non-negative real number, and $\Phi_y$ is a $T \times 1$ complex unitary vector. Since $\Phi$ is isotropically distributed, it is straightforward to show that $\Phi_y$ is also isotropically distributed.

### 2.4.3 Amplitude Distribution

The mutual information between $X$ and $Y$ is calculated as

$$I(X;Y) = h(Y) - h(Y|X),$$

where $h(Y|X)$ and $h(Y)$ are given in the following discussion.

$h(Y|X)$: Conditional on $X = [x_1, x_2, \cdots, x_T]^T$, $Y$ is jointly Gaussian distributed with the density:

$$p(Y|X) = \frac{1}{\pi^T \det(\Sigma)} \exp\left(-Y^\dagger \Sigma^{-1} Y\right) \tag{2.16}$$

where $\Sigma$ is the correlation matrix:

$$\Sigma = \begin{pmatrix} |x_1|^2 + 1 & \cdots & x_1 x_T^* \\ \vdots & \ddots & \vdots \\ x_T x_1^* & \cdots & |x_T|^2 + 1 \end{pmatrix}.$$

The entropy of multivariate complex Gaussian random variable is $\ln(\pi e)^T \det(\Sigma)$. Hence, the conditional entropy is given by

$$h(Y|X) = \ln\left[(\pi e)^T \det(\Sigma)\right]. \tag{2.17}$$

$h(Y)$: To obtain $h(Y)$, it is much easier for us to use the polar coordinates.

Define $R = |Y|^2$ and $r = |X|$.

Because $\Phi$ is an isotropically distributed unitary vector independent with $r$, the joint distribution of $r$ and $\Phi$, $p_{r,\Phi}(r, \Phi)$, can be written as

$$p_{r,\Phi}(r, \Phi) = p_r(r)\frac{1}{|\Phi|}. \tag{2.18}$$

Moreover, since the input $X$ is a function of $r$ and $\Phi$, by the Change of Variables Theorem, we have

$$p_{r,\Phi}(r, \Phi) = r p_X(X(r, \Phi)). \tag{2.19}$$

Combining (2.18) and (2.19) gives us

$$p_X(X(r, \Phi)) = p_r(r)\frac{1}{r\,|\Phi|}. \tag{2.20}$$

Since $\Phi_y$ are also isotropically distributed, we have

$$\begin{aligned} & p_{R,\Phi_y}(R, \Phi_y; p_X^*) \\ = \; & p_{R|\Phi_y}(R|\Phi_y; p_X^*) p_{\Phi_y}(\Phi_y; p_X^*) \\ = \; & p_R(R; p_X^*)\frac{1}{|\Phi_y|}. \end{aligned} \tag{2.21}$$

where $p_{R|\Phi_y}(R|\Phi_y; p_X^*) = p_R(R; p_X^*)$ due to the homogeneity of the output on direction. By the Change of Variables Theorem, $p_{R,\Phi_y}(R, \Phi_y; p_X^*)$ can be expressed in the form of $p_Y(Y; p_X^*)$:

$$p_{R,\Phi_y}(R, \Phi_y; p_X^*) = \frac{1}{2} p_Y(Y(R, \Phi_y); p_X^*). \tag{2.22}$$

Thus, the equation

$$p_Y(Y(R, \Phi_y); p_X^*) = p_R(R; p_X^*)\frac{2}{|\Phi_y|} \tag{2.23}$$

is obtained by combining (2.21) and (2.22).

By the Change of Variables Theorem, $p_{R,\Phi_y|r,\Phi}(R, \Phi_y|r, \Phi)$ can be derived from

$p_{Y|X}(Y|X)$:

$$p_{R,\Phi_y|r,\Phi}(R,\Phi_y|r,\Phi) = \frac{1}{2} p_{Y|X}(Y(R,\Phi_y)|X(r,\Phi)). \tag{2.24}$$

Further, according to the Total Probability Theorem and the definition of marginal probability, we have

$$p_{R|r}(R|r) = \int\int \frac{p_{R,\Phi_y|r,\Phi}(R,\Phi_y|r,\Phi)}{|\Phi|} d\Phi_y d\Phi. \tag{2.25}$$

Hence, by using (2.20), (2.23), (2.24), and (2.25), the entropy of $Y$ is

$$\begin{aligned} h(Y) &= -\int\int p_{Y|X}(Y|X) \ln p_Y(Y;p_X^*) dY p_X^*(X)\, dX \\ &= -\int\int p_{R|r}(R|r) \ln p_R(R;p_X^*) dR p_r(r) dr + \ln \frac{|\Phi_y|}{2}. \end{aligned} \tag{2.26}$$

After we obtain $h(Y)$ and $h(Y|X)$, we calculate $I(X,Y)$. By substituting (2.17) and (2.26) into (2.27), the mutual information is given by

$$\begin{aligned} I(X;Y) &= \int \Bigg[ -\int p_{R|r}(R|r) \ln p_R(R;p_r^*) dR \\ &\quad - \ln(1+r^2) - \ln(\pi e)^T + \ln \frac{|\Phi_y|}{2} \Bigg] p_r^*(r) dr. \end{aligned} \tag{2.27}$$

In (2.27), $p_{R|r}(R|r)$ can be derived by using the fact that $p_{R|r}(R|r)$ does not depend on $\Phi$, which is obvious by (2.25). As a result, we can choose an input with an arbitrary direction during the derivation of $p_{R|r}(R|r)$.

We pick the input $X_0 = [r, 0, \cdots, 0]^T$ with $\Phi = [1, 0, \cdots, 0]^T$, which give us

$$\Sigma = \begin{pmatrix} r^2+1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$$

and

$$\det(\Sigma) = r^2 + 1 \tag{2.28}$$

in the conditional density (2.16). The inverse matrix of $\Sigma$ is

$$\Sigma^{-1} = \begin{pmatrix} \frac{1}{r^2+1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}.$$

By substitution $\Sigma$ into (2.16), the probability of $Y$ conditional on $X_0$ is

$$p\left(Y|X_0\right) = \frac{1}{\pi^T\left(r^2+1\right)} \exp\left(-\frac{|y_1|^2}{r^2+1} - |y_2|^2 - \cdots - |y_T|^2\right).$$

The form of this joint distribution implies that $\{y_i\}$ are independently distributed with densities:

$$p\left(y_1|X_0\right) = \frac{1}{\pi\left(r^2+1\right)} \exp\left(-\frac{|y_1|^2}{r^2+1}\right),$$

and

$$p\left(y_i|X_0\right) = \frac{1}{\pi} \exp\left(-|y_i|^2\right) \quad (i \neq 1).$$

By changing the variables, $\eta_i = |y_i|^2$, the densities become

$$p\left(\eta_1|r\right) = \frac{1}{\left(r^2+1\right)} \exp\left(-\frac{\eta_1}{r^2+1}\right),$$

and

$$p\left(\eta_i|r\right) = \exp\left(-\eta_i\right) \quad (i \neq 1).$$

Since $R = \sum_{i=1}^{T} \eta_i$ is the sum of independent random variables, we can obtain its characteristic function from the characteristic functions of $\{\eta_i\}$, which are

$$\phi_1\left(\omega\right) = \frac{1}{1 - j\omega\left(r^2+1\right)},$$

and

$$\phi_i\left(\omega\right) = \frac{1}{1-j\omega} \quad (i \neq 1).$$

The characteristic function of $p_{R|r}(R|r)$ is the product of the characteristic functions

of $\{\eta_i\}$, given by

$$\phi_R(\omega) = \frac{1}{[1 - j\omega(r^2+1)](1-j\omega)^{T-1}}. \tag{2.29}$$

Expanding the rational form of $\phi_R(\omega)$ gives

$$\begin{aligned}\phi_R(\omega) &= \frac{v_1}{1 - j\omega(r^2+1)} + \frac{\mu_1}{1-j\omega} \\ &\quad + \frac{\mu_2}{(1-j\omega)^2} + \cdots + \frac{\mu_{T-1}}{(1-j\omega)^{T-1}}\end{aligned}$$

where

$$\mu_{T-1} = -\frac{1}{r^2},$$

$$\mu_{T-2} = \frac{\mu_{T-1}(r^2+1)}{r^2} = -\frac{r^2+1}{r^4},$$

$$\mu_{T-3} = \frac{\mu_{T-2}(r^2+1)}{r^2} = -\frac{(r^2+1)^2}{r^6},$$

$$\cdots$$

$$\mu_1 = \frac{\mu_2(r^2+1)}{r^2} = -\frac{(r^2+1)^{T-2}}{r^{2(T-1)}},$$

and

$$v_1 = -\mu_1\left(r^2+1\right) = \frac{(r^2+1)^{T-1}}{r^{2(T-1)}}.$$

After taking the inverse transform of $\phi_R(\omega)$, $p_{R|r}(R|r)$ is obtained:

$$\begin{aligned}p_{R|r}(R|r) &= \sum_{i=1}^{T-1} \frac{\mu_i R^{i-1}}{\Gamma(i)} \exp(-R) \\ &\quad + \frac{\nu_1}{r^2+1} \exp\left(-\frac{R}{r^2+1}\right).\end{aligned} \tag{2.30}$$

### Average Power Constraint

By substituting the mutual information obtained in (2.27) into (2.1), we get the KTC for the average power limited Rayleigh block fading channel:

$$\begin{aligned} k_a(r) &= \gamma(r^2 - P_{av}) + C + \underbrace{\int p_{R|r}(R|r) \ln p_R(R; p_r^*) dR}_{(b)} \\ &+ \ln(1+r^2) + \ln(\pi e)^T - \ln \frac{|\Phi_y|}{2} \geq 0. \end{aligned} \tag{2.31}$$

Notice that, the average power constraint, $E(r^2) \leq P_{av}$, imposed for the block fading channel, is an power constraint on the vector $X$. It is the second moment constraint for $r = |X|$. It is equivalent to an average power constraint on a single symbol with the upper bound $\frac{P_{av}}{T}$. Hence, results obtained for the first constraint case are applicable for the second constraint case.

By using the KTC, we would be able to prove that the optimal input amplitude distribution for the average power limited Rayleigh block fading channel is discrete with a finite number of mass points.

First, we prove that, the optimal input amplitude distribution is bounded.

**Lemma 2.4** *With the average power constraint, $E(r^2) \leq P_{av}$, and the time coherence, $T < \infty$, the optimal input amplitude distribution for Rayleigh block fading channel has a bounded support.*

**Proof:** Let's assume the support $\Re$ is not bounded, and later we will show that, the result derived based on this assumption will violate the KTC.

Based on (2.30), $p_{R|r}(R|r)$ can be rewritten as

$$p_{R|r}(R|r) = \exp\left(-\frac{R}{r^2+1}\right) f_1(R, r),$$

where

$$f_1(R, r) = \sum_{i=1}^{T-1} \frac{\mu_i R^{i-1}}{\Gamma(i)} \exp\left(-\frac{r^2 R}{r^2+1}\right) + \frac{\nu_1}{r^2+1}.$$

Substitute $\mu_i$'s and $\nu_1$ into $f_1(R,r)$:

$$\begin{aligned} f_1(R,r) &= -\sum_{i=1}^{T-1} \frac{(r^2+1)^{T-i-1} R^{i-1}}{\Gamma(i) r^{2(T-i)}} \exp\left(-\frac{r^2 R}{r^2+1}\right) + \frac{(r^2+1)^{T-2}}{r^{2(T-1)}} \\ &= \left[1 - \underbrace{\sum_{i=0}^{T-2} \frac{\left(\frac{r^2 R}{r^2+1}\right)^i}{i!} \exp\left(-\frac{r^2 R}{r^2+1}\right)}_{(c)}\right] \frac{(r^2+1)^{T-2}}{r^{2(T-1)}}. \end{aligned} \tag{2.32}$$

Notice that $(c)$ in (2.32) is the cumulative probability of a Poisson random variable. By definition, the cumulative probability of the Poisson distribution, $p_v(i) = \frac{v^i}{i!}\exp(-v)$, is

$$P_v(i<k) = \sum_{i=0}^{k-1} \frac{v^i}{i!} \exp(-v).$$

It can be rewritten as

$$\begin{aligned} P_v(i<k) &= \frac{1}{\Gamma(k)} \int_v^\infty x^{k-1} e^{-x} dx \\ &= \frac{\Gamma(k,v)}{\Gamma(k)}, \end{aligned} \tag{2.33}$$

where $\Gamma(k,v)$ is the upper incomplete gamma function given by $\Gamma(k,v) = \int_v^\infty x^{k-1}e^{-x}dx$.

By substituting (2.33) into (2.32), we obtain

$$f_1(R,r) = \frac{(r^2+1)^{T-2}}{r^{2(T-1)}\Gamma(T-1)} \int_0^{\frac{r^2 R}{r^2+1}} x^{T-2} e^{-x} dx,$$

which, after change of the variable in the integral, becomes

$$f_1(R,r) = \frac{(r^2+1)^{T-2}}{r^{2(T-1)}\Gamma(T-1)} \int_0^{\frac{r^2}{r^2+1}} R^{T-1} x^{T-2} e^{-xR} dx.$$

By using some real analysis arguments, it can be proven that, for any real number $\gamma$ satisfying $0 < \gamma \leq 1$, we can find a positive number $b$, such that for $\forall r > b$, the inequalities $\frac{r^2}{r^2+1} > \frac{\gamma}{2} > \frac{1}{b^2+1}$ hold. Based on the inequalities, $f_1(R,r)$ is lower

bounded by

$$\begin{aligned} f_1(R,r) &\geq \frac{(r^2+1)^{T-2}}{r^{2(T-1)}\Gamma(T-1)} \int_0^{\frac{\gamma}{2}} R^{T-1} x^{T-2} e^{-xR} dx \\ &\geq \frac{(r^2+1)^{T-2} e^{-\frac{\gamma R}{2}} R^{T-1}}{r^{2(T-1)}\Gamma(T-1)} \int_0^{\frac{\gamma}{2}} x^{T-2} dx. \end{aligned} \tag{2.34}$$

Since we assume that the support $\Re$ is not bounded, the following inequalities must be true:

$$\int_b^{\infty} p_r^*(r) dr > 0,$$

and

$$\int_0^{\infty} p_{R|r}(R|r) p_r^*(r) dr > \int_b^{\infty} p_{R|r}(R|r) p_r^*(r) dr. \tag{2.35}$$

Hence, for any $r > b$, the part denoted as $(b)$ in (2.31) can be lower bounded as

$$\begin{aligned} (b) &\geq \int_0^{\infty} p_{R|r}(R|r) \ln \left[ \int_b^{\infty} p_{R|r}(R|a_x) p_r^*(a_x) da_x \right] dR \\ &= \int_0^{\infty} p_{R|r}(R|r) \ln \left[ \int_b^{\infty} \exp\left( -\frac{R}{a_x^2+1} \right) f_1(R, a_x) p_r^*(a_x) da_x \right] dR \\ &\geq \int_0^{\infty} p_{R|r}(R|r) \ln \left[ \exp\left( -\frac{R}{b^2+1} - \frac{\gamma R}{2} \right) R^{T-1} \right] dR + D_3 \\ &= -\frac{r^2+T}{b^2+1} - \frac{\gamma(r^2+T)}{2} + D_3 + o\left(r^2\right). \end{aligned}$$

The first inequality is based on (2.35), and the second inequality is from (2.34). The term $D_3$ is given by

$$D_3 = \ln \left[ \int_0^{\frac{\gamma}{2}} x^{T-2} dx \int_b^{\infty} \frac{(a_x^2+1)^{T-2}}{a_x^{2(T-1)}\Gamma(T-1)} p_r^*(a_x) da_x \right]$$

and the term $o(r^2)$ represents a function growing much more slowly than $r^2$:

$$\int_0^{\infty} p_{R|r}(R|r) \ln R^{T-1} dR = o\left(r^2\right).$$

By substituting the lower bound we got for the part $(b)$ in (2.31), we obtain a

lower bound of $k_a(r)$ when $r > b$:

$$\begin{aligned} k_a(r) &\geq \gamma(r^2 - P_{av}) + C - \frac{r^2+T}{b^2+1} - \frac{\gamma(r^2+T)}{2} \\ &\quad + D_3 + o\left(r^2\right) + \ln(1+r^2) + \ln(\pi e)^T - \ln\frac{|\Phi_y|}{2}. \end{aligned} \tag{2.36}$$

In (2.36), $D_3$ is a finite real number determined by $p_r^*(r)$. Hence, the right hand side of (2.36) is of the order of $\left(\frac{\gamma}{2} - \frac{1}{b^2+1}\right) r^2$ as $r \to \infty$. Recall that $b$ is so picked that $\frac{\gamma}{2} > \frac{1}{b^2+1}$. Thus, the lower bound diverges as $r \to \infty$. This lower bound becomes active when the condition $r > b$ holds (which is true as $r \to \infty$). Hence, $k_a(r)$ diverges to infinity as $r \to \infty$, and $k_a(r) \neq 0$ for large enough $r$. This implies the support of the optimal input is bounded, contradicting the assumption that the support is not bounded. ■

Based on Lemma 2.4, we prove the main result of this section:

**Proposition 2.8** *With the average power constraint, $E(r^2) \leq P_{av}$, and the time coherence, $T < \infty$, the optimal input amplitude distribution of a Rayleigh block fading channel has a discrete support with a finite number of mass points.*

**Proof:** To show the optimal distribution has a finite number of mass points, we use a proof by contradiction.

Assume there are an infinite number of mass points in the distribution. Obviously, continuous supports are included in this case. Then by Lemma 2.4, the support is bounded. We have an infinite number of mass points in a bounded interval. It is easy to show that, out of these points, we can construct an infinite sequence with a limit. Since $k_a(r)$ is analytic and equals to zero on the infinite sequence, according to the Identity Theorem, $k_a(r) = 0$ on the whole domain $[0, \infty)$.

By substituting (2.30) into (2.31) and letting $k_a(r) = 0$, we attain

$$\begin{aligned} & -\sum_{i=1}^{T-1} \frac{\left(r^2+1\right)^{T-1-i}}{\Gamma(i)\, r^{2(T-i)}} \int R^{i-1} \exp(-R) \ln p_R(R; p_r^*) dR \\ + & \int \frac{\left(r^2+1\right)^{T-2}}{r^{2(T-1)}} \exp\left(-\frac{R}{r^2+1}\right) \ln p_R(R; p_r^*) dR \\ + & \gamma(r^2 - P_{av}) + C + \ln(1+r^2) + \ln(\pi e)^T - \ln \frac{|\Phi_y|}{2} = 0. \end{aligned} \tag{2.37}$$

We define a new variable $s$:

$$s = \frac{1}{r^2+1}.$$

Because $r \geq 0$, $s$ is located in the interval $(0, 1]$.

Substitute $r^2 = \frac{1}{s} - 1$ into (2.37):

$$\begin{aligned} & -\sum_{i=1}^{T-1} \frac{s}{\Gamma(i)\,(1-s)^{T-i}} \int R^{i-1} \exp(-R) \ln p_R(R; p_r^*) dR \\ & + \int \frac{s}{(1-s)^{T-1}} \exp(-Rs) \ln p_R(R; p_r^*) dR \\ & + \gamma\left(\frac{1}{s} - 1 - P_{av}\right) + C - \ln s + \ln(\pi e)^T - \ln \frac{|\Phi_y|}{2} = 0. \end{aligned} \tag{2.38}$$

Move all terms on the left hand side of (2.38) to the right hand side except the second term, and multiply $\frac{(1-s)^{T-1}}{s}$ on both sides:

$$\begin{aligned} & \int \exp(-Rs) \ln p_R(R; p_r^*) dR \\ = & -\frac{\gamma(1-s)^{T-1}}{s^2} + \frac{(1-s)^{T-1} \ln s}{s} \\ & + \frac{(1-s)^{T-1}}{s} \left[\gamma(1+P_{av}) - C - \ln(\pi e)^T + \ln \frac{|\Phi_y|}{2}\right] \\ & + \sum_{i=1}^{T-1} \frac{(1-s)^{i-1}}{\Gamma(i)} \int R^{i-1} \exp(-R) \ln p_R(R; p_r^*) dR. \end{aligned} \tag{2.39}$$

Notice that, on the right hand side of (2.39),

$$\gamma(1+P_{av}) - C - \ln(\pi e)^T + \ln \frac{|\Phi_y|}{2}$$

in the second term and

$$\int R^{i-1} \exp(-R) \ln p_R(R; p_r^*) dR,$$

in the third term are constants. Thus, we can rearrange and write the RHS of (2.39) as

$$\begin{aligned} \text{RHS} &= -\gamma s^{-2} + a_0 s^{-1} + a_1 + a_2 s + \cdots + a_{T-1} s^{T-2} \\ &\quad + \frac{\ln s}{s} + b_1 \ln s + b_2 s \ln s \\ &\quad + \cdots + b_{T-1} s^{T-2} \ln s \end{aligned}$$

where $a_i$ and $b_i$ are coefficients determined by the input distribution and channel parameters. We can treat them as constants here because their values will not change with $s$. In the follow derivation, we will only list the value of a coefficient whose value matters in our proof.

The LHS of (2.39) is the Laplace transform of $\ln p_R(R; p_r^*)$. We use inverse Laplace transform on both sides:

$$\begin{aligned} \ln p_R(R; p_r^*) &= a_0' + a_1' \delta(R) + a_2' \delta^{(1)}(R) + \cdots + a_{T-1}' \delta^{(T-2)}(R) \\ &\quad -\gamma R + (-C_E - \ln R) + b_1' \frac{d}{dR}(\ln R) \\ &\quad + b_2' \frac{d^2}{dR^2}(\ln R) + \cdots + b_{T-1}' \frac{d^{T-1}}{dR^{T-1}}(\ln R) \end{aligned}$$

where $C_E$ is Euler's constant, $a_i'$ and $b_i'$ are constant coefficients. Taking exponential on both sides yields:

$$\begin{aligned} p_R(R; p_r^*) &= \frac{1}{R} \exp\left[a_0' + a_1' \delta(R) + a_2' \delta^{(1)}(R) + \cdots \right. \\ &\quad + a_{T-1}' \delta^{(T-2)}(R) - \gamma R - C_E \\ &\quad \left. + \frac{b_1''}{R} + \frac{b_2''}{R^2} + \cdots + \frac{b_{T-1}''}{R^{T-1}}\right] \end{aligned}$$

where $b_i''$ are constants, and the value of $b_{T-1}''$ is

$$b_{T-1}'' = \Gamma(T-1).$$

Now, we calculate $\int_{0^+}^{1} p_R(R; p_r^*)dR$. We don't need to consider $\delta(R), \ldots, \delta^{(T-2)}(R)$ in the exponent of $p_R(R; p_r^*)$, because they are only nonzero at $R = 0$. In the rest of terms, $\frac{b_{T-1}''}{R^{T-1}}$ will dominate other terms when $R \to 0^+$. Since $b_{T-1}'' > 0$ and

$$\int_{0^+}^{1} \frac{1}{R} \exp\left[\frac{1}{R^{T-1}}\right] dR = \infty,$$

$\int_{0^+}^{1} p_R(R; p_r^*)dR$ diverges.

We have

$$\int_0^{\infty} p_R(R; p_r^*)dR \geq \int_{0^+}^{1} p_R(R; p_r^*)dR = \infty.$$

This conflicts with the fact that $p_R(R; p_r^*)$ is a probability density.

Hence, the assumption that there are an infinite number of mass points in the bounded support is not true. The only possible structure for the support is discrete with a finite number of mass points. ■

Note that, in the proof of the proposition, we assume the coherence time $T < \infty$. Otherwise, if $T = \infty$, the fading coefficients can be estimated precisely from previous transmissions, and both the transmitter and the receiver have the channel state information. Then the optimal input would be Gaussian distribution, the same as that in the average power limited AWGN channel.

### Peak Power Constraint

Peak power constraints for a Rayleigh block fading channel could be imposed because of a maximum amplitude limit for each symbol or a maximum norm limit for each symbol vector. The former limit is commonly used because transmitters and receivers have physical limits on the maximum amplitude they can handle. The latter limit makes sense when the transmitter does joint power detection over a coherence time and the power combiner has a maximum limit. In the first peak power limit case, the

optimal input phase distribution is no longer isotropic, and we do not have a tractable KTC. Hence, in the following discussion, we will use the second peak power limit. Although the peak power constraint on a vector of symbols is not as common as the peak power constraint on a symbol, it is still valuable for study because it provides us insights into transmissions in a channel imposed a limit on symbols' amplitude.

By substituting the mutual information obtained in (2.27) into (2.2), we get the KTC for the peak power limited Rayleigh block fading channel:

$$\begin{aligned} k_p(r) &= C + \int p_{R|r}(R|r) \ln p_R(R; p_r^*) dR \\ &+ \ln(1+r^2) + \ln(\pi e)^T - \ln \frac{|\Phi_y|}{2} \geq 0. \end{aligned} \tag{2.40}$$

Since there exists a peak power constraint, the amplitude is bounded. We prove the following proposition:

**Proposition 2.9** *With a peak power constraint, $r^2 \leq P_{pk}$, the capacity-achieving input density of a Rayleigh block fading channel has a discrete support with a finite number of mass points.*

**Proof:** To show the optimal distribution has a finite number of mass points, we use a proof by contradiction.

Assume there are an infinite number of mass points in the distribution. Clearly, continuous supports are included in this case. Owing to the peak power constraint, the support is bounded. We have an infinite number of mass points in a bounded interval. From these mass points, we can construct an infinite sequence with a limit. Since $k_p(r)$ is analytic and equals to zero on the infinite sequence, according to the Identity Theorem, $k_p(r) = 0$ on the whole domain $[0, \infty)$.

By substituting (2.30) into (2.40) and letting $k_p(r) = 0$, we obtain

$$\begin{aligned} &- \sum_{i=1}^{T-1} \frac{(r^2+1)^{T-1-i}}{\Gamma(i)\, r^{2(T-i)}} \int R^{i-1} \exp(-R) \ln p_R(R; p_r^*) dR \\ &+ \int \frac{(r^2+1)^{T-2}}{r^{2(T-1)}} \exp\left(-\frac{R}{r^2+1}\right) \ln p_R(R; p_r^*) dR \\ &+ C + \ln(1+r^2) + \ln(\pi e)^T - \ln \frac{|\Phi_y|}{2} = 0. \end{aligned} \tag{2.41}$$

We define a new variable $s$:

$$s = \frac{1}{r^2+1}.$$

Because $r \geq 0$, $s$ is located in the interval $(0, 1]$.

Substituting $r^2 = \frac{1}{s} - 1$ into (2.41), rearranging the terms, taking an inverse Laplace transform, and doing an exponentiation, we obtain:

$$\begin{aligned} &p_R(R; p_r^*) = \\ &\frac{1}{R} \exp\left[c_0 + c_1 \delta(R) + c_2 \delta^{(1)}(R) + \cdots + c_{T-1} \delta^{(T-2)}(R)\right. \\ &\left. -C_E + \frac{d_1}{R} + \frac{d_2}{R^2} + \cdots + \frac{d_{T-1}}{R^{T-1}}\right] \end{aligned}$$

where $c_i$ and $d_i$ are constants, and

$$d_{T-1} = \Gamma(T-1).$$

Now, we calculate $\int_{0^+}^{1} p_R(R; p_r^*) dR$. We do not need to consider $\delta(R)$, ... , $\delta^{(T-2)}(R)$ in the exponent of $p_R(R; p_r^*)$, because they are only nonzero at $R = 0$. In the rest of terms, $\frac{d_{T-1}}{R^{T-1}}$ will dominate other terms when $R \to 0^+$. Since $d_{T-1} > 0$ and

$$\int_{0^+}^{1} \frac{1}{R} \exp\left[\frac{1}{R^{T-1}}\right] dR = \infty,$$

$\int_{0^+}^{1} p_R(R; p_r^*) dR$ diverges.

We have

$$\int_0^{\infty} p_R(R; p_r^*) dR \geq \int_{0^+}^{1} p_R(R; p_r^*) dR = \infty.$$

This conflicts with the fact that $p_R(R;p_r^*)$ is a probability density. Hence, the assumption that there are an infinite number of mass points is not true. The only possible structure for the optimal input amplitude is discrete with a finite number of mass points. ■

For Rayleigh block fading channels with with an average power constraint $E(r^2) \leq P_{av}$ and a peak power $r \leq \sqrt{P_{pk}}$, we can also prove the optimal input amplitude distribution is discrete with a finite number of mass points. The proof is similar to that of the average power constraint case.

By substituting the mutual information obtained in (2.27) into (2.3), we get the KTC for the average power and peak power limited Rayleigh block fading channel:

$$\begin{aligned} k_{ap}(r) &= \gamma(r^2 - P_{av}) + C + \int p_{R|r}(R|r) \ln p_R(R;p_r^*) dR \\ &+ \ln(1+r^2) + \ln(\pi e)^T - \ln \frac{|\Phi_y|}{2} \geq 0. \end{aligned}$$

By assuming that the support has an infinite number of mass points, we construct a contradiction. Thus, the only possible structure for the optimal input amplitude is discrete with a finite number of mass points.

## 2.5 Conclusions

We have considered the capacity-achieving distributions of memoryless Rician fading channels and Rayleigh block fading channels. We have shown that, subject to an average power constraint, the optimal input amplitude distribution for Rician fading channels is discrete with a finite number of mass points in the low SNR regime. For memoryless Rician fading channels subject to a peak power constraint, or Rayleigh block fading channels subject to average power and peak power constraints, the optimal input amplitude distribution is discrete with a finite number of mass points. Moreover, for peak power limited channels, we provides a sufficient condition for the optimal input amplitude distribution to be discrete with a finite number of mass points. According to this rule, most channels subject to a peak power constraint

have discrete and finite optimal input amplitude distributions.

Our results, altogether with the results in [10 12, 15], lead to the conclusion that the continuity of the optimal distribution for AWGN channel is in fact a very special case. For practical wireless communications, the assumption that the channel has no fading and the signal have no peak power constraint is unrealistic. Thus, searching constellations with discrete and finite amplitude levels can actually lead to optimal solutions. In particular, combined with Gallager's result [19], our results show that the binary input distribution is optimal in the low SNR regime.

These results provide basis for two research areas: optimal input distribution searching and optimal signaling scheme designing. The discrete and finite characteristic of optimal input distributions makes numerical algorithm a possible candidate for optimal input searching. Also, knowledge on the optimal input distribution helps us in designing capacity-achieving signaling schemes. For example, for a wideband channel, degrees of freedom ($2WT$) is very large and energy per degree of freedom is very small (the degree of freedom of a channel is determined by how many samples we can obtain according to the Nyquest Sampling Theorem). By the results provided in this chapter, we know that the binary input distribution is optimal, which leads to on-off signaling on each sample (degree of freedom).

In the next chapter, we propose a family of capacity-achieving schemes in which binary amplitude distribution is used for each sample.

# Chapter 3

# Multi-tone Frequency-Shift Keying

## 3.1 Introduction

The study of communication in wideband fading channels dates back to the 1960's. In [1], Kennedy has shown that, for a given average received power, the capacity of a Rayleigh fading channel has the same limit as that of an additive white Gaussian noise channel as the bandwidth approaches infinity. Telatar and Tse [5] generalized this result and showed that Shannon's wideband capacity limit [2] is indeed achievable with any distribution of the channel fading coefficients.

A signaling scheme used in [4, §8.6] and [5] to achieve the capacity of fading channels at the infinite bandwidth limit, is a special *impulsive*, or "flash", FSK scheme, which transmits FSK signals with a low duty cycle and a high peak power. The flash FSK signaling does not spreads the signal energy evenly over the available bandwidth. Instead, its signal energy is peaky both in time and frequency. Because each symbol in the scheme has only one frequency, we call the scheme *Single-tone FSK*, compared to the Multi-tone FSK scheme that we will discuss later.

In [21], Verdú has shown that, in order to achieve the capacity of a wideband non-coherent fading channel, signaling must be peaky. In particular, let us consider, as in [21], the Taylor series expansion of capacity at vanishing signal-to-noise ratio

per degree of freedom, corresponding to the wideband scenarios,

$$C(\mathsf{SNR}) = \dot{C}(0)\mathsf{SNR} + \ddot{C}(0)\mathsf{SNR}^2 + o(\ \mathsf{SNR}^2),$$

where $\mathsf{SNR}$ is the SNR per degree of freedom, $\dot{C}(0)$ and $\ddot{C}(0)$ are the first and the second derivatives of the function $\mathsf{SNR} \mapsto C(\mathsf{SNR})$ at $\mathsf{SNR} = 0$. For a general class of channels, when CSI is not available at the receiver, flash signals are necessary in order to achieve the first order optimality, $\dot{C}(0)\mathsf{SNR} = \mathsf{SNR}$, which is the Shannon's wideband limit for coherent channels.

A number of flash signaling schemes are first order optimal. They achieve essentially the same capacity limit as the bandwidth approaches infinity. However, if one is interested in the performance at a large but finite bandwidth, it is not clear whether or not the above asymptotic analysis provides a close approximation. In fact, different signaling schemes may have different constraints at a finite bandwidth.

For Single-tone FSK, the data rate at a finite bandwidth, or equivalently as the SNR per degree of freedom increases from 0, is quickly limited by $\ln M$ (nats/symbol), where $M$ is the size of alphabet, i.e., the number of total frequency points. Some results concerning the practicality of Single-tone FSK in finite bandwidths are even more discouraging. From a peak power perspective, Lun et al. [27] [28] have shown that codeword error probability of the Single-tone FSK scheme decreases roughly inversely with bandwidth, leading to the need for very high peak power tones for a small codeword error probability or very large bandwidth. From a capacity perspective, Verdú [21] has shown that, for flash schemes, the second derivative in the Taylor series expansion of capacity goes to $-\infty$ as $\mathsf{SNR} \to 0$, so that the wideband capacity limit is approached extremely slowly as the bandwidth increases.

The Multi-tone FSK scheme discussed in this chapter can achieve the infinite bandwidth capacity limit, and more importantly, it provides flexibility on spectral efficiency such that the achievable data rate region for Multi-tone FSK is broader than that for Single-tone FSK. For the same bandwidth, the number of symbols in Multi-tone FSK is more than that in the Single-tone FSK scheme. Thus, at a finite

bandwidth. Multi-tone FSK can do better if the data rate is mainly limited by the size of alphabet.

Multi-tone FSK signaling schemes include Single-tone FSK as a special case, and the results derived for Multi-tone FSK are more generalized than those for Single-tone FSK. We characterize the performance of Multi-tone FSK at both the wideband limit and a finite bandwidth.

- We generalize the results of [29] to derive an upper bound and a lower bound of the codeword error probability for Multi-tone FSK. These two bounds lead to the error exponent of Multi-tone FSK, which depicts the effect of bandwidth, data rate, number of tones, and duty cycle of the scheme on the error probability.

- We present analytical performance evaluation and numerical results to show that simple Single- and Multi-tone FSK schemes with hard decision can yield rates approaching closely the wideband capacity limit at large but finite bandwidths, although we know that attaining rates arbitrarily close to capacity remains elusive.

The capacity of certain FSK schemes has been studied in other contexts. In [30], Stark determined the capacity of FSK schemes under non-selective Rician fading with receiver side information. In [31], non-coherent FSK was considered for Rayleigh fading channels with erasures. We consider instead an FSK scheme with duty cycle. The channel model we used is a frequency selective block fading model with no CSI.

This chapter is organized as follows: In Section 3.2, we introduce our channel model and the Multi-tone FSK scheme, calculate the bounds of error probability, and derive the capacity limit and the error exponent for Multi-tone FSK. Finally, we present numerical results to show the performance of single- and Multi-tone FSK at finite bandwidths. In Section 3.3, we introduce the Multi-tone FSK scheme with receiver diversity, calculate the bounds of error probability, and derive the capacity limit and the error exponent. In Section 3.4, we discuss Multi-tone FSK's bandwidth occupancy and its multiple access ability. Section 3.5 contains our conclusions.

## 3.2 Multi-tone Frequency-Shift Keying

### 3.2.1 Concept

In FSK systems, different symbols are represented by sinusoid waveforms tuned to different specific frequencies. For the Multi-tone FSK scheme, we use different combinations of multiple frequencies as symbols. For example, from a pool of $M$ tones, i.e., $M$ mutually orthogonal frequencies over a symbol time, we can use different combinations of $Q$ tones to represent different symbols. We fix the number $Q$ of concurrent tones that form a symbol. The scheme using $Q$ tones is called the *$Q$-tone FSK scheme*. Particularly, when $Q$ equals to 1, the scheme is Single-tone FSK, the flash FSK scheme used in [5] and [27].

In the Multi-tone FSK scheme, transmissions take place in a low duty cycle fashion. The transmitter concentrates power over a fraction $\theta$ ($0 < \theta \leq 1$) of time, and transmits on predetermined symbol slots. We refer to $\theta$ as the *duty cycle*. The main reason for using a low duty cycle is that transmissions with insufficient power over fading environments are not desirable, and that the channels we consider are power-limited with respect to the number of degrees of freedom. Therefore, we purposedly use a low proportion of the degrees of freedom for transmission, leading to a low duty cycle. A repetition code of length $N$ ($N \geq 1$) is employed, i.e., each symbol will be repeated $N$ times.

At the receiver side, we use a non-coherent receiver to demodulate and decode Multi-tone FSK signals. The receiver employs a bank of matched filters with their central frequencies tuned to each of the $M$ tones, respectively. The output of each matched filter is the correlation of its tuned frequency against the received Multi-tone FSK signal. Since the $M$ tones are mutually orthogonal over a symbol time, the output of a matched filter is proportional to the amplitude of the associated tone. Thus, we can view the $M$ tones as $M$ frequency-division subchannels. For each subchannel, the time average of received power over the $N$ consecutive symbol slots is obtained to compare with a threshold. If exactly $Q$ tones exceed the threshold, then the corresponding symbol is decoded; otherwise, the receiver declares an error.

Note that, we do not seek an optimal detection/decoding rule here, which may lead to a better performance but more elusive analysis. Our goal is to provide a feasible scheme with first order optimality, which achieves capacity in the wideband regime.

### 3.2.2 Mathematical Model

Consider a multipath fading channel with input $x(t)$. The output $y(t)$ is given by

$$y(t) = \sum_{l=1}^{L} a_l(t)x(t - d_l(t)) + z(t), \tag{3.1}$$

where $L$ is the total number of paths, the random processes $a_l(t)$ and $d_l(t)$ are the gain and the delay of the $l$th path, and $z(t)$ is complex white Gaussian noise with power spectral density $N_0/2$.

Wireless channels change both in time and frequency. The time coherence, $\mathcal{T}_c$, shows us how quickly the channel changes in time, and, similarly, the frequency coherence, $\mathcal{F}_c$, shows how quickly it changes in frequency. The selective fading in frequency is caused by the different delays of path. The frequency coherence is reciprocal to the multipath spread, $\mathcal{L}$. The major effect in determining time coherence is the Doppler shift, $\mathcal{D}$, which causes significant changes in channel gain. The relationship between the time coherence and the Doppler shift is reciprocal.

Assuming that the multipath spread $\mathcal{L}$ is much less than the time coherence $\mathcal{T}_c$ (an under-spread channel), and the gain and the delay of a path are constant within each coherence block and change independently from block to block (block-fading), we obtain a frequency-selective block fading channel model (see Figure 3-1).

If two symbols are transmitted within the same square, they have the same fading coefficient; otherwise, they experience independent fading.

For $Q$-tone FSK, since there are $\binom{M}{Q}$ possible $Q$-tone combinations from $M$ tones, the number of total symbols is $\binom{M}{Q}$. We let $\mathcal{S}$ denote the complete set of symbols, and $\mathcal{S}_m$ a symbol in the set, i.e., $\mathcal{S}_m \in \mathcal{S}$. That the $k$th tone is non-zero in a symbol $\mathcal{S}_m$ is denoted as $k \in \mathcal{S}_m$.

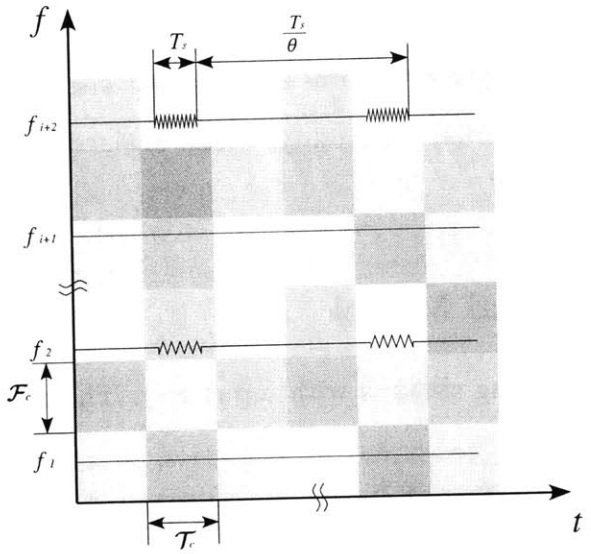


Figure 3-1: Block fading channel demonstrated in a frequency-time plane

Assume the symbol time is $T_s$, which satisfies $\mathcal{L} \ll T_s \leq \mathcal{T}_c$ (this is possible because of the underspread assumption). To send a symbol $\mathcal{S}_m$, $x(t)$ is given by

$$x(t) = \sum_{k \in \mathcal{S}_m} \exp(j2\pi f_k t), \qquad 0 \leq t \leq T_s, \tag{3.2}$$

where $j = \sqrt{-1}$ and $\{f_k\}$ are $M$ frequencies orthogonal over $[\mathcal{L}, T_s]$. Let $T_s'$ denote $T_s - \mathcal{L}$. For a system bandwidth $W$, there exist $M = WT_s'$ frequencies mutually orthogonal over $[\mathcal{L}, T_s]$ which are integer multiples of $1/T_s'$.

Let us consider the channel output over the interval $[\mathcal{L}, T_s]$. Owing to the block fading channel assumption, the gains $\{a_{l,k}(t)\}$ and the delays $\{d_{l,k}(t)\}$ are constant during this interval, denoted as $a_{l,k}$ and $d_{l,k}$, where the subscripts, $l$ and $k$, are the indices of path and tone, respectively. Hence, by (3.1), the received signal over $[\mathcal{L}, T_s]$ is

$$\begin{aligned} y(t) &= \sum_{k \in \mathcal{S}_m} \sum_{l=1}^{L} a_{l,k} \exp(j2\pi f_k (t - d_{l,k})) + z(t) \\ &= \sum_{k \in \mathcal{S}_m} g_k \sqrt{\frac{P}{Q\theta}} \exp(j2\pi f_k t) + z(t), \qquad \mathcal{L} \leq t \leq T_s, \end{aligned} \tag{3.3}$$

where $P$ denotes the average received signal power, and $g_k$ the aggregated channel gains for each tone,

$$g_k = \sum_{l=1}^{L} a_{l,k} \sqrt{\frac{Q\theta}{P}} \exp(-j2\pi f_k d_{l,k}).$$

We assume that the number of paths is very large and there is no dominant path. Based on this assumption, $g_k$ are zero-mean complex Gaussian random variables according to the Central Limit Theorem. Their variances have been normalized, i.e., $E[|g_k|^2] = 1$, by putting the average received power term alone in (3.3).

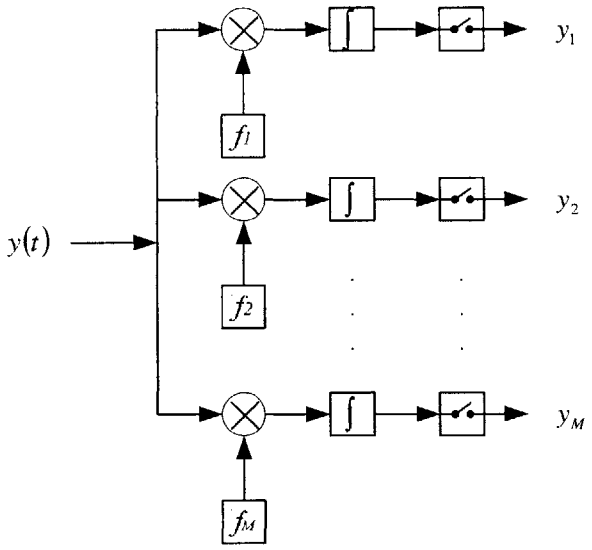


Figure 3-2: Matched filters used to demodulate Multi-tone FSK symbols

At the receiver, the matched filter (Figure 3-2) tuned to the $k$th tone outputs

$$y_k = \frac{1}{\sqrt{N_0 T_s'}} \int_{\mathcal{L}}^{T_s} \exp(-j2\pi f_k t) y(t) dt. \tag{3.4}$$

After manipulation, we obtain $y_k$,

$$y_k = \begin{cases} g_k \sqrt{\frac{PT_s'}{Q\theta N_0}} + w_k, & k \in \mathcal{S}_m, \\ w_k, & \text{otherwise;} \end{cases} \tag{3.5}$$

where the $w_k$, given by

$$w_k = \frac{1}{\sqrt{N_0 T_s'}} \int_{\mathcal{L}}^{T_s} \exp(-j2\pi f_k t) z(t) dt,$$

are mutually independent zero-mean complex Gaussian random variables with unity variance, because they are obtained by correlating white noise against orthogonal frequencies. With no requirement of phase information, the scheme is noncoherent. The phase error has already been considered in the calculation of the fading coefficients $\{g_k\}$.

According to (3.5), an equivalent channel can be obtained for the $k$th subchannel:

$$y_k = g_k x_k + w_k,$$

where the inputs $x_k$ are given by

$$x_k = \begin{cases} \sqrt{\frac{PT_s'}{Q\theta N_0}}, & k \in \mathcal{S}_m, \\ 0, & \text{otherwise.} \end{cases}$$

As $g_k$ and $w_k$ are independent zero-mean complex Gaussian random variables, the equivalent channel is a Rayleigh fading channel.

For the Rayleigh channel, the transition probability between input $x_k$ and sufficient statistic $r_k = |y_k|^2$ is

$$p_{r_k|x_k}(r|x) = \frac{1}{x^2+1} e^{-\frac{r}{x^2+1}}. \tag{3.6}$$

We sample the outputs of matched filters and thus obtain samples by subchannel for each FSK symbol. To distinguish samples at different time slots, we use one more subscript $n$ $(1 \le n \le N)$ to index the $N$ iterations. The symbols, $x_{k,n}$, $y_{k,n}$, $r_{k,n}$, $g_{k,n}$, and $w_{k,n}$, stand respectively for the input, output, sufficient statistic, fading coefficient, and additive noise of the $k$th subchannel at symbol slot $n$. Owing to the low duty cycle, successive symbols are far apart in time (see Figure 3-1). We thus assume different symbols experience independent fading, i.e., the gains $g_{k,n}$ $(1 \le n \le N)$ are independent, identically distributed (i.i.d.).

The decision variables obtained at the receiver are the averages of $r_{k,n}$ over the $N$

iterations, denoted as $S_k$,

$$S_k = \frac{1}{N}\sum_{n=1}^{N} r_{k,n}. \tag{3.7}$$

By substitution, we rewrite $S_k$ as

$$S_k = \begin{cases} \frac{1}{N}\sum_{n=1}^{N}\left|g_{k,n}\sqrt{\frac{PT_s'}{Q\theta N_0}} + w_{k,n}\right|^2, & k \in \mathcal{S}_m, \\ \frac{1}{N}\sum_{n=1}^{N}\left|w_{k,n}\right|^2, & \text{otherwise}, \end{cases} \tag{3.8}$$

which in either case are $\chi^2$ random variables with $2N$ degrees of freedom.

Since the average received power of the non-zero tones and the zero tones are $\frac{PT_s'}{Q\theta N_0} + 1$ and 1 respectively, we select the threshold $A$ to be

$$A = 1 + (1-\epsilon)\frac{PT_s'}{Q\theta N_0}, \tag{3.9}$$

where $\epsilon$ is chosen over $(0,1)$. If there are exact $Q$ of $\{S_k\}$ exceed the threshold, the corresponding $\mathcal{S}_m$ is decoded; otherwise, the receiver declares an error.

Since the scheme transmits $\ln\binom{M}{Q}$ nats of information in $NT_s/\theta$ seconds, the data rate $R$ is given by

$$R = \frac{\theta}{NT_s}\ln\binom{M}{Q}. \tag{3.10}$$

The value of $Q$ is chosen from the integers satisfying $Q \leq M/2$, because, for any integer $Q$ greater than $M/2$, we can find an integer $M-Q$ which leads to the same number of total symbols and a higher peak power per tone. In Multi-tone FSK, we pick the value of $Q$ to be much smaller than $M/2$.

### 3.2.3 Error Performance Analysis

As we know, an error occurs if $S_k \geq A$ for $k \notin \mathcal{S}_m$ (called *type I error*), or $S_k < A$ for $k \in \mathcal{S}_m$ (called *type II error*). Let $E_k$ be the event that an error occurs at the $k$th tone. For notational convenience, we define

$$p_{e,1} \triangleq \Pr(E_k) \quad \text{for} \quad k \notin \mathcal{S}_m,$$

and

$$p_{e,2} \triangleq \Pr(E_k) \quad \text{for} \quad k \in \mathcal{S}_m.$$

We will use the same techniques used in [27] and [28] to derive upper bounds and lower bounds of $p_{e,1}$ and $p_{e,2}$, denoted as $p_{e,1}^{(u)}$, $p_{e,2}^{(u)}$, $p_{e,1}^{(l)}$, and $p_{e,2}^{(l)}$.

### Derivation of $p_{e,1}^{(u)}$

To get the upper bound $p_{e,1}^{(u)}$, we use the Chernoff bound

$$\begin{aligned} p_{e,1} &= \Pr(NS_k \geq NA) \\ &\leq \inf_s E\left[e^{s(NS_k - NA)}\right] \quad \text{(Chernoff bound)} \\ &= e^{-N\Phi(A)}, \end{aligned} \tag{3.11}$$

where $s \in [0,1]$ and $\Phi(A)$ is given by

$$\Phi(A) = \sup_s \left[sA - \frac{1}{N}\ln\left(E\left[e^{sNS_k}\right]\right)\right]. \tag{3.12}$$

Further simplifying $\Phi(A)$, we obtain

$$\Phi(A) = \sup_s \left[sA - \frac{1}{N}\ln\left(E\left[e^{s\sum_{n=1}^{N}|w_{k,n}|^2}\right]\right)\right] \tag{3.13}$$

$$= \sup_s \left[sA + \ln(1-s)\right] \tag{3.14}$$

$$= A - 1 - \ln A, \tag{3.15}$$

where (3.13) is obtained by substituting the value of $S_k$ into (3.12); (3.14) is based on the fact that $|w_{k,n}|^2$ are independent, exponentially distributed, and the moment generating function of $\sum_{n=1}^{N}|w_{k,n}|^2$ is $(1-s)^{-N}$; (3.15) is derived by picking $s = (A-1)/A$. We have

$$p_{e,1}^{(u)} = \exp\left[-N(A - 1 - \ln A)\right]. \tag{3.16}$$

Note that, $A - 1 - \ln A > 0$ because $A > 1$. The upper bound $p_{e,1}^{(u)}$ decreases to zero as $N$ grows.

**Derivation of $p_{e,2}^{(u)}$**

We use the Chernoff bound again to get $p_{e,2}^{(u)}$. We have

$$\begin{aligned} p_{e,2} &= \Pr\left(NS_k \leq NA\right) \\ &\leq \inf_{s<0} E\left[e^{s(NS_k - NA)}\right] \quad \text{(Chernoff bound)}, \\ &= e^{-N\Phi(A')}, \end{aligned} \tag{3.17}$$

where

$$A' = \frac{A}{1 + \frac{PT_s'}{Q\theta N_0}}. \tag{3.18}$$

The derivation of (3.17) is as follows,

$$\begin{aligned} & \inf_{s<0} E\left[e^{s(NS_k - NA)}\right] \\ = & \inf_{s<0} \exp\left[-sNA + \ln\left(E\left[e^{s\sum_{n=1}^{N} r_{k,n}}\right]\right)\right] & (3.19) \\ = & \inf_{s<0} \exp\left(-N\left[sA + \ln\left(1 - \left[1 + \frac{P'T_s}{Q\theta N_0}\right]s\right)\right]\right) & (3.20) \\ = & \exp\left[-N\left(A' - 1 - \ln A'\right)\right], & (3.21) \end{aligned}$$

where (3.19) is obtained by substituting the value of $S_k$ and rearranging; (3.20) is due to the fact that $r_{k,n}$ are independent, exponentially distributed, and $E\left[\exp\left(s\sum_{n=1}^{N} r_{k,n}\right)\right]$ is the moment generating function of $\sum_{n=1}^{N} r_{k,n}$; (3.21) is the result of optimizing (3.20) over $s$. We have

$$p_{e,2}^{(u)} = \exp\left[-N\left(A' - 1 - \ln A'\right)\right]. \tag{3.22}$$

Since $A' \in (0,1)$ and $A' - 1 - \ln A' > 0$, the upper bound $p_{e,2}^{(u)}$ decreases to zero as $N \to \infty$.

**Derivation of $p_{e,1}^{(l)}$**

Since $S_k$ are $\chi^2$ random variables with $2N$ degrees of freedom, we can use the result in [32, §2.1.4] to evaluate the cumulative distribution functions. The probability $p_{e,1}$

is

$$p_{e,1} = \Pr\left(\sum_{n=1}^{N} |w_{k,n}|^2 > NA\right)$$
$$= \exp(-NA)\sum_{k=0}^{N-1}\frac{(NA)^k}{k!}, \tag{3.23}$$

by directly applying the formula in [32, §2.1.4]. Since $(NA)^k/k!$ is positive for all $k$, we have

$$\sum_{k=0}^{N-1}\frac{(NA)^k}{k!} \geq \frac{(NA)^{N-1}}{(N-1)!}, \tag{3.24}$$

which leads to

$$p_{e,1} \geq \exp\left[-NA + \ln\left(\frac{(NA)^{N-1}}{(N-1)!}\right)\right] \tag{3.25}$$

by using (3.24) in (3.23). Applying Stirling's approximation,

$$\sqrt{2\pi N}N^N e^{\left(-N+\frac{1}{12N+1}\right)} < N! < \sqrt{2\pi N}N^N e^{\left(-N+\frac{1}{12N}\right)},$$

we obtain

$$p_{e,1}^{(l)} = \exp\left(-N\left[A - 1 - \ln A + o_1(N)\right]\right), \tag{3.26}$$

where $o_1(N)$ is a vanishing term as $N$ increases, given by

$$o_1(N) = \frac{1}{2N}\ln\left(2\pi NA^2\right) + \frac{1}{12N^2}. \tag{3.27}$$

The lower bound $p_{e,1}^{(l)}$ decreases to zero as $N \to \infty$.

**Derivation of $p_{e,2}^{(l)}$**

For the probability $p_{e,2}$, we have

$$p_{e,2} = Pr\left(\sum_{n=1}^{N}\left|g_{k,n}\sqrt{\frac{P'T_s}{Q\theta N_0}} + w_{k,n}\right|^2 < NA\right)$$
$$= \exp(-NA')\sum_{k=N}^{\infty}\frac{(NA')^k}{k!}. \tag{3.28}$$

Since the inequality

$$\sum_{k=N}^{\infty} \frac{(NA')^k}{k!} \geq \frac{(NA')^N}{N!} \tag{3.29}$$

holds as $(NA')^k / k!$ is positive for all $k$, we can bound $p_{e,2}$ by applying (3.29) to (3.28),

$$p_{e,2} \geq \exp\left(-NA' + \ln\left[\frac{(NA')^N}{N!}\right]\right). \tag{3.30}$$

Using Stirling's approximation, we obtain

$$p_{e,2}^{(l)} = \exp\left[-N\left(A' - 1 - \ln A' + o_2\left(N\right)\right)\right], \tag{3.31}$$

where

$$o_2(N) = \frac{1}{2N}\ln(2\pi N) + \frac{1}{12N^2} \tag{3.32}$$

is a diminishing term as $N$ increases. The lower bound $p_{e,2}^{(l)}$ decreases to zero as $N \to \infty$.

Observe that, the difference between $p_{e,1}^{(u)}$ and $p_{e,1}^{(l)}$ is a vanishing term in the coefficient of $N$. Therefore, these two bounds decrease in the same order as $N \to \infty$. For the same reason, $p_{e,2}^{(u)}$ and $p_{e,2}^{(l)}$ also decrease in the same order as $N \to \infty$.

We represent $p_{e,1}^{(u)}$, $p_{e,2}^{(u)}$, $p_{e,1}^{(l)}$, and $p_{e,2}^{(l)}$ as functions of $N$. In fact, for any given data rate, $N$ is directly related to $M$, the number of total tones, according to (3.10). The value of $N$ is monotonically increasing with $M$. In the following discussion, we will frequently substitute $N$ with $M$ by using (3.10) to show the effect of $M$ on the scheme's performance. In particular, we are interested in writing the probability of error as a function of $M$. This way, we can characterize how the infinite bandwidth limit is approached as $M$ tends to infinity, and hence obtain the insights on how to design the optimal signaling at a finite but large bandwidth.

### 3.2.4 Upper Bound on Codeword Error Probability

Using the union bound, the codeword error probability is upper bounded by

$$p_e \leq (M-Q)\, p_{e,1} + Q p_{e,2}. \tag{3.33}$$

Since all symbols have the identical union bound, we need not to average the bound all over the symbol set.

Noticing that the inequality

$$M - Q \leq Q\binom{M}{Q}^{1/Q}$$

is always true for $M > Q$, we have

$$\begin{aligned}(M-Q)\, p_{e,1} &\leq Q\binom{M}{Q}^{1/Q} p_{e,1} \\ &\leq Q\binom{M}{Q}^{1/Q} p_{e,1}^{(u)}. \end{aligned} \tag{3.34}$$

Denote the upper bound of $(M-Q)\, p_{e,1}$ in (3.34) as $p_{e,\overline{\mathcal{S}}_m}^{(u)}$. By substituting the expressions of $p_{e,1}^{(u)}$, $A$, and $N$ into (3.34), we obtain

$$\begin{aligned} p_{e,\overline{\mathcal{S}}_m}^{(u)} &= Q \exp\left( -\ln\binom{M}{Q} \left[ \frac{(1-\epsilon)\, PT_s'}{QN_0RT_s} - \frac{1}{Q} \right.\right. \\ &\quad \left.\left. -\frac{\theta}{RT_s} \ln\left(1 + \frac{(1-\epsilon)\, PT_s'}{Q\theta N_0}\right)\right]\right). \end{aligned}$$

On the other hand, an upper bound of $Qp_{e,2}$ is given by $Qp_{e,2}^{(u)}$, denoted as $p_{e,\mathcal{S}_m}^{(u)}$. Substituting $A'$ and $N$ into $p_{e,2}^{(u)}$, we have the expression of $p_{e,\mathcal{S}_m}^{(u)}$,

$$\begin{aligned} p_{e,\mathcal{S}_m}^{(u)} &= Q \exp\left( -\frac{\theta \ln\binom{M}{Q}}{RT_s} \left[ \frac{-\epsilon PT_s'}{Q\theta N_0 + PT_s'} \right.\right. \\ &\quad \left.\left. - \ln\left(1 - \frac{\epsilon PT_s'}{Q\theta N_0 + PT_s'}\right)\right]\right). \end{aligned}$$

Thus, we can write an upper bound of the codeword error probability as

$$p_e \leq p_{e,\overline{\mathcal{S}}_m}^{(u)} + p_{e,\mathcal{S}_m}^{(u)}. \tag{3.35}$$

This upper bound holds for any $\epsilon \in (0,1)$. Since our goal is to derive an upper bound, picking any specific $\epsilon$ will serve this purpose. Instead of the optimal value of $\epsilon$ minimizing the sum, we pick a value simplifying our analysis. Yet, as we can see later, the corresponding upper bound is tight in the wideband regime.

Notice that $p_{e,\overline{\mathcal{S}}_m}^{(u)}$ strictly increases with $\epsilon$, while $p_{e,\mathcal{S}_m}^{(u)}$, on the contrary, strictly decreases. We choose the value of $\epsilon$ letting these two bounds equal, which is

$$\begin{aligned} \epsilon^* &= \frac{Q\theta N_0 + PT_s'}{PT_s'}\left[1 - \frac{RN_0T_s}{PT_s'}\right. \\ &\quad \left. - \frac{Q\theta N_0}{PT_s'}\ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right)\right]. \end{aligned} \tag{3.36}$$

To keep $\epsilon^* \in (0,1)$, the data rate $R$ must in the range

$$0 \leq R < \frac{PT_s'}{N_0T_s} - \frac{Q\theta}{T_s}\ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right). \tag{3.37}$$

With (3.37) holds, by substituting $\epsilon^*$ into $p_{e,\overline{\mathcal{S}}_m}^{(u)} + p_{e,\mathcal{S}_m}^{(u)}$, we obtain an upper bound of the error probability

$$p_e \leq 2Q\exp\left(-\ln\binom{M}{Q}E_r(M,Q,T_s,R,\theta)\right), \tag{3.38}$$

where

$$E_r(M,Q,T_s,R,\theta) = \frac{\theta}{RT_s}\left(\beta - 1 - \ln\beta\right), \tag{3.39}$$

and

$$\beta = \frac{RN_0T_s}{PT_s'} + \frac{Q\theta N_0}{PT_s'}\ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right). \tag{3.40}$$

By examining (3.40), we conclude that $\beta \geq 0$ for any feasible choice of $(M,Q,T_s,R,\theta)$. On the other hand, because the data rate should satisfy $R < \frac{PT_s'}{N_0T_s}$ for reliable communications according to (3.37), and the duty cycle $\theta$ decreases to zero with increasing

bandwidth as we suggest, we have $\beta < 1$ in the wideband regime.

### 3.2.5 Lower Bound on Codeword Error Probability

The codeword error probability could be lower bounded by

$$\begin{aligned} p_e &\geq \sum_{k=1}^{M} \Pr(E_k) - \sum_{k \neq j} \Pr(E_k \cap E_j) \\ &\geq (M-Q)\, p_{e,1}^{(l)} + Q p_{e,2}^{(l)} - (M-Q)\, Q p_{e,1}^{(u)} p_{e,2}^{(u)} \\ &\quad - \binom{M-Q}{2} p_{e,1}^{(u)^2} - \binom{Q}{2} p_{e,2}^{(u)^2}. \end{aligned} \tag{3.41}$$

Recall that, $p_{e,1}^{(u)}$, $p_{e,2}^{(u)}$, $p_{e,1}^{(l)}$, and $p_{e,2}^{(l)}$ are all decreasing functions of $N$. $(M-Q)\, p_{e,1}^{(l)}$ and $(M-Q)\, p_{e,1}^{(u)}$ decrease with $N$ in the same order, and so do $Q p_{e,2}^{(l)}$ and $Q p_{e,2}^{(u)}$. If $(M-Q)\, p_{e,1}^{(l)}$ decrease faster than $Q p_{e,2}^{(l)}$ in $N$, then the lower bound (3.41) is of the order of $Q p_{e,2}^{(l)}$, and all other four terms are dominated. Otherwise, the lower bound has the order of $(M-Q)\, p_{e,1}^{(l)}$. That is,

$$p_e \geq \max\left( (M-Q)\, p_{e,1}^{(l)}, Q p_{e,2}^{(l)} \right). \tag{3.42}$$

As $N$ approaches infinity, the coefficient of $N$ in the exponent of $(M-Q)\, p_{e,1}^{(l)}$ is given by

$$\begin{aligned} &\lim_{N\to\infty} \frac{\ln\left[(M-Q)\, p_{e,1}^{(l)}\right]}{N} \\ =\ &\lim_{N\to\infty} \left[ \frac{\ln p_{e,1}^{(l)}}{N} + \frac{\ln \binom{M}{Q}}{QN} \right] && (3.43) \\ =\ &\lim_{N\to\infty} \frac{\ln p_{e,1}^{(l)}}{N} + \frac{RT_s}{Q\theta} && (3.44) \\ =\ &-A + 1 + \ln A + \frac{RT_s}{Q\theta}. && (3.45) \end{aligned}$$

To obtain (3.43), we used

$$\lim_{N\to\infty} \frac{\ln(M-Q)}{N} = \lim_{N\to\infty} \frac{\ln \binom{M}{Q}}{QN},$$

where $Q$ is predetermined, and, for a given data rate, $M$ increases with $N$ according to (3.10).

As $N$ approaches infinity, the coefficient of $N$ in the exponent of $Qp_{e,2}^{(l)}$ is given by

$$\begin{aligned} & \lim_{N\to\infty} \frac{\ln\left(Qp_{e,2}^{(l)}\right)}{N} \\ = & \lim_{N\to\infty} \frac{\ln p_{e,2}^{(l)}}{N} \\ = & -A' + 1 + \ln A'. \end{aligned} \tag{3.46}$$

The values of (3.45) and (3.46) are both negative. The value of (3.45) strictly increases with $\epsilon$. Therefore, the larger $\epsilon$ is, the slower $(M-Q)\,p_{e,1}^{(l)}$ decreases with $N$. The value of (3.46), on the contrary, strictly decreases with $\epsilon$. The larger $\epsilon$ is, the faster $Qp_{e,2}^{(l)}$ decreases with $N$. Thus, the lower bound (3.42) depends on the value of $\epsilon$ picked.

If we pick the value $\epsilon^*$ given by (3.36), then the values of (3.45) and (3.46) are equal, which let $(M-Q)\,p_{e,1}^{(l)}$ and $Qp_{e,2}^{(l)}$ decrease with $N$ in the same order. Other values of $\epsilon$ will let one of these two terms decrease more slowly. Since (3.42) is determined by the larger term of $(M-Q)\,p_{e,1}^{(l)}$ and $Qp_{e,2}^{(l)}$, the value of $\epsilon$ other than $\epsilon^*$ will lift the lower bound upwards. Consequently, by picking $\epsilon^*$, we get a lower bound for $p_e$ in the wideband regime for all $\epsilon \in (0,1)$,

$$p_e \geq Q \exp\left(-\ln\binom{M}{Q} E_r(M, Q, T_s, R, \theta)\right), \tag{3.47}$$

where $E_r(M, Q, T_s, R, \theta)$ was given by (3.39).

### 3.2.6 Capacity-Achieving Property

In [5], Telatar and Tse have proven that, in multipath fading channels, Single-tone FSK can achieve the capacity of an AWGN channel with the same average received power at the wideband limit. We now prove that Multi-tone FSK has the same capacity-achieving property in the wideband limit.

**Proposition 3.1** *Multi-tone FSK can achieve any data rates that satisfy*

$$R < \left(1 - \frac{\mathcal{L}}{T_s}\right) \frac{P}{N_0} \tag{3.48}$$

*with an arbitrarily small probability of error over a multipath fading channel with average power constraint $P$, by using large enough bandwidth.*

**Proof:**

Since the inequality $\beta - 1 > \ln \beta$ holds for all $\beta \in [0, 1)$, we have $E_r(M, Q, T_s, R, \theta) > 0$ in (3.39). As a result, the upper bound (3.38) decreases to zero when $M$ increases to infinity, as long as (3.37) holds. That is, for the data rates in the range defined by (3.37), we can use Multi-tone FSK to transmit data with arbitrarily small error probability by using bandwidth large enough.

The data rate $R$ can get arbitrarily close to $\frac{PT_s'}{N_0 T_s}$, i.e., $\left(1 - \frac{\mathcal{L}}{T_s}\right) \frac{P}{N_0}$, by letting $\theta$ decrease to zero as the bandwidth increases to infinity. Hence, as long as the data rate satisfies (3.48), the Multi-tone FSK scheme can achieve arbitrarily small error probability by picking bandwidth large enough. ■

Recalling that the channel is assumed to be underspread and $\mathcal{L} \ll T_s$, we conclude that the capacity of Multi-tone FSK approaches Shannon's infinite bandwidth capacity limit $\frac{P}{N_0}$.

In the proof, we let $\theta$ decrease to zero as the bandwidth increases. However, $\theta$ must decrease more slowly than $1 \diagup \ln \binom{M}{Q}$ as $M$ increases. Otherwise, by substituting (3.39) into (3.38), we can verify that (3.38) will not decrease with $M$. Intuitively, it is because decreasing $\theta$ will reduce the information rate. To counteract this effect, we need to increase the information bits per symbol, i.e., increase $\ln \binom{M}{Q}$. If $\theta$ decreases too fast with $M$, the achievable data rate will be compromised.

At a particular finite but large bandwidth, the capacity of the Multi-tone FSK schemes with different $Q$ may vary. In such cases, considering the whole family of the Multi-tone FSK schemes, including Single-tone FSK, has the benefit of more flexibility in the tradeoff between the spectrum efficiency and the energy efficiency.

### 3.2.7 Error Exponent

In [27] and [28], Lun et al. discussed the error exponent for Single-tone FSK. We will discuss the error exponent of Multi-tone FSK in this subsection.

The merit of the study of the error exponents for these schemes is, as will become clear later on, a characterization of how the error performance is affected by the channel parameters including the SNR, rate, duty cycle, bandwidth, and $Q$.

**Proposition 3.2** *As bandwidth increases to infinity, Multi-tone FSK has the following relation between the error probability $p_e$ and the total number of symbol $\binom{M}{Q}$:*

$$\lim_{M\to\infty} \frac{-\ln p_e}{\ln \binom{M}{Q}} = E_r(M, Q, T_s, R, \theta), \tag{3.49}$$

*where $E_r(M, Q, T_s, R, \theta)$ is defined in (3.39).*

**Proof:** According to (3.47), we have

$$\lim_{M\to\infty} \frac{-\ln p_e}{\ln \binom{M}{Q}} \leq E_r(M, Q, T_s, R, \theta). \tag{3.50}$$

The reverse inequality follows from (3.38),

$$\lim_{M\to\infty} \frac{-\ln p_e}{\ln \binom{M}{Q}} \geq E_r(M, Q, T_s, R, \theta). \tag{3.51}$$

Combining the inequalities (3.50) and (3.51), we obtain (3.49), thus completing the proof. ■

Therefore, $E_r(M, Q, T_s, R, \theta)$ represents the true exponential dependence of the error probability on $\ln \binom{M}{Q}$ for $M$ sufficiently large. We call $E_r(M, Q, T_s, R, \theta)$ the error exponent of Multi-tone FSK which is analogous to the treatment of random block coding over discrete memoryless channels by Gallager [4, §5.8].

### 3.2.8 Numerical Results

In the previous sections, we have studied achievable rates and error probabilities of Multi-tone FSK. We now proceed to evaluate the capacity of Multi-tone FSK schemes with a simple hard-decision receiver for particular parameter choices. We show that, even with finite bandwidth, our scheme yields achievable rates that are for practical purposes very close to the wideband capacity limit.

Recall that, there are $\binom{M}{Q}$ possible symbols in a $Q$-tone FSK scheme. We can use an equivalent discrete memoryless channel (DMC) with $\binom{M}{Q}$ inputs and $\binom{M}{Q}$ outputs for capacity calculation. Owing to the symmetry of the DMC, the capacity-achieving input for the DMC is uniformly distributed. If we can obtain the transition probabilities between the input and the output, we can directly calculate the capacity. The transition probabilities depend on the detection rule, and are, for most of the cases, intractable. We restrict ourselves to the Single-tone case and the Two-tone case using a symbol-by-symbol detection scheme based on the maximum-a-posteriori (MAP) rule.

The symbol-by-symbol decision is optimal from a symbol detection point of view, since the channel is memoryless, but may not be optimal from a codeword decoding point of view. In effect, we are considering a simple hard-decision decoder instead of a soft-decision decoder. Hard-decision decoders have information loss during the symbol-by-symbol detection, while soft-decision decoders use these information to decode symbols jointly. Our results show that, even with a simple hard-decision receiver, Multi-tone FSK yields achievable rates very close to the wideband capacity limit.

As we know, the capacity $C$ of a fading channel is bounded by the AWGN capacity,

$$C \leq W \ln\left(1 + \frac{P}{N_0 W}\right),$$

where $W$ is the bandwidth of the system. We call this bound the *power-limited bound*, because the capacity of Multi-tone FSK is tightly bounded by this bound in the power-limited regime.

Another upper bound is derived from the DMC model. Because the entropy per input symbol is $\ln \binom{M}{Q}$, the achievable data rate is bounded by

$$C \leq \frac{1}{T_s} \ln \binom{M}{Q}.$$

This bound is tight when the system has a "less than enough" bandwidth. We denote this bound as the *bandwidth-limited bound.*

First, we consider a Single-tone FSK scheme and a Two-tone FSK scheme with the same bandwidth constraint. We compare their performance in terms of capacity versus signal-to-noise ratio. We assume the symbol time, $T_s = 10\mu s$, the multipath spread, $\mathcal{L} = 1\mu s$, and the system bandwidth, $W = 1MHz$. The SNR ($\frac{P}{N_0 W}$) ranges from $10^{-2}$ to $10^3$. We plot the capacity curves and the upper bounds in Figure 3-3.

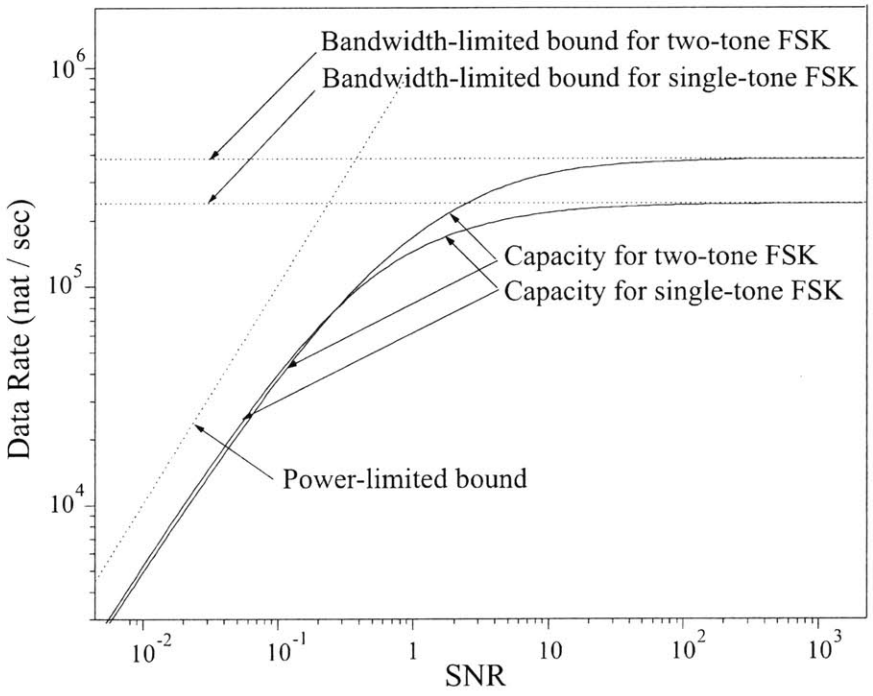


Figure 3-3: The capacity of Single-tone FSK and Two-tone FSK for $T_s = 10\mu s$, $\mathcal{L} = 1\mu s$, and $W = 1MHz$.

We can make the following observations from Figure 3-3:

- The capacity of the Single-tone FSK scheme and of the Two-tone FSK scheme are very close to the upper bounds both in the power-limited regime (the SNR

ranges from $10^{-2}$ to 1) and the bandwidth-limited regime (the SNR ranges from 1 to $10^3$).

- In the power-limited regime, the performances for these two schemes are very close, and the power-limited bound is tight in this regime, which is roughly 2 dB higher than the capacities. Comparing to the capacity of CDMA [7], OFDM [33], and PPM [34] in the wideband regime, the capacity of Multi-tone FSK is very close to the AWGN capacity. Notice that, this gap between the capacity curves and the power-limited bound can be closed up if we use a large enough bandwidth, althrough we use a hard-decision rule. This is because that, we use the cumulative received power in decoding the repetition code, which is essentially a soft-decision strategy although we do hard-decision block by block.

- In the bandwidth-limited regime, the performance of Two-tone FSK is better than that of Single-tone FSK because of its larger input set (higher spectrum efficiency).

In Figure 3-4, we assume the average power is fixed, where $T_s = 0.1s$, $\mathcal{L} = 1\mu s$, and $\frac{P}{N_0} = 40Hz$. The bandwidth ranges from $100Hz$ to $10KHz$. We observe that, in the bandwidth-limited regime (from $100Hz$ to $1KHz$), the performance of Two-tone FSK is better than that of Single-tone FSK. In the regime that the bandwidth is smaller than $100Hz$, however, Two-tone FSK shows no advantage at all. This is because that the number of total frequencies, $M$, is so small that $\binom{M}{1}$ and $\binom{M}{2}$ are comparable.

Our examples illustrate that, even for finite bandwidth, Multi-tone FSK schemes can yield rates that are appreciably close to the wideband capacity limit. The ability to vary the number of tones according to the transmission condition (average power, bandwidth constraint) provides a useful design parameter which can be adapted to different operating regimes.

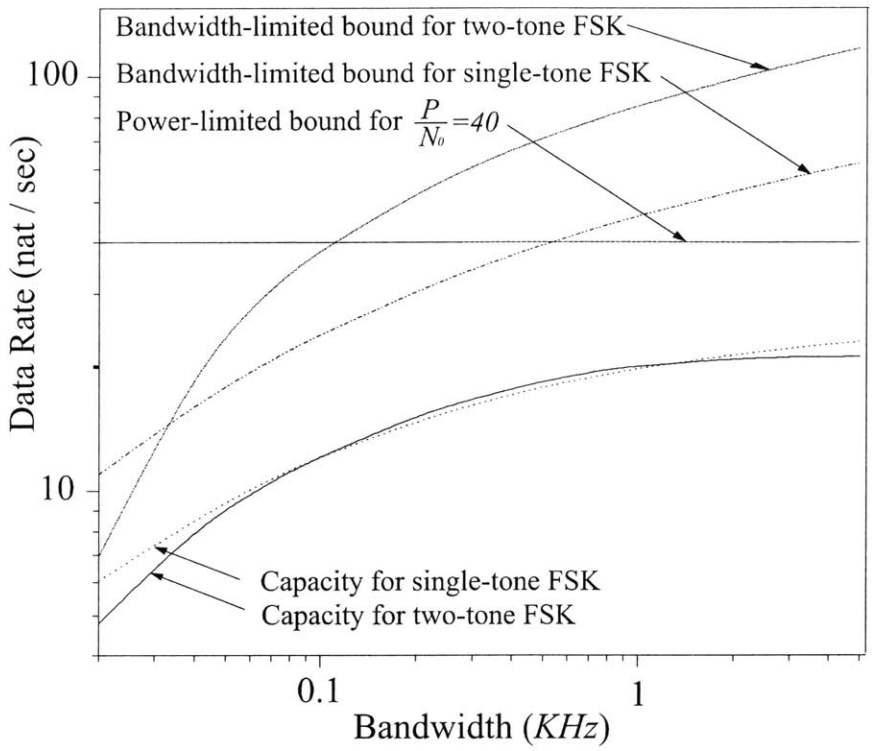


Figure 3-4: Capacity versus bandwidth curves for $T_s = 0.1s$, $\mathcal{L} = 1\mu s$, and $\frac{P}{N_0} = 40$.

## 3.3 Multi-tone FSK with Receiver Diversity

In this subsection, we discuss the performance of Multi-tone FSK when the receiver is equipped with multiple antennas (receiver diversity). The achievable data rate is improved by a factor of the order of diversity. As we know, the wideband capacity limit is also increased by a factor of the order of diversity. Hence, Multi-tone FSK can achieve the wideband capacity limit in the case that the receiver diversity is used.

### 3.3.1 Mathematical Model

Assume the receiver has $L$ antennas. Let $y_k^l$ be the output of the $l$th antenna, $g_k^l$ the fading coefficient, and $w_k^l$ the additive with Gaussian noise. An equivalent channel for the $k$th subchannel is described as

$$y_k^l = g_k^l x_k + w_k^l, \qquad \text{for } k = 1, ..., M,\ l = 1, ..., L.$$

The fading coefficients $g_k^l$ are assume to be mutually independent with standard complex normal distributions.

A sufficient statistic is then $r_k = \sum_{l=1}^{L} \left|y_k^l\right|^2$, whose probability conditional on $x_k$ is

$$p_{r_k|x_k}(r|x) = \frac{r^{L-1} e^{-\frac{r}{x^2+1}}}{\Gamma(L)\,(x^2+1)^L}. \tag{3.52}$$

As in the case without receiver diversity, the inputs $x_k$ are given by

$$x_k = \begin{cases} \sqrt{\frac{PT_s'}{Q\theta N_0}}, & k \in \mathcal{S}_m, \\ 0, & \text{otherwise.} \end{cases}$$

We still use a subscript $n$ $(1 \le n \le N)$ to index the $N$ iterations, such as $x_{k,n}$. We assume different symbols experience independent fading as before, i.e., the gains $g_{k,n}$ $(1 \le n \le N)$ are independent, identically distributed.

The decision variables obtained at the receiver are the averages of $r_{k,n}$ over the $N$ iterations, denoted as $S_k$,

$$S_k = \frac{1}{NL} \sum_{n=1}^{N} r_{k,n}. \tag{3.53}$$

The value of $S_k$ is

$$S_k = \begin{cases} \frac{1}{NL} \sum_{n=1}^{N} \sum_{l=1}^{L} \left| g_{k,n}^l \sqrt{\frac{PT_s'}{Q\theta N_0}} + w_{k,n}^l \right|^2, & k \in \mathcal{S}_m, \\ \frac{1}{NL} \sum_{n=1}^{N} \sum_{l=1}^{L} \left| w_{k,n}^l \right|^2, & \text{otherwise,} \end{cases} \tag{3.54}$$

which in either case are $\chi^2$ random variables with $2NL$ degrees of freedom. The average received power of the non-zero tones and the zero tones are $\frac{PT_s'}{Q\theta N_0} + 1$ and 1 respectively. The threshold $A$ is given by

$$A = 1 + (1-\epsilon)\frac{PT_s'}{Q\theta N_0}, \tag{3.55}$$

where $\epsilon$ is chosen over $(0, 1)$. The data rate $R$ is given by

$$R = \frac{\theta}{NT_s} \ln \binom{M}{Q}. \tag{3.56}$$

### 3.3.2 Error Performance Analysis

As we discussed before, detection errors can be classified as type I error and type II error. Let $p_{e,1}^{(u)}$, $p_{e,2}^{(u)}$, $p_{e,1}^{(l)}$, and $p_{e,2}^{(l)}$ be the upper bounds and lower bounds for the probabilities of type I and type II error.

**Derivation of $p_{e,1}^{(u)}$**

We use the Chernoff bound to upper bound $p_{e,1}$:

$$\begin{aligned} p_{e,1} &= \Pr\left(NLS_k \geq NLA\right) \\ &\leq \inf_s E\left[e^{s(NLS_k - NLA)}\right] \quad \text{(Chernoff bound)} \\ &= e^{-NL\Phi(A)}, \end{aligned} \tag{3.57}$$

where $s \in [0,1]$ and $\Phi\left(A\right)$ is given by

$$\Phi(A) = \sup_s \left[sA - \frac{1}{NL}\ln\left(E\left[e^{sNLS_k}\right]\right)\right]. \tag{3.58}$$

Further simplifying $\Phi(A)$, we obtain

$$\Phi(A) = \sup_s \left[sA - \frac{1}{NL}\ln\left(E\left[e^{s\sum_{n=1}^{N}\sum_{l=1}^{L}\left|w_{k,n}^{l}\right|^2}\right]\right)\right] \tag{3.59}$$

$$= \sup_s \left[sA + \ln\left(1-s\right)\right] \tag{3.60}$$

$$= A - 1 - \ln A, \tag{3.61}$$

where (3.59) is obtained by substituting the value of $S_k$ into (3.58); (3.60) is based on the fact that $|w_{k,n}|^2$ are i.i.d. exponential distributions, and the moment generating function for $\sum_{n=1}^{N}\sum_{l=1}^{L}\left|w_{k,n}^{l}\right|^2$ is $(1-s)^{-LN}$.

By substituting (3.61) into (3.57), we obtain the upper bound

$$p_{e,1}^{(u)} = \exp\left[-NL\left(A - 1 - \ln A\right)\right]. \tag{3.62}$$

Since $A - 1 - \ln A > 0$, the upper bound $p_{e,1}^{(u)}$ decreases to zero as $NL$ grows.

**Derivation of $p_{e,2}^{(u)}$**

The upper bound $p_{e,2}^{(u)}$ is derived as

$$\begin{aligned} p_{e,2} &= \Pr\left(NLS_k \leq NLA\right) \\ &\leq \inf_{s<0} E\left[e^{s(NLS_k - NLA)}\right] \quad \text{(Chernoff bound)}, \\ &= e^{-NL\Phi(A')}, \end{aligned} \tag{3.63}$$

where

$$A' = \frac{A}{1 + \frac{PT_s'}{Q\theta N_0}}. \tag{3.64}$$

The derivation of (3.63) is as follows,

$$\begin{aligned} & \inf_{s<0} E\left[e^{s(NLS_k - NLA)}\right] \\ = & \inf_{s<0} \exp\left[-sNLA + \ln\left(E\left[e^{s\sum_{n=1}^{N}\sum_{l=1}^{L} r_{k,n}^{l}}\right]\right)\right] & (3.65) \\ = & \inf_{s<0} \exp\left(-NL\left[sA + \ln\left(1 - \left[1 + \frac{P'T_s}{Q\theta N_0}\right]s\right)\right]\right) & (3.66) \\ = & \exp\left[-NL\left(A' - 1 - \ln A'\right)\right], & (3.67) \end{aligned}$$

where (3.65) is obtained by substituting the value of $S_k$; (3.66) is based on that $r_{k,n}^{l}$ are independent, exponentially distributed, and $E\left[\exp\left(s\sum_{n=1}^{N}\sum_{l=1}^{L} r_{k,n}^{l}\right)\right]$ is the moment generating function of $\sum_{n=1}^{N}\sum_{l=1}^{L} r_{k,n}^{l}$; (3.67) is the result of optimizing (3.66) over $s$.

The upper bound $p_{e,2}^{(u)}$ is thus given by (3.67):

$$p_{e,2}^{(u)} = \exp\left[-NL\left(A' - 1 - \ln A'\right)\right]. \tag{3.68}$$

Since $A' \in (0,1)$ and $A' - 1 - \ln A' > 0$, the upper bound $p_{e,2}^{(u)}$ decreases to zero as $N \to \infty$.

**Derivation of $p_{e,1}^{(l)}$**

Since $\sum_{n=1}^{N}\sum_{l=1}^{L}\left|w_{k,n}^{l}\right|^2$ are $\chi^2$ random variables, we use the formula in [32, §2.1.4] to evaluate the cumulative distribution functions of the $\chi^2$ distribution and rewrite

the probability $p_{e,1}$ as

$$\begin{aligned} p_{e,1} &= \Pr\left(\sum_{n=1}^{N}\sum_{l=1}^{L}\left|w_{k,n}^{l}\right|^2 > NLA\right) \\ &= \exp\left(-NLA\right)\sum_{k=0}^{NL-1}\frac{(NLA)^k}{k!}. \end{aligned} \tag{3.69}$$

Because $(NLA)^k / k!$ is positive for all $k$, we have the inequality

$$\sum_{k=0}^{NL-1}\frac{(NLA)^k}{k!} \geq \frac{(NLA)^{NL-1}}{(NL-1)!}, \tag{3.70}$$

which leads to

$$p_{e,1} \geq \exp\left[-NLA + \ln\left(\frac{(NLA)^{NL-1}}{(NL-1)!}\right)\right]. \tag{3.71}$$

Applying Stirling's approximation, we obtain

$$p_{e,1}^{(l)} = \exp\left(-NL\left[A - 1 - \ln A + o_1\left(NL\right)\right]\right), \tag{3.72}$$

where $o_1\left(NL\right)$ is a vanishing term as $NL$ increases.

The lower bound $p_{e,1}^{(l)}$ decreases to zero as $N \to \infty$.

**Derivation of $p_{e,2}^{(l)}$**

Using the formula in [32, §2.1.4] for the cumulative distribution functions of the $\chi^2$ distribution, we rewrite the probability $p_{e,2}$ as

$$\begin{aligned} p_{e,2} &= Pr\left(\sum_{n=1}^{N}\sum_{l=1}^{L}\left|g_{k,n}^{l}\sqrt{\frac{P'T_s}{Q\theta N_0}} + w_{k,n}^{l}\right|^2 < NLA\right) \\ &= \exp\left(-NLA'\right)\sum_{k=NL}^{\infty}\frac{(NLA')^k}{k!}. \end{aligned} \tag{3.73}$$

Since the inequality

$$\sum_{k=NL}^{\infty}\frac{(NLA')^k}{k!} \geq \frac{(NLA')^{NL}}{(NL)!} \tag{3.74}$$

holds as $(NLA')^k / k!$ is positive for all $k$, we can bound $p_{e,2}$ as

$$p_{e,2} \geq \exp\left(-NLA' + \ln\left[\frac{(NLA')^{NL}}{(NL)!}\right]\right). \tag{3.75}$$

Using Stirling's approximation, we obtain

$$p_{e,2}^{(l)} = \exp\left[-NL\left(A' - 1 - \ln A' + o_2\left(NL\right)\right)\right]. \tag{3.76}$$

where $o_2(NL)$ is a diminishing term as $NL$ increases. The lower bound $p_{e,2}^{(l)}$ decreases to zero as $N \to \infty$.

### 3.3.3 Upper Bound on Codeword Error Probability

Similar to the case without receiver diversity, we use the union bound to upper bound the codeword error probability

$$p_e \leq (M - Q)\, p_{e,1} + Q p_{e,2}. \tag{3.77}$$

For the data rate $R$ in the range

$$0 \leq R < \frac{LPT_s'}{N_0 T_s} - \frac{LQ\theta}{T_s} \ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right), \tag{3.78}$$

we obtain an upper bound for the error probability

$$p_e \leq 2Q \exp\left(-\ln\binom{M}{Q} E_r(M, Q, T_s, R, \theta, L)\right), \tag{3.79}$$

where

$$E_r(M, Q, T_s, R, \theta, L) = \frac{L\theta}{RT_s}\left[\beta - 1 - \ln\beta\right], \tag{3.80}$$

and

$$\beta = \frac{RN_0 T_s}{LPT_s'} + \frac{Q\theta N_0}{PT_s'} \ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right). \tag{3.81}$$

Note that $\beta \geq 0$ for any feasible choice of $(M, Q, T_s, R, \theta, L)$. Thus, $\beta - 1 - \ln\beta > 0$ and

$E_r(M, Q, T_s, R, \theta, L)$ is always greater than zero. The upper bound (3.79) decreases to zero as the bandwidth grows.

### 3.3.4 Lower Bound on Codeword Error Probability

The codeword error probability could be lower bounded by

$$p_e \geq \sum_{k=1}^{M} \Pr(E_k) - \sum_{k \neq j} \Pr(E_k \cap E_j)$$

Consequently, using the similar technology we used to derive (3.47), we get a lower bound for $p_e$ in the wideband regime,

$$p_e \geq Q \exp\left( - \ln \binom{M}{Q} E_r(M, Q, T_s, R, \theta, L) \right), \tag{3.82}$$

where $E_r(M, Q, T_s, R, \theta, L)$ was given by (3.80).

### 3.3.5 Capacity-Achieving Property

The Multi-tone FSK with receiver diversity can also achieve the wideband capacity limit.

**Proposition 3.3** *Multi-tone FSK can achieve any data rates that satisfy*

$$R < \left(1 - \frac{\mathcal{L}}{T_s}\right) \frac{LP}{N_0} \tag{3.83}$$

*with an arbitrarily small probability of error over a multipath fading channel with average power constraint $P$, by using bandwidth large enough, when the order of receiver diversity is $L$.*

**Proof:** Since the inequality $\beta - 1 > \ln \beta$ holds for all $\beta \in [0, 1)$, we have $E_r(M, Q, T_s, R, \theta, L) > 0$ in (3.80). As a result, the upper bound (3.79) decreases to zero when $M$ increases to infinity, as long as (3.78) holds. That is, the data rate $R$

can get arbitrarily close to $\frac{LPT_s'}{N_0 T_s}$, i.e., $\left(1 - \frac{\mathcal{L}}{T_s}\right)\frac{LP}{N_0}$, by letting $\theta$ decrease to zero as the bandwidth increases to infinity. ■

As in the case without receiver diversity, we can also show that, the upper bound and the lower bound are tight, in the sense that, they are the same order in $\ln\binom{M}{Q}$ in the wideband limit. Therefore, $E_r(M, Q, T_s, R, \theta, L)$ is the error exponent and represents the true exponential dependence of the error probability on $\ln\binom{M}{Q}$ for $M$ sufficiently large.

## 3.4 Resource Usage

In this section, we characterize the ability of Multi-tone FSK to co-exist with pre-existing systems and its multiple access compatibility.

### 3.4.1 Flexible Bandwidth Occupation

Different from the PPM scheme, the Multi-tone FSK scheme doesn't have low spectral emission. Instead, on the bandwidth occupied by Multi-tone FSK, we have high peak power, low duty cycle, narrow bandwidth occupancy symbols. When the scheme overlaps with other systems in bandwidth, it will cause strong interference to other systems. However, Multi-tone FSK still have very good compatibility with pre-existing systems due to its flexible bandwidth occupation. Since the only requirement for the frequencies used in Multi-tone FSK is that they are orthogonal over a symbol time, we can use frequencies apart in bandwidth. The Multi-tone FSK signaling scheme can thus be fitted into discontinuous bandwidth while most other systems require a trunk of continuous bandwidth.

In Figure 3-5, we show how the Multi-tone FSK scheme coexists with two pre-existing systems, A and B. The peaky columns distributed alongside the systems A and B represent the tones in the Multi-tone FSK scheme and the columns with a different color stand for nonzero tones. They occupy a discontinuous frequency band without interfering existing systems.

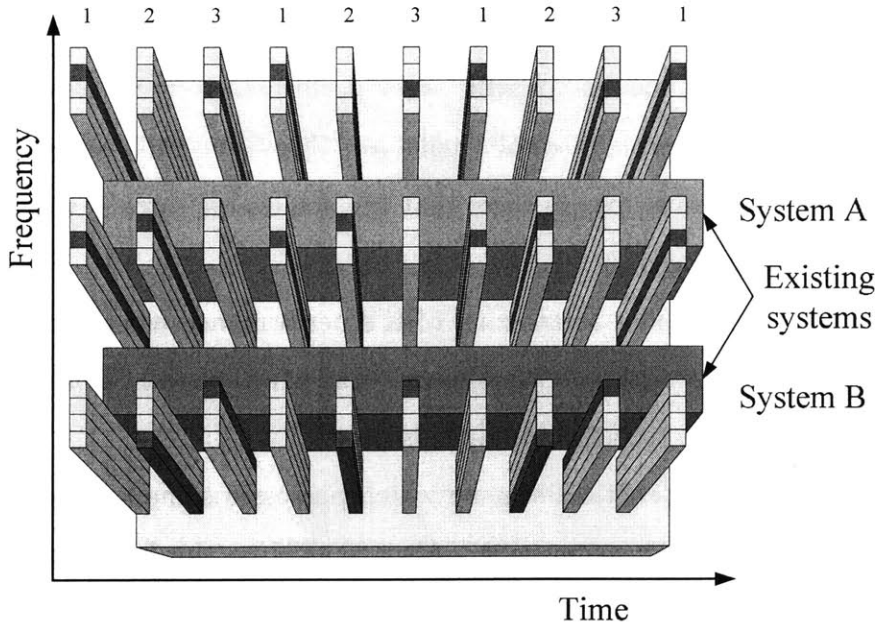


Figure 3-5: Resource Usage of Multi-tone FSK

### 3.4.2 Multiple Access

From the multiple access perspective, the resource usage of Multi-tone FSK is also very good. The Multi-tone FSK symbols are transmitted in predetermined time slots with a low duty cycle. This feature enables Multi-tone FSK to be extended to the multiple access case easily. By using time division multiple access, we can incorporate different users in the same frequency band without interfering with each other. Each user can achieve the wideband capacity limit. Figure 3-5 demonstrates the multiple access scheme for three users, 1, 2, and 3.

## 3.5 Conclusions

In this chapter, we introduced a class of Multi-tone FSK schemes for non-coherent communications in wideband Rayleigh fading channels. For a Rayleigh fading channel with infinite bandwidth, the AWGN capacity limit can be achieved by using Multi-tone FSK schemes.

We derived an upper bound and a lower bound for the codeword error probability. These two bounds lead to the error exponent of Multi-tone FSK, which describes

how the error probability depends on the different parameters. According to our results, Multi-tone FSK has comparable error performance as Single-tone FSK. We also calculated the achievable rates of Single- and Two-tone FSK schemes with finite bandwidth. The numerical results show that the achievable rates is very close to the capacity in fading channels.

Furthermore, we consider Multi-tone FSK's performance when receiver diversity is incorporated into the system. We showed that Multi-tone FSK is still capacity-achieving when multiple receiver antennas is used.

Multi-tone FSK is a practical scheme which possesses some good characteristics, such as flexible bandwidth occupancy. Also, thanks to the low duty cycle in the transmission, Multi-tone FSK can be easily extend to multiple access cases by using time division multiple access.

Here, we do not claim that the Multi-tone FSK signaling scheme is optimal. For example, if we consider soft decoding/partially soft decoding in Multi-tone FSK, by tracking all/several multiple-symbol codewords over a certain period of time, we could further close the gap between the achievable data rates and the capacity limit. Instead, Multi-tone FSK is a robust and practical scheme which performs very closely to the optimal limit. In the next chapter, we study how feedback can help to improve Multi-tone FSK's error performance and propose a feedback scheme which reduces the error probability significantly.

# Chapter 4

# Multi-tone Frequency-Shift Keying with Feedback

## 4.1 Introduction

In Chapter 3, we used the Multi-tone FSK scheme in a wideband fading channel to achieve the wideband capacity limit. We noticed that, the codeword error probability decays slowly with bandwidth. In this chapter, we consider a modified Multi-tone FSK scheme which employs a feedback link. We show that, the feedback scheme improves the error performance significantly.

For AWGN channels without bandwidth constraint, Schalkwijk and Kailath have shown in [40] that feedback can significantly reduce the codeword error probability, a "double exponential" decay in code length compared to an exponential decay otherwise. Although the capacity of a memoryless noisy channel cannot be increased by feedback according to [39, Theorem 6], we greatly improve error performance by using feedback.

In the following discussion, we present an Multi-tone FSK scheme with feedback for memoryless wideband Rayleigh fading channels. We call this scheme as the *Feedback Multi-tone FSK* scheme. The purpose of feedback is to reduce transmission error, and not for channel measurement and tracking.

Although both Multi-tone FSK and Feedback Multi-tone FSK have codeword

error probability decreasing roughly inversely with bandwidth, the latter makes the codeword error probability vanish much faster as bandwidth increases, leading to a lower requirement for bandwidth at the same codeword error probability and data rate.

This chapter is organized as follows: In Section 4.2, we introduce the concept of the Feedback Multi-tone FSK scheme. In Section 4.3, we study the power consumption of this scheme, and prove the feedback mechanism will not increase the power consumption in the forward link. In Section 4.4, we provide an upper bound and a lower bound of the codeword error probability to derive Feedback Multi-tone FSK's error exponent. In Section 4.5, we prove that Feedback Multi-tone FSK achieves the wideband capacity limit. In Section 4.6, we compare Feedback Multi-tone FSK with Multi-tone FSK on their performance. In Section 4.7, we discuss the peakiness requirement of Feedback Multi-tone FSK and Multi-tone FSK. The conclusion follows in Section 4.8.

## 4.2 Feedback Multi-tone FSK

We assume a frequency-selective time-varying block fading channel model with time coherence $\mathcal{T}_c$, frequency coherence $\mathcal{F}_c$, and multipath spread $\mathcal{L}$ as in Chapter 3. Symbols not in the same coherent bandwidth or time interval experience independent fading.

The forward link of the Feedback Multi-tone FSK scheme still uses Multi-tone FSK symbols and repeats each symbol $N$ times as the Multi-tone FSK scheme does. That is, we pick $Q$ tones out of the total $M$ tones and use the combination of frequencies to represent a symbol. At the receiver side, we use a non-coherent receiver to demodulate and decode Multi-tone FSK symbols. The receiver employs a bank of match filters with their central frequencies tuned to each of the $M$ frequencies. The output of each match filter is the correlation of its tuned frequency against the received Multi-tone FSK signal. Since all the $M$ frequencies are mutually orthogonal on a symbol time, the output of a match filter is proportional to the amplitude of the associated tone.

We can treat the $M$ frequencies as $M$ frequency-division subchannels.

At the $n$th symbol time, the input and output of the $k$th subchannel, $\widetilde{x}_{k,n}$ and $\widetilde{y}_{k,n}$, have the following relation:

$$\widetilde{y}_{k,n} = h_{k,n}\widetilde{x}_{k,n} + w_{k,n}, \qquad \text{for } 1 \leq k \leq M,$$

where $h_{k,n}$ is the fading coefficient, and $w_{k,n}$ is a zero-mean complex additive white Gaussian noise with a variance $\sigma_w^2$, $CN(0, \sigma_w^2)$. The variables $\{h_{k,n}\}$ and $\{w_{k,n}\}$ are mutually independent. Under the Rayleigh fading assumption, $h_{k,n}$ is a complex Gaussian random variable, $CN(0, \sigma_h^2)$. The elements in $\{h_{k,n}\}_n$ can be considered to be independent if two symbols are far apart in time.

Letting $x_{k,n} = \widetilde{x}_{k,n}\sigma_h/\sigma_w$ and $r_{k,n} = |\widetilde{y}_{k,n}/\sigma_w|^2$, we obtain an equivalent channel with the transition probability

$$p_{r_{k,n}|x_{k,n}}(r|x) = \frac{1}{x^2+1} e^{-\frac{r}{x^2+1}}. \tag{4.1}$$

Similarly to the Multi-tone FSK scheme, the Feedback Multi-tone FSK scheme transmits in a low duty cycle fashion. The transmitter concentrates power over a fraction $\theta$ $(0 < \theta \leq 1)$ of time, and transmits on predetermined time intervals. The parameter $\theta$ is the duty cycle. A Feedback Multi-tone FSK symbol is also repeated $N$ times to obtain diversity. However, the amplitudes of tones are no more constant. The receiver will adjust the amplitudes of different tones to counteract the fading effect, which will be introduced in detail later.

The reverse link is, as the forward link, a fading channel subject to additive noise. Here, we assume that we have enough power (e.g., power consumption is not crucial for basestations) and enough degrees of freedom such that the capacity of the reverse link is much larger than the feedback data rate. Thus, we can receive feedback information with negligible error probability. Throughout this chapter, we assume that the feedback is error-free.

Note that, in wideband channels, we have sufficient degrees of freedom to establish feedback links from the receiver to the transmitter. Figure 4-1 shows a possible scheme

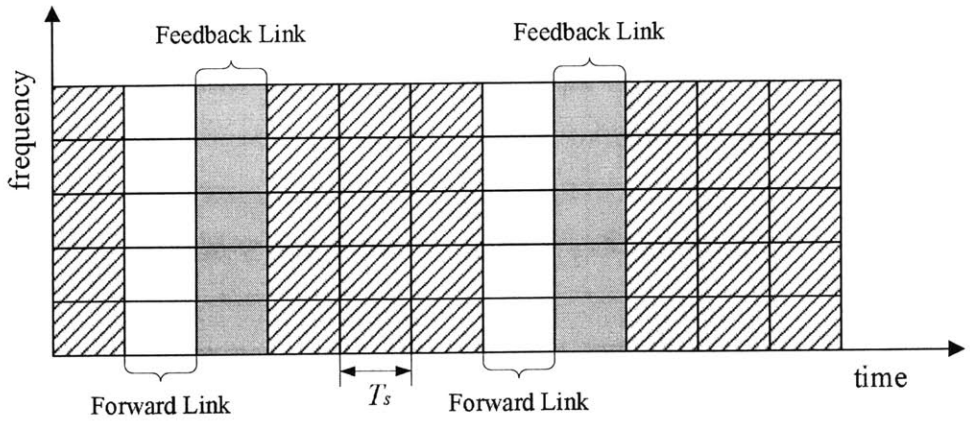


Figure 4-1: The forward link and the feedback link

when the duty cycle $\theta$ is low. That is, the feedback link uses the same $M$ frequencies in the forward link and transmits in the time slots right after the forward slots.

As we mentioned before, one major difference between the Multi-tone FSK scheme and the Feedback Multi-tone FSK scheme is that, for an Feedback Multi-tone FSK symbol, the amplitude of each tone is repeatedly adjusted according to the feedback. After each forward transmission, the receiver sends 1 bit of information back for each tone to indicate the reception quality. The transmitter adjusts each tone's amplitude accordingly in the next iteration to reduce decoding errors.

We use $\mathcal{S}$ to denote the complete set of symbols, and $\mathcal{S}_i \in \mathcal{S}$ $(i = 1, 2, 3, ..., |\mathcal{S}|)$ for a symbol. We use $k \in \mathcal{S}_i$ to indicate the $k$th tone is one of the non-zero tones in $\mathcal{S}_i$. For a $Q$-tone Feedback Multi-tone FSK scheme with total $M$ tones, there are $\binom{M}{Q}$ combinations of $Q$ tones available, i.e., the size of $\mathcal{S}$, $|\mathcal{S}|$, is $\binom{M}{Q}$.

Suppose we send a symbol $\mathcal{S}_m$ from the receiver to the transmitter. In the forward link, if a tone $k \notin \mathcal{S}_m$, i.e., the $k$th tone is set to be "0" in the symbol $\mathcal{S}_m$, then $x_{k,n} = 0$ for all the $N$ iterations regardless of the feedback. Otherwise, $k$ is one of the $Q$ non-zero tones and $x_{k,n}$ is determined as follows.

For the first iteration, let $x_{k,1} = \overline{x}$, where $\overline{x}$ is determined by average power $P$:

$$\overline{x} = \sqrt{\frac{PT_s}{Q\theta}},$$

where $T_s$ is the time length of a symbol and $\theta$ is the duty cycle. The receiver obtains

$r_{k,1}$. If $r_{k,1}$ is greater than $2\overline{x}^2+1$, then the receiver feeds back "1" in the $k$th feedback subchannel; otherwise, it feeds back "0".

At the $l$th iteration ($l \geq 2$), the receiver compares the cumulative received power

$$V_{k,l} = \sum_{n=1}^{l} r_{k,n}, \tag{4.2}$$

with the threshold

$$A_l = (l+1)\,\overline{x}^2 + l. \tag{4.3}$$

If $V_{k,l}$ is greater, the receiver feeds back "1"; otherwise, "0".

To decide $x_{k,l+1}$, we use the following power control algorithm: if the feedback is "1", $x_{k,l+1} = 0$; otherwise,

$$x_{k,l+1} = \begin{cases} \overline{x}, & \text{if } x_{k,l} = 0, \\ \sqrt{x_{k,l}^2 + \overline{x}^2 + 1}, & \text{otherwise.} \end{cases}$$

After the $N$ iterations, the receiver forms the average received power $S_{k,N}$

$$S_{k,N} = \frac{1}{N} V_{k,N},$$

which are compared with the threshold

$$A = 1 + (1-\epsilon)\overline{x}^2, \tag{4.4}$$

where the parameter $\epsilon$ is chosen over $(0, 1)$.

For each of the $M$ tones, the same procedure is executed. The time averages of $r_{k,n}$ over the $N$ iterations, $S_{k,N}$, are obtained to compare with a threshold. If exactly $Q$ tones exceed the threshold, then the corresponding message $\mathcal{S}_m$ is decoded; otherwise, the receiver declares an error.

In Figure 4-2, we illustrate the underlying mechanism of the feedback scheme. Note that, since we use power detection for each tone, and for zero tones, we cannot cancel out the noise's power after we receive it, we do not use adaptive transmission on zero tones. The purpose of the feedback scheme is to improve the transmission quality

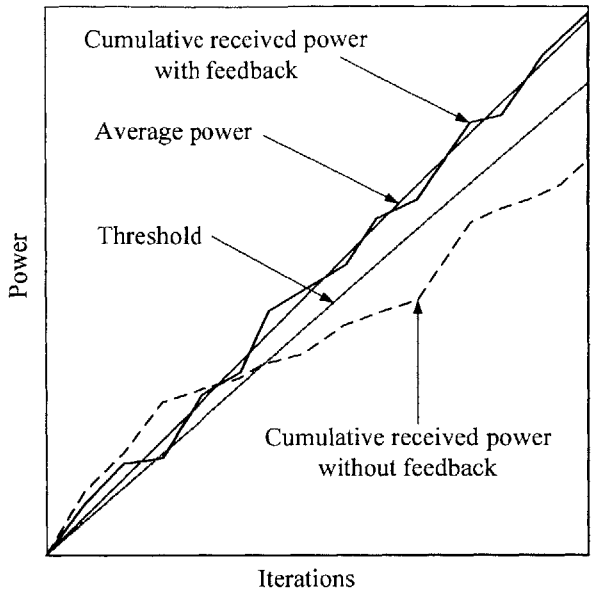


Figure 4-2: The cumulative received power with and without feedback

on nonzero tones by using adaptive transmission based on feedback. For a nonzero tone, we plot cumulative received power ($V_{k,N}$) curves in terms of repetition length $N$, alongside the threshold curve ($NA$) and the expected received power ($N(1+\overline{x}^2)$). The dashed line is a sample path of the cumulative received power without feedback. In this case, the cumulative received power has an increasing variance because it is a sum of i.i.d. random variables. The solid zigzaging line is a sample path of the cumulative received power with feedback. Since transmission power is adjusted according to the received signal quality, the curve is bounded around the expected received power. The boundedness is proved in Appendix C. Clearly, the feedback scheme can significantly reduce the detection error.

## 4.3 Power Consumption

Since the amplitudes of tones, $x_{k,n}$, are determined in run time, power consumption in the Feedback Multi-tone FSK scheme is a random variable. In this section, we prove that the average power consumption in the forward link converges to $P$ as $N$ grows. For a large $N$, the average power approximates $P$.

Notice

$$E\left[S_{k,N}\right] = \frac{1}{N}\sum_{n=1}^{N} E\left[r_{k,n}\right] \tag{4.5}$$

and

$$E\left[r_{k,n}\right] = x_{k,n}^2 + 1. \tag{4.6}$$

By substituting (4.6) into (4.5), we obtain

$$E\left[S_{k,N}\right] = \frac{1}{N}\sum_{n=1}^{N} x_{k,n}^2 + 1. \tag{4.7}$$

It can be proven that, for a non-zero tone,

$$\lim_{N\to\infty} E\left[\left|V_{k,N} - A_N\right|\right] < \infty \tag{4.8}$$

for arbitrary $N \geq 1$. The proof is available in Appendix C.

Based on (4.8), we have the limit

$$\lim_{N\to\infty} \frac{1}{N} E\left[\left|V_{k,N} - A_N\right|\right] = 0. \tag{4.9}$$

On the other hand,

$$\begin{aligned} & \lim_{N\to\infty} \frac{1}{N} E\left[\left|V_{k,N} - A_N\right|\right] \\ = & \lim_{N\to\infty} \frac{1}{N} E\left[\left|NS_{k,N} - (1+N)\,\overline{x}^2 - N\right|\right] \\ = & \lim_{N\to\infty} E\left[\left|S_{k,N} - \left(\overline{x}^2 + 1\right)\right|\right], \end{aligned}$$

which, combined with (4.9), leads to

$$\lim_{N\to\infty} E\left[\left|S_{k,N} - \left(\overline{x}^2 + 1\right)\right|\right] = 0. \tag{4.10}$$

Equation (4.10) implies

$$\lim_{N\to\infty} E\left[S_{k,N}\right] = 1 + \overline{x}^2. \tag{4.11}$$

Substituting (4.7) into (4.11), we have, for a non-zero tone $k$,

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} x_{k,n}^2 = \overline{x}^2. \tag{4.12}$$

Considering $\overline{x}^2$ occurs with a duty cycle $\theta$ in a symbol time $T_s$, the average power in the forward link converges to $Q\theta\overline{x}^2/T_s = P$, as $N$ grows.

## 4.4 Error Probability

During transmission, an error occurs if $S_{i,N} \geq A$ for some $i \notin \mathcal{S}_m$ (type I error), or $S_{k,N} < A$ for some $k \in \mathcal{S}_m$ (type II error).

Using the union bound, we upper bound the codeword error probability by

$$\begin{aligned} p_e &\leq \Pr\left(\cup_{i\notin\mathcal{S}_m} S_{i,N} \geq A\right) + \Pr\left(\cup_{k\in\mathcal{S}_m} S_{k,N} < A\right) \\ &\leq (M-Q)\Pr\left(S_{i,N} \geq A\right) + Q\Pr\left(S_{k,N} < A\right). \end{aligned} \tag{4.13}$$

The first term in (4.13) has been bounded in Chapter 3. We cite the result here:

$$\begin{aligned} & (M-Q)\Pr\left(S_{i,N} \geq A\right) \\ \leq\ & Q\exp\left(-\ln\binom{M}{Q}\left[\frac{(1-\epsilon)PT_s'}{QN_0RT_s} - \frac{1}{Q}\right.\right. \\ & \left.\left. -\frac{\theta}{RT_s}\ln\left(1+\frac{(1-\epsilon)PT_s'}{Q\theta N_0}\right)\right]\right), \end{aligned} \tag{4.14}$$

where $R$ is the data rate,

$$R = \frac{\theta}{NT_s}\ln\binom{M}{Q}, \tag{4.15}$$

$N_0$ is the noise power density, and $T_s'$ is the effective symbol time which equals to the symbol time $T_s$ minus the channel delay spread $\mathcal{L}$. The reason that we call $T_s'$ the effective symbol time is because we setup a guard time which equals to $\mathcal{L}$ at the head of each symbol to avoid intersymbol interference and only detect signal power over the time interval $[\mathcal{L}, T_s]$. Note that, in the derivation of the upper bound (4.14), we have

considered the channel delay spread $\mathcal{L}$ which is not incorporated into the model that we used for Feedback Multi-tone FSK in this chapter. However, it is straightforward to extend our model to the case considering the channel delay spread. All conclusions will remain the same. Hence, in the following discussion, we will use (4.14) directly, and thus consider the channel delay spread in our results.

As for the second term in (4.13), we derive an upper bound as follows:

Define $m$ as

$$m = \left\lfloor \frac{\epsilon N \overline{x}^2 - 1}{1 + \overline{x}^2} \right\rfloor. \tag{4.16}$$

Since $\epsilon$ is chosen in $(0, 1)$ and $\overline{x}^2$ is positive, we can pick $N$ large enough such that $m$ is a positive integer. Based on (4.16), we have

$$\epsilon N \overline{x}^2 \geq m\overline{x}^2 + m + 1. \tag{4.17}$$

Multiplying the two sides of (4.4) by $N$, we obtain $NA = N + N\overline{x}^2 - \epsilon N\overline{x}^2$, which, combined with (4.17), generates the inequality

$$NA \leq (N - m)\,\overline{x}^2 + (N - m - 1) = A_{N-m-1}. \tag{4.18}$$

where $A_l$ is defined in (4.3).

The inequality $V_{k,N} < NA$, derived from $S_{k,N} < A$, yields $V_{k,N} < A_{N-m-1}$ by applying (4.18). Noticing that $V_{k,i}$ is monotonically increasing with $i$, we have

$$V_{k,i} < A_{N-m-1} \text{ for } i = N - m - 1, ..., N. \tag{4.19}$$

Since $S_{k,N} < A$ is a sufficient condition for (4.19), its probability is smaller than that of (4.19), i.e.,

$$\begin{aligned} &\Pr\left(S_{k,N} < A\right) \leq \Pr\left(V_{k,N-m-1} < A_{N-m-1}, ..., V_{k,N} < A_{N-m-1}\right) \\ = &\int_0^{A_{N-m-1}} \prod_{i=N-m}^{N} \Pr\left(V_{k,i} < A_{N-m-1} | V_{k,i-1} < A_{N-m-1}, V_{k,N-m-1}\right) p_{V_{k,N-m-1}}(v)\, dv \end{aligned} \tag{4.20}$$

where (4.20) is obtained by applying the Total Probability Theorem, the chain rule of conditional probability and Markov property of $\{V_{k,i}\}$. It can be proven that, for $i = 0, 1, ..., m$,

$$\begin{aligned}&\Pr\left(V_{k,N-m+i} < A_{N-m-1} | V_{k,N-m+i-1} < A_{N-m-1}, V_{k,N-m-1}\right)\\&\leq 1 - \exp\left(-\frac{A_{N-m-1} - V_{k,N-m-1}}{A_{N-m+i-1} - V_{k,N-m-1}}\right).\end{aligned}$$

Substituting the bound into (4.20), we have

$$\Pr\left(S_{k,N} < A\right) \leq \int_0^{A_{N-m-1}} P_{V_{k,N-m-1}}(v) \prod_{i=N-m-1}^{N-1} \left[1 - \exp\left(-\frac{A_{N-m-1} - v}{A_i - v}\right)\right] dv. \tag{4.21}$$

Since $1 - \exp\left(-\frac{a-x}{b-x}\right)$ is a concave function of $x$ when $0 \leq x \leq a < b$, the function

$$f(v) = \prod_{i=N-m-1}^{N-1} \left[1 - \exp\left(-\frac{A_{N-m-1} - v}{A_i - v}\right)\right]$$

is also concave. By applying Jensen's Inequality to (4.21), we obtain

$$\Pr\left(S_{k,N} < A\right) \leq \xi_1 \prod_{i=N-m-1}^{N-1} \left[1 - \exp\left(-\frac{A_{N-m-1} - \xi_2}{A_i - \xi_2}\right)\right] \tag{4.22}$$

where $\xi_1 = \Pr\left(V_{k,N-m-1} < A_{N-m-1}\right)$ and $\xi_2 = E\left[V_{k,N-m-1} | V_{k,N-m-1} < A_{N-m-1}\right]$.

Define a function

$$E_f(x) = -\frac{1}{x} \sum_{i=0}^{x} \ln\left[1 - \exp\left(-\frac{\xi_3}{i + \xi_3}\right)\right], \tag{4.23}$$

where the constant

$$\xi_3 = \frac{A_{N-m-1} - \xi_2}{\overline{x}^2 + 1} > 0.$$

We rewrite (4.22) as

$$\Pr\left(S_{k,N} < A\right) \leq \xi_1 \exp\left[-m E_f(m)\right]. \tag{4.24}$$

For $i \gg 1$, we have the following approximation:

$$-\ln\left(1-e^{-\frac{\xi_3}{i+\xi_3}}\right) \approx \ln\left(i+\xi_3\right) - \ln\left(\xi_3\right).$$

Hence, $E_f\left(x\right)$ is of the order of $\ln\left(x\right)$ as $x$ increases.

Based on (4.16), we have

$$m \geq \frac{\epsilon N \overline{x}^2 - 1}{1+\overline{x}^2}. \tag{4.25}$$

Rearranging (4.15) yields the equation

$$N = \frac{\theta}{RT_s} \ln \binom{M}{Q}. \tag{4.26}$$

By substituting (4.26) into (4.25), we obtain

$$m \geq \ln \binom{M}{Q} \frac{\theta\epsilon\overline{x}^2}{RT_s\left(1+\overline{x}^2\right)} - \frac{1}{1+\overline{x}^2}. \tag{4.27}$$

Applying (4.27) to (4.24), we upper bound the second term in (4.13) by

$$Q \Pr\left(S_{k,N} < A\right) \leq Q\xi_1 \exp\left[-\mathfrak{m}\left(M,R,\theta,\epsilon\right) E_f\left(\mathfrak{m}\left(M,R,\theta,\epsilon\right)\right)\right]. \tag{4.28}$$

where $E_f\left(x\right)$, defined in (4.23), is of the order of $\ln\left(x\right)$ as $x$ increases; $\xi_1$ is a constant satisfying $0 < \xi_1 \leq 1$; and

$$\mathfrak{m}\left(M,R,\theta,\epsilon\right) = \ln \binom{M}{Q} \frac{\theta\epsilon\overline{x}^2}{RT_s\left(1+\overline{x}^2\right)} - \frac{1}{1+\overline{x}^2}.$$

As discussed in Chapter 3, $\theta$ and $\epsilon$ need to approach 0 to achieve the AWGN capacity in the wideband limit. Since $E_f\left(x\right)$ is of the order of $\ln\left(x\right)$ as $x$ increases, we may make $\mathfrak{m}\left(M,R,\theta,\epsilon\right) E_f\left(\mathfrak{m}\left(M,R,\theta,\epsilon\right)\right)$ grow faster than $\ln\binom{M}{Q}$ as $M$ increases while keeping $\theta$ and $\epsilon$ decreasing. Therefore, the second term in (4.13) is dominated by the first term and the codeword error probability is upper bounded by

$$p_e \leq Q \exp\left(-\ln\binom{M}{Q}\left[\frac{\left(1-\epsilon\right)PT_s'}{QN_0RT_s} - \frac{1}{Q} - \frac{\theta}{RT_s}\ln\left(1+\frac{\left(1-\epsilon\right)PT_s'}{Q\theta N_0}\right)\right]\right). \tag{4.29}$$

in the wideband regime.

On the other hand, a lower bound of $p_e$ is derived by using

$$\begin{aligned} p_e &\geq \sum Pr(\text{one error occurs}) - \sum Pr(\text{two errors occur}) \\ &= (M-Q)\,\Pr\left(S_{i,N} \geq A\right) + Q\Pr\left(S_{k,N} < A\right) \\ &\quad -(M-Q)Q\Pr\left(S_{i,N} \geq A\right)\Pr\left(S_{k,N} < A\right) \\ &\quad -(M-Q)(M-Q-1)\Pr\left(S_{i,N} \geq A\right)^2/2 \\ &\quad -Q(Q-1)\Pr\left(S_{k,N} < A\right)^2/2. \end{aligned} \tag{4.30}$$

Notice that, among all the terms in (4.30), the term $(M-Q)\ \Pr\left(S_{i,N} \geq A\right)$ decreases slowest as $M$ increases. Thus, the lower bound (4.30) is dominated by the term $(M-Q)\ \Pr\left(S_{i,N} \geq A\right)$ for a large $M$. We omit other terms in the lower bound calculation.

According to Chapter 3, $\Pr\left(S_{i,N} \geq A\right)$ is lower bounded by

$$\begin{aligned} \Pr\left(S_{i,N} \geq A\right) \geq \exp\Bigg(-\ln\binom{M}{Q}\Bigg[&\frac{(1-\epsilon)\,PT_s'}{QN_0RT_s} - \frac{1}{Q} \\ &-\frac{\theta}{RT_s}\ln\left(1+\frac{(1-\epsilon)\,PT_s'}{Q\theta N_0}\right) + o\left(N\right)\Bigg]\Bigg), \end{aligned} \tag{4.31}$$

where $o(N)$ is a term that vanishes as $N$ increases,

$$o(N) = \frac{1}{2N}\ln(2\pi NA^2) + \frac{1}{12N^2},$$

and $R$ stands for the data rate of the forward link. Noticing

$$M - Q \geq \frac{(M-Q)\binom{M}{Q}^{1/Q}}{M-Q+1},$$

we derive a lower bound of $p_e$ based on (4.31),

$$\begin{aligned} p_e \geq \frac{M-Q}{M-Q+1} \times \\ \exp\Bigg(-\ln\binom{M}{Q}\Bigg[\frac{(1-\epsilon)\,PT_s'}{QN_0RT_s} - \frac{1}{Q} - \frac{\theta}{RT_s}\ln\left(1+\frac{(1-\epsilon)\,PT_s'}{Q\theta N_0}\right) + o_1\left(N\right)\Bigg]\Bigg). \end{aligned} \tag{4.32}$$

Comparing the upper bound (4.29) with the lower bound (4.32), we conclude that, in the wideband regime,

$$p_e \sim \exp\left(-\ln\binom{M}{Q} E_{r,FMFSK}(R)\right), \tag{4.33}$$

where

$$\begin{aligned} E_{r,FMFSK}(R) &= \frac{(1-\epsilon)\,PT_s'}{QN_0RT_s} - \frac{1}{Q} \\ &\quad -\frac{\theta}{RT_s}\ln\left(1+\frac{(1-\epsilon)\,PT_s'}{Q\theta N_0}\right). \end{aligned} \tag{4.34}$$

The function $E_{r,FMFSK}(R)$ describes the true asymptotic dependence of the error probability on $M$. We call it the error exponent of Feedback Multi-tone FSK which is analogous to the treatment of random block coding over discrete memoryless channels by Gallager [4, §5.8].

## 4.5 Capacity-Achieving Property

In Chapter 3, we have proven that, in multipath fading channels, Multi-tone FSK can achieve the capacity of an AWGN channel with the same average received power at the wideband limit. We now prove that Feedback Multi-tone FSK has the same capacity-achieving property in the wideband limit.

**Proposition 4.1** *Feedback Multi-tone FSK can achieve any data rates that satisfy*

$$R < \left(1-\frac{\mathcal{L}}{T_s}\right)\frac{P}{N_0} \tag{4.35}$$

*with an arbitrarily small probability of error over a multipath fading channel with average power constraint $P$, by using large enough bandwidth.*

**Proof:** For the Feedback Multi-tone FSK scheme, the capacity is the maximal data rate that keeps $E_r(R) \geq 0$ in (4.33), which is

$$C = \frac{(1-\epsilon)\,PT_s'}{N_0 T_s} - \frac{Q\theta}{T_s} \ln\left(1 + \frac{(1-\epsilon)\,PT_s'}{Q\theta N_0}\right). \tag{4.36}$$

For any data rate less than (4.36), the upper bound (4.29) decreases to zero when $M$ increases to infinity. That is, we can use Feedback Multi-tone FSK to transmit data with arbitrarily small error probability by using bandwidth large enough when the data rate is less than the capacity defined by (4.36).

The capacity $C$ can get arbitrarily close to $\frac{PT_s'}{N_0 T_s}$, i.e., $\left(1 - \frac{\mathcal{L}}{T_s}\right)\frac{P}{N_0}$, by letting $\theta$ and $\epsilon$ decrease to zero as the bandwidth increases to infinity. Hence, as long as the data rate satisfies (4.35), the Feedback Multi-tone FSK scheme can achieve arbitrarily small error probability by picking bandwidth large enough. ■

Recalling that the channel is assumed to be underspread and $\mathcal{L} \ll T_s$, we conclude that the capacity of Multi-tone FSK approaches Shannon's infinite bandwidth capacity limit $\frac{P}{N_0}$.

## 4.6 Gain from Feedback

In Chapter 3, we derived the error exponent of Multi-tone FSK. In the following, we compare the performance of the two schemes using their error exponents.

First, we define two parameters to rewrite the expressions of error exponents such that they can be compared under a unified framework. We define a parameter $\kappa$ as

$$\kappa = 1 - \frac{\text{Capacity}}{\text{Wideband Capacity Limit}},$$

which is the normalized difference between the capacity and the wideband capacity limit. The wideband capacity limit refers to the capacity achieved by using an infinite bandwidth. As we derived in Chapter 3, the wideband capacity limit for the Multi-tone FSK scheme is $\frac{PT_s'}{N_0 T_s}$. The Feedback Multi-tone FSK scheme has the same wideband capacity limit as Multi-tone FSK.

Define another parameter $\tau$ as

$$\tau = 1 - \frac{\text{Data Rate}}{\text{Wideband Capacity Limit}},$$

which is the normalized difference between the data rate and the wideband capacity limit. Note that $\tau \geq \kappa$, since data rates should be lower than capacity in order to achieve reliable communication.

In Chapter 3, we have obtained the capacity of the $Q$-tone FSK scheme,

$$C = \frac{PT_s'}{N_0 T_s} - \frac{Q\theta}{T_s} \ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right),$$

and its error exponent,

$$E_{r,MFSK}(R) = \frac{\theta}{RT_s}\left(\beta - 1 - \ln\beta\right),$$

where

$$\beta = \frac{RN_0T_s}{PT_s'} + \frac{Q\theta N_0}{PT_s'} \ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right).$$

The asymptotic dependence of the codeword error probability on $M$ is described by

$$p_e \sim \exp\left(-\ln\binom{M}{Q} E_{r,MFSK}(R)\right). \tag{4.37}$$

Thus, for $Q$-tone FSK,

$$\kappa = \frac{Q\theta N_0}{PT_s'} \ln\left(1 + \frac{PT_s'}{Q\theta N_0}\right), \tag{4.38}$$

and $\tau = 1 - \frac{RN_0T_s}{PT_s'}$. By defining $\vartheta(x)$ as

$$\vartheta(x) = x \ln\left(\frac{1}{x} + 1\right),$$

and $\vartheta^{-1}(x)$ its inverse function, we obtain

$$E_{r,MFSK}(\kappa,\tau) = \frac{\vartheta^{-1}(\kappa)}{Q(1-\tau)}\left[\kappa - \tau - \ln(1+\kappa-\tau)\right] \tag{4.39}$$

$$\approx \frac{\vartheta^{-1}(\kappa)(\tau-\kappa)^2}{2Q(1-\tau)} \quad \text{(Taylor expansion)}. \tag{4.40}$$

Since $\vartheta^{-1}(\kappa)$ and the first-order derivative of $\vartheta^{-1}(\kappa)$ are both 0 at $\kappa = 0$, $\vartheta^{-1}(\kappa) = O(\kappa^2)$ by a Taylor expansion.

For the Feedback Multi-tone FSK scheme, the capacity is the maximal data rate that keeps $E_r(R) \geq 0$, which is

$$C = \frac{(1-\epsilon)\,PT_s'}{N_0 T_s} - \frac{Q\theta}{T_s} \ln\left(1 + \frac{(1-\epsilon)\,PT_s'}{Q\theta N_0}\right).$$

Thus,

$$\kappa = \epsilon + \frac{Q\theta N_0}{PT_s'} \ln\left(1 + \frac{(1-\epsilon)\,PT_s'}{Q\theta N_0}\right), \tag{4.41}$$

and $\tau = 1 - \frac{RN_0T_s}{PT_s'}$. The error exponent of Feedback Multi-tone FSK is

$$E_{r,FMFSK}(\kappa, \tau) = \frac{\tau - \kappa}{Q(1-\tau)}. \tag{4.42}$$

After we write $E_{r,MFSK}$ and $E_{r,FMFSK}$ as functions of $\tau$ and $\kappa$, we compare the two schemes' performance.

### 4.6.1 Improved Error Performance

In both Multi-tone FSK and Feedback Multi-tone FSK cases, the capacity converges to $\frac{PT_s'}{N_0T_s}$ as bandwidth increases, and thus $\kappa$ approaches 0 as $M$ grows, while $\tau$ is kept constant as long as the data rate is unchanged.

When $\kappa$ is very small, the order of (4.39) in terms of $\kappa$ is determined by $\vartheta^{-1}(\kappa)$, because, according to (4.40), the remainder of the error exponent is approximated by $\frac{(\tau-\kappa)^2}{2Q(1-\tau)}$ which is a value determined by $\tau$ when $\kappa$ goes to zero. Hence, when $\tau$ is fixed, the error exponent of Multi-tone FSK $E_{r,MFSK}(\kappa, \tau)$ is a function of $\kappa$ in the order of $O(\kappa^2)$, and decreases to zero as bandwidth increases. To let the codeword error probability decrease with an increasing bandwidth, the decreasing rate of the error exponent should be slower than the increasing rate of $\ln\binom{M}{Q}$ such that $\ln\binom{M}{Q} E_{r,MFSK}(R) \to \infty$.

For Feedback Multi-tone FSK, the error exponent (4.42) converges to a limit $\frac{\tau}{Q(1-\tau)}$ as $\kappa$ goes to zero. The limit is always greater than zero if $\tau$ is greater than

$\kappa$. Hence, for Feedback Multi-tone FSK, as long as the data rate is less than the capacity, the exponent in (4.33) satisfies $\ln \binom{M}{Q} E_{r,FMFSK}(R) \to \infty$, and the codeword error probability decreases with bandwidth.

The ratio of $E_{r,FMFSK}(\kappa,\tau)$ to $E_{r,MFSK}(\kappa,\tau)$ is $\vartheta^{-1}(\kappa)(\tau-\kappa)/2$. Clearly, the error exponent of Multi-tone FSK is much smaller than that of Feedback Multi-tone FSK,

$$E_{r,MFSK}(\kappa,\tau) \ll E_{r,FMFSK}(\kappa,\tau), \tag{4.43}$$

for a small $\kappa$ and some $\tau$ greater than $\kappa$, because $\vartheta^{-1}(\kappa) \ll 1$ and $\tau-\kappa < 1$. Since the dependence of error probability on bandwidth is characterized by the error exponent, which is a function of data rate (recall that $\tau$ is a parameter determined by the data rate), we can find the tradeoff between error probability and data rate by substituting the error exponent into the error probability. Here, by (4.43), the Feedback Multi-tone FSK scheme has a better error probability-data rate tradeoff than the Multi-tone FSK scheme when any other conditions are equivalent. Or, equally, Feedback Multi-tone FSK provides a better error performance when both schemes have the same data rate.

By substituting (4.42) into (4.33) and noticing $\binom{M}{Q}^{\frac{1}{Q}} \sim M$ when $Q \ll M$, we conclude that the error probability of the Feedback Multi-tone FSK scheme decreases roughly inversely with bandwidth:

$$p_e \sim \frac{1}{M} \exp\left(-QE_{r,FMFSK}(R)\right).$$

The error probability of the Multi-tone FSK scheme decreases in bandwidth much more slowly than the inverse of $M$, owing to the scheme's vanishing error exponent. In Figure 4-3, we plot the error probability curves of Multi-tone FSK and Feedback Multi-tone FSK in terms of the number of total tones $M$. Note that $M$ increases linearly with bandwidth.

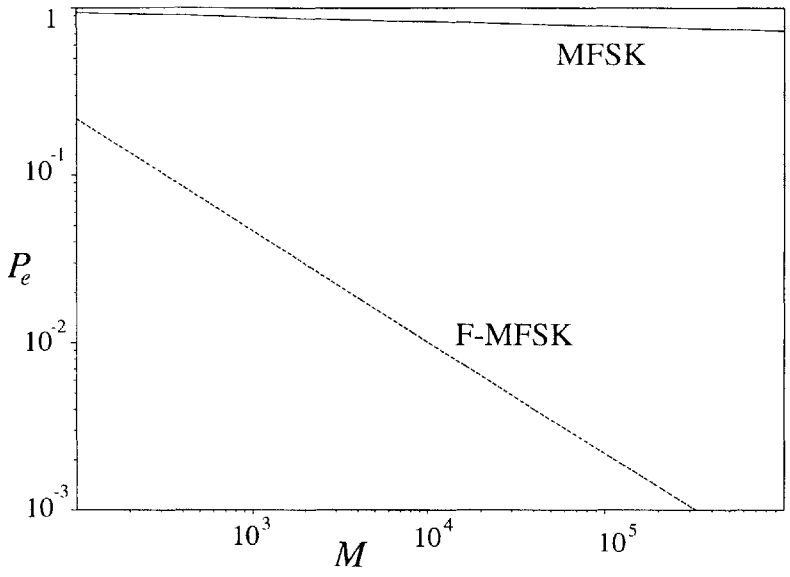


Figure 4-3: The error probability curves of Multi–tone FSK and Feedback Multi-tone FSK for $\kappa = 0.001$ and $\tau = 0.4$

### 4.6.2 Smaller Bandwidth

As we pointed out, the vanishing error exponent of Multi-tone FSK causes a very slow decaying rate of the codeword error probability as bandwidth increases, while the error exponent of Feedback Multi-tone FSK tends to a constant value greater than zero, causing a much faster decay rate. From the bandwidth perspective, Feedback Multi-tone FSK needs a smaller bandwidth to achieve the same error performance as Multi-tone FSK when all other conditions are equal.

Let Multi-tone FSK and Feedback Multi-tone FSK have the same error performance when their data rates are the same, i.e., (4.33) and (4.37) have the same exponent when $\tau_{FMFSK} = \tau_{MFSK} = \tau$. Then, the following equation can be obtained by substituting (4.40) and (4.42) into (4.33) and (4.37) respectively, and equalling them:

$$\ln \binom{M_{MFSK}}{Q} \frac{\vartheta^{-1}(\kappa_{MFSK})(\tau - \kappa_{MFSK})^2}{2(\tau - \kappa_{FMFSK})} = \ln \binom{M_{FMFSK}}{Q}.$$

In the wideband regime, the capacities of Multi-tone FSK and Feedback Multi-tone FSK approach the same wideband capacity limit, i.e., $\kappa_{FMFSK} \approx \kappa_{MFSK} \approx 0$. Thus, considering $\tau_{FMFSK} = \tau_{MFSK} = \tau$, we roughly have $\tau - \kappa_{FMFSK} \approx \tau - \kappa_{MFSK} \approx \tau$

which lead to

$$\ln\binom{M_{MFSK}}{Q}\frac{\vartheta^{-1}\left(\kappa_{MFSK}\right)\left(\tau-\kappa_{MFSK}\right)}{2}\approx\ln\binom{M_{FMFSK}}{Q},$$

where $\vartheta^{-1}\left(\kappa_{MFSK}\right)\ll 1$ and $\tau-\kappa_{MFSK}<1$ as discussed previously. Hence, $M_{MFSK}$ and $M_{FMFSK}$ satisfy the inequality

$$\ln\binom{M_{MFSK}}{Q}\gg\ln\binom{M_{FMFSK}}{Q},$$

and as a result, $M_{MFSK}\gg M_{FMFSK}$ because logarithm is a monotonically increasing function. That is, Multi-tone FSK need much more bandwidth to achieve the same error performance for the same data rate.

### 4.6.3 Lower Average Power

From the average power perspective, Feedback Multi-tone FSK needs a lower average power than Multi-tone FSK to achieve the same error performance because its better performance.

Let the Feedback Multi-tone FSK scheme and the Multi-tone FSK scheme have the same bandwidth, duty cycle, and data rate. Substituting (4.40) and (4.42) into (4.33) and (4.37), equalling them, and cancelling out the equivalent terms, we obtain

$$\frac{\vartheta^{-1}\left(\kappa_{MFSK}\right)\left(\tau-\kappa_{MFSK}\right)^{2}}{2}=\tau-\kappa_{FMFSK}. \tag{4.44}$$

Noticing that $\vartheta^{-1}\left(\kappa_{MFSK}\right)=O\left(\kappa_{MFSK}^{2}\right)$ and $\kappa_{MFSK}<\tau$, we derive an inequality from (4.44) for $\kappa_{MFSK}$ small enough:

$$\frac{\tau^{3}\left(\tau-\kappa_{MFSK}\right)}{2}>\tau-\kappa_{FMFSK}. \tag{4.45}$$

By rearranging the terms in (4.45), we obtain

$$\frac{\tau-\kappa_{FMFSK}}{\tau-\kappa_{MFSK}}<\frac{\tau^{3}}{2}.$$

Considering $\tau < 1$, the above inequality leads to

$$\frac{\tau - \kappa_{FMFSK}}{\tau - \kappa_{MFSK}} < 1,$$

which is equivalent to

$$\kappa_{MFSK} < \kappa_{FMFSK}.$$

We use $P_{MFSK}$ and $P_{FMFSK}$ to denote the average power for Multi-tone FSK and Feedback Multi-tone FSK.

From (4.38), we obtain $\kappa_{MFSK}$ as

$$\kappa_{MFSK} = \frac{QN_0\theta}{P_{MFSK}T_s'} \ln\left(1 + \frac{P_{MFSK}T_s'}{QN_0\theta}\right). \tag{4.46}$$

Similarly, by using (4.41), $\kappa_{FMFSK}$ is obtained:

$$\kappa_{FMFSK} = \epsilon + \frac{QN_0\theta}{P_{FMFSK}T_s'} \ln\left(1 + (1-\epsilon)\frac{P_{FMFSK}T_s'}{QN_0\theta}\right). \tag{4.47}$$

Note that $\epsilon$ is a parameter chosen to be $\epsilon \ll \max\left(1, \frac{QN_0\theta}{P_{FMFSK}T_s'}\right)$. (4.47) can be approximated as

$$\kappa_{FMFSK} = \frac{QN_0\theta}{P_{FMFSK}T_s'} \ln\left(1 + \frac{P_{FMFSK}T_s'}{QN_0\theta}\right). \tag{4.48}$$

Comparing (4.46) with (4.48), and noticing that $\frac{1}{x}\ln(1+x)$ is a monotonically decreasing function of $x$, we have

$$P_{MFSK} > P_{FMFSK}, \tag{4.49}$$

viz., when bandwidth, duty cycle, and data rates of Multi-tone FSK and Feedback Multi-tone FSK are equal, Feedback Multi-tone FSK needs a lower average power to achieve the same error probability.

## 4.7 Requirement on Peakiness

In this section, we consider the Multi-tone FSK scheme and the Feedback Multi-tone FSK scheme's requirements on peakiness.

We use a parameter $a$ to denote the ratio of $Q$ to $M$, i.e., $a = \frac{Q}{M}$, which characterizes the peakiness of Multi-tone FSK and Feedback Multi-tone FSK in frequency. The smaller $a$, the peakier the schemes in frequency. In the following, we prove that, both the Multi-tone FSK scheme and the Feedback Multi-tone FSK scheme need peakiness for vanishing error probability. Moreover, the requirement on peakiness is less strict for Feedback Multi-tone FSK, owing to its better error exponent.

For the Multi-tone FSK signaling scheme and the Feedback Multi-tone FSK scheme, the codeword error probabilities are

$$p_e \sim \exp\left(-\ln\binom{M}{Q}\frac{\vartheta^{-1}(\kappa)}{Q(1-\tau)}\left[\kappa-\tau-\ln(1+\kappa-\tau)\right]\right), \tag{4.50}$$

and

$$p_e \sim \exp\left(-\ln\binom{M}{Q}\frac{\tau-\kappa}{Q(1-\tau)}\right), \tag{4.51}$$

respectively.

Applying Stirling's approximation,

$$\sqrt{2\pi N}N^N e^{\left(-N+\frac{1}{12N+1}\right)} < N! < \sqrt{2\pi N}N^N e^{\left(-N+\frac{1}{12N}\right)},$$

we obtain an upper bound for $\binom{M}{Q}^{\frac{1}{Q}}$ given $Q = aM$:

$$\binom{M}{Q}^{\frac{1}{Q}} \leq \frac{\exp\left(\frac{1}{12a^2M^2}-\frac{1}{(12aM+1)aM}-\frac{1}{[12(M-aM)+1]aM}\right)}{a(1-a)^{\frac{1-a}{a}}}\left(\frac{1}{\sqrt{2\pi Ma(1-a)}}\right)^{\frac{1}{aM}},$$

and a lower bound for $\binom{M}{Q}^{\frac{1}{Q}}$:

$$\binom{M}{Q}^{\frac{1}{Q}} \geq \frac{\exp\left(\frac{1}{aM(12aM+1)}-\frac{1}{(12aM)aM}-\frac{1}{[12(M-aM)]aM}\right)}{a(1-a)^{\frac{1-a}{a}}}\left(\frac{1}{\sqrt{2\pi Ma(1-a)}}\right)^{\frac{1}{aM}}.$$

Note that $\lim_{M\to\infty}\left(\frac{1}{M}\right)^{\frac{1}{M}}=1$. We obtain the limit

$$\begin{aligned}\lim_{M\to\infty}\frac{1}{Q}\ln\binom{M}{Q} &= -\ln a\,(1-a)^{\frac{1-a}{a}}\\ &= \frac{1}{a}H_b(a),\end{aligned}\tag{4.52}$$

where

$$H_b(a)=-a\ln a-(1-a)\ln(1-a)$$

is the entropy of a Bernoulli distribution with the parameter $a$. By substituting (4.52) into (4.50) and (4.51), we get

$$p_e\sim\exp\left(-\frac{1}{a}H_b(a)\,\frac{\vartheta^{-1}(\kappa)}{Q(1-\tau)}\left[\kappa-\tau-\ln(1+\kappa-\tau)\right]\right),\tag{4.53}$$

and

$$p_e\sim\exp\left(-\frac{1}{a}H_b(a)\,\frac{\tau-\kappa}{(1-\tau)}\right).\tag{4.54}$$

In both (4.53) and (4.54), to let $p_e\to 0$ in the wideband domain, we must let $a\to 0$. Hence, in order to have a decreasing error probability as the bandwidth increases, both the Multi-tone FSK scheme and the Feedback Multi-tone FSK scheme need peakiness (a vanishing ratio of $Q$ to $M$). Another observation based on (4.53) and (4.54) is that, to achieve the same error performance, the Multi-tone FSK scheme need a smaller $a$ than the Feedback Multi-tone FSK scheme. In the exponents of (4.53) and (4.54), we have the inequality between the two error exponents:

$$\frac{\vartheta^{-1}(\kappa)}{Q(1-\tau)}\left[\kappa-\tau-\ln(1+\kappa-\tau)\right]\ll\frac{\tau-\kappa}{(1-\tau)}.$$

Thus, the value of $\frac{1}{a}H_b(a)$ in (4.53) should be much larger than that in (4.54) to have the same error probability in the two schemes.

## 4.8 Conclusions

In this chapter, we proposed a Feedback Multi-tone FSK scheme which achieves the wideband capacity limit. We derived an upper bound and a lower bound for the codeword error probability. These two bounds lead to the error exponent of Feedback Multi-tone FSK, which describes how the error probability depends on the different parameters. According to our results, in the wideband regime, Feedback Multi-tone FSK enjoys a better error performance than Multi-tone FSK, although the capacity is not improved by using feedback. From the resource usage perspective, Feedback Multi-tone FSK requires a smaller bandwidth or a lower average power to achieve the same performance as Multi-tone FSK.

Our results show that, in power limited scenarios, even if channel measurement and tracking are not feasible, feedback could still significantly improve the error performance.

Note that, our results introduced in this chapter are based on the error-free feedback assumption which is unrealistic. In the future research, we can investigate the feedback scheme by incorporating a noisy feedback link and study the impact of feedback link quality on the scheme's performance.

# Chapter 5

# Conclusions

This dissertation has investigated the information theoretic aspects of the communication over wideband fading channels where neither the transmitter nor the receiver has channel state information. First, the structure of the optimal input signals has been explored for the discrete-time memoryless Rician fading channel and the Rayleigh block fading channel. We have proven that the optimal input amplitude of the fading channels has certain discrete structure when the input is subject to an average power constraint or a peak power constraint. In particular, it has been shown that binary inputs are optimal in the low SNR regime.

We proposed a Multi-tone FSK scheme, which is capacity achieving in the wideband limit and has a vanishing codeword error probability when the bandwidth increases. Furthermore, considering the slow decay of the codeword error probability of Multi-tone FSK in bandwidth, we proposed the Feedback Multi-tone FSK scheme to improve the error performance. We have been shown that, with a small amount of feedback, Feedback Multi-tone FSK has a much better error probability than Multi-tone FSK.

Although using peaky signaling for transmission of information over wideband fading channels is justified from the theoretic point of view, there exist doubts regarding these schemes' practicality (peakiness, bandwidth, etc.). Compared with other peaky signaling schemes, such as the PPM and the flash FSK, Multi-tone FSK has many desirable features.

First, Multi-tone FSK can achieve the wideband capacity limit in general multi-path fading channels regardless the increase of the number of channel paths with the bandwidth, while many other schemes, such as the PPM, perform greatly depending on the channel uncertainty. As we have proven, the achievable data rate of Multi-tone FSK is $\left(1-\frac{\mathcal{L}}{T_s}\right)\frac{P}{N_0}$, which is related to the multi-path effect by the channel delay spread $\mathcal{L}$. Since $\mathcal{L}$ is determined by the time difference between the first arrival path and the last arrival path which is almost invariant with respect to the bandwidth, the increase of the number of resolvable paths has no impact on its performance.

Also, in the case that the number of channel paths increases very fast, signaling schemes requiring multi-path tracking and combining become infeasible due to the increasing receiver complexity. Multi-tone FSK, on the contrary, is still practical because it requires no channel tracking.

From the symbol generating point of view, Multi-tone FSK symbols are easy to generate. When the bandwidth is very large, schemes like the PPM requires extremely short pulses, which is difficult to generate and catch. In Multi-tone FSK, the symbol time is much larger than the inverse of the bandwidth, and the increasing bandwidth only leads to an increasing number of tones, which poses little difficulty for symbol generation.

Conceptually, Single-tone FSK shares many characteristics with Multi-tone FSK. In practice, however, Multi-tone FSK presents advantage of providing the choice of using more than one tone per symbol. For example, when the duty cycle is fixed, Multi-tone FSK can provide a higher achievable data rate than Single-tone FSK, or alleviate the need of large bandwidth, owing to its larger alphabet set. Also, in this case, with the same average power constraint, Multi-tone FSK has lower peak power per tone than Single-tone FSK, which reduces the schemes' requirement on peakiness. As we have shown in Chapter 3, owing to its flexibility on spectral efficiency, Multi-tone FSK enjoys a better performance in many occasions.

Feedback Multi-tone FSK scheme provides an even better performance if feedback links are available between the receiver and the transmitter. By using a small fraction of the available degrees of freedom for the feedback channel (e.g. Figure 4-1) and

power from the receiver, we can improve the error performance of the forward link greatly. This kind of transmission is desirable if the transmitter is highly power-constrained and the receiver has an abundant power supply.

The concepts of Multi-tone FSK and Feedback Multi-tone FSK combine theoretical principles with practical considerations in providing promising signaling schemes for wideband fading channels. By choosing more sophisticated coding schemes, detection rules, or other designs, the performance of the schemes can be further improved beyond the performance we provided in this thesis. These are topics for future study, which are out of the scope of this thesis. Rather than seek optimality, we provide robust and practical schemes which perform very close to the optimal limit. Also, in future research, we can study the feedback scheme with a noisy feedback link.

Besides theoretically studying efficient signaling schemes, experimental research is also necessary because a theoretically sound idea may not be implementable. In Appendix D, we include some experimental results to endorse the Multi-tone FSK scheme. To fully study the performance of Multi-tone FSK in practice, further experimental investigation is needed.

# Appendix A

# Existence and Uniqueness of the Optimal Input Distribution

In this Appendix, we establish the existence of the optimal amplitude distributions for the discrete-time memoryless Rician fading channels and the Rayleigh block fading channels. We also prove the uniqueness of the optimal amplitude distribution for the Rayleigh block fading channels. The method that we use in the following is similar to the approach used in [10] and [15].

If $X$ is a normed linear space, then the set of continuous linear functionals on $X$ is called the dual (or conjugate) space of $X$ and is denoted $X^*$. In our optimization problem, we consider distributions as elements of the dual of the space of the bounded continuous functions. A distribution $D(x) \in X^*$ acts on a bounded continuous function $i(x)$ as $\langle i(x), D(x) \rangle = \int i(x)\, dD(x)$. For discrete random variables, the density includes mass points, which can be represented as $\sum_j p_j \delta(x - x_j)$. The distribution $\sum_j p_j u(x - x_j)$ is a linear functional which acts on $i(x)$ as $\sum_j p_j i(x_j)$. This approach includes not only the continuous random variables but also the discrete ones into the consideration. Optimization results are then obtained using the weak-star topology on $X^*$ [22, Sec. 5.10].

Existence and uniqueness can be showed by using the following theorem, which was proven in [15]. The remainder of this Appendix establishes that the sets of distribution functions and the mutual informations meet the conditions of the theorem.

**Theorem A.1** *(Abou-Fayçal et al. [15, Theorem 1]) If $I$ is a real-valued, weak-star continuous functional on a weak-star compact set $\Omega$, then $I$ achieves its maximum on $\Omega \subseteq X^*$. If furthermore $\Omega$ is convex, and $I$ is strictly concave, then the maximum*

$$C = \max_{D \in \Omega} I(D)$$

*is achieved by a unique $D^*$ in $\Omega$.*

## A.1 Convexity and Weak-Star Compact of the Support

Let $\mathcal{D}$ denote the set of all distribution functions, and let $\Omega_a$, $\Omega_p$, $\Omega_{ap} \subseteq \mathcal{D}$ be the distribution functions of nonnegative random variables with an average power constraint, an peak power constraint, and both, respectively. That is,

$$\Omega_a = \left\{ D \in \mathcal{D} | D\left(0^-\right) = 0 \text{ and } \int_0^\infty x^2 dD(x) \leq P_{av} \right\},$$

$$\Omega_p = \left\{ D \in \mathcal{D} | D\left(0^-\right) = 0 \text{ and } D\left(P_{peak}^+\right) = 1 \right\},$$

and

$$\begin{aligned} \Omega_{ap} \;=\; & \left\{ D \in \mathcal{D} | D\left(0^-\right) = 0, \right. \\ & \left. \int_0^\infty x^2 dD(x) \leq P_{av}, \text{ and } D\left(P_{peak}^+\right) = 1 \right\}. \end{aligned}$$

$\Omega_a$ was shown to be convex and weak-star compact in [15, Appendix 1.A]. We prove $\Omega_p$, and $\Omega_{ap}$ are convex and weak-star compact as follows.

$\Omega_p$: For any $D_1$, $D_2 \in \Omega_p$ and $\lambda \in [0,1]$, the linear combination $D = \lambda D_1 + (1-\lambda) D_2$ is nondecreasing, right continuous, with $D(0^-) = 0$ and $D\left(P_{peak}^+\right) = 1$. Hence, $D$ is still a distribution function in $\Omega_p$ and $\Omega_p$ is convex.

To prove that $\Omega_p$ is weak-star compact, we first show it is *tight*. Then, because the weak-star topology on distribution functions can be metrized by a metric (such as

the Lévy metric), Prokhorov's Theorem [25, Sec. III.2, Theorem 1] can be applied to show that $\Omega_p$ is *relatively compact*, i.e., every sequence of probability measures from $\Omega_p$ contains a subsequence which converges weakly to a probability measure. By proving that the limit probability measure belong to the original class, we establish that $\Omega_p$ is *sequentially compact*, which leads to that $\Omega_p$ is weak-star compact.

For $K > 0$, the interval $[0, K]$ is a compact set on the real line, the probability of the complimentary set is $1 - D(K) + D(0^-)$. For every $\varepsilon > 0$, if there is a $K > 0$ such that

$$\sup_{D \in \Omega_p} \left(1 - D(K) + D\left(0^-\right)\right) \leq \varepsilon,$$

the set $\Omega_p$ is tight. By setting $K = P_{peak}$, it is straightforward to show that $\Omega_p$ is tight.

Because $D(0^-) = 0$ and $D\left(P_{peak}^+\right) = 1$ which follows from [25, Sec. III.1, Theorem 1], the limit distribution function $D^*$ is in $\Omega_p$. Hence, $\Omega_p$ is weak-star compact.

$\Omega_{ap}$: For any $D_1$, $D_2 \in \Omega_{ap}$, the linear combination $D = \lambda D_1 + (1 - \lambda) D_2$ satisfies

$$\begin{aligned} & \int_0^\infty x^2 dD(x) \\ = & \int_0^\infty x^2 d\lambda D_1 + \int_0^\infty x^2 d(1-\lambda) D_2(x) \\ \leq & \lambda P_{av} + (1-\lambda) P_{av} \\ = & P_{av}. \end{aligned}$$

With $D(0^-) = 0$ and $D\left(P_{peak}^+\right) = 1$, $D$ is still a distribution function in $\Omega_{ap}$. Hence, $\Omega_{ap}$ is convex.

We use the same argument as that of proving that $\Omega_p$ is weak-star compact to show that $\Omega_{ap}$ is weak-star compact.

The set $\Omega$ is tight if, for every $\varepsilon > 0$, there is a $K > 0$ such that

$$\sup_{D \in \Omega_{ap}} \left(1 - D(K) + D\left(0^-\right)\right) \leq \varepsilon.$$

By setting $K = P_{peak}$, we show that $\Omega_{ap}$ is tight.

It remains to prove that the limit distribution function $D^*$ is in $\Omega_{ap}$.

$D(0^-) = 0$ and $D\left(P_{peak}^+\right) = 1$ follow from [25, Sec. III.1, Theorem 1]. Moreover, it is clear that $D^*$ satisfies the peak power constraint. To show that $D^*$ satisfies the average power constraint, we use the same argument used in [15]

$$\int_0^\infty x^2 dD^*(x) \leq \lim_{n\to\infty} \inf \int_0^\infty x^2 dD(x) \leq P_{av}.$$

Hence, the limit distribution function $D^*$ is in $\Omega_{ap}$ and $\Omega_{ap}$ is weak-star compact.

## A.2 Weak-star Continuity of Mutual Information

Weak-star continuity of a function $f$ is defined as

$$D_n \xrightarrow{w*} D \Rightarrow f(D_n) \to f(D).$$

We prove that the mutual informations for Rician fading channel and Rayleigh block fading channel are weak-star continuous. In the following discussion, the mutual information resulting from an input distribution function $D_n$ is denoted by $I_n$.

### A.2.1 Rician Fading Channels

As we derived before, with an input distribution $D_n$, the mutual information in a Rician fading channel can be written as

$$\begin{aligned} I_n(x;y) &= -\int_0^\infty \left[\ln\left(\sigma_h^2 r_x^2 + \sigma_w^2\right) + 1\right] dD_n(r_x) \\ &\quad -\int_0^\infty p_R(R;D_n) \ln p_R(R;D_n) dR \end{aligned} \tag{A.1}$$

where

$$p_R(R;D_n) = \int_0^\infty g(R|r_x) dD_n(r_x) \tag{A.2}$$

and $g(R|r_x)$ is defined in (2.7).

Let $D_n \xrightarrow{w*} D$. If we can show that the first term in (A.1) satisfies

$$\int_0^\infty \left[\ln\left(\sigma_h^2 r_x^2 + \sigma_w^2\right) + 1\right] dD_n\left(r_x\right) \to \int_0^\infty \left[\ln\left(\sigma_h^2 r_x^2 + \sigma_w^2\right) + 1\right] dD\left(r_x\right) \tag{A.3}$$

and the second term satisfies

$$\int_0^\infty p_R(R; D_n) \ln p_R(R; D_n) dR \to \int_0^\infty p_R(R; D) \ln p_R(R; D) dR, \tag{A.4}$$

then, as the difference of these two terms, the mutual information can be proven to be weak-star continuous.

The proof of (A.3) is as follows:

We must show that

$$\int_0^\infty \ln\left(\frac{\sigma_h^2}{\sigma_w^2} r_x^2 + 1\right) dD_n\left(r_x\right) \to \int_0^\infty \ln\left(\frac{\sigma_h^2}{\sigma_w^2} r_x^2 + 1\right) dD\left(r_x\right).$$

This follows if the integrand is uniformly integrable in $D_n$. That is, if the function

$$q_1\left(r_x\right) = \begin{cases} 0 & \text{for } r_x < 0, \\ \ln\left(\frac{\sigma_h^2}{\sigma_w^2} r_x^2 + 1\right), & \text{for } r_x \geq 0, \end{cases}$$

is continuous, $\int_0^\infty |q_1| \, dD_n < \infty$, and

$$\lim_{b \to \infty} \int_b^\infty \ln\left(\frac{\sigma_h^2}{\sigma_w^2} r_x^2 + 1\right) dD_n\left(r_x\right) = 0 \tag{A.5}$$

uniformly in $n$.

The function $q_1$ is continuous. Moreover, since $\ln\left(\frac{\sigma_h^2}{\sigma_w^2} r_x^2 + 1\right) \leq \frac{\sigma_h^2}{\sigma_w^2} r_x^2$ for $r_x \geq 0$

$$\int_0^\infty |q_1| \, dD_n \leq \begin{cases} \frac{\sigma_h^2}{\sigma_w^2} P_{av} & \text{in the average power limited case} \\ \frac{\sigma_h^2}{\sigma_w^2} P_{peak} & \text{in the peak power limited case} \end{cases}$$

for every $n$. To verify (A.5), note that, for $r_x$ large enough, we have $\ln\left(\frac{\sigma_h^2}{\sigma_w^2} r_x^2 + 1\right) \leq r_x$. Therefore, for $b$ large enough,

$$\begin{aligned}\int_b^\infty \ln\left(\frac{\sigma_h^2}{\sigma_w^2}r_x^2+1\right)dD_n\left(r_x\right) &\leq \int_b^\infty r_x dD_n\left(r_x\right)\\ &\leq \frac{1}{b}\int_b^\infty r_x^2 dD_n\left(r_x\right)\\ &\leq \frac{P_{av}}{b} \text{ or } \frac{P_{peak}}{b}.\end{aligned}$$

The bound is independent of $n$ and converges to 0 as $b \to \infty$ in either the average power limited case or the peak power limited case. Hence, $q_1\left(r_x\right)$ is uniformly integrable in $D_n$, and the first term in (A.1) is weak-star continuous.

To show (A.4), we need to establish

$$\lim_{n\to\infty}\int_0^\infty p_R(R;D_n)\ln p_R(R;D_n)dR \to \int_0^\infty \lim_{n\to\infty} p_R(R;D_n)\ln p_R(R;D_n)dR \quad \text{(A.6)}$$

and

$$\lim_{n\to\infty} p_R(R;D_n)\ln p_R(R;D_n) = p_R(R;D)\ln p_R(R;D). \quad \text{(A.7)}$$

To prove (A.7), note that the integrand in (A.2) is a bounded continuous function of $r_x$. Hence, $p_R(R;D)$ is a continuous function of $D$ for all $R>0$. Since $x\ln x$ is continuous, $p_R(R;D)\ln p_R(R;D)$ is continuous in $D$. Thus, (A.7) holds.

To show (A.6) is more complicated. We need to find an integrable function $q_2\left(R\right)$ such that

$$\left|p_R(R;D_n)\ln p_R(R;D_n)\right| \leq q_2\left(R\right). \quad \text{(A.8)}$$

First, we consider the average power limited Rician fading channel:

Since $I_0\left(x\right) < \exp(x)$, for $\forall r_x > 0$

$$g(R|r_x) \leq \frac{1}{\sigma_h^2 r_x^2+\sigma_w^2}\exp\left[-\frac{R-2\sqrt{R}mr_x+m^2r_x^2}{\sigma_h^2r_x^2+\sigma_w^2}\right]. \quad \text{(A.9)}$$

We first establish an upper bound of $g(R|r_x)$ in the form of $\left(\sigma_h^2r_x^2+\sigma_w^2\right)f_2\left(R\right)$

where

$$f_2(R) \geq \frac{1}{\left(\sigma_h^2 r_x^2 + \sigma_w^2\right)^2} \exp\left[-\frac{R - 2\sqrt{R}mr_x + m^2 r_x^2}{\sigma_h^2 r_x^2 + \sigma_w^2}\right]. \tag{A.10}$$

$f_2(R)$ can be obtained in different forms. A straightforward way to get a form of $f_2(R)$ is to use the first order condition (FOC) and obtain the maximum of the RHS in (A.10). The value of $r_x$ maximizing the RHS in (A.10) is determined by the equation

$$\frac{\sqrt{R}m\sigma_w^2}{r_x} = 2r_x^2\sigma_h^4 + \sqrt{R}mr_x\sigma_h^2 + m^2\sigma_w^2 + 2\sigma_h^2\sigma_w^2 - R\sigma_h^2. \tag{A.11}$$

The closed form solution $r_x^*(R)$ for (A.11) is very complicated. However, if we can find a lower bound $f_3(R)$ to $r_x^*(R)$, then the inequality

$$\frac{1}{\left(\sigma_h^2 f_3(R)^2 + \sigma_w^2\right)^2} \geq \frac{1}{\left(\sigma_h^2 r_x^{*2} + \sigma_w^2\right)^2} \exp\left[-\frac{R - 2\sqrt{R}mr_x^* + m^2 r_x^{*2}}{\sigma_h^2 r_x^{*2} + \sigma_w^2}\right]$$

clearly holds. Hence, instead, we can define $f_2(R)$ as

$$f_2(R) = \frac{1}{\left(\sigma_h^2 f_3(R)^2 + \sigma_w^2\right)^2}.$$

To find a lower bound $f_3(R)$ to $r_x^*$, we assume $\frac{\sqrt{R}m\sigma_w^2}{r_x} = u$ in (A.11). Then $r_x^*$ must satisfy the following equation

$$2r_x^2\sigma_h^4 + \sqrt{R}mr_x\sigma_h^2 + m^2\sigma_w^2 + 2\sigma_h^2\sigma_w^2 - R\sigma_h^2 = u. \tag{A.12}$$

Since $r_x^*$ is a non-negative real number, (A.12) can be rewritten as

$$r_x^* = \frac{1}{4\sigma_h^2}\left(\sqrt{8u + Rm^2 + 8R\sigma_h^2 - 8m^2\sigma_w^2 - 16\sigma_h^2\sigma_w^2} - \sqrt{R}m\right) \tag{A.13}$$

by solving (A.12) and keeping the non-negative solution. Note that $u$ is non-negative. A lower bound for $r_x^*$ is

$$r_x^* \geq \frac{1}{4\sigma_h^2}\left(\sqrt{Rm^2 + 8R\sigma_h^2 - 8m^2\sigma_w^2 - 16\sigma_h^2\sigma_w^2} - \sqrt{R}m\right).$$

Define the RHS as $f_3(R)$:

$$f_3(R) = \frac{1}{4\sigma_h^2}\left(\sqrt{Rm^2 + 8R\sigma_h^2 - 8m^2\sigma_w^2 - 16\sigma_h^2\sigma_w^2} - \sqrt{R}m\right).$$

Since $f_3(R) > 0$ over $R \in \left(\frac{m^2\sigma_w^2}{\sigma_h^2} + 2\sigma_w^2, \infty\right)$, we only use $f_3(R)$ to get $f_2(R)$ on this range.

For $R \in \left(\frac{m^2\sigma_w^2}{\sigma_h^2} + 2\sigma_w^2, \infty\right)$, we can obtain an upper bound for $g(R|r_x)$ as

$$\begin{aligned} g(R|r_x) &\leq \left(\sigma_h^2 r_x^2 + \sigma_w^2\right) f_2(R) \\ &= \frac{\sigma_h^2 r_x^2 + \sigma_w^2}{\left(\sigma_h^2 f_3(R)^2 + \sigma_w^2\right)^2}. \end{aligned}$$

For $R \in \left[0, \frac{m^2\sigma_w^2}{\sigma_h^2} + 2\sigma_w^2\right]$, we get an upper bound for $g(R|r_x)$ directly by omitting the exponential function in (A.9) and letting $r_x = 0$: $g(R|r_x) \leq \frac{1}{\sigma_w^2}$. Hence, we get

$$\begin{aligned} p_R(R; D_n) &\leq \begin{cases} \frac{1}{\sigma_w^2} & \text{if } R \in \left[0, \frac{m^2\sigma_w^2}{\sigma_h^2} + 2\sigma_w^2\right], \\ \frac{\sigma_h^2 P_{av} + \sigma_w^2}{\left(\sigma_h^2 f_3(R)^2 + \sigma_w^2\right)^2} & \text{if } R \in \left(\frac{m^2\sigma_w^2}{\sigma_h^2} + 2\sigma_w^2, \infty\right). \end{cases} \\ &\triangleq q_3(R) \end{aligned} \tag{A.14}$$

by simply integrating the two upper bounds with respect to the probability measure.

Since $x \log x$ satisfies the inequalities

$$\begin{aligned} |x \log x| &\leq 4x^{3/4}, \qquad \text{if } x \in [0, 1], \\ |x \log x| &\leq x^2, \qquad \text{if } x \in (1, \infty), \end{aligned} \tag{A.15}$$

we define $q_2(R) = \max\left(4\left(q_3(R)\right)^{3/4}, q_3(R)^2\right)$ which is an integrable function. Thus, (A.6) holds for average power limited Rician fading channels.

Now, we prove (A.6) for peak power limited Rician fading channels:

Note that, for $R \in [0, \infty)$,

$$g(R|r_x) \leq \frac{1}{\sigma_w^2} \exp\left[-\frac{R}{\sigma_h^2 P_{peak} + \sigma_w^2}\right] I_0\left(\frac{2\sqrt{R}mP_{peak}}{\sigma_w^2}\right). \tag{A.16}$$

It is straightforward to show that

$$\begin{aligned} p_R(R;D_n) &\leq \frac{1}{\sigma_w^2}\exp\left[-\frac{R}{\sigma_h^2 P_{peak}+\sigma_w^2}\right] I_0\left(\frac{2\sqrt{R}mP_{peak}}{\sigma_w^2}\right) \\ &\triangleq q_4(R). \end{aligned}$$

Using (A.15), we have

$$|p_R(R;D_n)\ln p_R(R;D_n)| \leq \max\left(4q_4(R)^{\frac{3}{4}}, q_4(R)^2\right).$$

Define $q_2(R) = \max\left(4q_4(R)^{\frac{3}{4}}, q_4(R)^2\right)$ which is an integrable function. Hence, (A.6) holds for peak power limited Rician fading channels.

As a conclusion, the mutual information of a Rician fading channel under an average power constraint or a peak power constraint is weak-star continuous.

### A.2.2 Rayleigh Block Fading Channels

As derived before, with an input distribution $D_n$, the mutual information in a Rayleigh block fading channel is

$$\begin{aligned} I_n(X,Y) &= \int\left[-\ln(1+r^2) - \ln(\pi e)^T + \ln\frac{|\Phi_y|}{2}\right] dD_n(r) \\ &\quad - \int p_R(R;D_n)\ln p_R(R;D_n)dR \end{aligned} \tag{A.17}$$

where

$$p_R(R;D_n) = \int_0^\infty p_{R|r}(R|r)\, dD_n(r) \tag{A.18}$$

and $p_{R|r}(R|r)$ is defined in (2.30).

Let $D_n \xrightarrow{w^*} D$. We must show that the first term in (A.17) satisfies

$$\begin{aligned} &\int_0^\infty \left[-\ln(1+r^2) - \ln(\pi e)^T + \ln\frac{|\Phi_y|}{2}\right] dD_n(r) \\ \rightarrow\ &\int_0^\infty \left[-\ln(1+r^2) - \ln(\pi e)^T + \ln\frac{|\Phi_y|}{2}\right] dD(r) \end{aligned} \tag{A.19}$$

and the second term satisfies

$$\int_0^\infty p_R(R;D_n)\ln p_R(R;D_n)dR \to \int_0^\infty p_R(R;D)\ln p_R(R;D)dR. \tag{A.20}$$

For average power limited Rician channels and peak power limited Rician channels, the proof of (A.19) is similar to that of (A.3), which will be omitted here.

To show (A.20), we need to establish

$$\lim_{n\to\infty}\int_0^\infty p_R(R;D_n)\ln p_R(R;D_n)dR \to \int_0^\infty \lim_{n\to\infty} p_R(R;D_n)\ln p_R(R;D_n)dR \tag{A.21}$$

and

$$\lim_{n\to\infty} p_R(R;D_n)\ln p_R(R;D_n) = p_R(R;D)\ln p_R(R;D). \tag{A.22}$$

In order to prove (A.22), we first show that the integrand $p_{R|r}(R|r)$ in (A.18) is a bounded continuous function of $r$.

Rewrite $p_{R|r}(R|r)$ by substituting all coefficients into (2.30),

$$\begin{aligned} p_{R|r}(R|r) &= -\sum_{i=1}^{T-1}\left[\frac{\left(r^2+1\right)^{T-i-1}R^{i-1}}{\Gamma(i)\,r^{2(T-i)}}\exp(-R)\right] \\ &\quad +\frac{\left(r^2+1\right)^{T-2}}{r^{2(T-1)}}\exp\left(-\frac{R}{r^2+1}\right). \end{aligned} \tag{A.23}$$

It is straightforward to show that $p_{R|r}(R|r)$ is continuous over $r\in(0,\infty)$ for all $R\geq 0$. Now, we show that $p_{R|r}(R|r)$ is also continuous at $r=0$. That is,

$$\lim_{r\to 0} p_{R|r}(R|r) = p_{R|r}(R|r=0).$$

The characteristic function of $p_{R|r}(R|r)$ is

$$\phi(\omega) = \frac{1}{\left[1-j\omega\left(r^2+1\right)\right](1-j\omega)^{T-1}},$$

which is a continuous function of $r$ at 0. For any $R\geq 0$, we establish the following

inequality

$$\begin{aligned} & \left|p_{R|r}\left(R|\varepsilon\right)-p_{R|r}\left(R|0\right)\right| \\ = & \left|\frac{1}{2\pi}\int_{-\infty}^{\infty}\left(\frac{j\omega\varepsilon^{2}}{\left[1-j\omega\left(\varepsilon^{2}+1\right)\right]\left(1-j\omega\right)^{T}}\right)e^{-j\omega R}d\omega\right| \\ \leq & \frac{\varepsilon^{2}}{2\pi}\int_{-\infty}^{\infty}\left|\frac{j\omega}{\left[1-j\omega\left(\varepsilon^{2}+1\right)\right]\left(1-j\omega\right)^{T}}\right|d\omega, \end{aligned} \tag{A.24}$$

where the integral in (A.24) is integrable. For arbitrary $\zeta > 0$, there exists an $\varepsilon > 0$ such that $\left|p_{R|r}\left(R|\varepsilon\right)-p_{R|r}\left(R|0\right)\right| < \zeta$ for all $R \geq 0$. Hence, $p_{R|r}\left(R|r\right)$ is continuous on $[0,\infty)$.

To prove that $p_{R|r}\left(R|r\right)$ is bounded, we notice that, at $r = 0$,

$$p_{R|r}\left(R|r=0\right) = \frac{R^{T-1}\exp\left(-R\right)}{\Gamma\left(T-1\right)} \tag{A.25}$$

is bounded for all $R \geq 0$. From (A.23), we know that, for every $R \geq 0$, as $r \to \infty$, the limit of $p_{R|r}\left(R|r\right)$ is zero. It is then straightforward to prove that $p_{R|r}\left(R|r\right)$ is bounded.

Now, since $p_{R|r}\left(R|r\right)$ is a bounded continuous function of $r$, $p_R(R;D)$ is a bounded continuous function of $D$ for all $R > 0$. Furthermore, because $x\ln x$ is continuous, $p_R(R;D)\ln p_R(R;D)$ is continuous in $D$. That is, (A.22) holds.

In order to prove (A.21), it suffices to find an integrable function $q_5\left(R\right)$ such that

$$\left|p_R(R;D_n)\ln p_R(R;D_n)\right| \leq q_5\left(R\right). \tag{A.26}$$

First, we consider the average power limited Rayleigh block fading channels:

Note that, if $\frac{1}{r^2+1}p_{R|r}\left(R|r\right)$ is viewed as a function of $r$ with a parameter $R$, then there exists a $R_0 > 0$ which satisfies the following statement: for all $R > R_0$, we can find a $r_0 > 0$, the value of $r$ to reach the maximum of the function is in $(r_0,\infty)$. We

denote the value of $r$ to reach the maximum as $r^*$. Thus, for $R > R_0$,

$$p_R(R; D_n) = \int_0^\infty \left(r^2+1\right) \frac{1}{r^2+1} p_{R|r}\left(R|r\right) dD_n\left(r\right) \tag{A.27}$$

$$\leq \left( \underbrace{-\sum_{i=1}^{T-1} \left[ \frac{\left(r^{*2}+1\right)^{T-i-2} R^{i-1}}{\Gamma\left(i\right) r^{*2(T-i)}} \exp\left(-R\right) \right]}_{(e)} + \frac{\left(r^{*2}+1\right)^{T-3}}{r^{*2(T-1)}} \exp\left(-\frac{R}{r^{*2}+1}\right) \right)$$

$$\times \int_0^\infty \left(r^2+1\right) dD_n\left(r\right) \tag{A.28}$$

$$\leq \frac{\left(r_0^2+1\right)^{T-1}}{r_0^{2(T-1)}} \frac{\exp\left(-\frac{R}{r^{*2}+1}\right)}{\left(r^{*2}+1\right)^2} \left(P_{av}+1\right) \tag{A.29}$$

$$\leq \frac{\left(r_0^2+1\right)^{T-1}}{r_0^{2(T-1)}} \frac{4e^{-2}}{R^2} \left(P_{av}+1\right). \tag{A.30}$$

(A.27) holds by definition. (A.28) is from point-wise maximization of the integrand. (A.29) is obtained in two steps: First, we note that $(e)$ in (A.28) is negative and we can omit $(e)$ to get an upper bound; second, $\left(\frac{x+1}{x}\right)^{T-1}$ is a monotonically decreasing function of $x$, we substitute $r_0^2$ for $r^{*2}$ where $r_0^2 < r^{*2}$. (A.30) is obtained by maximizing $\frac{1}{\left(r^{*2}+1\right)^2} \exp\left(-\frac{R}{r^{*2}+1}\right)$ over $[0, \infty)$.

Define a function $q_6\left(R\right)$ as

$$q_6\left(R\right) = \frac{\left(r_0^2+1\right)^{T-1}}{r_0^{2(T-1)}} \frac{4e^{-2}}{R^2} \left(P_{av}+1\right).$$

Based on (A.15), we define the upper bound of $|p_R(R; D_n) \ln p_R(R; D_n)|$ as

$$q_5\left(R\right) = \begin{cases} \max\left(p_R(R; D_n) \ln p_R(R; D_n)\right) & \text{if } R \in [0, R_0], \\ \max\left(4q_6\left(R\right)^{\frac{3}{4}}, q_6\left(R\right)^2\right) & \text{if } R \in (R_0, \infty). \end{cases}$$

Since $p_R(R; D)$ is a bounded continuous function for all $R > 0$, $\max\left(p_R(R; D_n) \ln p_R(R; D_n)\right)$ is finite. $q_5\left(R\right)$ is an integrable function. As a result, (A.21) holds for average power limited Rayleigh block fading channels.

The proof of (A.21) for peak power limited Rayleigh block fading channels is similar. The only difference is that $P_{av}$ in (A.29) is changed to $P_{peak}$. Define a

function $q_7(R)$ as

$$q_7(R) = \frac{\left(r_0^2+1\right)^{T-1}}{r_0^{2(T-1)}} \frac{4e^{-2}}{R^2} \left(P_{peak}+1\right).$$

Then the upper bound $q_5(R)$ is given by

$$q_5(R) = \begin{cases} \max\left(p_R(R;D_n)\ln p_R(R;D_n)\right) & \text{if } R \in [0,R_0], \\ \max\left(4q_7(R)^{\frac{3}{4}}, q_7(R)^2\right) & \text{if } R \in (R_0,\infty). \end{cases}$$

Since $q_5(R)$ is an integrable function, (A.21) holds for peak power limited Rayleigh block fading channels.

As a conclusion, the mutual information of a Rayleigh block fading channel under an average power constraint or a peak power constraint is weak-star continuous.

## A.3 Strict Concavity of Mutual Information

In this section, we prove that the mutual information of the Rayleigh block fading channel is a strictly concave function over $\Omega$. Recall that the mutual information

$$I(X;Y) = h(Y) - h(Y|X)$$

where $h(Y)$ is a strictly concave function of $p(Y;D)$ and the second term $h(Y|X)$ is linear in $D$. If $p(Y;D)$ is an injective linear function of $D$ , then $h(Y)$ is a strictly concave function of $D$ which implies $I(X;Y)$ is a strictly concave function of $D$. In the following discussion, we will prove the operator $D \to p(Y;D)$ is injective. That is, given $p_R(R;D) = p_R(R;D')$, we prove $D = D'$.

Let $p_R(R;D) = p_R(R;D')$. Then the characteristic functions of these two PDFs are equivalent:

$$\int \exp(-Rs)\, p_R(R;D)dR = \int \exp(-Rs)\, p_R(R;D')dR.$$

Change the order of integrals, and note the characteristic function of $p_{R|r}(R|r)$ is

given by (2.29). We obtain

$$\begin{aligned} & \int \frac{1}{\left[1-j\omega\left(r^2+1\right)\right]\left(1-j\omega\right)^{T-1}} dD\left(r\right) \\ = & \int \frac{1}{\left[1-j\omega\left(r^2+1\right)\right]\left(1-j\omega\right)^{T-1}} dD'\left(r\right). \end{aligned} \tag{A.31}$$

Cancel out $\frac{1}{(1-j\omega)^{T-1}}$ on both sides of (A.31):

$$\int \frac{1}{1-j\omega\left(r^2+1\right)} dD\left(r\right) = \int \frac{1}{1-j\omega\left(r^2+1\right)} dD'\left(r\right). \tag{A.32}$$

Because PDFs and characteristic functions are invertible, we inverse transform both sides of (A.32) and have

$$\int \frac{e^{-\frac{R}{r^2+1}}}{r^2+1} dD\left(r\right) = \int \frac{e^{-\frac{R}{r^2+1}}}{r^2+1} dD'\left(r\right). \tag{A.33}$$

The equation (A.33) means that, for a channel whose transition probability is $p_{R|r}(R|r) = \frac{1}{r^2+1} e^{-\frac{R}{r^2+1}}$, the output $R$'s distribution under the input distribution $D\left(r\right)$ and $D'\left(r\right)$ is the same.

Let $w = r^2 + 1$ be a function of $r$. Then $R$ can be described as an output of a multiplicative channel $R = wz$, where $w$ is the channel input determined by $D\left(r\right)$ and $z$ is independent of $r$ and has a probability density function $p\left(z\right) = e^{-z}$.

Let $w$ and $w'$ be the inputs corresponding to $D\left(r\right)$ and $D'\left(r\right)$, respectively. The outputs $R = wz$ and $R' = w'z$ are equal in distribution according to (A.33). Therefore, the functions $\ln\left(R\right) = \ln\left(w\right) + \ln\left(z\right)$ and $\ln\left(R'\right) = \ln\left(w'\right) + \ln\left(z\right)$ are equal in distribution. Equivalently, $\ln\left(R\right)$ and $\ln\left(R'\right)$ have equal characteristic functions

$$\phi_{\ln(w)}\phi_{\ln(z)} = \phi_{\ln(w')}\phi_{\ln(z)}.$$

This implies that $\phi_{\ln(w)} = \phi_{\ln(w')}$ everywhere except at isolated zero points of $\phi_{\ln(z)}$. Given that characteristic functions are continuous, $\phi_{\ln(w)} = \phi_{\ln(w')}$ everywhere. Therefore, $w$ and $w'$ are equal in distribution and hence so are $r$ and $r'$, which proves $D\left(r\right) = D'\left(r\right)$.

We have proven that, $p_R(R;D) = p_R(R;D') \Rightarrow D = D'$. Hence, $I(X;Y)$ is a strictly concave function of $D$ in Rayleigh block fading channels.

## A.4 Existence and Uniqueness of the Optimal Input Distribution

$\Omega_a$, $\Omega_p$, and $\Omega_{ap}$ are convex and weak-star compact. We have proven that, under an average power constraint or a peak power constraint, the mutual informations of Rician fading channels and Rayleigh block fading channels are weak-star continuous. Furthermore, we have proven that, the mutual informations of Rayleigh block fading channels are strictly concave functions with respect to $D$.

According to Theorem A.1, the optimal distributions exist in Rician fading channels and Rayleigh block fading channels under the three power constraints. In Rayleigh block fading channels, the optimal distribution is unique.

# Appendix B

# Kuhn-Tucker Condition for Peak Power Limited Channels

First, we give the definition of *weakly differentiable,* which is used in the proof of the theorem.

Definition: Let $p^*$ be a fixed element in a convex set $\mathcal{P}$, and $\lambda$ a real number in $[0,1]$. A functional $f$ on $\mathcal{P}$ is said to be weakly differentiable at $p^*$ if the limit exists

$$f'_{p^*}(p) = \lim_{\lambda \to 0} \frac{f\left((1-\lambda)p^* + \lambda p\right) - f(p^*)}{\lambda}, \quad \forall p \in \mathcal{P}.$$

If is $f$ weakly differentiable in $\mathcal{P}$ at $p$ for all $p \in \mathcal{P}$, $f$ is said to be weakly differentiable in $\mathcal{P}$.

For a weakly differentiable funtional which achieves its maximum in a convex set $\mathcal{P}$, we have following theorem.

**Theorem B.1** *(Smith. [10]) Assume $f$ is a weakly differentiable funtional and achieves its maximum in a convex set $\mathcal{P}$.*

*1) If $f$ is concave and $f'_{p^*}(p) \leq 0$ for all $p \in \mathcal{P}$, then $f$ achieves its maximum at $p^*$.*

*2) If $f$ achieves its maximum at $p^*$, then $f'_{p^*}(p) \leq 0$ for all $p \in \mathcal{P}$.*

We have proven that the set, $\Omega_p$, of possible distribution functions with the

bounded support $|x|^2 \leq P_{peak}$ is a convex set. Then the mutual information $I$ is a functional defined on $\Omega_p$. In [15], $I$ is proved to be weakly differentiable on the set of possible distribution functions with no peak power constraint. Using the same argument, we can establish that $I$ is weakly differentiable in $\Omega_p$ and the weak derivative is

$$I'_{p^*}(p) = \int i(x; p^*)\, p(x)dx - I(p^*)$$

where

$$i(x; p^*) = \int p(y \mid x) \ln \left[\frac{p(y \mid x)}{p(y; p^*)}\right] dy.$$

Hence, according to Theorem A.1, a necessary and sufficient condition for $p^*$ to achieve the maximum of $I$ is

$$I'_{p^*}(p) \leq 0, \quad \forall p \in \Omega_p$$

that is

$$\int i(x; p^*)\, p(x)dx \leq C. \tag{B.1}$$

Now let us prove that (B.1) is true if and only if

$$i(x; p^*) \leq C, \quad \forall |x|^2 \leq P_{peak} \tag{B.2}$$

and

$$i(x; p^*) = C, \quad \forall x \in S_0 \tag{B.3}$$

where $S_0$ is the support of $p^*$.

The proof from (B.2) to (B.1) is straightforward. For the converse, assume (B.2) is false. Then there must exist an $x_0$ such that

$$i(x_0; p^*) > C.$$

If $p(x)$ is a delta function at $x_0$, then

$$\int i(x; p^*)\, \delta(x - x_0)dx = i(x_0; p^*) > C$$

which contradicts (B.1).

Assume (B.2) is true and (B.3) is false, then there exists $x_0 \in S_0$ such that

$$i(x_0; p^*) < C.$$

The inequality is strictly satisfied over a non-zero measure neighborhood $S'$ of $x_0$, since $i(x; p^*)$ is continuous in $x$.

$$\begin{aligned} C &= \int_{S_0} i(x; p^*)\, p^*(x) dx \\ &= \int_{S_0 - S'} i(x; p^*)\, p^*(x) dx + \int_{S'} i(x; p^*)\, p^*(x) dx \\ &< C \int_{S_0 - S'} p^*(x) dx + C \int_{S'} p^*(x) dx \\ &< C \end{aligned}$$

which is a contradiction.

Hence, (B.2) and (B.3) are the necessary and sufficient conditions for an optimal distribution.

# Appendix C

# Boundedness of the difference between $V_{k,l}$ and $A_l$

In this Appendix, we prove the inequality $\lim_{l\to\infty} E\left[|V_{k,l} - A_l|\right] < \infty$. Recall that $\{A_l\}$ are the thresholds

$$A_l = (l+1)\,\overline{x}^2 + l.$$

The inequality means the difference between the random variable $V_{k,l}$ and the threshold $A_l$ is bounded.

We first prove a lemma.

**Lemma C.1** *Let the kth tone be a non-zero tone. For any integer $l > 1$, there exists a function $\beta\left(\overline{x}^2\right) < 1$ such that the conditional probability $\Pr\left(V_{k,l} < A_l \mid V_{k,l-1} < A_{l-1}\right)$ is bounded by the function:*

$$\Pr\left(V_{k,l} < A_l \mid V_{k,l-1} < A_{l-1}\right) \leq \beta\left(\overline{x}^2\right).$$

**Proof:** For notational convenience, we use $\{I_l\}$ to denote the intervals

$$I_l \triangleq [A_{l-1}, A_l).$$

Assume $V_{k,l-1} \in I_{l-1-n}$. Since $\{V_{k,l}\}$ is monotonically increasing with $l$, we have the

inequalities

$$V_{k,l-n-1} \leq \ldots \leq V_{k,l-2} \leq V_{k,l-1} < A_{l-1-n}.$$

According to the scheme, the inputs must satisfy $x_{k,l-n}^2 \geq \overline{x}^2$, $x_{k,l-n+1}^2 \geq 2\overline{x}^2$, ... , $x_{k,l-1}^2 \geq n\overline{x}^2$, and $x_{k,l}^2 \geq (n+1)\,\overline{x}^2$. Since the output of the channel is exponentially distributed conditional on the input, the probability $\Pr\left(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}\right)$ is bounded as

$$\begin{aligned} & \Pr\left(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}\right) \\ \leq\ & 1 - \exp\left(-\frac{A_l - V_{k,l-1}}{(n+1)\,\overline{x}^2 + 1}\right) \\ \leq\ & 1 - \exp\left(-\frac{(n+2)\,\overline{x}^2 + n + 2}{(n+1)\,\overline{x}^2 + 1}\right). \end{aligned} \tag{C.1}$$

The derivative of the bound (C.1) with respect to $n$ is given by

$$\begin{aligned} & \frac{d\left[1 - \exp\left(-\frac{(n+2)\overline{x}^2+n+2}{(n+1)\overline{x}^2+1}\right)\right]}{dn} \\ =\ & \frac{(1-\overline{x}^4)\exp\left(-\frac{(n+2)\overline{x}^2+n+2}{\overline{x}^2+n\overline{x}^2+1}\right)}{\left(\overline{x}^2 + n\overline{x}^2 + 1\right)^2}. \end{aligned} \tag{C.2}$$

When $\overline{x}^2 > 1$, the derivative (C.2) is negative for any $n$, the maximum value of the bound (C.1) occurs at $n = 0$, and the probability $\Pr\left(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}\right)$ is bounded as

$$\Pr\left(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}\right) \leq 1 - \exp\left(-2\right);$$

When $\overline{x}^2 \leq 1$, the derivative (C.2) is positive for any $n$, the maximum value of the bound (C.1) occurs at $n = \infty$, and the probability $\Pr\left(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}\right)$ satisfies

$$\Pr\left(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}\right) \leq 1 - \exp\left(-\frac{\overline{x}^2 + 1}{\overline{x}^2}\right).$$

Define $\beta(\overline{x}^2)$ as

$$\beta\left(\overline{x}^2\right) = \begin{cases} 1 - \exp(-2), & \overline{x}^2 > 1; \\ 1 - \exp\left(-\frac{\overline{x}^2+1}{\overline{x}^2}\right), & 0 < \overline{x}^2 \leq 1. \end{cases}$$

Then $\Pr(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}) \leq \beta(\overline{x}^2)$ for $\overline{x}^2 > 0$ and the bound does not depend on the value of $n$. Hence,

$$\begin{aligned} & \Pr(V_{k,l} < A_l \mid V_{k,l-1} < A_{l-1}) \\ = & \sum_{n=0}^{l-1} \Pr(V_{k,l} < A_l \mid V_{k,l-1} \in I_{l-1-n}) \Pr(V_{k,l-1} \in I_{l-1-n}) \\ \leq & \beta\left(\overline{x}^2\right). \end{aligned}$$

■

**Proposition C.1** *Let the kth tone be a non-zero tone. Then the difference between the random variable $V_{k,l}$ and the threshold $A_l$ is bounded*

$$\lim_{l\to\infty} E[|V_{k,l} - A_l|] < \infty. \tag{C.3}$$

**Proof:** Using the Total Expectation Theorem, we rewrite $E[|V_{k,l} - A_l|]$ as

$$\begin{aligned} & E[|V_{k,l} - A_l|] = \\ & E[|V_{k,l} - A_l| \mid V_{k,l} < A_{l-1}] \Pr(V_{k,l} < A_{l-1}) \\ & + E[|V_{k,l} - A_l| \mid V_{k,l} \in I_l] \Pr(V_{k,l} \in I_l) \\ & + E[|V_{k,l} - A_l| \mid V_{k,l} \geq A_l] \Pr(V_{k,l} \geq A_l). \end{aligned}$$

The inequality (C.3) is true if the following inequalities hold:

$$\lim_{l\to\infty} E[|V_{k,l} - A_l| \mid V_{k,l} < A_{l-1}] \Pr(V_{k,l} < A_{l-1}) < \infty, \tag{C.4}$$

$$\lim_{l\to\infty} E[|V_{k,l} - A_l| \mid V_{k,l} \in I_l] \Pr(V_{k,l} \in I_l) < \infty, \tag{C.5}$$

and

$$\lim_{l\to\infty} E\left[|V_{k,l} - A_l| \mid V_{k,l} \geq A_l\right] \Pr\left(V_{k,l} \geq A_l\right) < \infty. \tag{C.6}$$

We first prove (C.4).

By definition,

$$\begin{aligned} & E\left[|V_{k,l} - A_l| \mid V_{k,l} < A_{l-1}\right] \\ = & \int_{V_{k,l} < l\overline{x}^2 + l - 1} \left(V_{k,l} - A_l\right) p\left(V_{k,l} \mid V_{k,l} < A_{l-1}\right) dV_{k,l} \\ \leq & \sum_{n=0}^{l-1} (n+2)\left(\overline{x}^2 + 1\right) \Pr\left(V_{k,l} \in I_{l-1-n} \mid V_{k,l} < A_{l-1}\right), \end{aligned} \tag{C.7}$$

where (C.7) is obtained by upper bounding $|V_{k,l} - A_l|$ on subsets $V_{k,l} \in I_{l-1-n}$. In (C.7), the conditional probability $\Pr\left(V_{k,l} \in I_{l-1-n} \mid V_{k,l} < A_{l-1}\right)$ is equivalent to

$$\begin{aligned} & \Pr\left(V_{k,l} \in I_{l-1-n} \mid V_{k,l} < A_{l-1}\right) \\ = & \frac{\Pr\left(V_{k,l} \in I_{l-1-n}, V_{k,l} < A_{l-1}\right)}{\Pr\left(V_{k,l} < A_{l-1}\right)} \\ = & \frac{\Pr\left(V_{k,l} \in I_{l-1-n}\right)}{\Pr\left(V_{k,l} < A_{l-1}\right)}, \end{aligned} \tag{C.8}$$

where (C.8) is obtained because the event $V_{k,l} < A_{l-1}$ is implied by $V_{k,l} \in I_{l-1-n}$. Further, the probability $\Pr\left(V_{k,l} \in I_{l-1-n}\right)$ in (C.8) is bounded as

$$\begin{aligned} & \Pr\left(V_{k,l} \in I_{l-1-n}\right) \\ \leq & \Pr\left(V_{k,l} < A_{l-1-n}\right) \\ \leq & \Pr\left(V_{k,l-n-1} \leq A_{l-1-n}, \ldots, V_{k,l-1} < A_{l-1-n}\right) && \text{(C.9)} \\ \leq & \Pr\left(V_{k,l-n-1} \leq A_{l-1-n}, \ldots, V_{k,l-1} < A_{l-1}\right) && \text{(C.10)} \\ \leq & \sum_{i=1}^{n} \Pr\left(V_{k,l-i} < A_{l-i} \mid V_{k,l-1-i} \leq A_{l-1-i}\right) && \text{(C.11)} \\ \leq & \beta\left(\overline{x}^2\right)^n, && \text{(C.12)} \end{aligned}$$

where (C.9) and (C.10) are derived by using $A \subseteq B \Rightarrow \Pr(A) \leq \Pr(B)$; (C.11) is based on the Markov property of $V_{k,l}$; (C.12) is attained by using Lemma C.1.

Hence, by substituting (C.8) and (C.12) into (C.7), we have

$$\begin{aligned} & \lim_{l\to\infty} E\left[|V_{k,l} - A_l| \mid V_{k,l} < A_{l-1}\right] \Pr\left(V_{k,l} < A_{l-1}\right) \\ \leq & \lim_{l\to\infty} \sum_{n=0}^{l-1} (n+2)\left(\overline{x}^2+1\right)\beta\left(\overline{x}^2\right)^n \\ < & \infty, \end{aligned}$$

which proves (C.4).

The inequality (C.5) is obviously true based on $V_{k,l} \in I_l$.

Finally, we prove (C.6).

Again, we use the Total Expectation Theorem:

$$\begin{aligned} & E\left[|V_{k,l} - A_l| \mid V_{k,l} \geq A_l\right] \\ = & \sum_{n=-1}^{l} E\left(|V_{k,l} - A_l| \mid x_{k,l}^2 = (n+1)\,\overline{x}^2, V_{k,l} \geq A_l\right) \\ & \times \Pr\left(x_{k,l}^2 = (n+1)\,\overline{x}^2 \mid V_{k,l} \geq A_l\right), \end{aligned} \tag{C.13}$$

where we consider all possibilities of the input $x_{k,l}^2$. In (C.13), the conditional expectation $E\left(|V_{k,l} - A_l| \mid x_{k,l}^2 = (n+1)\,\overline{x}^2, V_{k,l} \geq A_l\right)$ is equivalent to

$$\begin{aligned} & E\left(|V_{k,l} - A_l| \mid x_{k,l}^2 = (n+1)\,\overline{x}^2, V_{k,l} \geq A_l\right) \\ = & E\left(V_{k,l} \mid x_{k,l}^2 = (n+1)\,\overline{x}^2, V_{k,l} \geq A_l\right) - A_l \\ = & (n+1)\,\overline{x}^2 + 1, \end{aligned} \tag{C.14}$$

and the conditional probability $\Pr\left(x_{k,l}^2 = (n+1)\,\overline{x}^2 \mid V_{k,l} \geq A_l\right)$ is given by

$$\begin{aligned} & \Pr\left(x_{k,l}^2 = (n+1)\,\overline{x}^2 \mid V_{k,l} \geq A_l\right) \\ = & \frac{\Pr\left(x_{k,l}^2 = (n+1)\,\overline{x}^2, V_{k,l} \geq A_l\right)}{\Pr\left(V_{k,l} \geq A_l\right)}. \end{aligned} \tag{C.15}$$

The probability $\Pr\left(x_{k,l}^2 = (n+1)\,\overline{x}^2, V_{k,l} \geq A_l\right)$ in (C.15) is bounded as

$$\begin{aligned}
& \Pr\left(x_{k,l}^2 = (n+1)\,\overline{x}^2, V_{k,l} \geq A_l\right) \\
\leq\ & \Pr\left(V_{k,l-n-1} \leq A_{l-1-n}, ..., V_{k,l-1} < A_{l-1}\right) && \text{(C.16)} \\
\leq\ & \sum_{i=1}^{n} \Pr\left(V_{k,l-i} < A_{l-i} \mid V_{k,l-i-1} \leq A_{l-1-i}\right) && \text{(C.17)} \\
\leq\ & \beta\left(\overline{x}^2\right)^n, && \text{(C.18)}
\end{aligned}$$

where (C.16) is derived by using $A \subseteq B \Rightarrow \Pr(A) \leq \Pr(B)$; (C.17) is based on the Markov property of $V_{k,l}$; (C.18) is obtained by using Lemma C.1.

By substituting (C.18) and (C.15) into (C.13), we obtain

$$\begin{aligned}
& \lim_{l\to\infty} E\left[|V_{k,l} - A_l| \mid V_{k,l} \geq A_l\right] \Pr\left(V_{k,l} \geq A_l\right) \\
\leq\ & \lim_{l\to\infty} \sum_{n=0}^{l} \left[(n+1)\,\overline{x}^2 + 1\right] \beta\left(\overline{x}^2\right)^n \\
<\ & \infty,
\end{aligned}$$

which proves (C.6).

Hence, (C.4), (C.5), and (C.6) are true and hence (C.3) holds. ■

# Appendix D

# Experimental Results

In this thesis, we have discussed theoretic principles underlying the wideband capacity-achieving signaling schemes and proposed a family of Multi-tone FSK schemes designed to achieve the wideband capacity limit. By using theoretical analysis and numerical results, we have shown that the achievable data rate of Multi-tone FSK can be very close to the capacity limit.

However, in general, a theoretically sound idea may not be implementable owing to problems of over-simplified analysis, rough approximations, inappropriate assumptions, or unrealistic requirements on the system implementation.

In this appendix, we show some experimental results to endorse the Multi-tone FSK scheme. By using the ultra-wideband platform provided by the Ultra-wideband project in Microsystems Technology Laboratories at MIT [45], we have realized the Multi-tone FSK transmissions in an indoor environment. Since this thesis is not an experimental thesis, exploring performance of Multi-tone FSK to a full extend is out of the scope of this thesis. The experimental results shown should be deemed as demonstrations supporting our theoretical proposal.

## D.1 Ultra-Wideband Platform

The ultra-wideband platform is divided into three distinct sections: the transmitter, the receiver, and the baseband processing. The transmitter and receiver communicate

through air using a wideband horn antenna.

The transmitted ultra-wideband signal is synthesized by using a programmable 4 GS/s arbitrary waveform generator (AWG) and a vector signal generator (VSG). The arbitrary waveform generator generates a baseband modulated signal, and the vector signal generator up-converts the signal to a center frequency of 5.355 GHz. The analog bandwidth is 500 MHz, therefore the synthesized Multi-tone FSK signals can at most occupy a bandwidth of 500 MHz. Figure D-1 is a simplified diagram of the transmitter.

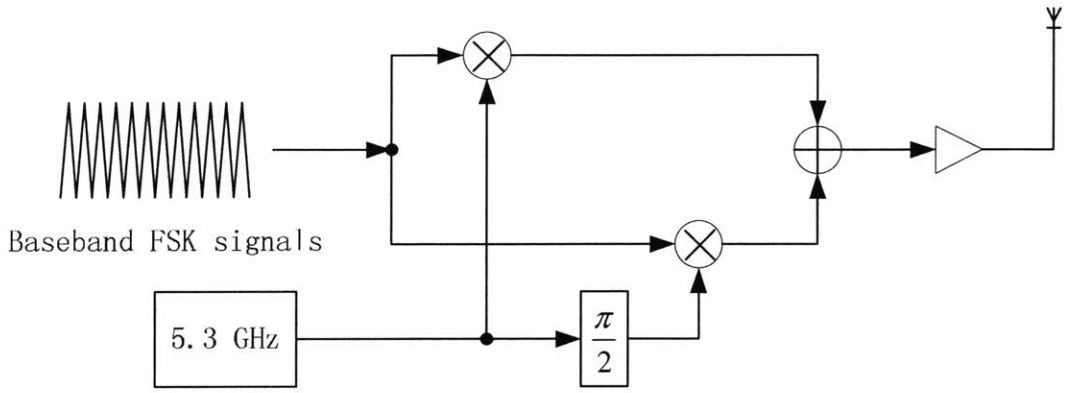


Figure D-1: The transmitter used to test the Multi-tone FSK scheme

At the receiver, the received signal from the RF front-end is amplified by two cascaded LNAs. Then the signal is split and applied to two identical passive mixers which perform I/Q direct conversion. The baseband signal is filtered by low-pass filters and amplified with an adjustable gain.

The baseband I and Q signals from the front-end are sampled by a dual-channel, 8-bit, 1 GS/s Analog to Digital Conversion (ADC) board that interfaces to a computer directly through the PCI bus. The received baseband samples are buffered and captured to a waveform file, later to be processed by baseband processing programs. Figure D-2 is a diagram of the receiver. The gray block in the figure is the baseband processing section implemented by software in a computer.

We use Matlab for the baseband processing. Once the sampled waveform files are read into Matlab, virtually any baseband algorithm not requiring real-time control of the system may be tested.

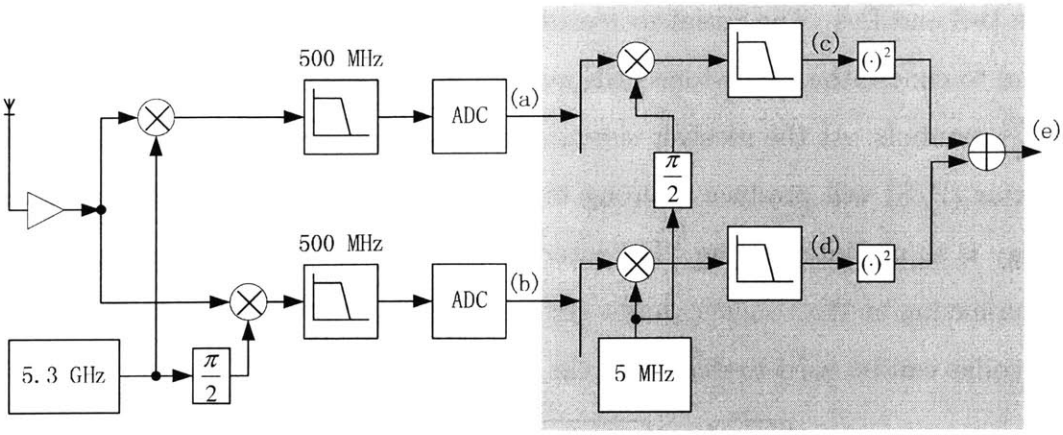


Figure D-2: The receiver used to test the Multi-tone FSK scheme

The baseband processing part for the Multi-tone FSK scheme is a bank of matched filters as shown in the gray box in Figure D-2. The received baseband waveform is correlated with the central frequency of each tone, and integrated over a symbol time. Since all tones are orthogonal over a symbol time, the outputs of all matched filters will be close to zero except the $Q$ nonzero tones. The outputs are sampled and decoded using a threshold detection rule.

Since the Multi-tone FSK signals are short sinusoid waveform pulses with a low duty cycle, we don't do phase synchronization at the matched filters. The scheme is noncoherent.

## D.2 Coding and Decoding Scheme

The input of the arbitrary waveform generator at the transmitter is from the channel coder: the channel coder generates a bit string which selects one of the $\binom{M}{Q}$ symbols and the arbitrary waveform generator produces the baseband waveform accordingly.

One important aspect in designing the channel coder for the Multi-tone FSK scheme is the code scheme's ability to combat burst errors. Since $\binom{M}{Q}$ is usually a large number ($M$ is very large for wideband applications), one symbol error is likely to cause a large number of bit errors clustered together.

We give an example of the coding and decoding scheme for Multi-tone FSK in

Figure D-3 and D-4. The Serial to Parallel converter (S/P) in the diagram generates a signal to choose the Multi-tone FSK symbol by mapping the bit sequence to one of the $\binom{M}{Q}$ symbols. At the receiver side, if a symbol error occurs, the Parallel to Serial converter (P/S) will produce a wrong bit string which may cause burst errors. Our strategy is to use interleaving. The interleaver scrambles coded bits such that, after de-interleaving at the receiver, burst errors will be scattered. Then trellis codes and CRC codes can be used to deal with the non-burst errors.

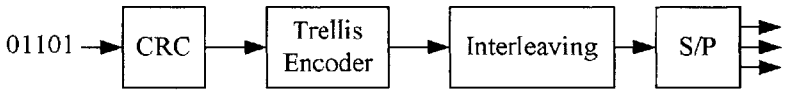


Figure D-3: An encoder for the Multi-tone FSK scheme

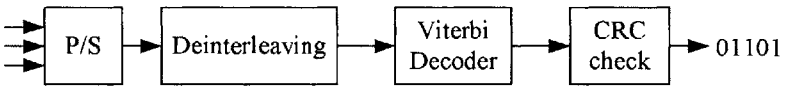


Figure D-4: The corresponding decoder for the Multi-tone FSK scheme

## D.3 Experimental Results

In the following section, we show some of the experiment results attained in our tests. These results illustrate some typical settings of the Multi-tone FSK scheme. From the observations, we find that different tones experience different fading and the fading coefficients dramatically change in time. It justifies our frequency selective time varying fading channel assumption. Moreover, we caught some of the Multi-tone FSK signals corrupted by unknown strong interference. The demodulation result shows that the Multi-tone FSK waveform survived the severe scenario and was decoded properly.

### D.3.1 Multi-tone FSK Waveform

In Figure D-5, we plot baseband waveforms which were captured at the output of the ADC (Point (a) and (b) in Figure D-2) when the receiver is demodulating Two-tone

FSK symbols. The top waveform is from the I-Channel and the bottom waveform is from the Q-channel.

In the Two-tone FSK scheme, the symbol time $T_s$ is 1000 ns and the duty cycle $\theta$ is 1/6. The baseband bandwidth is 500 MHz, and we use frequencies 5 MHz apart (i.e., 5 MHz, 10 MHz, ..., 500 MHz). The value of $M$ is 500 MHz/5 MHz = 100 in the scheme and because it is a Two-tone FSK scheme, the value of $Q$ is 2. The total number of symbols is $\binom{Q}{M} = 4950$. The Two-tone FSK scheme uses a repetition code with length $N = 10$.

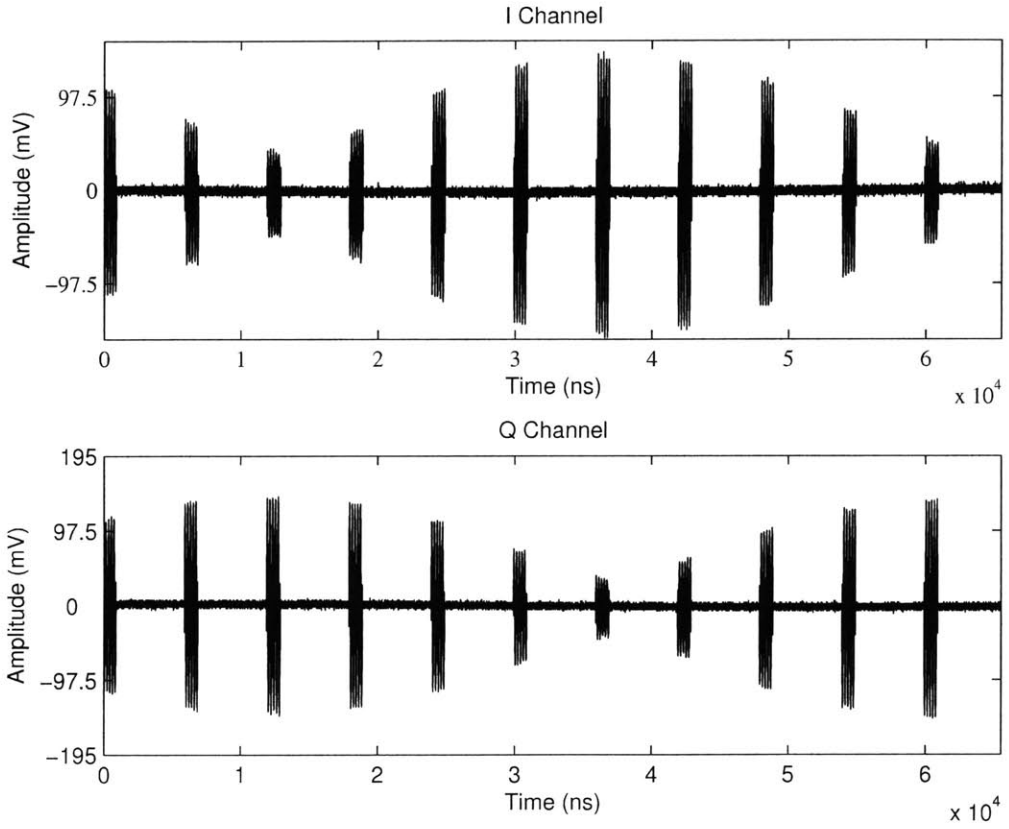


Figure D-5: Baseband waveform of Two-tone FSK symbols

The Two-tone FSK waveform is showed in Figure D-5. A Two-tone FSK symbol is a short pulse. Each pulse includes sinusoid waveforms with different frequencies. The pulses in Figure D-5 have two frequencies: 5 MHz and 100 MHz. One symbol is repeated over time to get time diversity. We magnify one pulse in Figure D-6 so that we can see the 5 MHz component (the large wave) and the 100 MHz component (the

short zigzags on the wave).

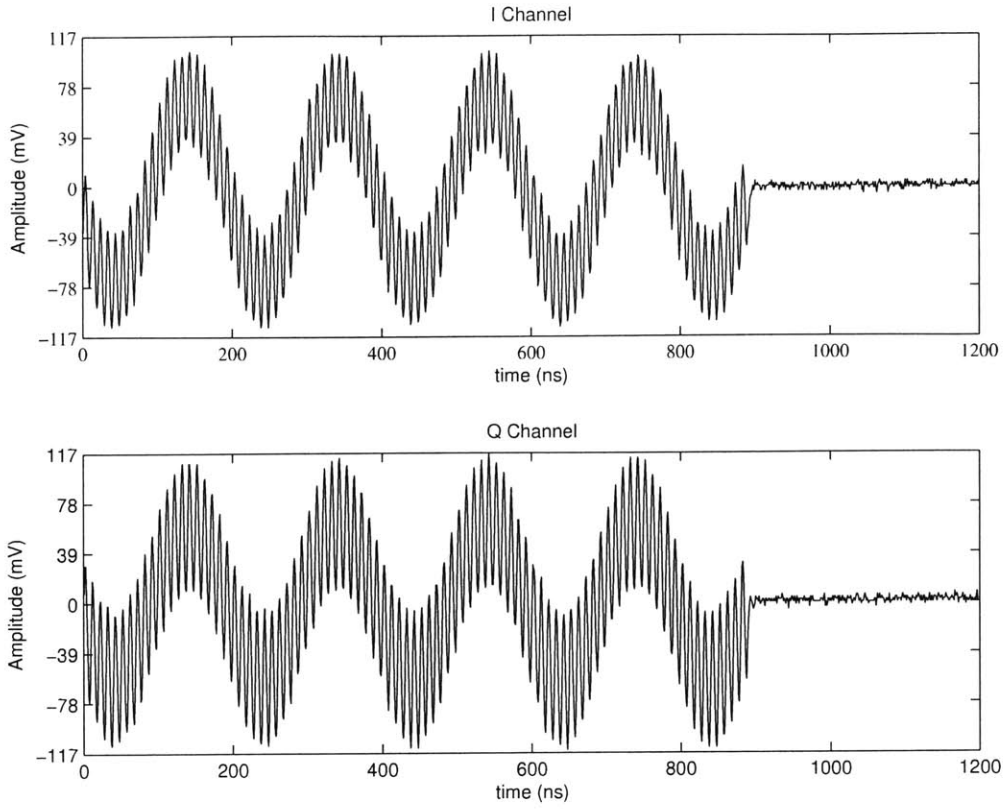


Figure D-6: A Two-tone FSK symbol

Multi-tone FSK symbols are easy to generate: pick sinusoid waveforms corresponding to the central frequencies and add them together. There are many off-the-shelf components available on the market which make the implementation of the Multi-tone FSK scheme very easy.

### D.3.2 Frequency-Selective Time-Varying Fading

To demodulate symbols, the I and Q signals are correlated with central frequencies and filtered. After that, the I and Q outputs are squared and power-combined. The combined outputs (as shown in Figure D-7) are sampled with a period of symbol time $T_s$. Then the sampled values are cumulated over $N$ repetitive symbols and the averaged value is compared with a threshold to decode the symbol.

The sampled outputs of the combiners are the received power of faded tones. Their magnitudes are determined by the channel fading coefficients. In Figure D-7, the sampled outputs (corresponding to the peaks) indicate that the 5 MHz tone and the 100 MHz tone experienced time varying fading, because the output magnitudes change very fast with time. Also, when we compare the values of these two tones at the same time, we find the fading coefficients are different for these two tones, which is an indication of frequency selective fading.

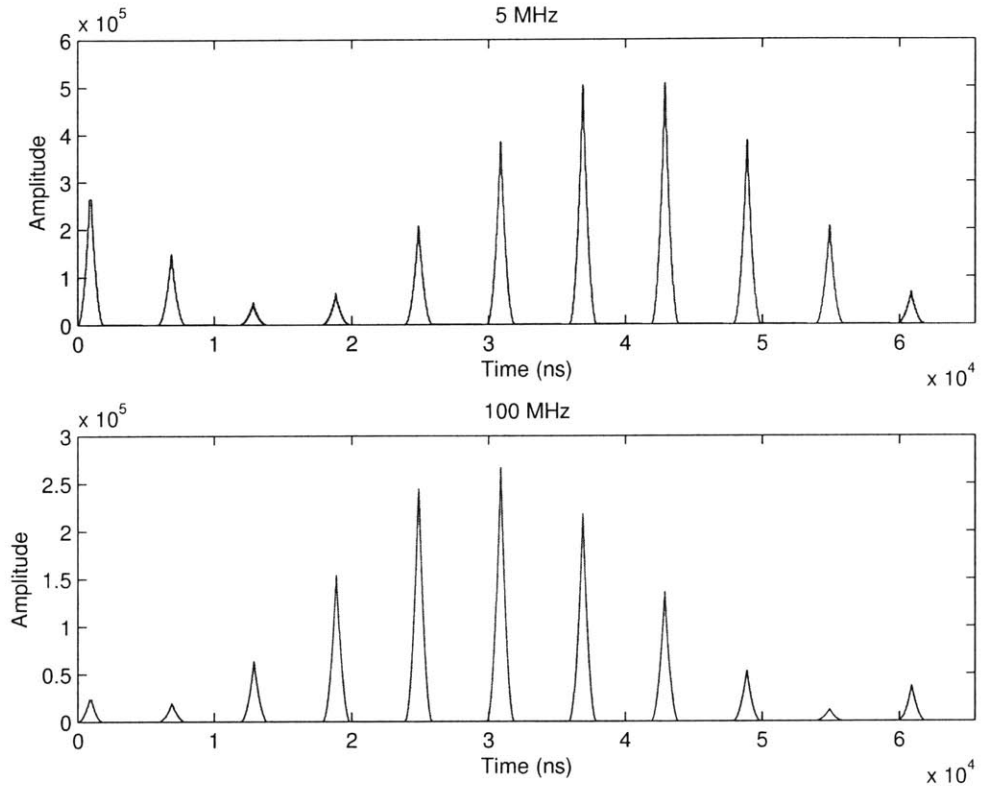


Figure D-7: Frequency selective time varying fading

Hence, it appears that our frequency selective time varying fading channel assumption in the previous chapters is reasonable.

### D.3.3 Strong Interference

Besides Two-tone FSK, we also tested Single-tone FSK, Three-tone FSK, and Four-tone FSK on the platform. During the tests, we occasionally observed very strong

unknown interference alongside the Multi-tone FSK waveforms (as shown in Figure D-8). Not surprisingly, the interfered waveforms can still be demodulated properly. This is because the Multi-tone FSK signals are peaky in time and frequency and not easily to be submerged by interference. Even very strong interference is encountered and the receiver is saturated so that the Multi-tone FSK waveforms are partially cut off, we can still demodulate the tones correctly. The reason inhabits in the characteristic of sinusoid waveforms: when a sinusoid waveform is cut off by saturation, the remainder will still exhibit a very strong frequency component with the original frequency.

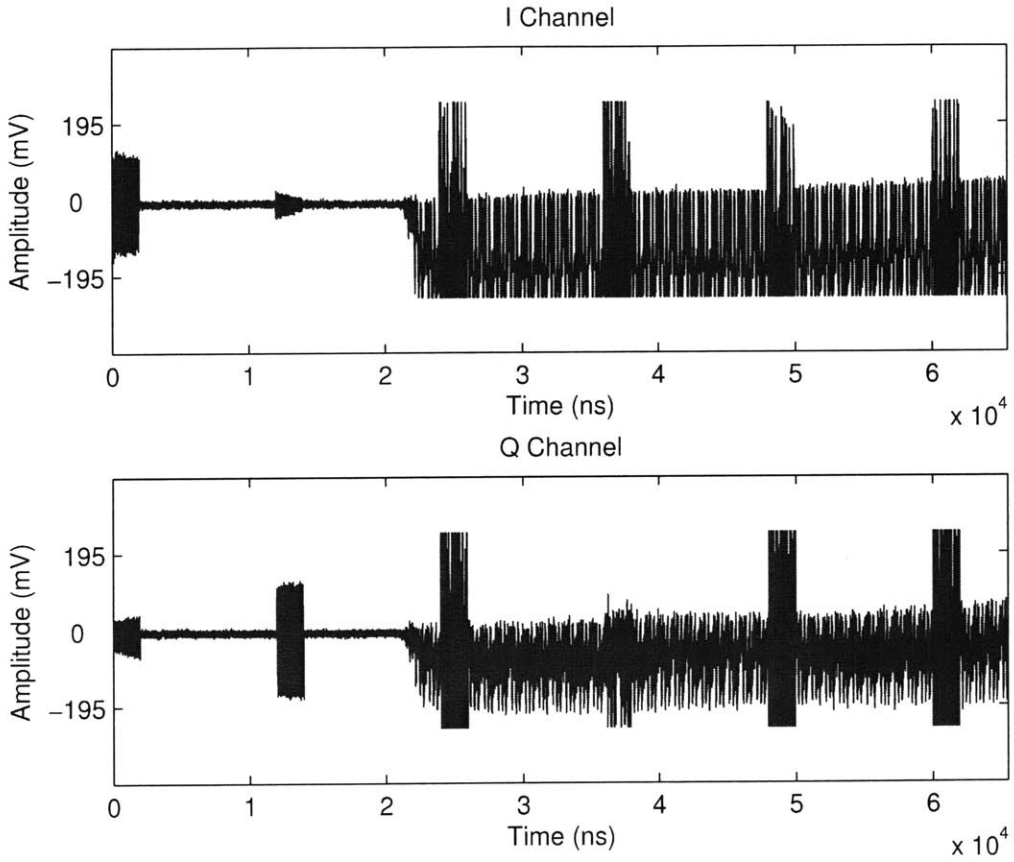


Figure D-8: Multi-tone FSK waveform corrupted by strong interference

Figure D-8 shows an interfered Four-tone FSK signal, the symbol time $T_s$ is 1000 ns and the duty cycle $\theta$ is 1/6. The central frequencies remain the same (5 MHz, 10 MHz, ..., 500 MHz). The four nonzero tones are 10 MHz, 40 MHz, 60 MHz and 100 MHz.

The waveform deteriorates badly due to the interference. The signals are almost jammed. However, the demodulated waveforms (Figure D-9) are still good over the period in which the interference prevails.

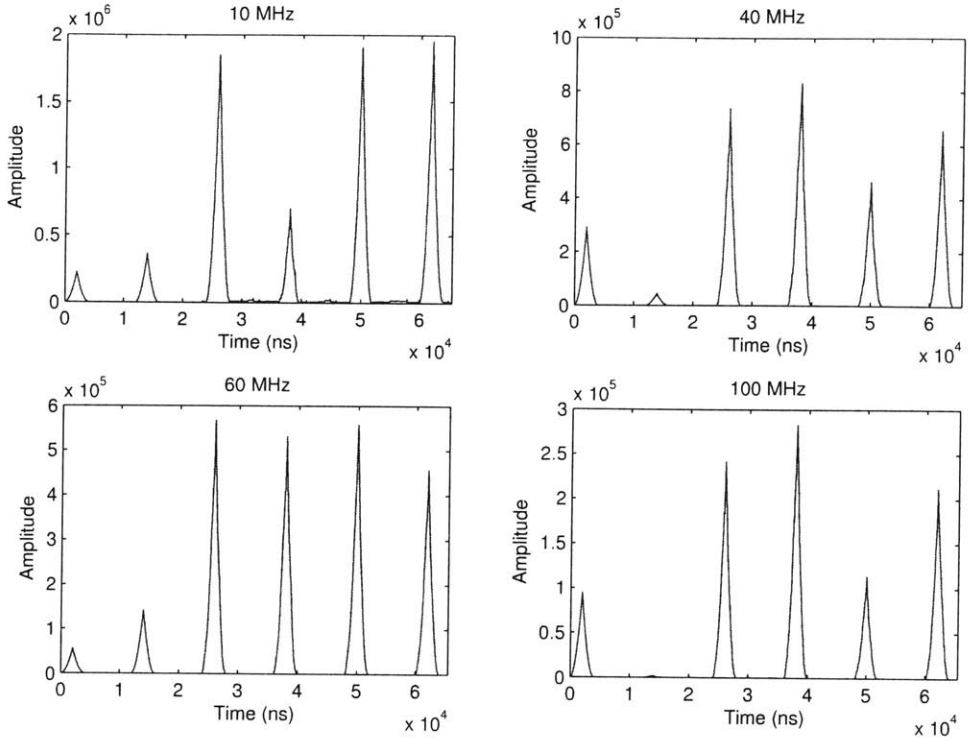


Figure D-9: The demodulation results after the Multi-tone FSK waveform was corrupted by strong interference

# Bibliography


[1] R. S. Kennedy, *Fading Dispersive Communication Channels.* New York, NY: Wiley Interscience, 1969.

[2] C. E. Shannon, *The Mathematical Theory of Information.* Urbana, IL:University of Illinois Press, 1949

[3] C. E. Shannon, "Communication in the presence of noise," *Proc. IRE*, vol. 37, pp. 10-21, Jan. 1949.

[4] R. G. Gallager, *Information Theory and Reliable Communication.* New York, NY: John Wiley & Sons, 1968.

[5] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inform. Theory*, vol. 46, no. 4, pp. 1384-1400, July 2000.

[6] M. Médard and R. G. Gallager, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 840-852, April 2002.

[7] V. G. Subramanian and B. Hajek, "Broad-band fading channels: signal burstiness and capacity," *IEEE Trans. Inform. Theory*, vol. 48, no. 4, pp. 809-827, April 2002.

[8] D. Porrat, D. N. C. Tse and S. Nacu, "Channel uncertainty in ultra wideband communication systems." working paper, 2005.

[9] L. Rusch, C. Prettie, D. Cheung, Q. Li, and M. Ho, "Characterization of UWB propagation from 2 to 8 GHz in a residential environment," *submitted to IEEE Journal on Selected Areas in Communications.*

[10] J. G. Smith, "The information capacity of amplitude and variance-constrained scalar Gaussian channels," *Inform. Contr.*, 18, pp. 203-219, 1971.

[11] S. Shamai and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inform. Theorey*, 41(4), pp. 1060-1071, 1995.

[12] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time non-coherent additive white Gaussian noise channel," *Proc. 2002 IEEE Int'l. Symp. Inform. Theory*, Lausanne, Switzerland, June 30 - July 5, 2002, p. 165.

[13] R. Nuriyev and A. Anastasopoulos, "Capacity and coding for the block-independent noncoherent AWGN channel," *IEEE Trans. Information Theory*, vol. 51, pp. 866-883, Mar 2005.

[14] J. S. Richters, "Communication over fading dispersive channels," *MIT' Res. Lab. Electronics*, Cambridge, MA, Tech. Rep. 464, Nov. 30, 1967.

[15] I. C. Abou-Fayçal, M. D. Trott, and S. Shamai, "The capcity of discrete-time memoryless Rayleigh-fading channels," *IEEE Trans. on Information Theory*, vol. 47, No. 4, pp. 1290-1301, May 2001.

[16] M. C. Gursoy, H. V. Poor, and S. Verdú, "The noncoherent rician fading channel - Part I: structure of the capacity achieving input," accepted for publication in *IEEE Transactions on Wireless Communications.*

[17] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inform. Theorey* , Vol. 45, No. 1, pp. 139-157, Jan 1999.

[18] R. Palanki, "On the capacity achieving distributions of some fading channels." *The 40th Annual Allerton Conference on Communication, Control, and Computing*, October 2002.

[19] R. G. Gallager, "Energy limited channels: coding, multiaccess and spread spectrum," *M.I.T. LIDS Report LIDS-P-1714*, November 1987.

[20] K. Knopp, *Theory of Functions*, New York: Dover, 1945.

[21] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inform. Theory*, vol. 48, no. 6, pp. 1319-1343, June 2002.

[22] D. G. Luenberger, *Optimization by Vector Space Methods*, New York: Wiley, 1969.

[23] J. Huang and S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," To appear on *IEEE Trans. Info. Theory*.

[24] E. Telatar, "Envelope detected complex gaussian noise channel," *Personal Communication*, Bell Laboratories, Lucent Technologies, Murray-Hill, NJ, USA, Jan 21, 1999.

[25] A. N. Shiryayev, *Probability*, New York: Springer-Verlag, 1984.

[26] K. L. Chung, *A Course in Probability Theory*, New York: Harcourt, Brace & World, Inc, 1968.

[27] D. S. Lun, M. Médard, and I. C. Abou-Fayçal, "Error exponents for wideband multipath fading channels - a strong coding theorem," In *Proc. Conference on Information Sciences and Systems*, March 2002.

[28] D. S. Lun, M. Médard, and I. C. Abou-Fayçal, "Error exponents for capacity-achieving signaling on wideband rayleigh fading channels," In *Proc. International Symposium on Information Theory and its Applications*, October 2002.

[29] D. S. Lun, M. Médard, and I. C. Abou-Faycal, "On the performance of peaky capacity-achieving signaling on multipath fading channels," *IEEE Trans. Commun.*, vol. 52, no. 6, pp. 931-938, June 2004.

[30] W. E. Stark, "Capacity and cutoff rate of noncoherent FSK with nonselective Rician fading," *IEEE Trans. Commun.*, vol. COM-33, no. 11, pp. 1153-1159, November 1985.

[31] A. Matache and J. A. Ritcey, "Optimum code rates for noncoherent MFSK with errors and erasures decoding over rayleigh fading channels," In *Proc. The Thirty-First Asilomar Conference on Signals, Systems and Computers*, pp. 62-66, 1997.

[32] J. G. Proakis, *Digital Communications*. New York, NY: McGraw-Hill, 1995.

[33] D. Schafhuber, H. Boleskei, and G. Matz, "System capacity of wideband OFDM communications over fading channels without channel knowledge," In *Proc. International Symposium on Information Theory*, pp. 389-389, June 2004.

[34] Y. Souilmi and R. Knopp, "On the achievable rates of ultra-wideband PPM with non-coherent detection in multipath environments," In *Proc. IEEE International Conference on Communications*, vol. 5, pp. 3530-3534, May 2003.

[35] C. Luo and M. Médard, "Frequency-shift Keying for ultrawideband - achieving rates of the order of capacity," In *Proc. The 40th Annual Allerton Conference on Communication, Control, and Computing*, October 2002.

[36] C. Luo and M. Médard, "Performance of single-tone and two-tone - Frequency-Shift Keying for ultrawideband," In *Proc. The 36th Asilomar Conference on Signals, Systems and Computers*, Nov. 2002.

[37] C. Luo, M. Médard, and L. Zheng, "Error Exponents for Multi-tone Frequency Shift Keying on Wideband Rayleigh Fading Channels," *GLOBECOM '03. IEEE*, Vol: 2, pp. 779 783, 1-5 Dec. 2003.

[38] C. Luo, M. Médard. and L. Zheng, "On capacity-achieving distribution in Rician fading channel," *Conference on Information Sciences and Systems*, March 2005.

[39] C. E. Shannon, "The zero-error capacity of a noisy channel," *IRE Trans. on Inform. Theory*, vol. IT-2, pp. 8-19, September 1956.

[40] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," IEEE Trans. Inform. Theory, 12(2): 172–182, Apr. 1966.

[41] M. Nakagami, "The m-distribution, a general formula of intensity distribution of rapid fading," *Statistical Methods in Radio Wave Propagation*, pp. 3-36, Pergamon Press, Oxford, England, 1960.

[42] G. L. Turin, et al., "A statistical model of urban multipath propagation," *IEEE Trans. Vehicul. Technol.*, pp. 1-8, Feb. 1972.

[43] C. Luo, M. Médard, L. Zheng, and D. S. Lun, "Multi-tone FSK with feedback," *IEEE International Symposium on Information Theory*, September 2005.

[44] C. Luo, M. Médard, and L. Zheng, "On approaching wideband capacity using multi-tone FSK," *IEEE Journal on Selected Areas in Communications*, vol. 23, No. 9, pp. 1830-1838, Sep. 2005.

[45] MIT Ultra-wideband Project, *http: //www-mtl.mit.edu /researchgroups / icsystems /uwb /about.html*