A DECOMPOSITION FOR THE TILTED CHANNEL OF THE FAST FADING CHANNELS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MÜCAHIT FURKAN YILDIZ

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

SEPTEMBER 2022

Approval of the thesis:

A DECOMPOSITION FOR THE TILTED CHANNEL OF THE FAST FADING CHANNELS

submitted by **MÜCAHIT FURKAN YILDIZ** in partial fulfillment of the requirements for the degree of **Master of Science in Electrical and Electronics Engineering Department, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ______________

Prof. Dr. İlkay Ulusoy
Head of Department, **Electrical and Electronics Engineering** ______________

Assist. Prof. Dr. Barış Nakiboğlu
Supervisor, **Electrical and Electronics Engineering, METU** ______________

Examining Committee Members:

Assoc. Prof. Dr. Ayşe Melda Yüksel Turgut
Electrical and Electronics Engineering, METU ______________

Assist. Prof. Dr. Barış Nakiboğlu
Electrical and Electronics Engineering, METU ______________

Assist. Prof. Dr. Hao-Chung Cheng
Electrical and Electronics Engineering, National Taiwan U. ______________

Assist. Prof. Dr. Gökhan Muzaffer Güvensen
Electrical and Electronics Engineering, METU ______________

Assist. Prof. Dr. Elif Tuğçe Ceran Arslan
Electrical and Electronics Engineering, METU ______________

Date: 01.09.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Surname: Mücahit Furkan Yıldız

Signature :

ABSTRACT

A DECOMPOSITION FOR THE TILTED CHANNEL OF THE FAST FADING CHANNELS

Yıldız, Mücahit Furkan
M.S., Department of Electrical and Electronics Engineering
Supervisor: Assist. Prof. Dr. Barış Nakiboğlu

September 2022, 64 pages

A decomposition property for the tilted channel of the fast fading channels with the channel state information at the receiver is proposed. A necessary and sufficient condition for the decomposition and the resulting expressions for the sphere packing exponent and the random coding exponent are determined. These expressions are used to calculate the error exponents for the discrete memoryless channels with certain symmetries. The existence of a similar decomposition property for Gaussian channels under Gaussian input distributions is analyzed.

Keywords: Augustin Information Measures, Sphere Packing Exponent, Random Coding Exponent, Fast Fading, Discrete Channels, Gaussian Channels

ÖZ

HIZLI SÖNÜMLEMELİ KANALLARIN EĞİK KANALLARI İÇİN BİR DEKOMPOSİZYON

Yıldız, Mücahit Furkan
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi: Dr. Öğr. Üyesi. Barış Nakiboğlu

Eylül 2022 , 64 sayfa

Kanal durum bilgisinin alıcıda olduğu hızlı sönümlü kanalların eğik kanalları için bir dekompozisyon özelliği önerildi. Böyle bir dekompozisyonun olabilmesi içim hem gerekli hem de yeterli olan bir koşul ve bu koşulu sağlayan durumlar için küre sıkıştırma üssü ve rassal kodlama üssü ifadeleri elde edildi. Bu ifadeleri kullanarak bazı simetrilere sahip kanalların hata üssü fonksiyonları belirlendi. Gauss kanallarda girdi dağılımı Gauss olduğunda benzer bir dekompozisyonun var olup olmadığı incelendi.

Anahtar Kelimeler: Augustin Bilgi Ölüçüleri, Küre Sıkıştırma Üssü, Rassal Kodlama Üssü, Ayrık Kanallar, Gauss Kanallar

To my family and Feyza

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor Assist. Prof. Dr. Barış Nakiboğlu, for his guidance, encouragement, and endless patience throughout the completion of this thesis. I have learned a lot during our discussions.

I would like to thank Prof. Dr. Sinan Gezici, Assoc. Prof. Dr. Ayşe Melda Yüksel Turgut, Assist. Prof. Dr. Hao-Chung Cheng, Assist. Prof. Dr. Gökhan Muzaffer Güvensen, Assist. Prof. Dr. Elif Tuğçe Ceran Aslan for accepting to read this thesis. I would like to thank Assist. Prof. Dr. Serkan Sarıtaş for his valuable reviews of this thesis.

I would also like to thank TÜBİTAK for financially supporting this thesis under Grant 119E053.

I would like to extend my sincere thanks to all my dear friends. They never wavered in their support and never let me down.

I could not have undertaken this journey without my precious wife, Feyza. I am thankful for her constant support, encouragement, patience, and understanding throughout the development of this thesis.

Finally, this endeavor would not have been possible without my beloved mother Şahsenem, father Güngör, and brother Ömer. I have always felt and will continue to feel your support. I am fortunate to have you and owe you everything I have.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

As a result of the rapid developments in communication technologies, people started to realize the different usage areas of communication systems. At first, it was only possible to talk to landline phones. Then cellular phones replaced landline phones because people wanted to be connected and instantly reachable while walking outside or sitting in a cafe. When the internet was first presented, computers required cable to connect to the internet, and people used it for simple tasks such as sending an e-mail and browsing. However, people want to watch a movie, do online shopping, and handle their bank accounts on their phones, leading to the development of fast and robust wireless connections. The development process is accelerating with new ideas such as machine-to-machine (M2M) connection without human interference, self-driving cars, and smart homes. The high mobility, high data rate, and low latency requirements are growing more than ever.

Despite all the exciting ideas of wireless communication, it inherits a genuine problem that is inevitable and does not exist in wired communication. Wires have characteristics of channel responses, and they do not change over time. However, environmental circumstances highly affect wireless channels, and rapid shifts can cause a reduction in the signal quality, called the fading effect.

To overcome the fading effect, researchers model the wireless channels with some parameters, i.e., coherence time and bandwidth. Coherence time indicates the duration in which the wireless channels nearly have the same channel response, and the coherence bandwidth indicates the bandwidth in which wireless channels nearly have the same channel response for the frequencies apart from each other. Under these parameters, different models with different assumptions are used in the literature. Some models assume that the channel characteristics can be tracked in the coherence time, and this response information, i.e., the channel state information (CSI), is known at the receiver side. This model is known as the fast-fading channel with CSI at the receiver. We will use this model in the rest of the thesis.

When the high-reliability demands are the primary considerations of the communication, error exponents and their refinements are used. They give us the relation between the codelength and the exponential decay rate of the probability of error. Analyzing the error exponents, one can know the minimum codelength that achieves the required exponential decay of the error probability in a coding scheme. We aim to understand the fading effect by deriving the relationship between the corresponding non-fading channels and the fast-fading channels with CSI at the receiver.

1.1 Related Works

In literature, fast-fading channels with CSI at the receiver is a widely used model by researchers. In [1], the author calculates the sphere packing exponent (SPE) and the random coding exponent (RCE) of the fast-fading discrete memoryless channels with CSI at the receiver by using the conditioned non-fading discrete memoryless channels for each fading parameter for given input distribution. Then, the author showed that the calculated SPE and the RCE are always less than the expected value of the SPE and the RCE of the non-fading discrete memoryless channels conditioned on the fading parameter. Another significant channel model is the fast-fading Gaussian channel model with CSI at the receiver because the distribution of the thermal noise is modeled as the Gaussian distribution. Researchers widely adopt this model with various fading distributions and calculate the RCE using Gallager's idea [2]. In [3], the author calculates the RCE of fast-fading Gaussian channels for capacity-achieving input distribution without showing that the chosen input distribution is optimum. Although it is optimum for the rate equal to the channel capacity, it may not be the optimum input distribution for the rates below the channel capacity. We summarized the works on the fast-fading Gaussian channels in Table 1.1. Generally, authors adopt the massive input - massive output (MIMO) channel models.

Article	Channel Model	Power Allocation	Used Techniques
[4]	Rayleigh-Fading with Diversity and CSI	Capacity-Achieving Gaussian Distribution	RCE
[5]	Rayleigh-Fading Memoryless Correlated Channels with Full or Partial CSI	Equal Power Case	RCE
[6]	Rayleigh and Nakagami-Fading with Diversity and CSI	Capacity-Achieving Gaussian Distribution	RCE
[7]	Rayleigh and Nakagami-Fading and Diversity		RCE
[8]	Rayleigh-Fading with MIMO and CSI	Capacity-Achieving Gaussian Distribution	RCE
[9]	Rayleigh-Fading with MIMO and CSI	Capacity-Achieving Gaussian Distribution	RCE
[10]	Rician-Fading with MIMO and CSI		RCE and Upper-bound
[11]	Rayleigh-Fading Channels with CSI	Approximations for Special Cases	Expurgated Bound, RCE and SPE
[12]	Block Fading MIMO with CSI	Capacity-Achieving Gaussian Distribution	Expurgated Bound and RCE
[13]	Block Fading MIMO with CSI	Binary and Capacity-Achieving Gaussian Distribution	RCE
[14]	Nakagami-m Fading MIMO with CSI	Equal Power	RCE
[15]	Multi-keyhole MIMO with CSI	Equal Power	RCE
[16]	Generalized K-Fading MIMO with CSI	Capacity-Achieving Gaussian	RCE

[17]	Rayleigh-Fading Product MIMO with CSI	Capacity-Achieving Gaussian	RCE
[18]	Multiple-Rayleigh Scattering MIMO with CSI	Capacity-Achieving Gaussian	RCE
[18]	Multiple-Rayleigh Scattering MIMO with CSI	Uniform Power Allocated Gaussian Input	RCE
[19]	Multiple-Rayleigh Scattering MIMO with CSI	Uniform Power Allocated Gaussian Input	RCE and Expurgated Bound
[20]	η-μ Fading MIMO with CSI	Capacity-Achieving Gaussian	RCE
[21]	Block Fading MIMO with CSI		RCE

Table 1.1: The table of the related works

1.2 Contributions and Novelties

In Chapter 3, we proved the Lemma 1. This Lemma explains the required and sufficient conditions to decompose tilted channels of fast-fading discrete memoryless channels for an input distribution into a tilted fading distribution and tilted channels of the corresponding non-fading discrete memoryless channels for a given CSI with the same input distribution. Then we proved that the Lemma 1 conditions are satisfied for a particular class of channels. We calculated the parametric expressions of SPEs and the rates. Then we analyzed fading binary symmetric channels (FBSC) and fading binary erasure channels (FBEC) as examples. We found that the SPEs of FBSCs are always greater than SPEs of non-fading binary symmetric channels (BSC) under the same-capacity condition for all rates lower than the channel capacity. We found that the fading distribution that takes only two values, 0 and 0.5, maximizes the SPEs of the FBSCs. We showed that SPEs of all FBECs are equal to the SPEs of non-fading binary erasure channels (BEC) under the same-capacity condition for all rates lower than the channel capacity. We explain how to generalize the Lemma 1 to the infinite sets and the probability density functions in Chapter 4. Then, we checked the Gaussian channels to find that the decomposition described in the Lemma 1 is not applicable for the fading Gaussian channels. Lastly, we proved that the conditional output distribution of tilted fading channels conditioned on the fading parameter could not be a Gaussian channel for the Gaussian input distribution. This result also implies that the conditional output distribution of the Augustin mean can not be a Gaussian distribution.

1.3 Thesis Outline

We organized this thesis as follows. We define the preliminary concepts and the Augustin information measures on the finite sets and the probability mass function in Chapter 2. Then we explain the notation of the thesis, discrete channels, and the coding. This chapter discusses the error exponent functions, i.e., the SPE and the RCE. Chapter 3 explains fading discrete channels, Lemma 1, and its applications with examples. Then we compare the results of the examples. In Chapter 4, we discuss the fading Gaussian channels in the aspect of the decomposition of their tilted channels. Finally, Chapter 5 concludes the thesis.

CHAPTER 2

AUGUSTIN INFORMATION MEASURES AND ERROR EXPONENT FUNCTIONS

In this Chapter, we describe the notation of the thesis, which we will use hereafter, and give preliminary definitions. We describe the Augustin information measures, discrete memoryless channels, and the error exponent functions. Then, we explain the connections between these concepts and their operational features. Lastly, we discuss the advantages of the Augustin information measures over the more commonly used Gallager's function.

2.1 Notation and Preliminaries

2.1.1 Notation

Throughout the thesis, we will use the word *alphabet* to denote the finite sets. We show all the probability mass functions (PMFs) on the alphabet $\mathcal{Y}$, with $\mathcal{P}(\mathcal{Y})$. A *Channel* W is a function satisfying $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$. Thus, for any $x \in \mathcal{X}$, $W(x)$ is a PMF in $\mathcal{P}(\mathcal{Y})$. We show values of $W(x)$ for any $y \in \mathcal{Y}$ with $W(y|x)$. We denote the set of all strings with length *n*, which we can construct on the alphabet $\mathcal{Y}$ with $\mathcal{Y}^n$. Note that $\mathcal{Y}^n$ is also a finite set. Therefore, all the previous notations are also valid for $\mathcal{Y}^n$. A PMF $w \in \mathcal{P}(\mathcal{Y})$ is said to be absolutely continuous in a probability mass function $q \in \mathcal{P}(\mathcal{Y})$, denoted by $w \prec q$, if $w(y) = 0$ for all y where $q(y) = 0$.

Example 1. For PMFs q_1, q_2, q_3

$$q_1(y) = \begin{cases} 0 & \text{if } y = 1 \\ 1 & \text{if } y = 2 \\ 0 & \text{if } y = 3 \end{cases} \quad q_2(y) = \begin{cases} 0.5 & \text{if } y = 1 \\ 0.5 & \text{if } y = 2 \\ 0 & \text{if } y = 3 \end{cases} \quad q_3(y) = \begin{cases} 0 & \text{if } y = 1 \\ 0.5 & \text{if } y = 2 \\ 0.5 & \text{if } y = 3 \end{cases}$$

Only $q_1 \prec q_2$ and $q_1 \prec q_3$. However, $q_2 \not\prec q_1$, $q_2 \not\prec q_3$, $q_3 \not\prec q_1$, and $q_3 \not\prec q_2$.

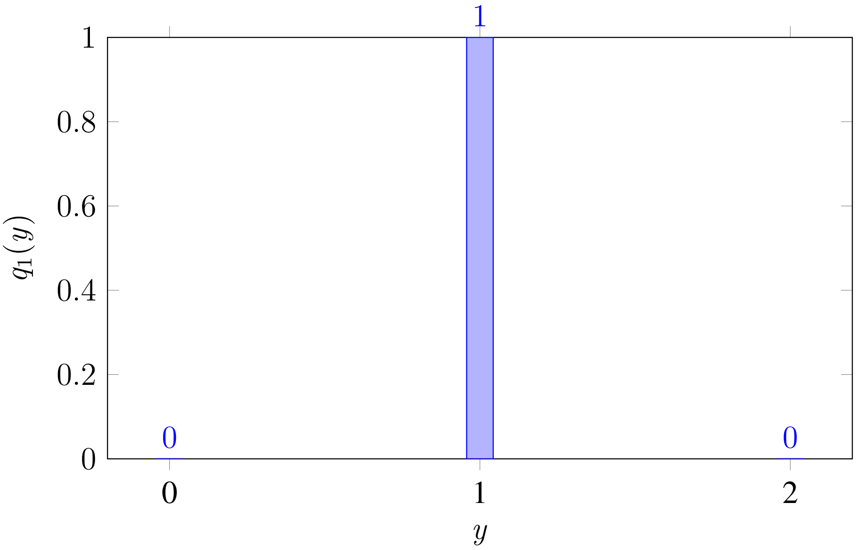

Figure 2.1: The plot of the probability mass function q_1 in Example 1

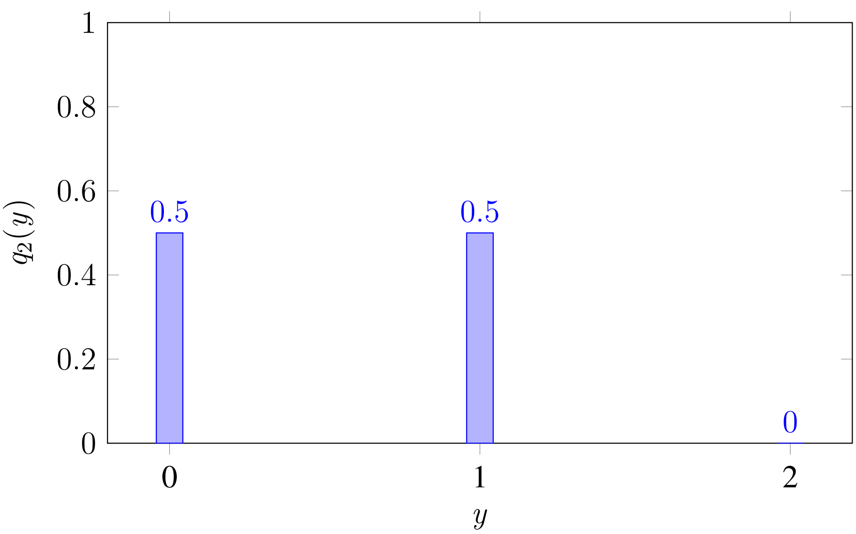

Figure 2.2: The plot of the probability mass function q_2 in Example 1

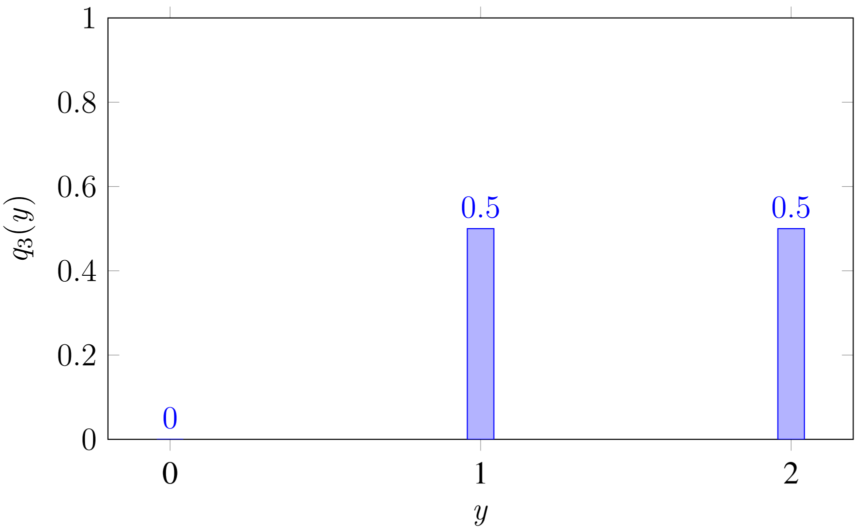

Figure 2.3: The plot of the probability mass function q_3 in Example 1

The absolutely continuity relation has no commutative property. Two PMFs $w, q \in \mathcal{P}(\mathcal{Y})$ are equivalent if and only if $w \prec q$ and $q \prec w$ and we show the equivalence relation with $w \sim q$. The equivalence relation has the commutative property.

2.1.2 Preliminary Definitions

Definition 1. For any $\alpha \in \mathbb{R}_+$ and $w, q \in \mathcal{P}(\mathcal{Y})$, the order α Rényi divergence between w and q, denoted by $D_\alpha(w\| q)$, is

$$D_\alpha(w\| q) := \begin{cases} \frac{1}{\alpha-1} \ln \sum_y [w(y)]^\alpha [q(y)]^{1-\alpha} & \text{if } \alpha \in \mathbb{R}_+ \setminus \{1\} \\ \sum_y w(y) \ln \frac{w(y)}{q(y)} & \text{if } \alpha = 1 \end{cases}. \tag{2.1}$$

The order one Rényi divergence is equal to the Kullback-Leibler divergence. By [22, Lemma 2], for any $\alpha \in \mathbb{R}_+$ and $w, q \in \mathcal{P}(\mathcal{Y})$,

$$D_\alpha(w\| q) \geq \tfrac{\min(1,\alpha)}{2} \sum_y (w(y) - q(y))^2. \tag{2.2}$$

(2.2) is known as the Pinsker's inequality for $\alpha \in (0,1)$. As a result of (2.2), for any $w, q \in \mathcal{P}(\mathcal{Y})$ and any $\alpha \in \mathbb{R}_+$, we get

$$0 \leq D_\alpha(w\| q), \tag{2.3}$$

and the equality in (2.3) holds only when $w(y) = q(y)$ for all $y \in \mathcal{Y}$ by (2.2). The Rényi divergence has no commutative property for $\alpha \neq \frac{1}{2}$, i.e., for any $\alpha \neq \frac{1}{2}$, there exist $w, q \in \mathcal{P}(\mathcal{Y})$ s.t.

$$D_{\alpha}(w\| q) \neq D_{\alpha}(q\| w)\,. \tag{2.4}$$

The Rényi divergence has commutative property for $\alpha = \frac{1}{2}$, i.e., for all $w, q \in \mathcal{P}(\mathcal{Y})$

$$D_{\frac{1}{2}}(w\| q) = D_{\frac{1}{2}}(q\| w)\,. \tag{2.5}$$

We can interpret the Rényi divergence as a measure of the distance between the probability mass functions. However, it is not a metric because it has no commutative property.

Definition 2. For any $\alpha \in \mathbb{R}_{+}$, $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, $Q:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$, and $p\in\mathcal{P}(\mathcal{X})$, the order α conditional Renyi divergence, denoted by $D_{\alpha}(W\| Q| p)$, is

$$D_{\alpha}(W\| Q| p) := \sum\nolimits_{x} p(x) D_{\alpha}(W(x)\| Q(x))\,. \tag{2.6}$$

In the case when the second argument of the conditional Rényi divergence is independent of the parameter x, i.e., $Q(x) = q$ for some $q \in \mathcal{P}(\mathcal{Y})$, $D_{\alpha}(W\| q| p)$ can be used instead of $D_{\alpha}(W\| Q| p)$.

Definition 3. For any $\alpha \in \mathbb{R}_{+}$, $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$, and $q \in \mathcal{P}(\mathcal{Y})$, the order α tilted channel, denoted by $W_{\alpha}^{q}(y|x)$, is

$$W_{\alpha}^{q}(y|x) := [W(y|x)]^{\alpha}[q(y)]^{1-\alpha} e^{(1-\alpha)D_{\alpha}(W(x)\|q)}, \qquad \forall x \in \mathcal{X}. \tag{2.7}$$

Using (2.1) in (2.7) , we get

$$W_{\alpha}^{q}(y|x) = \begin{cases} \frac{[W(y|x)]^{\alpha}[q(y)]^{1-\alpha}}{\sum_{\tilde{y}}[W(\tilde{y}|x)]^{\alpha}[q(\tilde{y})]^{1-\alpha}} & \text{if } \alpha\in\mathbb{R}_{+}\setminus\{1\} \\ W(y|x) & \text{if } \alpha=1 \end{cases}. \tag{2.8}$$

The conditional Rényi divergence and the tilted channel play a significant role in characterizing the Augustin mean in Section 2.2.1.

2.2 Augustin Information Measures

This section explains the Augustin information measures, including the Augustin information, the Augustin mean, the Augustin capacity, the Rényi radius, and the Augustin center.

2.2.1 Augustin Mean and Information

Definition 4. For any $\alpha \in \mathbb{R}_+$, channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and input distribution $p \in \mathcal{P}(\mathcal{X})$, the order α Augustin information is

$$I_\alpha(p; W) := \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(W \| q | p) . \tag{2.9}$$

For the order $\alpha = 1$

$$D_1(W \| q | p) = D_1(W \| q_p | p) + D_1(q_p \| q) \tag{2.10}$$

where

$$q_p := \sum_x p(x) W(x). \tag{2.11}$$

Since $D_1(q_p \| q) \geq 0$ by (2.3), infimum in (2.9) is achieved by q_p. Therefore,

$$q_{1,p} = q_p \tag{2.12}$$

$$I_1(p; W) = D_1(W \| q_p | p) . \tag{2.13}$$

Hence, the order one Augustin information is equal to the mutual information. The probability distribution that achieves the infimum in (2.9) has a closed-form expression for the order one. However, a closed-form expression of the probability distribution that achieves the infimum in (2.9) is not available for orders other than one. Nevertheless, the existence of the unique output distribution that achieves the infimum in (2.9) for any PMF p on arbitrary classical channels is proved for orders $\alpha \in (0, 1)$ in [23] and for orders $\alpha \in (1, \infty)$ in [22]. The output distribution that achieves the infimum in (2.9) is called as the order α *Augustin mean*, and denoted by $q_{\alpha,p}$. Thus,

$$I_\alpha(p; W) = D_\alpha(W \| q_{\alpha,p} | p) . \tag{2.14}$$

Although the closed-form of the Augustin mean does not exist for orders other than one, the Augustin mean can be characterized with the Augustin operator [22, Lemma 13].

Definition 5. The Augustin operator $\mathtt{T}_{\alpha,p}(\cdot) : \mathcal{Q}_{\alpha,p} \to \mathcal{P}(\mathcal{Y})$ is

$$\mathtt{T}_{\alpha,p}(q) := \sum_x p(x) W_\alpha^q(x) \qquad\qquad \forall q \in \mathcal{Q}_{\alpha,p}, \tag{2.15}$$

where $\mathcal{Q}_{\alpha,p} := \{q \in \mathcal{P}(\mathcal{Y}) : D_\alpha(W\| q| p) < \infty\}$.

The order α Augustin mean satisfies the fixed point property, i.e., $\mathtt{T}_{\alpha,p}(q_{\alpha,p}) = q_{\alpha,p}$, which also satisfies $q_{\alpha,p} \sim q_p$. Furthermore, any probability distribution q which is a fixed point of the Augustin operator, i.e., $\mathtt{T}_{\alpha,p}(q) = q$, that also satisfies $q_p \prec q$ is the Augustin mean by [22, Lemma 13]. In addition to the proof of the existence of the unique Augustin mean for any PMF p on arbitrary classical channels in [22], it also proved for arbitrary input distribution p on arbitrary classical channels in [24]. Furthermore, the existence of the unique Augustin mean is proved for any PMF p on the classical-quantum channels in [25].

2.2.2 Rényi Mean and Information

We can define other information measures using the Rényi divergence. One of them is the Rényi information which is implicitly defined by the Gallager[1] in [2].

Definition 6. For any $\alpha \in \mathbb{R}_+$, channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and input distribution $p \in \mathcal{P}(\mathcal{X})$, the order α Rényi information is

$$I_\alpha^g(p; W) := \inf_{q \in \mathcal{P}(\mathcal{Y})} D_\alpha(p \circledast W \| p \otimes q), \tag{2.16}$$

where $p \circledast W$ denotes the PMF whose marginal on $\mathcal{X}$ is p and whose conditional distribution given x is $W(x)$. $p \otimes q \in \mathcal{P}(\mathcal{X} \otimes \mathcal{Y})$ denotes the PMF defined on the product set of $\mathcal{X} \otimes \mathcal{Y}$ whose marginal on $\mathcal{X}$ is p and marginal on $\mathcal{Y}$ is q.

[1] With different parametrization and scaling and a restriction s.t. $\alpha \in (0, 1)$.

Definition 7. For any $\alpha \in \mathbb{R}_+$, channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and input distribution $p \in \mathcal{P}(\mathcal{X})$, the order α power mean is

$$\mu_{\alpha,p}(y) = \left[\sum_x p(x)(W(y|x))^\alpha\right]^{\frac{1}{\alpha}}. \tag{2.17}$$

Definition 8. For any $\alpha \in \mathbb{R}_+$, channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ and input distribution $p \in \mathcal{P}(\mathcal{X})$, the order α Rényi mean is

$$q^g_{\alpha,p} = \frac{\mu_{\alpha,p}}{||\mu_{\alpha,p}||}. \tag{2.18}$$

We can write

$$D_\alpha(p \circledast W \| p \otimes q) = D_\alpha\big(p \circledast W \| p \otimes q^g_{\alpha,p}\big) + D_1\big(q^g_{\alpha,p} \| q\big). \tag{2.19}$$

Since $D_1\big(q^g_{\alpha,p} \| q\big) \geq 0$ by (2.3), infimum in (2.16) is achieved by $q^g_{\alpha,p}$. Therefore,

$$I^g_\alpha(p; W) = D_\alpha\big(p \circledast W \| p \otimes q^g_{\alpha,p}\big). \tag{2.20}$$

The order one Rényi information is equal to the order one Augustin information, and the order one Rényi mean is equal to the order one Augustin mean for any input distribution p. However, unlike the Augustin mean, the Rényi mean has a closed-form expression for orders other than one. Thus, characterizing the Augustin information in terms of the Rényi information is helpful. The corresponding characterization is done in [22]. As a result of this characterization, we get the inequalities

$$I_\alpha(W;p) \geq I^g_\alpha(W;p) \qquad \forall \alpha \in (0,1), \tag{2.21}$$

$$I_\alpha(W;p) \leq I^g_\alpha(W;p) \qquad \forall \alpha \in (1,\infty). \tag{2.22}$$

2.2.3 Augustin Center and Capacity

The Augustin capacity directly appears in the expressions of the SPE and the RCE. For this reason, it is the pivotal quantity in the SPE and the RCE analysis with Augustin information measures.

Definition 9. The order α Augustin capacity of a channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ for a constraint set $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ is the supremum of the order α Augustin information over all the possible input distributions $p \in \mathcal{A}$. Thus,

$$C_{\alpha,W,\mathcal{A}} := \sup_{p \in \mathcal{A}} I_\alpha(p; W). \tag{2.23}$$

If the constraint set is the whole $\mathcal{P}(\mathcal{X})$, we denote the Augustin capacity with the $C_{\alpha,W}$. Thus

$$C_{\alpha,W,\mathcal{P}(\mathcal{X})} = C_{\alpha,W}. \tag{2.24}$$

Definition 10. The order α Rényi radius of a channel $W:\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ is

$$S_{\alpha,W} := \inf_{q\in\mathcal{P}(\mathcal{Y})} \max_{x\in\mathcal{X}} D_{\alpha}(W(x)\| q). \tag{2.25}$$

Using the definition of the Augustin information in the definition of the Augustin capacity, we get

$$C_{\alpha,W,\mathcal{A}} = \sup_{p\in\mathcal{A}} \inf_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(W\| q| p). \tag{2.26}$$

We can change the order of the supremum and infimum for convex input sets by [22, Theorem 1], and we get

$$C_{\alpha,W,\mathcal{A}} = \inf_{q\in\mathcal{P}(\mathcal{Y})} \sup_{p\in\mathcal{A}} D_{\alpha}(W\| q| p). \tag{2.27}$$

If the constraint set is the whole $\mathcal{P}(\mathcal{X})$, then the Augustin capacity is equal to the Rényi radius.

$$C_{\alpha,W} = \inf_{q\in\mathcal{P}(\mathcal{Y})} \max_{x\in\mathcal{X}} D_{\alpha}(W(x)\| q), \tag{2.28}$$

$$= S_{\alpha,W}. \tag{2.29}$$

Suppose $C_{\alpha,W,\mathcal{A}}$ is finite, which is always the case for DMC's. In that case, there exists a unique output distribution that achieves the infimum in (2.27) by [22, Theorem 1]. It is called as the order α *Augustin center* of the constrained set $\mathcal{A}$, and denoted by $q_{\alpha,W,\mathcal{A}}$. Thus,

$$C_{\alpha,W,\mathcal{A}} = \sup_{p\in\mathcal{A}} D_{\alpha}(W\| q_{\alpha,W,\mathcal{A}}| p). \tag{2.30}$$

Similar to the Augustin capacity, if the constraint set is the whole $\mathcal{P}(\mathcal{X})$, we denote the Augustin center with the $q_{\alpha,W}$. Thus

$$q_{\alpha,W,\mathcal{P}(\mathcal{X})} = q_{\alpha,W}. \tag{2.31}$$

In [22], the existence of the unique order α Augustin center is generalized for the arbitrary input distribution p.

2.2.4 Rényi Center and Capacity

Similar to the Augustin capacity, Rényi capacity with the constraint set $\mathcal{A} \subset \mathcal{P}(\mathcal{X})$ can be defined as the supremum of the Rényi information for the possible input distributions $p \in \mathcal{A}$.

Definition 11. The order α Rényi capacity of a channel $W : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ with constraint set $\mathcal{A}$ is the supremum of the order α Rényi information over all the possible input distributions $p \in \mathcal{A}$. Thus,

$$C^{g}_{\alpha,W,\mathcal{A}} := \sup\nolimits_{p\in\mathcal{A}} I^{g}_{\alpha}(p;W) , \tag{2.32}$$

$$= \sup\nolimits_{p\in\mathcal{A}} \inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} D_{\alpha}(p \circledast W \| \, p \otimes q) . \tag{2.33}$$

If the constraint set is the whole $\mathcal{P}(\mathcal{X})$, we denote the Rényi capacity with the $C^{g}_{\alpha,W}$. Thus,

$$C^{g}_{\alpha,W,\mathcal{P}(\mathcal{X})} = C^{g}_{\alpha,W}. \tag{2.34}$$

Again, we can change the order of the supremum and infimum for convex $\mathcal{A}$ by [26].

$$C^{g}_{\alpha,W,\mathcal{A}} = \inf\nolimits_{q\in\mathcal{P}(\mathcal{Y})} \sup\nolimits_{p\in\mathcal{A}} D_{\alpha}(p \circledast W \| \, p \otimes q) . \tag{2.35}$$

If the constraint set is the whole $\mathcal{P}(\mathcal{X})$, then the Rényi capacity is equal to the Rényi radius.

$$C^{g}_{\alpha,W} = \inf\nolimits_{q\in\mathcal{A}} \max\nolimits_{x\in\mathcal{X}} D_{\alpha}(W(x) \| \, q) , \tag{2.36}$$

$$= S_{\alpha,W}. \tag{2.37}$$

If $C^{g}_{\alpha,W}$ is finite, which is always the case for DMCs, then there exists a unique output distribution that achieves the infimum in (2.35) by [26]. It is called as the order α *Rényi center* and denoted by the $q^{g}_{\alpha,W,\mathcal{A}}$. Thus,

$$C^{g}_{\alpha,W,\mathcal{A}} = \sup\nolimits_{p\in\mathcal{A}} D_{\alpha}\big(p \circledast W \| \, p \otimes q^{g}_{\alpha,W}\big) . \tag{2.38}$$

Similar to the Rényi capacity, if the constraint set is the whole $\mathcal{P}(\mathcal{X})$, we denote the Rényi center with the $q^{g}_{\alpha,W}$. Thus,

$$q^{g}_{\alpha,W,\mathcal{A}} = q^{g}_{\alpha,W}. \tag{2.39}$$

Because the order one Augustin information is equal to the Rényi information for any input p, we can say that the order one constrained Augustin capacity is equal to the order one constrained Rényi capacity. However, for the orders other than one, the Augustin information is greater than the Rényi information for $\alpha \in (0,1)$ and the Augustin information is lower or equal than the Rényi information for $\alpha \in (1,\infty)$ by (2.21) and (2.22). Therefore the Augustin capacity is greater than the Rényi capacity for $\alpha \in (0,1)$ and the Augustin capacity is lower or equal to the Rényi capacity for $\alpha \in (1,\infty)$, as well.

$$C_{\alpha,W,\mathcal{A}} \geq C^{g}_{\alpha,W,\mathcal{A}} \qquad \forall\alpha \in (0,1), \tag{2.40}$$

$$C_{\alpha,W,\mathcal{A}} \leq C^{g}_{\alpha,W,\mathcal{A}} \qquad \forall\alpha \in (1,\infty). \tag{2.41}$$

The equality in (2.40) and (2.41) holds only if $D_{\alpha}(W\| q_{\alpha,W,\mathcal{A}})$ equal to each other for all x s.t. $p(x) > 0$ for the Augustin capacity-achieving input distribution $p \in \mathcal{A}$. Furthermore, the Augustin capacity is equal to the Rényi capacity for the case $\mathcal{A} = \mathcal{P}(\mathcal{X})$ because they are both equal to the Rényi radius.

2.3 Discrete Memoryless Channels and Coding

Before introducing the error exponent functions, we describe the discrete memoryless channels and coding, which are the concepts that should be known beforehand. A discrete channel, shown in Figure 2.4, is a channel that takes an input symbol from the finite input set $\mathcal{X}$ and produces a probability mass function on the finite output set $\mathcal{Y}$. We consider the discrete channel memoryless if the output distribution on the

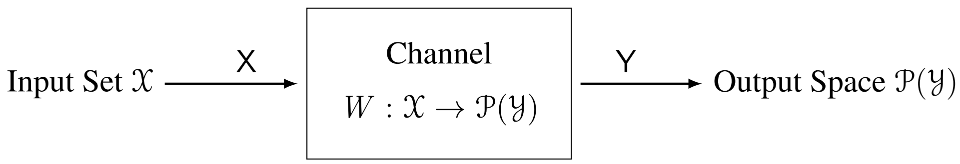

Figure 2.4: The block diagram of the discrete channel.

finite output set $\mathcal{Y}$ is independent of the previously transmitted input symbols. We can characterize this channel with the equation

$$W(y^n|x^n) = \prod\nolimits_{i=1}^{n} W(y_i|x_i) \tag{2.42}$$

for any $n \in \mathbb{Z}_{+}, y^n \in \mathcal{Y}^n, x^n \in \mathcal{X}^n$, and x_i denotes the i-th element in the sequence x^n. An encoder $\psi : \mathcal{M} \to \mathcal{X}^n$ is a function that maps the integers $m \in \mathcal{M}$, where $\mathcal{M} = \{1, \ldots, M\}$, into the strings $x^n \in \mathcal{X}^n$ of length n. Each of these strings $\psi(1), \ldots, \psi(M)$ are called the codeword, their length is called the blocklength, and $\mathcal{M}$ is called the message set containing M number of messages. A code is a collection of the M codewords, and a code ensemble is the group of the codes. A list decoder $\Theta : \mathcal{Y}^n \to \widehat{\mathcal{M}}$ is a function that maps the strings $y^n \in \mathcal{Y}^n$ into the set of messages $\mathcal{L} \in \widehat{\mathcal{M}}$, where $\widehat{\mathcal{M}} = \{\mathcal{L} : \mathcal{L} \subset \mathcal{M}\}$. Figure 2.5 shows the encoder and the decoder blocks combined with the discrete channel.

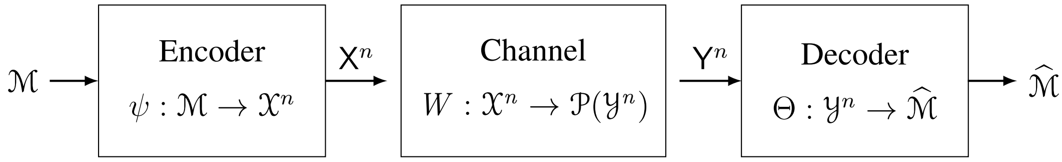

Figure 2.5: The block diagram of the discrete channel with the coding.

We can define the code rate R

$$R = \frac{\ln(M)}{n}. \tag{2.43}$$

An error occurs when the decoder decodes string y^n to the message list, which does not contain m. We can indicate the probability of error of the message m

$$\mathbf{P}_e^m = \mathbf{P}(m \notin \Theta(y^n) | x^n = \psi(m)), \tag{2.44}$$

and we can calculate the average error probability

$$\mathbf{P}_e^{av} = \frac{1}{M} \sum_{m=1}^{M} \mathbf{P}_e^m. \tag{2.45}$$

2.4 Error Exponent Funtions

In his seminal work, E. C. Shannon [27] showed that for a broad class of channels, there is a threshold rate at which we can decrease the probability of error as much as we want if we communicate at rates below a threshold. This threshold rate is called the channel capacity. However, this theorem does not say how fast the error probability decreases with increasing the block-length n. It is essential to know the relation between the decay rate of error probability and the block-length if the high-reliability constraints are the primary concerns for the application. Error exponent functions establish an exponential connection between the error probability, the code-length, and transmission rates and show how the error probability decreases by increasing the block-length. Error exponent functions fully characterize the channel for the fixed rates between the critical rate and the channel capacity. Furthermore, error exponents provide a better opportunity for more comprehensive comparisons between the channels than the channel capacity. RCE, derived by Gallager [2] in 1965, is an upper bound, and the SPE, derived by Shannon, Gallager, and Berlekamp [28] in 1967, is a lower bound for the error probability.

$$K_L(n)e^{-nE_{sp}(R)} \leq \mathbf{P}_e \leq K_U(n)e^{-nE_r(R)}, \tag{2.46}$$

where the coefficients K_L and K_U satisfy

$$\lim_{n\to\infty} \tfrac{\ln K_L(n)}{n} = 0, \tag{2.47}$$

$$\lim_{n\to\infty} \tfrac{\ln K_u(n)}{n} = 0. \tag{2.48}$$

We reproduce and explain the proof of the RCE in Appendix A.

2.4.1 Exponent Function and Augustin Information Measures

Gallager's function characterizes the SPE and the RCE; when they are first found, see [2], [28]. However, they can be characterized by the constrained Augustin capacity, see [29], [25].

$$E_{sp}(R, W, \mathcal{A}) := \sup_{\alpha\in(0,1)} \tfrac{1-\alpha}{\alpha} \left(C_{\alpha,W,\mathcal{A}} - R\right), \tag{2.49}$$

$$E_r(R, W, \mathcal{A}) := \sup_{\alpha\in(0.5,1)} \tfrac{1-\alpha}{\alpha} \left(C_{\alpha,W,\mathcal{A}} - R\right). \tag{2.50}$$

For the case when constraint set is equal to whole $\mathcal{P}(\mathcal{X})$, i.e., $\mathcal{A} = \mathcal{P}(\mathcal{X})$, we get,

$$E_{sp}(R, W) := \sup_{\alpha\in(0,1)} \tfrac{1-\alpha}{\alpha} \left(C_{\alpha,W} - R\right), \tag{2.51}$$

$$E_r(R, W) := \sup_{\alpha\in(0.5,1)} \tfrac{1-\alpha}{\alpha} \left(C_{\alpha,W} - R\right). \tag{2.52}$$

For the cost constraint case, Rényi-Gallager capacity, see [22], can be used as pointed out by Gallager in [2]. It includes an additional non-standard use of the Lagrange multiplier. The primary advantage of Augustin information measures over Gallager's functions (hence Renyi information [26] measures) is that when a cost constraint or a convex composition constraint is imposed on the codes, the resulting exponent functions can be obtained by using the constrained Augustin capacity with the same constraint in the input set. Furthermore, for the constant composition codes, we can write the parametric expression of the SPE in terms of the tilted channel by [29]. For any $R \in (\lim_{\alpha \downarrow 0} I_{\alpha}(p; W), I_1(p; W))$, there exists a unique order $\alpha \in (0, 1)$ satisfying

$$E_{sp}(R, W, p) = D_1(W_{\alpha}^{q_{\alpha,p}} \| W | p), \tag{2.53}$$

$$R = D_1(W_{\alpha}^{q_{\alpha,p}} \| q_{\alpha,p} | p), \tag{2.54}$$

$$= I_1(p; W_{\alpha}^{q_{\alpha,p}}). \tag{2.55}$$

CHAPTER 3

FADING DISCRETE MEMORYLESS CHANNELS

3.1 Introduction

In this Chapter, we explain the fading discrete memoryless channel model with CSI at the receiver. Secondly, we prove the main lemma of the thesis. Lemma shows how to decompose tilted channels of the fast-fading discrete memoryless channels using the tilted fading distribution and tilted channels of the discrete memoryless channels conditioned on fading parameters. Thirdly, we describe one particular class of channels that satisfies the requirements of the decomposition lemma. Afterward, we apply the decomposition lemma to some example channels from this class, i.e., fast-fading BSCs, BECs, and Gallager symmetric channels. Then, we calculate the resulting Augustin capacity, SPE, and RCE for any fading distribution. Lastly, we compare the results with the corresponding non-fading channels under the equal capacity condition.

3.1.1 Fading Discrete Memoryless Channels

Fading discrete channels, shown by the Figure 3.1, are described over the discrete memoryless channels with an additional fading parameter $h \in \mathcal{H}$. Like discrete channels, we define the fading discrete channel as a channel that takes an input symbol from the finite input set $\mathcal{X}$ and produces a PMF on the finite output set $\mathcal{Y}$ and the finite fading set $\mathcal{H}$. We say that the fading discrete channel is a fast-fading discrete memoryless channel if the output distribution on the finite output set $\mathcal{Y}$ is independent of the previously used input symbols, and the fading parameter is independent and identically distributed for every input letter for every use of the channel. The equation

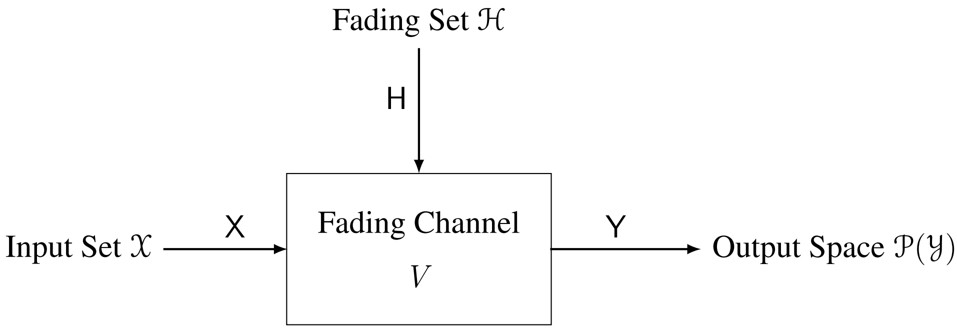

Figure 3.1: The block diagram of the fading discrete channel.

characterizes a fast fading discrete memoryless channel

$$V(h^n, y^n|x^n) = \prod_{i=1}^{n} V(h_i, y_i|x_i), \tag{3.1}$$

$$= \prod_{i=1}^{n} g(h_i|x_i)\, W(y_i|x_i, h_i). \tag{3.2}$$

where $g \in \mathcal{P}(\mathcal{H})$ is the fading distribution and for any $n \in \mathbb{Z}_+, h^n \in \mathcal{H}^n, y^n \in \mathcal{Y}^n, x^n \in \mathcal{X}^n$, and x_i denotes the i-th element in the sequence x^n. We assume that fading parameters are independent of transmitted letters x_i.

$$V(h^n, y^n|x^n) = \prod_{i=1}^{n} g(h_i)\, W(y_i|x_i, h_i) \tag{3.3}$$

If receivers knows the CSIs, resulting channels are called a fast-fading discrete memoryless channels with CSI at the receiver. We will use this model used in the thesis.

Remark 1. Notice that the discrete memoryless channels are a special case of the fast fading discrete memoryless channels, in which the fading distribution has a single fading parameter h in the $\mathcal{H}$ satisfying $g(h) = 1$. More interestingly, the fast-fading discrete memoryless channels are also discrete memoryless channels. Consider an input set $\mathcal{X}$ and a new output set $\mathcal{Z} = \mathcal{H} \times \mathcal{Y}$. $\mathcal{Z}$ is the cartesian product of the fading set ant the previous output set. Then the fast fading discrete memoryless channel $V : \mathcal{X} \to \mathcal{Z}$ is a discrete memoryless channel, and its probability distributions satisfy the equality

$$V(z|x) = g(h)\, W(y|x, h), \qquad \forall x, h, y. \tag{3.4}$$

The idea of the discrete memoryless channels and fast fading discrete memoryless channels covering each other has significant operational importance in the discussions of the following sections.

3.2 Decomposition Property of the Fast-Fading Discrete Memoryless Channels with CSI at the Receiver

By [22, Lemma 13], the unique order α Augustin mean exists for the discrete memoryless channels and by Remark 1, fast fading discrete memoryless channels are also discrete memoryless channels. Consequently, the unique order α Augustin mean exists for fast fading discrete memoryless channels, satisfying the properties of the order α Augustin mean of discrete memoryless channels. Explicitly, the order α Augustin mean of the fast fading discrete memoryless channel $s_{\alpha,p} \sim s_p$ satisfies,

$$s_{\alpha,p}(z) = \sum\nolimits_{x} p(x)\, V_{\alpha}^{s_{\alpha,p}}(z|x). \tag{3.5}$$

For the order $\alpha = 1$

$$D_1(V\| \, s| \, p) = D_1(W\| \, s_p| \, p) + D_1(s_p\| \, s)\,, \tag{3.6}$$

where

$$s_p := \sum\nolimits_{x} p(x)\, V(x), \tag{3.7}$$

since $D_1(s_p\| \, s) \geq 0$ by (2.3), infimum is achieved by s_p. Therefore,

$$s_{1,p} = s_p, \tag{3.8}$$

$$I_1(p; V) = D_1(V\| \, s_p| \, p)\,. \tag{3.9}$$

Tilted fading channel $V_{\alpha}^{s} : \mathcal{X} \to \mathcal{P}(\mathcal{Z})$ is

$$V_{\alpha}^{s}(z|x) := [\, V(z|x)]^{\alpha}[s(z)]^{1-\alpha} e^{(1-\alpha)D_{\alpha}(V(x)\|s)}. \tag{3.10}$$

Furthermore, for each realization of the fading parameter $h \in \mathcal{H}$, conditioned channels $W(\cdot \mid \cdot, h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ are a discrete memoryless channels and by [22, Lemma 13] has their own unique Augustin mean $q_{\alpha,p}(h) \sim q_p(h)$ satisfying

$$q_{\alpha,p}(y|h) = \sum\nolimits_{x} p(x)\, W_{\alpha}^{q_{\alpha,p}(h)}(y|x, h), \tag{3.11}$$

where q_p is defined as

$$q_p(h) := \sum\nolimits_{x} p(x)\, W(x, h), \tag{3.12}$$

and tilted channel $W_{\alpha}^{q(h)}(y|x,h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ is defined in (2.7) for each $h \in \mathcal{H}$. By (2.12), order one Augustin mean of the channel $W(\cdot \mid \cdot, h)$ is equal to the $q_p(h)$

$$q_{1,p}(h) = q_p(h) \tag{3.13}$$

Using (3.4), (3.7), and (3.12)

$$\begin{aligned} s_{1,p}(h,y) &= s_p(h,y) &(3.14)\\ &= \sum\nolimits_x p(x)\, V(h,y|x), &(3.15)\\ &= \sum\nolimits_x p(x) g(h)\, W(y|x,h), &(3.16)\\ &= g(h) \sum\nolimits_x p(x)\, W(y|x,h), &(3.17)\\ &= g(h) q_p(y|h) &(3.18)\\ &= g(h) q_{1,p}(y|h). &(3.19) \end{aligned}$$

We can decompose the order one Augustin mean, $s_{1,p}(h,y)$, of the fast fading discrete memoryless channel V, into the order one Augustin mean, $q_{1,p}(y|h)$, of the corresponding non-fading discrete memoryless channels $W(\cdot \mid \cdot, h)$, and the original fading distribution, $g(h)$, of the original fast fading memoryless channel V.
We ask the question of that, is there any similar relationship between the order α Augustin mean of the fast-fading discrete memoryless channel and conditioned discrete memoryless channels $W(\cdot \mid \cdot, h)$ for the orders other than one or not. We answer this question in Lemma 1.

Lemma 1. *For a given discrete channel $V : \mathcal{X} \to \mathcal{P}(\mathcal{H} \times \mathcal{Y})$ of the form* (3.4)*, input distribution $p \in \mathcal{P}(\mathcal{X})$, and order $\alpha \in \mathbb{R}_{+}$, the following two statements are equivalent*

(i) There exist $g_\alpha \in \mathcal{P}(\mathcal{H})$ satisfying

$$V_{\alpha}^{s_{\alpha,p}}(h,y|x) = g_\alpha(h)\, W_{\alpha}^{q_{\alpha,p}(h)}(y|x,h) \qquad \forall h, y, \tag{3.20}$$

and for all x s.t. $p(x) > 0$, for $V_{\alpha}^{s_{\alpha,p}}$ defined in (3.10) *and $W_{\alpha}^{q_{\alpha,p}}$ defined in* (2.7).

(ii) There exist $a_\alpha : \mathcal{X} \to \mathbb{R}$ and $b_\alpha : \mathcal{H} \to \mathbb{R}$ satisfying

$$D_\alpha(W(x,h) \| \, q_{\alpha,p}(h)) = a_\alpha(x) + b_\alpha(h) \qquad \forall h, \tag{3.21}$$

and for all x s.t. $p(x) > 0$.

Furthermore, if either statement holds, then

$$s_{\alpha,p}(h,y) = g_\alpha(h) q_{\alpha,p}(y|h). \tag{3.22}$$

Proof. If the (3.20) is true, then by taking the expectation of both sides over the input distribution p and using the Augustin operator given in (2.15),

$$\sum_x p(x)\, V_\alpha^{s_{\alpha,p}}(h,y|x) = \sum_x p(x) g_\alpha(h)\, W_\alpha^{q_{\alpha,p}(h)}(y|x,h), \tag{3.23}$$

$$= g_\alpha(h) \sum_x p(x)\, W_\alpha^{q_{\alpha,p}(h)}(y|x,h), \tag{3.24}$$

$$s_{\alpha,p}(h,y) = g_\alpha(h) q_{\alpha,p}(y|h). \tag{3.25}$$

Therefore, statement (i) implies the decomposition of the Augustin mean of the fast fading discrete memoryless channel.

If we insert (3.4), (3.22), and the statement (i) into (3.10) and use (2.7),

$$g_\alpha(h)\, W_\alpha^{q_{\alpha,p}(h)}(y|x,h) = [g(h)\, W(y|x,h)]^\alpha [g_\alpha(h) q_{\alpha,p}(y|h)]^{1-\alpha} e^{(1-\alpha) D_\alpha(V(x)\| s_{\alpha,p})} \tag{3.26}$$

$$\frac{W_\alpha^{q_{\alpha,p}(h)}(y|x,h)}{[W(y|x,h)]^\alpha [q_{\alpha,p}(y|h)]^{1-\alpha}} = \frac{[g(h)]^\alpha}{[g_\alpha(h)]^\alpha} e^{(1-\alpha) D_\alpha(V(x)\| s_{\alpha,p})} \tag{3.27}$$

$$e^{(1-\alpha) D_\alpha(W(x)\| q_{\alpha,p})} = \frac{[g(h)]^\alpha}{[g_\alpha(h)]^\alpha} e^{(1-\alpha) D_\alpha(V(x)\| s_{\alpha,p})} \tag{3.28}$$

$$D_\alpha(W(x)\|\, q_{\alpha,p}) = \tfrac{\alpha}{1-\alpha} \ln \tfrac{g(h)}{g_\alpha(h)} + D_\alpha(V(x)\|\, s_{\alpha,p}) \tag{3.29}$$

$$= \tfrac{\alpha}{1-\alpha} \ln \tfrac{g(h)}{g_\alpha(h)} + D_\alpha(V(x)\|\, g_\alpha q_{\alpha,p}) \tag{3.30}$$

Then, for the expressions

$$a_\alpha(x) = D_\alpha(V(x)\|\, g_\alpha q_{\alpha,p}), \tag{3.31}$$

$$b_\alpha(h) = \tfrac{\alpha}{\alpha-1} \ln \tfrac{g_\alpha(h)}{g(h)}. \tag{3.32}$$

statement (i) implies statement (ii).

Let's assume that the statement (ii) holds and choose $s(h,y)$ to be

$$s(h,y) := \frac{g(h) e^{\frac{\alpha-1}{\alpha} b_\alpha(h)}}{\sum_{\widetilde{h}} g(\widetilde{h}) e^{\frac{\alpha-1}{\alpha} b_\alpha(\widetilde{h})}} q_{\alpha,p}(y|h). \tag{3.33}$$

Then, we can write (3.10) as

$$
\begin{aligned}
V_\alpha^s(z|x) &= [V(h,y|x)]^\alpha [s(z)]^{1-\alpha} e^{(1-\alpha)D_\alpha(V(x)\|s)} && (3.34)\\
&= [g(h)\,W(y|x,h)]^\alpha \left[\frac{g(h)e^{\frac{\alpha-1}{\alpha}b_\alpha(h)}}{\sum_{\widetilde{h}} g(\widetilde{h})e^{\frac{\alpha-1}{\alpha}b_\alpha(\widetilde{h})}} q_{\alpha,p}(y|h)\right]^{1-\alpha} e^{(1-\alpha)a_\alpha(x)} && (3.35)\\
&= [g(h)]^\alpha \left[\frac{g(h)e^{\frac{\alpha-1}{\alpha}b_\alpha(h)}}{\sum_{\widetilde{h}} g(\widetilde{h})e^{\frac{\alpha-1}{\alpha}b_\alpha(\widetilde{h})}}\right]^{1-\alpha} \frac{1}{e^{(1-\alpha)b_\alpha(h)}}\, W_\alpha^{q_{\alpha,p}(h)}(y|x,h) && (3.36)\\
&= \frac{g(h)e^{\frac{\alpha-1}{\alpha}b_\alpha(h)}}{\sum_{\widetilde{h}} g(\widetilde{h})e^{\frac{\alpha-1}{\alpha}b_\alpha(\widetilde{h})}}\, W_\alpha^{q_{\alpha,p}(h)}(y|x,h) && (3.37)
\end{aligned}
$$

In addition, s satisfies the fixed point property and $s \sim s_p$. For this reason, s is the Augustin mean. Then the statement (ii) holds for the tilted fading distribution

$$
g_\alpha(h) = \left(\sum_{\widetilde{h}} g(\widetilde{h})e^{\frac{\alpha-1}{\alpha}b_\alpha(\widetilde{h})}\right)^{-1} g(h)e^{\frac{\alpha-1}{\alpha}b_\alpha(h)} \tag{3.38}
$$

□

Remark 2. Lemma 1 is important for the fading channels because the only thing that is necessary to understand if the fast fading discrete memoryless channels have the decomposition property or not is looking at the Rényi divergence between the discrete memoryless channels $W(\cdot|\cdot,h)$ and their order α Augustin means $q_{\alpha,p}(h)$ for every fading parameter h by (3.32). One can ask if the decomposition of the Augustin mean of the fading DMC V implies the decomposition of the tilted channel. In our opinion, it is not valid. However, it is still an open problem to solve.

For the channels, which satisfy the decomposition lemma, we can write parametric expressions of SPE and the rate as,

$$
\begin{aligned}
E_{sp}(R, V, p) =& D_1(V_\alpha^{s_{\alpha,p}} \| V | p), & (3.39)\\
=& \sum_x p(x) \left[\sum_{y,h} V_\alpha^{s_{\alpha,p}}(h, y|x) \ln \frac{V_\alpha^{s_{\alpha,p}}(h,y|x)}{V(h,y|x)} \right] & (3.40)\\
=& \sum_x p(x) \left[\sum_{y,h} g_\alpha(h) W_\alpha^{q_{\alpha,p}(h)}(y|x,h) \ln \frac{g_\alpha(h) W_\alpha^{q_{\alpha,p}(h)}(y|x,h)}{g(h) W(y|x,h)} \right] & (3.41)\\
=& \sum_x p(x) \left[\sum_{y,h} g_\alpha(h) W_\alpha^{q_{\alpha,p}(h)}(y|x,h) \frac{g_\alpha(h)}{g(h)} \right] &\\
&+ \sum_x p(x) \left[\sum_{y,h} g_\alpha(h) W_\alpha^{q_{\alpha,p}(h)}(y|x,h) \frac{W_\alpha^{q_{\alpha,p}(h)}(y|x,h)}{W(y|x,h)} \right] & (3.42)\\
=& D_1(g_\alpha \| g) + D_1\left(W_\alpha^{q_{\alpha,p}(h)} \| W \middle| p g_\alpha\right), & (3.43)\\
R =& D_1(V_\alpha^{s_{\alpha,p}} \| s_{\alpha,p} | p) & (3.44)\\
=& \sum_x p(x) \left[\sum_{y,h} V_\alpha^{s_{\alpha,p}}(h, y|x) \ln \frac{V_\alpha^{s_{\alpha,p}}(h,y|x)}{s_{\alpha,p}(h,y)} \right] & (3.45)\\
=& \sum_x p(x) \left[\sum_{y,h} g_\alpha(h) W_\alpha^{q_{\alpha,p}(h)}(y|x,h) \ln \frac{g_\alpha(h) W_\alpha^{q_{\alpha,p}(h)}(y|x,h)}{g_\alpha(h) q_{\alpha,p}(y|h)} \right] & (3.46)\\
=& D_1\left(W_\alpha^{q_{\alpha,p}(h)} \| q_{\alpha,p} \middle| p g_\alpha\right). & (3.47)
\end{aligned}
$$

3.2.1 Fast Fading DMCs with Common Capacity Achieving Input Distribution

Lemma 1 is not applicable for all fast-fading discrete memoryless channels with CSI at the receiver and all the input probability distributions. However, for a special class of the channels, i.e., fast-fading discrete channels with CSI at the receiver with the common Augustin capacity-achieving input distribution for the channels $W(\cdot, h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ for all h, Lemma 1 holds. Augustin capacity of the channel $W(\cdot, h)$ will be denoted as $C_{\alpha,h}$ instead of $C_{\alpha, W(\cdot,h)}$ to make the notation more easy to follow. This subsection will prove that the channels in this class satisfy the Lemma 1, and the resulting Augustin mean, Augustin capacity, and error exponent functions are derived.

For channels in this class, there exist a $p \in \mathcal{P}(\mathcal{X})$,

$$
I_\alpha(p; W(\cdot, h)) = C_{\alpha,h} \qquad \forall h : g(h) > 0. \qquad (3.48)
$$

Necessary condition to (3.48) be true is the Augustin mean that corresponding to the p is the Augustin center of the channel $W(\cdot, h)$ by [22, Theorem 1]. Consequently, it also implies

$$D_\alpha(W(x,h)\| \, q_{\alpha,p}(h)) = C_{\alpha,h}, \tag{3.49}$$

for all x s.t. $p(x) > 0$, and h s.t. $g(h) > 0$. Because if the L.H.S of the (3.49) is strictly less than $C_{\alpha,h}$ for at least one $\{x : p(x) > 0\}$, by (2.14)

$$\begin{aligned} I_\alpha(p; W(\cdot,h)) &= D_\alpha(W(x,h)\| \, q_{\alpha,p}(h)| \, p) && (3.50)\\ &= \sum_x p(x) D_\alpha(W(x,h)\| \, q_{\alpha,p}(h)) && (3.51)\\ &< C_{\alpha,h}. && (3.52) \end{aligned}$$

which is a contradiction with the (3.48). Thus (3.48) holds and the channels in this class satisfy the Lemma (1) through (3.21) where

$$\begin{aligned} b_\alpha(h) &= C_{\alpha,h}, && (3.53)\\ a_\alpha(x) &= 0. && (3.54) \end{aligned}$$

Tilted fading distribution can be calculated by using (3.38) as

$$g_\alpha(h) = \left(\sum_{\widetilde{h}} g(\widetilde{h}) e^{\frac{\alpha-1}{\alpha} C_{\alpha,\widetilde{h}}} \right)^{-1} g(h) e^{\frac{\alpha-1}{\alpha} C_{\alpha,h}}. \tag{3.55}$$

Derivations until here are enough to satisfy the requirements of the Lemma 1 for this class of channels. However, we can show that the p is also the Augustin capacity-achieving input distribution for channel V. We show that the Augustin information between the channel V and input distribution p is equal to the maximum value of the conditional Rényi divergence between the channel V and the Augustin mean $s_{\alpha,p}$ for any input distribution $p' \in \mathcal{P}(\mathcal{X})$. Using (2.14), (2.6), (3.22), the Augustin information between the channel V and the input distribution p is

$$
\begin{aligned}
I_\alpha(p;V) &= D_\alpha(V(x)\| \, s_{\alpha,p} | \, p) \,, && (3.56)\\
&= \sum_x p(x) \tfrac{1}{\alpha-1} \ln \sum_{h,y} \left[V(h,y|x) \right]^\alpha \left[s_{\alpha,p}(h,y) \right]^{1-\alpha} , && (3.57)\\
&= \sum_x p(x) \tfrac{1}{\alpha-1} \ln \sum_{h,y} \left[g(h) \, W(y|x,h) \right]^\alpha \left[g_\alpha(h) q_{\alpha,p}(y|h) \right]^{1-\alpha} , && (3.58)\\
&= \sum_x p(x) \tfrac{1}{\alpha-1} \ln \sum_{h} \left[g(h) \right]^\alpha \left[g_\alpha(h) \right]^{1-\alpha} e^{(\alpha-1) C_{\alpha,h}}, && (3.59)\\
&= \sum_x p(x) \tfrac{1}{\alpha-1} \ln \sum_{h} \left(\sum_{\widetilde{h}} g(\widetilde{h}) e^{\frac{\alpha-1}{\alpha} C_{\alpha,\widetilde{h}}} \right)^{\alpha-1} g(h) e^{\frac{\alpha-1}{\alpha} C_{\alpha,h}}, && (3.60)\\
&= \tfrac{\alpha}{\alpha-1} \ln \sum_h g(h) e^{\frac{\alpha-1}{\alpha} C_{\alpha,h}}, && (3.61)
\end{aligned}
$$

where

$$
\sum_y \left[W(y|x,h) \right]^\alpha \left[q_{\alpha,p}(y|h) \right]^{1-\alpha} = e^{(\alpha-1) C_{\alpha,h}} \qquad (3.62)
$$

by (3.49). However, (3.62) is valid for the values of x s.t. $p(x) > 0$. For other values of x,

$$
\sum_y \left[W(y|x,h) \right]^\alpha \left[q_{\alpha,p}(y|h) \right]^{1-\alpha} \geq e^{(\alpha-1) C_{\alpha,h}} \qquad (3.63)
$$

since $C_{\alpha,h}$ is the Augustin capacity. For this reason, the conditional Rényi divergence between the channel V and the Augustin mean $s_{\alpha,p}$ for any $p' \in \mathcal{P}(\mathcal{X})$

$$
D_\alpha(V(x)\| \, s_{\alpha,p} | \, p') \leq \tfrac{\alpha}{\alpha-1} \ln \sum_h g(h) e^{\frac{\alpha-1}{\alpha} C_{\alpha,h}}. \qquad (3.64)
$$

Thus, the equality in (3.64) happens when $p' = p$, and it indicates that the input distribution p is the Augustin capacity-achieving input distribution for the channel V

$$
I_\alpha(p;V) = C_{\alpha,V}. \qquad (3.65)
$$

Another result of the input distribution p being the Augustin capacity-achieving input distribution of the fast fading discrete memoryless channel V is,

$$
E_{sp}(R, V) = E_{sp}(R, V, p) \,. \qquad (3.66)
$$

if (3.48) holds for every $\alpha \in (0,1)$. We drop the input distribution p because the SPE corresponding to the input distribution p becomes the SPE of the channel due to optimal, i.e., Augustin capacity-achieving, input distribution p and by the parametric expression of the error exponent functions (3.39), and (3.44),

$$E_{sp}(R, V) = D_1(g_\alpha \| g) + D_1(W_{\alpha,p} \| W | p g_\alpha), \qquad \alpha \in (0,1) \tag{3.67}$$

$$R = D_1(W_{\alpha,p} \| q_{\alpha,p} | p g_\alpha). \tag{3.68}$$

For the tilted fading distribution g_α given in (3.55).

3.3 Examples

In this section, we applied the Lemma 1 to the fast-fading BSC and the fast-fading BEC because these channels are type of channels described in Section 3.2.1. We determine the tilted channel and the tilted fading distribution for these channels. Augustin capacity and the parametric expression of the sphere packing exponent function are derived. Then, these channels' sphere packing exponents are analyzed and compared with the non-fading binary symmetric channel and non-fading binary erasure channels under the equal Augustin capacity condition. Lastly, we expand our examples to the Fading DMC's with Gallager Symmetric channels.

3.3.1 Fading Binary Erasure Channel (FBEC)

FBEC $V : \mathcal{X} \rightarrow \mathcal{P}(\mathcal{H}, \mathcal{Y})$ is a fading channel satisfying (3.4) for $\mathcal{X} = \{0,1\}, \mathcal{Y} = \{0,1,e\}$ and $\mathcal{H}$, which is the finite subset of the interval $[0,1]$. For a given input symbol x, and the fading parameter h, the distribution of the output symbol y is equal to

$$W(y|x,h) = \begin{cases} 1-h & \text{if } y = x, \\ h & \text{if } y = e, \\ 0 & \text{else,} \end{cases} \tag{3.69}$$

Figure 3.2 shows the fading binary erasure channel diagram.

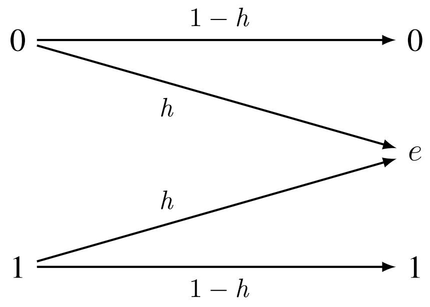

Figure 3.2: FBEC

Now, let p be the uniform distribution on $\mathcal{X}$ and $q_\alpha(\cdot|h)$ be

$$q_\alpha(y|h) = \begin{cases} \dfrac{1-h}{2-2h+2^{1/\alpha}h} & \text{if } y \in \{0,1\} \\ \dfrac{2^{1/\alpha}h}{2-2h+2^{1/\alpha}h} & \text{if } y = e \end{cases}. \tag{3.70}$$

on $\mathcal{Y}$ for all $h \in [0,1]$. Then, by (2.8), the order α tilted channel of the channel $W(\cdot,h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ for $q_\alpha(\cdot|h)$ is

$$W_\alpha^{q_\alpha(h)}(y|x,h) = \frac{[W(y|x,h)]^\alpha [q_\alpha(y|h)]^{1-\alpha}}{\sum_{\widetilde{y}} [W(\widetilde{y}|x,h)]^\alpha [q_\alpha(\widetilde{y}|h)]^{1-\alpha}} \tag{3.71}$$

$$= \begin{cases} \dfrac{2-2h}{2-2h+2^{1/\alpha}h} & \text{if } y = x, \\ \dfrac{2^{1/\alpha}h}{2-2h+2^{1/\alpha}h} & \text{if } y = e, \\ 0 & \text{else,} \end{cases} \tag{3.72}$$

If we take the expectation of both sides over the p, we get the Augustin operator of $q_\alpha(\cdot|h)$ equal to

$$\mathtt{T}_{\alpha,p}(q_\alpha(y|h)) = \sum_x p(x) W_\alpha^{q_\alpha(h)}(y|x,h)$$
$$= \begin{cases} \dfrac{1-h}{2-2h+2^{1/\alpha}h} & \text{if } y \in \{0,1\} \\ \dfrac{2^{1/\alpha}h}{2-2h+2^{1/\alpha}h} & \text{if } y = e \end{cases},$$

which equals $q_\alpha(\cdot|h)$.

$$\mathtt{T}_{\alpha,p}(q_\alpha(y|h)) = q_\alpha(y|h), \tag{3.73}$$

Thus, q_α satisfies the fixed point property. Moreover, $q_\alpha \prec q_p$, where q_p is defined in (3.12). Therefore, q_α is the order α Augustin mean of the channel $W(\cdot,h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ for the input distribution p for any fading parameter $h \in [0,1]$, by [22, Lemma 13].

$$q_\alpha(h) = q_{\alpha,p}(\cdot \mid h) \tag{3.74}$$

$$I_\alpha(p; W(\cdot,h)) = D_\alpha(W(\cdot,h) \| \, q_\alpha(h) | \, p) \,. \tag{3.75}$$

Moreover, if we calculate the Rényi divergence between the distribution $W(x,h)$ and the distribution $q_{\alpha,p}(h)$, we get

$$D_\alpha(W(x,h)\| \, q_{\alpha,p}(h)) = \begin{cases} \frac{\alpha}{\alpha-1}\ln\left[h + 2^{\frac{\alpha-1}{\alpha}}(1-h)\right] & \text{if } \alpha \neq 1, \\ (1-h)\ln 2, & \text{if } \alpha = 1, \end{cases} \tag{3.76}$$

Thus, $D_\alpha(W(x,h)\| \, q_{\alpha,p}(h))$ is independent of x. Then, for all $x \in \mathcal{X}$, Augustin information for the input distribution p is equal to $D_\alpha(W(x,h)\| \, q_{\alpha,p}(h))$

$$I_\alpha(p; W(\cdot,h)) = \sum_x p(x) D_\alpha(W(x,h)\| \, q_{\alpha,p}(h)) \,, \tag{3.77}$$

$$= D_\alpha(W(x,h)\| \, q_{\alpha,p}(h)) \qquad \forall x \in \mathcal{X}. \tag{3.78}$$

Therefore, $I_\alpha(p; W(\cdot,h)) = C_{\alpha,h}$, and (3.49) hold. Consequently,

$$C_{\alpha,h} = \begin{cases} \frac{\alpha}{\alpha-1}\ln\left[h + 2^{\frac{\alpha-1}{\alpha}}(1-h)\right] & \text{if } \alpha \neq 1, \\ (1-h)\ln 2, & \text{if } \alpha = 1, \end{cases} \tag{3.79}$$

$$C_{\alpha,V} = \begin{cases} \frac{\alpha}{\alpha-1}\ln\left[\bar{h} + 2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})\right] & \text{if } \alpha \neq 1, \\ \left(1-\bar{h}\right)\ln 2, & \text{if } \alpha = 1, \end{cases} \tag{3.80}$$

$$g_\alpha(h) = \frac{g(h)}{\bar{h} + 2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})}\left[h + 2^{\frac{\alpha-1}{\alpha}}(1-h)\right], \tag{3.81}$$

where

$$\bar{h} = \sum_h g(h)h. \tag{3.82}$$

Remark 3. Tilted channel of the fading binary erasure channel using the order α Augustin center, i.e., $q_{\alpha,h}$ is also a fading binary erasure channel with the erasure probability

$$W(e|x,h) = \frac{2^{1/\alpha}h}{2 - 2h + 2^{1/\alpha}h}. \tag{3.83}$$

Inserting (3.81) in (3.67), and (3.68), parametric expressions of the SPE and the rate are

$$E_{sp}(R, V) = \ln \frac{1}{\bar{h}+2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})} + \frac{\alpha-1}{\alpha} \frac{2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})}{\bar{h}+2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})} \ln 2, \tag{3.84}$$

$$R = \frac{2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})}{\bar{h}+2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})} \ln 2. \tag{3.85}$$

Lastly, we compare the sphere packing exponent of the non-fading binary erasure channels and the fading binary erasure channels to understand the fading effect on the sphere packing exponent. However, to make a fair comparison, we include the equal channel capacity condition. By comparing the two different fading channels having the same channel capacity, the difference in the sphere packing exponent of the channels only originates from the fading effect. We derive the following lemma.

Lemma 2. *For a fading binary erasure channel V and a non-fading binary erasure channel W, described in Section 3.3.1,*

$$E_{sp}(V, R) = E_{sp}(W, R) \tag{3.86}$$

if and only if

$$C_{\alpha,V} = C_{\alpha,W} \tag{3.87}$$

Proof. By (3.80),

$$C_{\alpha,V} = \begin{cases} \frac{\alpha}{\alpha-1} \ln \left[\bar{h} + 2^{\frac{\alpha-1}{\alpha}}(1-\bar{h})\right] & \text{if } \alpha \neq 1 \\ \left(1-\bar{h}\right) \ln 2 & \text{if } \alpha = 1 \end{cases}, \tag{3.88}$$

where

$$\bar{h} = \sum_{\widetilde{h}} g(\widetilde{h})\widetilde{h} \tag{3.89}$$

where g is the fading distribution of the fading binary erasure channels V. Similarly,

$$C_{\alpha,W} = \begin{cases} \frac{\alpha}{\alpha-1} \ln \left[h + 2^{\frac{\alpha-1}{\alpha}}(1-h)\right] & \text{if } \alpha \neq 1 \\ (1-h) \ln 2 & \text{if } \alpha = 1 \end{cases}, \tag{3.90}$$

Therefore, the expected value of the fading distribution g and the erasure probability of the non-fading BEC must be the same for equal channel capacity conditions, and we get

$$\bar{h} = h. \tag{3.91}$$

Because the expected value of the fading parameter of g and the erasure probability h of the non-fading BEC are the same, by (3.84) and (3.85), their sphere packing exponent functions are the same for all rates.

$$E_{sp}(V, R) = E_{sp}(W, R) \tag{3.92}$$

Furthermore, the expected value of the fading parameter of g and the erasure probability $\bar{h}$ of the non-fading binary erasure channel must be the same as the SPEs to be same with each other. Then, their capacities are the same. □

Interestingly, fading does not affect SPEs of the FBECs under the same channel capacity condition, and fading binary erasure channels have the same sphere packing exponents as the non-fading binary erasure channels under the same channel capacity condition. Recall that the binary erasure channel has the largest sphere packing exponent among all the discrete memoryless channels with the same channel capacity by [30–32]. Therefore, we know that the SPE of FBEC cannot be greater than the SPE of the non-fading BEC with equal channel capacity; however, all discussions in terms of the fading binary erasure channels and the Lemma 2 are new.

3.3.2 Fading Binary Symmetric Channel (FBSC)

FBSC $V : \mathcal{X} \to \mathcal{P}(\mathcal{H}, \mathcal{Y})$ is a fading channel satisfying (3.4) for $\mathcal{X} = \{0, 1\}, \mathcal{Y} = \{0, 1\}$ and $\mathcal{H}$, which is the finite subset of the interval $[0, 0.5]$. For a given input symbol x, and the fading parameter h, the distribution of the output symbol y is equal to

$$W(y|x, h) = \begin{cases} 1 - h & \text{if } y = x \\ h & \text{if } y \neq x \end{cases}. \tag{3.93}$$

Figure 3.3 shows the fading binary symmetric channel diagram.

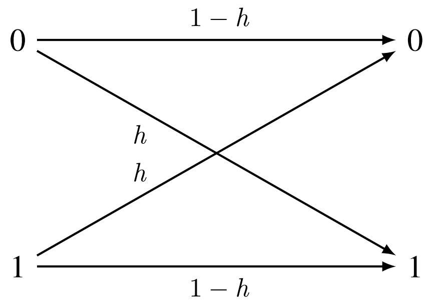

Figure 3.3: FBSC

Now, let p be the uniform distribution on $\mathcal{X}$ and $q_\alpha(\cdot|h)$ be the uniform distribution on $\mathcal{Y}$ for all $h \in [0, 0.5]$. Then, by (2.8), the order α tilted channel of the channel $W(\cdot, h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ for $q_\alpha(\cdot|h)$ is

$$W_\alpha^{q_\alpha(h)}(y|x,h) = \frac{[W(y|x,h)]^\alpha [q_\alpha(y|h)]^{1-\alpha}}{\sum_{\widetilde{y}} [W(\widetilde{y}|x,h)]^\alpha [q_\alpha(\widetilde{y}|h)]^{1-\alpha}} \tag{3.94}$$

$$= \begin{cases} \dfrac{(1-h)^\alpha}{h^\alpha + (1-h)^\alpha} & \text{if } y = x, \\ \dfrac{h^\alpha}{h^\alpha + (1-h)^\alpha} & \text{if } y \neq x. \end{cases} \tag{3.95}$$

If we take the expectation of both sides over the p, we get the Augustin operator of q_α equal to

$$\mathtt{T}_{\alpha,p}\left(q_\alpha(y|h)\right) = \sum_x p(x)\, W_\alpha^{q_\alpha(h)}(y|x,h)$$

$$= \begin{cases} \frac{1}{2} & \text{if } y = 0, \\ \frac{1}{2} & \text{if } y = 1. \end{cases}$$

which is the uniform distribution on the $\mathcal{Y}$.

$$\mathtt{T}_{\alpha,p}\left(q_\alpha(y|h)\right) = q_\alpha(y|h) \tag{3.96}$$

Thus, q_α satisfies the fixed point property. Moreover, $q_\alpha \prec q_p$, where q_p is defined in (3.12). Therefore, q_α is the Augustin mean of the channel $W(\cdot, h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ for the input distribution p for any fading parameter $h \in [0, 0.5]$, by [22, Lemma 13].

$$q_\alpha(h) = q_{\alpha,p}(\cdot \mid h) \tag{3.97}$$

$$I_\alpha(p; W(\cdot, h)) = D_\alpha(W(\cdot, h) \| \, q_\alpha(h) | \, p)\,. \tag{3.98}$$

Furthermore, if we calculate the Rényi divergence between the distribution $W(x,h)$ and the distribution $q_{\alpha,p}(h)$, we get

$$D_\alpha(W(x,h)\|\, q_{\alpha,p}(h)) = \begin{cases} \frac{1}{\alpha-1}\ln\left[h^\alpha\left(\frac{1}{2}\right)^{1-\alpha}+(1-h)^\alpha\left(\frac{1}{2}\right)^{1-\alpha}\right] & \text{if } \alpha\in\mathbb{R}_+\setminus\{1\} \\ h\ln\frac{h}{0.5}+(1-h)\ln\frac{1-h}{0.5} & \text{if } \alpha=1 \end{cases} \tag{3.99}$$

Thus, $D_\alpha(W(x,h)\|\, q_{\alpha,p}(h))$ is independent of x. Then, for all $x\in\mathcal{X}$, order α Augustin information for the input distribution p is equal to $D_\alpha(W(x,h)\|\, q_{\alpha,p}(h))$,

$$I_\alpha(p;W(\cdot,h)) = \sum_x p(x)D_\alpha(W(x,h)\|\, q_{\alpha,p}(h))\,, \tag{3.100}$$

$$= D_\alpha(W(x,h)\|\, q_{\alpha,p}(h)) \qquad \forall x\in\mathcal{X}. \tag{3.101}$$

Therefore, $I_\alpha(p;W(\cdot,h)) = C_{\alpha,h}$ and (3.49) hold. Consequently,

$$C_{\alpha,h} = d_\alpha(h\|\, 1/2) \tag{3.102}$$

$$C_{\alpha,V} = \begin{cases} \frac{\alpha}{\alpha-1}\ln\sum_h g(h)e^{\frac{\alpha}{\alpha-1}d_\alpha(h\|1/2)} & \text{if } \alpha\neq 1 \\ \sum_h g(h)d_1(h\|\, 1/2) & \text{if } \alpha=1 \end{cases} \tag{3.103}$$

$$g_\alpha(h) = \frac{g(h)(h^\alpha+(1-h)^\alpha)^{\frac{1}{\alpha}}}{\sum_{\widetilde{h}} g(\widetilde{h})(\widetilde{h}^\alpha+(1-\widetilde{h})^\alpha)^{\frac{1}{\alpha}}} \tag{3.104}$$

where $d_\alpha(\cdot\|\,\cdot):[0,1]\times[0,1]\to[0,\infty]$ is defined as

$$d_\alpha(\varepsilon\|\,\tau) := \begin{cases} \frac{\ln\left(\varepsilon^\alpha\tau^{1-\alpha}+(1-\varepsilon)^\alpha(1-\tau)^{1-\alpha}\right)}{\alpha-1} & \text{if } \alpha\neq 1 \\ \varepsilon\ln\frac{\varepsilon}{\tau}+(1-\varepsilon)\ln\frac{1-\varepsilon}{1-\tau} & \text{if } \alpha=1 \end{cases}. \tag{3.105}$$

Remark 4. Tilted channel of FBSCs corresponding to the order α Augustin center, i.e., $q_{\alpha,h}$ is also a FBSC with the cross-over probability

$$W(y\neq x|x,h) = \frac{(1-h)^\alpha}{h^\alpha+(1-h)^\alpha}. \tag{3.106}$$

Inserting (3.104) in (3.67), and (3.68), we get the parametric expressions of the SPE and the rate.

$$E_{sp}(R,V) = D_1(g_\alpha\|\, g)+\sum_h g_\alpha(h)d_1\left(\frac{h^\alpha}{h^\alpha+(1-h)^\alpha}\middle\|\, h\right), \tag{3.107}$$

$$R = \sum_h g_\alpha(h)d_1\left(\frac{h^\alpha}{h^\alpha+(1-h)^\alpha}\middle\|\, \frac{1}{2}\right), \tag{3.108}$$

where g_α is given in (3.104). Lastly, we proved the following lemma to determine the fading effect on the SPE of the FBSC.

Lemma 3. *For a FBSC V, described in Section 3.3.2, the fading distribution that maximizes the SPE, among all fading distributions which result in the same channel capacity, is*

$$g(h) = \begin{cases} p, & \textit{if } h = 0, \\ (1-p), & \textit{if } h = \frac{1}{2}, \end{cases} \tag{3.109}$$

.

Proof. To prove the Lemma 3, we first show that the channel capacity of the V is same with the channel capacity of the FBEC V_1 described in the Subsection 3.3.2 with the fading distribution g_1 which is

$$g_1(h) = \begin{cases} p, & \text{if } h = 0, \\ (1-p), & \text{if } h = 1, \end{cases} \tag{3.110}$$

and their sphere packing error exponents are the same. By (3.102),

$$C_{\alpha,V} = \ln 2(1-p), \tag{3.111}$$

and by (3.80),

$$C_{\alpha,V_1} = \ln 2(1-p). \tag{3.112}$$

. Thus, their channel capacities are the same. Then, by (3.104) the tilted fading distribution of the V is

$$g_\alpha(h) = \begin{cases} \frac{p}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}}, & \text{if } h = 0, \\ \frac{(1-p)2^{\frac{1-\alpha}{\alpha}}}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}}, & \text{if } h = \frac{1}{2}. \end{cases} \tag{3.113}$$

Furthermore, for $h = 0$ and $h = \frac{1}{2}$, we get

$$d_1\left(\tfrac{h^\alpha}{h^\alpha+(1-h)^\alpha} \middle\| h\right) = 0. \tag{3.114}$$

Therefore,

$$\sum_h g_\alpha(h) d_1\left(\tfrac{h^\alpha}{h^\alpha+(1-h)^\alpha} \middle\| h\right) = 0, \tag{3.115}$$

for the fading distribution g. Thus by (3.107),

$$E_{sp}(V,R) = D_1(g_\alpha \| g) \tag{3.116}$$

$$= \ln \frac{1}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}} + \frac{1-\alpha}{\alpha} \frac{(1-p)2^{\frac{1-\alpha}{\alpha}}}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}} \ln 2 \tag{3.117}$$

For the fading binary erasure channel V_1, the expected value of the fading parameter h is

$$\overline{h} = \sum_h g_1(h)h \tag{3.118}$$

$$= 1-p \tag{3.119}$$

Thus, by (3.84) the SPE of the channel FBEC V_1 is

$$E_{sp}(V_1,R) = \ln \frac{1}{(1-p)+p2^{\frac{\alpha-1}{\alpha}}} + \frac{\alpha-1}{\alpha} \frac{2^{\frac{\alpha-1}{\alpha}}p}{(1-p)+p2^{\frac{\alpha-1}{\alpha}}} \ln 2 \tag{3.120}$$

$$= \ln \frac{2^{\frac{1-\alpha}{\alpha}}}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}} + \frac{\alpha-1}{\alpha} \frac{2^{\frac{\alpha-1}{\alpha}}p}{(1-p)+p2^{\frac{\alpha-1}{\alpha}}} \ln 2 \tag{3.121}$$

$$= \ln \frac{1}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}} + \frac{1-\alpha}{\alpha} \left(1 - \frac{2^{\frac{\alpha-1}{\alpha}}p}{(1-p)+p2^{\frac{\alpha-1}{\alpha}}}\right) \ln 2 \tag{3.122}$$

$$= \ln \frac{1}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}} + \frac{1-\alpha}{\alpha} \frac{(1-p)2^{\frac{1-\alpha}{\alpha}}}{p+(1-p)2^{\frac{1-\alpha}{\alpha}}} \ln 2 \tag{3.123}$$

$$= E_{sp}(V,R) \tag{3.124}$$

Therefore their sphere packing exponents are the same. We know that the SPE of the FBECl V_1 is equal to the SPE of the non-fading BEC with the same channel capacity by Lemma 2 and the non-fading BEC has the greatest SPE among all the binary input DMCs by [30–32]. Therefore the g maximizes the SPE of the FBSC among the FBSCs with the same Augustin capacity. □

The intuition behind the Lemma 3 is the following. The fading binary symmetric channel V with the fading distribution g given in the Lemma 3 can be thought as the combination of the two non-fading binary symmetric channel $W_1 : \{0,1\} \to \{0,1\}$ and $W_2 : \{0,1\} \to \{0,1\}$ which the channel transition probabilities are

$$W_1(y \mid x) = \begin{cases} 1 & \text{if } y = x, \\ 0 & \text{if } y \neq x, \end{cases} \tag{3.125}$$

and

$$W_2(y \mid x) = \begin{cases} \frac{1}{2} & \text{if } y = x, \\ \frac{1}{2} & \text{if } y \neq x. \end{cases} \tag{3.126}$$

and the fading binary symmetric channel V switches between the non-fading binary symmetric channels W_1 and W_2 according to the fading distribution. The fading binary erasure channel V_1 with the fading distribution g_1 given in the Lemma 3 can be thought as the combination of the two non-fading binary erasure channel $W_3 : \{0,1\} \to \{0,e,1\}$ and $W_4 : \{0,1\} \to \{0,e,1\}$ which the channel transition probabilities are

$$W_3(y \mid x) = \begin{cases} 1 & \text{if } y = x, \\ 0 & \text{if } y = e, \\ 0 & \text{else}, \end{cases} \tag{3.127}$$

and

$$W_4(y \mid x) = \begin{cases} 0 & \text{if } y = x, \\ 1 & \text{if } y = e, \\ 0 & \text{else}, \end{cases} \tag{3.128}$$

and the fading binary erasure channel V_1 switches between the non-fading binary erasure channels W_3 and W_4 according to the fading distribution. The non-fading BSC W_1 and the non-fading BEC W_3 are the same channels. In addition, the non-fading binary symmetric channel W_2 can not be distinguished from the non-fading binary symmetric channel W_4 combined with the coin flip. Their contributions to channel capacity and error exponent are zero. Therefore the fading channels V and the V_1 defined in the Lemma 3 should have the same SPEs. After proving the Lemma 3, we calculated the SPE function of the non-fading BEC with the cross-over probability equal to 0.1 and the FBSC with g given in (3.109) with the same capacity of the non-fading channel. Figure 3.4 represents the resulting sphere packing exponents.

The SPE of the FBSC is much greater than the SPE function of the non-fading BSC for every rate lower than the channel capacity. Lastly, we calculate ratio of SPE of non-fading BSC with the cross-over probability equal to 0.1 and the FBSCs having

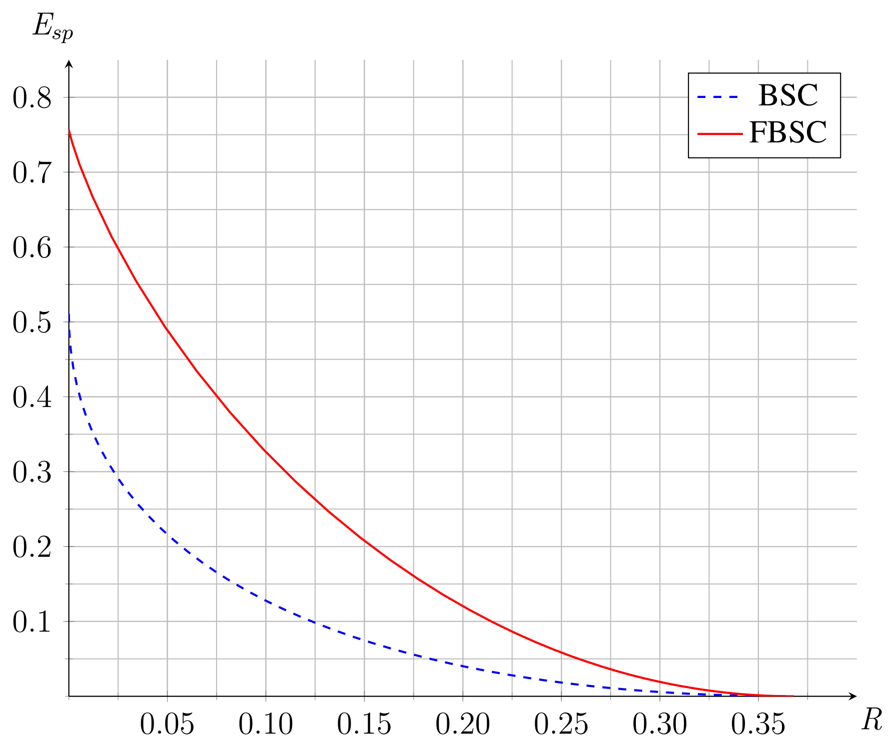

Figure 3.4: The SPEs of BSC with crossover probability 0.1 and FBSC with $g(0) = 0.531$ and $g(0.5) = 0.469$.

two fading parameters and the same capacity with the non-fading BSC at one reference rate equal to 0.130812 to see how fading effects the SPE of the FBSC. The plot of resulting ratios can be seen as a surf plot in Figure 3.5, and as a heat map in Figure 3.6. As expected, the SPE is maximized for the fading distribution g which is given in (3.109) and minimized when the channel has no fading. We know that the binary symmetric channel has the smallest sphere packing exponent among all the discrete memoryless channels with the same channel capacity by [30–32].

3.3.3 Fading Discrete Memoryless Channels with Gallager Symmetry

A fading discrete memoryless channel $V : \mathcal{X} \to \mathcal{P}(\mathcal{H}, \mathcal{Y})$ satisfying (3.4) is Gallager-symmetric if the output set $\mathcal{Y}$ of each channel $W(\cdot, h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ can be partitioned into disjoint subsets $\mathcal{Y}_{h,1}, \cdots, \mathcal{Y}_{h,k_h}$ in such a way that within each $|\mathcal{X}| \times |\mathcal{Y}_{h,k_h}|$ transition matrix all rows are permutations of each other and all columns are permutations of each other. Previously given example channels (FBEC, FBSC) are examples of the

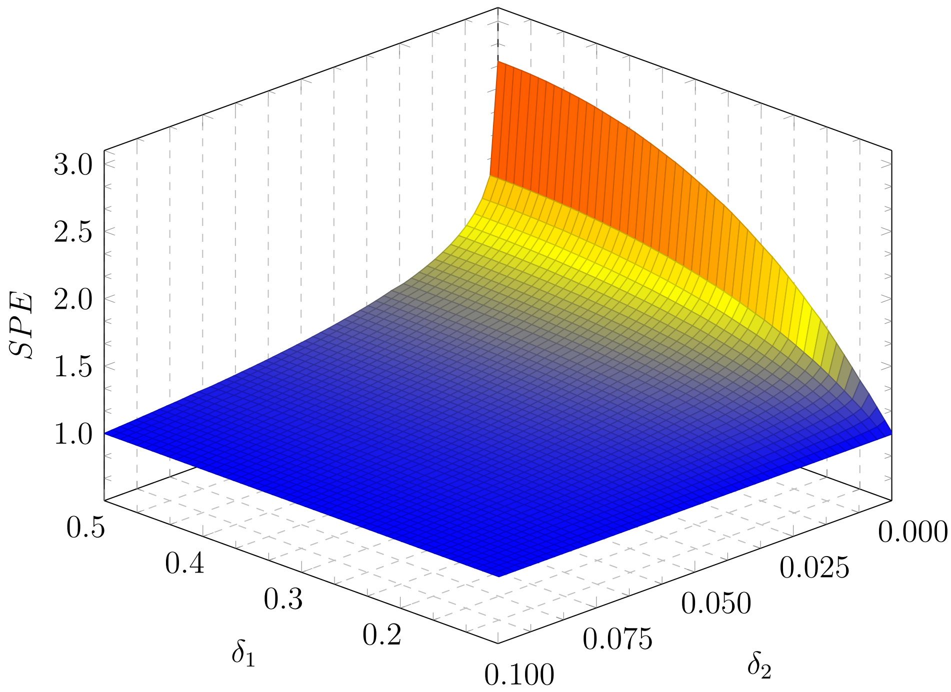

Figure 3.5: The surf plot of the ratios of the SPEs of the non-fading BSC with the cross-over probability equal to $\delta = 0.1$, and the FBSCs having 2 different crossover-probability and same channel capacities (C = 0.368064) at the reference rate (R = 0.130812)

Gallager-symmetric channels. For Gallager-symmetric FDMCs, we can define a α output norm for each output symbol, i.e.,

$$||W(y \mid \cdot, h)||_\alpha = \left[\sum_{x\in\mathcal{X}} W(y \mid x, h)^\alpha\right]^{\frac{1}{\alpha}}. \tag{3.129}$$

If two output symbols $y, \widetilde{y} \in \mathcal{Y}_{h,k}$, their α output norms are the same because the columns in the same partitioning are permutations of each other.

$$||W(y \mid \cdot, h)||_\alpha = ||W(\widetilde{y} \mid \cdot, h)||_\alpha. \tag{3.130}$$

We can denote the α output norms of the outputs symbols in the same partitioning i for a given fading parameter h with $r_{\alpha,h,i}$.

$$r_{\alpha,h,i} = ||W(y \mid \cdot, h)||_\alpha \qquad \forall y \in \mathcal{Y}_{h,i}, \tag{3.131}$$

$$= \left[\sum_{x\in\mathcal{X}} W(y \mid x, h)^\alpha\right]^{\frac{1}{\alpha}} \qquad \forall y \in \mathcal{Y}_{h,i}. \tag{3.132}$$

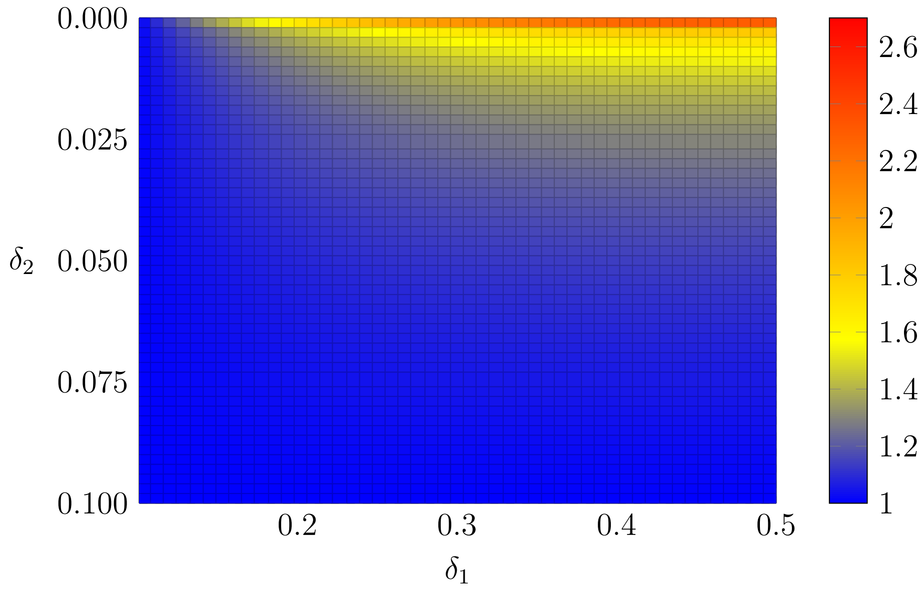

Figure 3.6: The heat map of the ratios of the SPEs of the non-fading BSC with the cross-over probability equal to $\delta = 0.1$, and the FBSCs having 2 different crossover-probability and same channel capacities (C = 0.368064) at the reference rate (R = 0.130812)

Now, let p be the uniform distribution on $\mathcal{X}$ and $q_\alpha(\cdot|h)$ be

$$q_\alpha(y|h) = \frac{r_{\alpha,h,i}}{\sum_{j=1}^{k_h} r_{\alpha,h,j}|\mathcal{Y}_{h,j}|} \qquad \forall y \in \mathcal{Y}_{h,i}, \tag{3.133}$$

on $\mathcal{Y}_h$ for all $h \in \mathcal{H}$. Then, by (2.8), the order α tilted channel of the channel $W(\cdot,h) : \mathcal{X} \to \mathcal{P}(\mathcal{Y})$ for $q_\alpha(\cdot|h)$ is

$$W_\alpha^{q_\alpha(h)}(y|x,h) = \frac{[W(y|x,h)]^\alpha [q_\alpha(y|h)]^{1-\alpha}}{\sum_{\tilde{y}}[W(\tilde{y}|x,h)]^\alpha [q_\alpha(\tilde{y}|h)]^{1-\alpha}}, \tag{3.134}$$

$$= \frac{W(y \mid x)^\alpha r_{\alpha,h,i}^{1-\alpha}}{\sum_{j=1}^{k_h} r_{\alpha,h,j} \frac{|\mathcal{Y}_{h,j}|}{|\mathcal{X}|}} \qquad \forall y \in \mathcal{Y}_{h,i}. \tag{3.135}$$

If we take the expectation of both sides over the p, we get the Augustin operator of $q_\alpha(\cdot|h)$ equal to

$$\mathtt{T}_{\alpha,p}\left(q_\alpha(y|h)\right) = \sum_x p(x)\, W_\alpha^{q_\alpha(h)}(y|x,h),$$

$$= \frac{r_{\alpha,h,i}}{\sum_{j=1}^{k_h} r_{\alpha,h,j}|\mathcal{Y}_{h,j}|} \qquad \forall y \in \mathcal{Y}_{h,i},$$

which is equal to the $q_\alpha(\cdot|h)$.

$$\mathtt{T}_{\alpha,p}\left(q_\alpha(y|h)\right)=q_\alpha(y|h), \tag{3.136}$$

Thus, q_α satisfies the fixed point property. Moreover, $q_\alpha \prec q_p$, where q_p is defined in (3.12). Therefore, q_α is the order α Augustin mean of the channel $W(\cdot,h):\mathcal{X}\to\mathcal{P}(\mathcal{Y})$ for the input distribution p for any fading parameter $h\in\mathcal{H}$, by [22, Lemma 13].

$$q_\alpha(h)=q_{\alpha,p}(\cdot\mid h) \qquad\qquad \forall y\in\mathcal{Y} \tag{3.137}$$

$$I_\alpha(p;W(\cdot,h))=D_\alpha(W(\cdot,h)\|\,q_\alpha(h)|\,p)\,. \tag{3.138}$$

Furthermore, if we calculate the Rényi divergence between the distribution $W(x,h)$ and the distribution $q_{\alpha,p}(h)$, we get

$$D_\alpha(W(x,h)\|\,q_{\alpha,p}(h))=\begin{cases}\frac{\alpha}{\alpha-1}\ln\left[\frac{1}{|\mathcal{X}|}\left(\sum_{j=1}^{k_h}r_{\alpha,h,j}|\mathcal{Y}_{h,j}|\right)^{\alpha}\right]^{\frac{1}{\alpha}} & \text{if } \alpha\neq 1\\ \ln|\mathcal{X}|+\sum\limits_{j=1}^{k_h}\sum\limits_{y\in\mathcal{Y}_{h,j}}W(y\mid x,h)\ln\frac{W(y|x,h)}{r_{1,h,j}}, & \text{if } \alpha=1\end{cases}. \tag{3.139}$$

Thus, $D_\alpha(W(x,h)\|\,q_{\alpha,p}(h))$ is independent of x. Then, for all $x\in\mathcal{X}$, Augustin information for the input distribution p is equal to $D_\alpha(W(x,h)\|\,q_{\alpha,p}(h))$

$$I_\alpha(p;W(\cdot,h))=\sum_x p(x)D_\alpha(W(x,h)\|\,q_{\alpha,p}(h))\,, \tag{3.140}$$

$$=D_\alpha(W(x,h)\|\,q_{\alpha,p}(h)) \qquad\qquad \forall x\in\mathcal{X}. \tag{3.141}$$

Therefore, $I_\alpha(p;W(\cdot,h))=C_{\alpha,h}$, and (3.49) hold. Consequently,

$$C_{\alpha,h}=\begin{cases}\frac{\alpha}{\alpha-1}\ln\left[\frac{1}{|\mathcal{X}|}\left(\sum_{j=1}^{k_h}r_{\alpha,h,j}|\mathcal{Y}_{h,j}|\right)^{\alpha}\right]^{\frac{1}{\alpha}} & \text{if } \alpha\neq 1\\ \ln|\mathcal{X}|+\sum\limits_{j=1}^{k_h}\sum\limits_{y\in\mathcal{Y}_{h,j}}W(y\mid x,h)\ln\frac{W(y|x,h)}{r_{1,h,j}}, & \text{if } \alpha=1\end{cases}, \tag{3.142}$$

$$C_{\alpha,V}=\begin{cases}\frac{\alpha}{\alpha-1}\ln\sum_h g(h)\left[\frac{1}{|\mathcal{X}|}\left(\sum_{j=1}^{k_h}r_{\alpha,h,j}|\mathcal{Y}_{h,j}|\right)^{\alpha}\right]^{\frac{1}{\alpha}}, & \text{if } \alpha\neq 1\\ \ln|\mathcal{X}|+\sum_h g(h)\sum\limits_{j=1}^{k_h}\sum\limits_{y\in\mathcal{Y}_{h,j}}W(y\mid x,h)\ln\frac{W(y|x,h)}{r_{1,h,j}}, & \text{if } \alpha=1\end{cases}, \tag{3.143}$$

$$g_\alpha(h)=\frac{g(h)}{\sum_{\widetilde{h}}g(\widetilde{h})\left[\frac{1}{|\mathcal{X}|}\left(\sum_{j=1}^{k_h}r_{\alpha,h,j}|\mathcal{Y}_{h,j}|\right)^{\alpha}\right]^{\frac{1}{\alpha}}}\left[\frac{1}{|\mathcal{X}|}\left(\sum_{j=1}^{k_h}r_{\alpha,h,j}|\mathcal{Y}_{h,j}|\right)^{\alpha}\right]^{\frac{1}{\alpha}}. \tag{3.144}$$

Parametric expressions of the sphere packing exponent and the rate can be found by inserting (3.144), (3.135), and (3.133) in (3.67), and (3.68).

CHAPTER 4

GAUSSIAN CHANNELS

4.1 Introduction

The most commonly used model in communication systems is the Gaussian channel model. Because we can model the thermal noise as a zero-mean Gaussian random variable with some variance σ^2. A Gaussian random variable with variance σ^2 is denoted with the symbol φ_{σ^2}, and its probability density function is

$$\varphi_{\sigma^2}(x) = \tfrac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}, \qquad \forall x \in \mathbb{R}. \tag{4.1}$$

We can say that the output of the Gaussian channel $y \in \mathbb{R}$ is the sum of the input of the channel $x \in \mathbb{R}$ and a Gaussian random variable k,

$$y = x + k. \tag{4.2}$$

Thus, the probability density function (PDF) of the output of W is

$$W(y|x) = \varphi_{\sigma^2}(y - x), \tag{4.3}$$

$$= \tfrac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-x)^2}{2\sigma^2}} \qquad \forall x, y \in \mathbb{R}. \tag{4.4}$$

Often, Gaussian channels are considered with the constraints on the input to model the real world more accurately. If the constraint function ρ is

$$\rho(x) = x^2, \qquad \forall x \in \mathbb{R}, \tag{4.5}$$

and the constraint is

$$\mathbf{E}[\rho] \le \varrho, \tag{4.6}$$

then, the resulting constraint is called the average power constraint. In Chapter 2, we define the Rényi divergence, the conditional Rényi divergence, the tilted channel, and the Augustin information measures for the finite sets and the probability mass functions on finite sets. We can generalize these definitions to the infinite sets and the probability density functions on infinite sets by replacing the sum operators with the integral operators. The resulting definitions are presented in [22]. Furthermore, we can use them in the same way with the finite case because all the proofs, including the existence of the unique Augustin mean with the fixed point property and the unique Augustin center for arbitrary input distribution on the arbitrary channels in [24] and [22], respectively.

4.1.1 Fading Gaussian Channels

Fading Gaussian channels are described over the Gaussian channels with an additional fading parameter h, also known as the channel state information (CSI). A fading Gaussian channel V is said to be a fast fading Gaussian channel if the produced output probability density function is independent of the previously transmitted inputs, and the fading parameter is independent and identically distributed for every input letter for every use of the channel.

$$V(h^n, y^n|x^n) = \prod_{i=1}^{n} V(h_i, y_i|x_i) \tag{4.7}$$

$$= \prod_{i=1}^{n} g(h_i|x_i)\, W(y_i|x_i, h_i) \tag{4.8}$$

$$= \prod_{i=1}^{n} g(h_i|x_i) \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i - x_i h_i)^2}{2\sigma^2}} \tag{4.9}$$

for any $n \in \mathbb{Z}_+, h^n \in \mathbb{R}^n, y^n \in \mathbb{R}^n, x^n \in \mathbb{R}^n$ where g is the PDF of the fading parameter. x_i denotes the ith element in the sequence x^n. If the fading parameter is independent of the input letter x_i, we can rewrite the channel equation (4.9) as

$$V(h^n, y^n|x^n) = \prod_{i=1}^{n} g(h_i) \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y_i - x_i h_i)^2}{2\sigma^2}} \tag{4.10}$$

Suppose receivers know the value of the CSI. In that case, the resulting channels are called a fast-fading Gaussian channel with CSI at the receiver. This is the model used in this Chapter with the average power constraint.

4.2 An Analysis on the Tilted Channel of the Fading Gaussian Channels

Assume the fading Gaussian channel V satisfying (4.10), with the variance σ^2, cost function ρ given in (4.5) and the cost constraint ϱ given in (4.6). For Gaussian channels, zero-mean Gaussian input distribution with the variance ϱ is the Augustin capacity-achieving, i.e., optimum, input distribution of the channel, and the order α Augustin mean is a zero-mean Gaussian distribution with the variance $\theta_{\alpha,\sigma,\varrho}$ by [22].

$$p = \varphi_{\varrho} \tag{4.11}$$

$$q_{\alpha,p} = \varphi_{\theta_{\alpha,\sigma,\varrho}} \tag{4.12}$$

$$\theta_{\alpha,\sigma,\varrho} = \sigma^2 + \tfrac{\varrho}{2} - \tfrac{\sigma^2}{2\alpha} + \sqrt{\tfrac{\varrho}{2} - \tfrac{\sigma^2}{2\alpha} + \varrho\sigma^2} \tag{4.13}$$

Therefore, for the fixed value of the fading parameter h, the optimum input distribution of the channel $W(\cdot, h)$, which achieves the order α Augustin capacity, is also a zero-mean Gaussian input distribution with the variance ϱ and the order α Augustin mean of the channel $W(\cdot, h)$ is also a zero mean Gaussian distribution with the variance $\theta_{\alpha,\sigma,\varrho,h}$,

$$p = \varphi_{\varrho} \tag{4.14}$$

$$q_{\alpha,p}(h) = \varphi_{\theta_{\alpha,\sigma,\varrho,h}} \tag{4.15}$$

$$\theta_{\alpha,\sigma,\varrho,h} = \sigma^2 + \tfrac{\varrho h^2}{2} - \tfrac{\sigma^2}{2\alpha} + \sqrt{\tfrac{\varrho h^2}{2} - \tfrac{\sigma^2}{2\alpha} + \varrho\sigma^2} \tag{4.16}$$

The order α Rényi divergence between the channel $W(\cdot, h)$ and its order α Augustin mean is

$$D_\alpha\big(W(\cdot, h) \| \varphi_{\theta_{\alpha,\sigma,\varrho,h}}\big) = \begin{cases} \frac{\alpha h^2 x^2}{2\left(\alpha\theta_{\alpha,\sigma,\varrho,h} + (1-\alpha)\sigma^2\right)} + \frac{1}{\alpha-1} \ln \frac{\theta_{\alpha,\sigma,\varrho,h}^{\frac{\alpha}{2}} \sigma^{(1-\alpha)}}{\sqrt{\alpha\theta_{\alpha,\sigma,\varrho,h} + (1-\alpha)\sigma^2}}, & \alpha \neq 1 \\ \frac{\sigma^2 + h^2x^2 - \theta_{\alpha,\sigma,\varrho,h}}{2\theta_{\alpha,\sigma,\varrho,h}} + \frac{1}{2} \ln \frac{\theta_{\alpha,\sigma,\varrho,h}}{\sigma^2}, & \alpha = 1 \end{cases} \tag{4.17}$$

Remark 5. Because the existence of the unique Augustin mean is proved in [24] for arbitrary input distribution on the arbitrary channels, we can generalize the Lemma 1 to use in the Gaussian channels by going through the same set of steps of the proof of the Lemma 1 with the Augustin information measures for infinite sets and probability density functions.

We cannot write the (4.17) in the form of the statement (ii) of the Lemma 1.

$$D_\alpha\big(W(\cdot,h\|\,\varphi_{\theta_{\alpha,\sigma,\varrho,h}}\big) = a_\alpha(x)c_\alpha(h) + d_\alpha(h), \tag{4.18}$$

$$\neq a_\alpha(x) + b_\alpha(h). \tag{4.19}$$

Therefore, we immediately conclude that the tilted channel of the fading Gaussian channel cannot be decomposed into the tilted fading distribution and the tilted channel, i.e., $W_\alpha^{q_{\alpha,\varphi_\varrho}(h)}$, of the Gaussian channels $W(\cdot,h)$. However, the tilted channel of the fading Gaussian channels can be written as a tilted fading distribution time a conditional distribution of y for a given h. Let us determine if the conditional distribution of y of the tilted channel can still be a Gaussian distribution with properly chosen mean γ and the variance Θ. Our assumption for the following analysis is

$$V_\alpha^{s_\alpha,\varphi_\varrho}(h,y\mid x) = g_\alpha(h)\varphi_\Theta(y-\gamma x|h). \tag{4.20}$$

Then, if we take the expected value of the tilted channel over the input distribution φ_ϱ,

$$\mathbf{E}_{\varphi_\varrho}\big[V_\alpha^{s_\alpha,\varphi_\varrho}(h,y|x)\big] = \mathbf{E}_{\varphi_\varrho}[g_\alpha(h)\varphi_\Theta(y-\gamma x|h)] \tag{4.21}$$

$$= g_\alpha(h)\mathbf{E}_{\varphi_\varrho}[\varphi_\Theta(y-\gamma x|h)] \tag{4.22}$$

$$= g_\alpha(h)\varphi_{\gamma^2\varrho+\Theta}(y|h) \tag{4.23}$$

Because the order α Augustin mean of the fading Gaussian channel should satisfy the fixed point property, it should be

$$s_{\alpha,\varphi_\varrho}(h,y) = g_\alpha(h)\varphi_{\gamma^2\varrho+\Theta}(y|h). \tag{4.24}$$

By the tilted channel definition,

$$V_\alpha^{s_\alpha,\varphi_\varrho}(h,y\mid x) = \frac{[g(h)\,V(y\mid h,x)]^\alpha\,\big[g_\alpha(h)\varphi_{\gamma^2\varrho+\Theta}(y\mid h)\big]^{1-\alpha}}{\iint\Big[g(\widetilde{h})\,V(\widetilde{y}\mid\widetilde{h},x)\Big]^\alpha\,\Big[g_\alpha(\widetilde{h})\varphi_{\gamma^2\varrho+\Theta}(\widetilde{y}\mid\widetilde{h})\Big]^{1-\alpha}\,\mathrm{d}\widetilde{h}\mathrm{d}\widetilde{y}}, \tag{4.25}$$

and if we put (4.20) in (4.25) and rearrange, we get

$$\frac{[g(h)\,V(y|h,x)]^\alpha\,\big[\varphi_{\gamma^2\varrho+\Theta}(y|h)\big]^{1-\alpha}}{[g_\alpha(h)]^\alpha\,\varphi_\Theta(y-\gamma x)} = \iint\Big[g(\widetilde{h})\,V(\widetilde{y}\mid\widetilde{h},x)\Big]^\alpha\,\Big[g_\alpha(\widetilde{h})\varphi_{\gamma^2\varrho+\Theta}(\widetilde{y}\mid\widetilde{h})\Big]^{1-\alpha}\,\mathrm{d}\widetilde{h}\mathrm{d}\widetilde{y}. \tag{4.26}$$

The R.H.S of (4.26) only depends on x. Therefore, the L.H.S of (4.26) must be independent of the output y and the fading parameter h for all α and all x. If we write the left-hand side of (4.26) explicitly, we get

$$\frac{[g(h)\,V(y|h,x)]^\alpha\,\big[\varphi_{\gamma^2\varrho+\Theta}(y|h)\big]^{1-\alpha}}{[g_\alpha(h)]^\alpha\,\varphi_\Theta(y-\gamma x)} = \left[\frac{g(h)}{g_\alpha(h)}\right]^\alpha\frac{\Theta^{\frac{1}{2}}}{\sigma^\alpha(\gamma^2\varrho+\Theta)^{\frac{1-\alpha}{2}}}f(y), \tag{4.27}$$

where

$$f(y) = \exp\left(\frac{1}{2}\left(-\frac{(y-hx)^2\alpha}{\sigma^2} - \frac{y^2(1-\alpha)}{\gamma^2\varrho+\Theta} + \frac{(y-\gamma x)^2}{\Theta}\right)\right), \tag{4.28}$$

$$= \frac{1}{2}\left[y^2\left(\frac{1}{\Theta} - \frac{\alpha}{\sigma^2} - \frac{1-\alpha}{\gamma^2\varrho+\Theta}\right) + 2y\left(\frac{hx\alpha}{\sigma^2} - \frac{\gamma x}{\Theta}\right) + \left(\frac{\gamma^2x^2}{\Theta} - \frac{h^2x^2\alpha}{\sigma^2}\right)\right]. \tag{4.29}$$

In (4.27), the only thing that depends on the output y is the function f. Thus, it must be independent of the output y. To make f independent of the output y, the coefficient of y^2 and the coefficient of y must be zero individually. Then, if we make the coefficient of y^2 equal to zero, we get

$$0 = \frac{1}{\Theta} - \frac{\alpha}{\sigma^2} - \frac{1-\alpha}{\gamma^2\varrho+\Theta}, \tag{4.30}$$

$$\gamma^2 = \frac{\alpha\Theta(\Theta-\sigma^2)}{\varrho(\sigma^2-\alpha\Theta)}. \tag{4.31}$$

For this expression to be valid, Θ must satisfy the equation

$$\Theta \geq \sigma^2 \geq \alpha\Theta. \tag{4.32}$$

If we make the coefficient of y equal to zero, we get

$$0 = \frac{hx\alpha}{\sigma^2} - \frac{\gamma x}{\Theta}, \tag{4.33}$$

$$\gamma = \frac{h\alpha\Theta}{\sigma^2}. \tag{4.34}$$

By (4.31) and (4.34), we get the quadratic equation for the Θ,

$$\Theta\left(\Theta^2 - \Theta\left(\frac{\sigma^2}{\alpha} - \frac{\sigma^4}{\alpha^2h^2\varrho}\right) - \frac{\sigma^6}{\alpha^2h^2\varrho}\right) = 0. \tag{4.35}$$

(4.31) and (4.34) are satisfied if Θ is the root of the (4.35). The only root of the (4.35) which is greater than the σ^2, i.e., satisfying the (4.32), is

$$\widehat{\Theta} = \frac{1}{2}\left[\frac{\sigma^2}{\alpha} - \frac{\sigma^4}{\alpha^2h^2\varrho} + \sqrt{\left(\frac{\sigma^2}{\alpha} - \frac{\sigma^4}{\alpha^2h^2\varrho}\right)^2 + 4\frac{\sigma^6}{\alpha^2h^2\varrho}}\right] \tag{4.36}$$

If we put the $\widehat{\Theta}$ into the (4.27), we get

$$\frac{[g(h)\,V(y|h,x)]^\alpha\left[\varphi_{\gamma^2\varrho+\widehat{\Theta}}(y|h)\right]^{1-\alpha}}{[g_\alpha(h)]^\alpha\,\varphi_{\widehat{\Theta}}(y-\gamma x)} = j(h)\exp\left(k(h)x^2\right), \tag{4.37}$$

where

$$j(h) = \left[\frac{g(h)}{g_\alpha(h)}\right]^\alpha \frac{\Theta^{\frac{1}{2}}}{\sigma^\alpha(\gamma^2\varrho+\Theta)^{\frac{1-\alpha}{2}}}, \tag{4.38}$$

$$k(h) = \frac{1}{2}\left(\frac{\gamma^2x^2}{\Theta} - \frac{h^2x^2\alpha}{\sigma^2}\right). \tag{4.39}$$

Until here, we find the conditions to the left-hand side of (4.26) to be independent of the output y and get the expression (4.37) when the conditions are met. However, the left hand side of (4.26) must be also independent of fading parameter h for all α and all x. Thus, the functions j and k must be independent of the fading parameter h, because for the $x = 0$ case, function j must be independent of the fading parameter h and if function j is independent of the fading parameter h, function k must be independent of the fading parameter h to make the left hand side of (4.26) be independent of the fading parameter h for all the values of inputs x. Using (4.31), we get

$$k(h) = \frac{1}{2}h^2\frac{\alpha}{\sigma^2}\left[\frac{1}{2}\left(1 - \frac{\sigma^2}{\alpha h^2 \varrho} + \sqrt{\left(1 - \frac{\sigma^2}{\alpha h^2 \varrho}\right)^2 + 4\frac{\sigma^2}{h^2 \varrho}}\right) - 1\right]. \tag{4.40}$$

If we draw function k for $\alpha = 1$, $\sigma^2 = 1$, and the $\varrho = 2$, we get

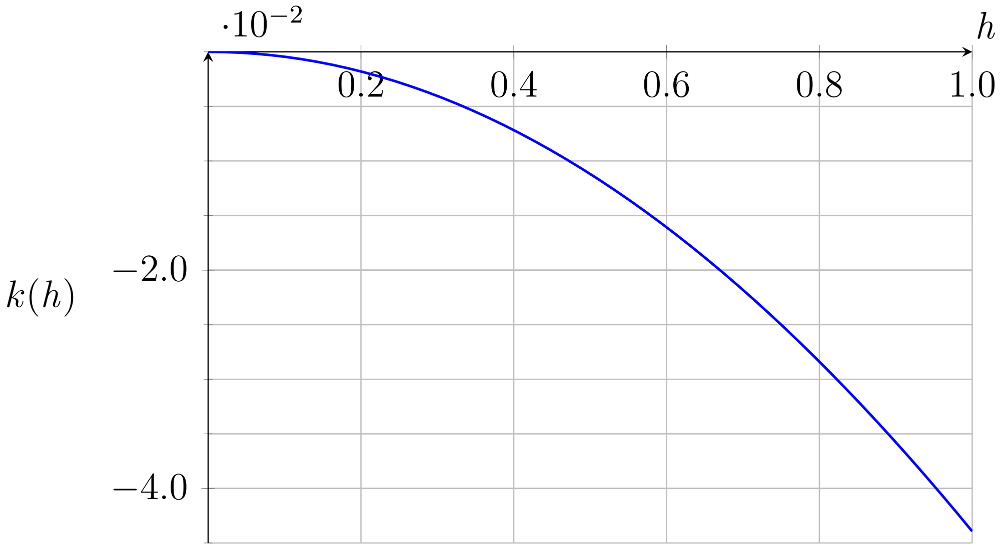

Figure 4.1: The function $k(h)$ with $\alpha = 0.1, \sigma^2 = 1$ and $\varrho = 2$.

Therefore, the function k is not independent of the fading parameter h, which is a contradiction. Thus the conditional output distribution of the tilted channel cannot be a Gaussian distribution.

Remark 6. In contrast to the DMC results, the conditional output distribution of the Augustin mean of the fading Gaussian channel must be the Gaussian distribution to the conditional output distribution of the tilted channel of the fading Gaussian channel be a Gaussian distribution by the definition of the tilted channel. Therefore, by showing the impossibility of the conditional output distribution of the tilted channel of the fading Gaussian channel being Gaussian distribution, we also showed that the conditional output distribution of the Augustin mean cannot be a Gaussian distribution for fading Gaussian channels.

CHAPTER 5

CONCLUSIONS

In this thesis, we proved a decomposition for the tilted channel of the fast-fading DMCs with the CSI at the receiver as a product of a tilted fading distribution and the tilted channels of the non-fading channels for every channel state. We determined the necessary and sufficient conditions. We showed that the fading discrete channels with the common Augustin capacity achieving distributions for every channel state satisfy the conditions of the decomposition lemma, i.e., Lemma 1. We calculated the parametric expressions of the SPE and the rate of channels in the class. We used our calculations and determined the results of the FBSCs, FBECs, and fading channels with Gallager symmetry. Then we compare these channels with the corresponding non-fading channels under the equal Augustin capacity circumstances. We had the observations that are

1. By Lemma 2, the fading binary erasure channels always have the same SPEs with the corresponding non-fading BEC. The only thing to characterize the SPE of the FBEC is the expected value of the fading parameter, which is the erasure probability in the non-fading BEC. Furthermore, if the sphere packing exponents of the two binary erasure channels are the same, then the expected value of the fading distribution must be the same.

2. By Lemma 3, the fading binary symmetric channels have greater sphere packing exponents than the non-fading binary symmetric channels for every rate and fading distribution. Furthermore, the sphere packing exponent of the fading binary symmetric channels is maximized when the fading distribution only has two fading parameters, which are 0 and 0.5.

Then we turn our attention to the Gaussian channels. We used the optimal input distribution of the non-fading Gaussian channel, i.e., the zero-mean Gaussian distribution with the variance equal to the cost constraint, as the input distribution. Then, we showed that the Lemma 1 is not applicable for fast-fading Gaussian channels with the CSI at the receiver for this input distribution. Lastly, we assumed that the conditional output distribution of the tilted channel is Gaussian distributed and proved that the assumption could not be valid. By the observation that the conditional output distribution of the tilted channel cannot be a Gaussian distribution, we found out that the conditional output distribution of the Augustin mean also cannot be a Gaussian distribution.

APPENDIX A

Let $\mathcal{X}^n$ is the input alphabet and $\mathcal{Y}^n$ be the output alphabet. We consider a code consisting of M code word with the encoder function ψ that maps the integers from 1 to M into the codewords $\psi(1), \ldots, \psi(M)$, where $\psi(m) \in \mathcal{X}^n$ and $m \in \{1, \ldots, M\}$. We assume that the Maximum Likelihood (ML) decoding is used at the receiver. ML decoder decodes the output sequence y^n into the integer $\widehat{m}$ if

$$\mathbf{P}(y^n | \psi(\widehat{m})) > \mathbf{P}(y^n | \psi(m)) \qquad \forall m \neq \widehat{m}. \tag{A.1}$$

In the equality case, decoder chooses randomly between the possible messages. Decoder makes an error if the output sequence is more probable for an input message $\widehat{\mathsf{M}}$ other than the $\psi(m)$. Then the probability of error for a message m is

$$\mathbf{P}_e^m = \sum_{y^n \in \mathcal{Y}^n} \mathbf{P}(y^n | \psi(m)) \, \phi_m(y^n), \tag{A.2}$$

where the indicator function $\phi_m(y^n)$ indicates if the decoder makes an error when the output y^n is received or not.

$$\phi_m(y^n) = \begin{cases} 1 & \text{if } \mathbf{P}(y^n | \psi(m)) \leq \mathbf{P}(y^n | \psi(\overline{m})) \text{ for some } \overline{m} \neq m \\ 0 & \text{otherwise} \end{cases}. \tag{A.3}$$

We can upper bound the indicator function with

$$\phi_m(y^n) \leq \left[\frac{\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}}{\mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}}} \right]^{\rho} \qquad \rho > 0. \tag{A.4}$$

Then, the probability of error of the message m can be upper bounded by

$$\mathbf{P}_e^m \leq \sum_{y^n \in \mathcal{Y}^n} \mathbf{P}(y^n | \psi(m)) \left[\frac{\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}}{\mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}}} \right]^{\rho} \qquad \rho > 0, \tag{A.5}$$

$$\leq \sum_{y^n \in \mathcal{Y}^n} \mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}} \left[\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}} \right]^{\rho} \qquad \rho > 0. \tag{A.6}$$

Upper bound on the indicator function is valid because it is always greater than 0, which makes it true for the case $\phi_m(y^n) = 0$. When it is equal to one, at least one other codeword has more probability of generating the output sequence than the transmitted codeword. It makes the right hand side of (A.4) greater than one. This bound is true for any codes and codewords; however, it is not useful for many codes, particularly the codes, including a large number of the codewords. Therefore, calculating another simple bound on the probability of error is required. We can simplify (A.6) by averaging. Assume that the $\mathbf{P}(x^n)$ is defined on the input alphabet $\mathcal{X}^n$. As an ensemble, let us consider all the codes generated by the randomly chosen codewords according to the $\mathbf{P}(x^n)$. The code consisting $\psi(1), \ldots, \psi(M)$ has the probability $\prod_{m=1}^{M} \mathbf{P}(\psi(m))$ and at least one code has a lower probability of error than the ensemble-average probability of error. If we take the expected value of the (A.6),

$$\mathbf{E}[\mathbf{P}_e^m] \leq \mathbf{E}\left[\sum_{y^n \in \mathcal{Y}^n} \mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}} \left(\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}\right)^{\rho}\right] \qquad \rho > 0, \quad \text{(A.7)}$$

$$\leq \sum_{y^n \in \mathcal{Y}^n} \mathbf{E}\left[\mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}} \left(\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}\right)^{\rho}\right] \qquad \rho > 0, \quad \text{(A.8)}$$

$$\leq \sum_{y^n \in \mathcal{Y}^n} \mathbf{E}\left[\mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}}\right] \mathbf{E}\left[\left(\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}\right)^{\rho}\right] \quad \rho > 0, \quad \text{(A.9)}$$

$$\leq \sum_{y^n \in \mathcal{Y}^n} \mathbf{E}\left[\mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}}\right] \left(\mathbf{E}\left[\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}\right]\right)^{\rho} \quad 0 < \rho \leq 1, \tag{A.10}$$

$$\leq \sum_{y^n \in \mathcal{Y}^n} \mathbf{E}\left[\mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}}\right] \left(\sum_{\overline{m} \neq m} \mathbf{E}\left[\mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}\right]\right)^{\rho} \quad 0 < \rho \leq 1. \tag{A.11}$$

We get (A.8) because the position of the expectation and the summation can be changed.

$$\mathbf{P}(y^n | \psi(m))^{\frac{1}{1+\rho}} \tag{A.12}$$

and

$$\left[\sum_{\overline{m} \neq m} \mathbf{P}(y^n | \psi(\overline{m}))^{\frac{1}{1+\rho}}\right]^{\rho} \tag{A.13}$$

are independent random variables in (A.8) and for the independent random variables A and B, we have

$$\mathbf{E}[AB] = \mathbf{E}[A]\,\mathbf{E}[B]\,.$$

Thus, we can split (A.8) up to (A.9). Random variable inside of the brackets in (A.13) is a concave function for the exponent $\rho \in (0, 1]$ at the outside of the bracket. For this reason, by Jensen's inequality,

$$\mathbf{E}\left[\left(\sum_{\overline{m}\neq m} \mathbf{P}(y^n|\,\psi(\overline{m}))^{\frac{1}{1+\rho}}\right)^{\rho}\right] \leq \left(\mathbf{E}\left[\sum_{\overline{m}\neq m} \mathbf{P}(y^n|\,\psi(\overline{m}))^{\frac{1}{1+\rho}}\right]\right)^{\rho}. \tag{A.14}$$

Again, by using the interchange property of the expectation and summation, we attain (A.11). Moreover, since the codewords are selected independently,

$$\mathbf{E}\Big[\mathbf{P}(y^n|\,\psi(m))^{\frac{1}{1+\rho}}\Big] = \sum_{x^n\in\mathcal{X}^n} \mathbf{P}(x^n)\,\mathbf{P}(y^n|\,x^n)^{\frac{1}{1+\rho}}\,. \tag{A.15}$$

Furthermore, the expected values of the random variables in (A.11) are independent of the message m, and we can replace them with the right hand side of the (A.15).

$$\mathbf{E}[\mathbf{P}_e^m] \leq (M-1)^{\rho} \sum_{y^n\in\mathcal{Y}^n} \left[\sum_{x^n\in\mathcal{X}^n} \mathbf{P}(x^n)\,\mathbf{P}(y^n|\,x^n)^{\frac{1}{1+\rho}}\right]^{1+\rho} \qquad 0<\rho\leq 1. \tag{A.16}$$

This bound is applicable for any discrete channel, for any selection of $\mathbf{P}(x^n)$, and any $\rho \in (0, 1]$. Provided that the discrete channel is also memoryless, we can simplify (A.16) more. A discrete channel is memoryless if

$$\mathbf{P}(y^n|\,x^n) = \prod_{i=1}^{n} \mathbf{P}(y_i|\,x_i)\,. \tag{A.17}$$

where $x_i \in \mathcal{X}$ is the i-th letter of the codeword x^n and similarly $y_i \in \mathcal{Y}$ is the i-th letter of the output sequence $\mathcal{Y}^n$. Now, we restrict the code ensemble with the codes in which the codewords are constructed according to the $\mathbf{P}(x)$ randomly to simplify the bound of the error probability. In that case, the probability of the codeword in a code is equal to,

$$\mathbf{P}(x^n) = \prod_{i=1}^{n} \mathbf{P}(x_i)\,. \tag{A.18}$$

If we insert (A.17) and (A.18) into (A.16),

$$\mathbf{E}[\mathbf{P}_e^m] \leq (M-1)^{\rho} \sum_{y^n\in\mathcal{Y}^n} \left[\sum_{x^n\in\mathcal{X}^n} \prod_{i=1}^{n} \mathbf{P}(x_i)\,\mathbf{P}(y_i|\,x_i)^{\frac{1}{1+\rho}}\right]^{1+\rho} \qquad 0<\rho\leq 1. \tag{A.19}$$

If we rearrange the terms inside the summation and product operators,

$$\mathbf{E}[\mathbf{P}_e^m] \leq (M-1)^\rho \prod_{i=1}^{n} \sum_{y_i \in \mathcal{Y}} \left[\sum_{x_i \in \mathcal{X}} \mathbf{P}(x_i)\, \mathbf{P}(y_i|\, x_i)^{\frac{1}{1+\rho}} \right]^{1+\rho} \qquad 0 < \rho \leq 1. \tag{A.20}$$

Because the multiplied values are independent of the operator of the product, i,

$$\mathbf{E}[\mathbf{P}_e^m] \leq (M-1)^\rho \left[\sum_{y_i \in \mathcal{Y}} \left[\sum_{x_i \in \mathcal{X}} \mathbf{P}(x_i)\, \mathbf{P}(y_i|\, x_i)^{\frac{1}{1+\rho}} \right]^{1+\rho} \right]^{n} \qquad 0 < \rho \leq 1. \tag{A.21}$$

Transmission rate R, is defined in (2.43) and if the $M-1$ is upper bounded by the

$$M = e^{nR}. \tag{A.22}$$

Then, (A.21) is turned to be

$$\mathbf{E}[\mathbf{P}_e^m] \leq e^{n[E_0(\rho, \mathbf{P}(x)) - \rho R]} \qquad 0 < \rho \leq 1, \tag{A.23}$$

where

$$E_0(\rho, \mathbf{P}(x)) = -\ln \sum_{y_i \in \mathcal{Y}} \left[\sum_{x_i \in \mathcal{X}} \mathbf{P}(x_i)\, \mathbf{P}(y_i|\, x_i)^{\frac{1}{1+\rho}} \right]^{1+\rho}. \tag{A.24}$$

This upper bound is true for any discrete memoryless channel, $0 < \rho \leq 1$, and $\mathbf{P}(x)$. Furthermore, it should be satisfied if the error probability is minimized over the ρ and $\mathbf{P}(x)$. Because the right hand side of the (A.23) is independent of the m, it is bound to the average probability of decoding error for the ensemble of the codes. Because at least one code must have a lower probability of error than the average, then there should be a code in which the probability of decoding error is upper bounded by

$$\mathbf{P}_e^{av} \leq e^{-nE_r(R)}, \tag{A.25}$$

where the $E_r(R)$, called as the random coding exponent [2, Theorem 1] function, is

$$E_r(R) = \max_{\rho, \mathbf{P}(x)} \left[E_0(\rho, \mathbf{P}(x)) - \rho R \right]. \tag{A.26}$$

where the maximization is over all $0 < \rho \leq 1$, and all possible $\mathbf{P}(x)$.
Important properties of the Random Coding Exponent function is

- It is a non-negative and non-increasing function of rate R.
- It is equal to the zero at the rate R equal to the channel capacity.

- ρ that maximizes the $E_r(R)$ is non-increasing function for rate R.
- Because the maximizing ρ is non-increasing function, maximum rate for which the ρ that maximizes the $E_r(R)$ is named as *critical rate*, and denoted as R_{crit}.
- For every rate below the critical rate, $E_r(R)$ is a linear line with the slope -1, and maximization happens at the $\rho = 1$.

For the lower rates, an improved version of the Random Coding Exponent function, i.e., Expurgated Random Coding Exponent function $E_x(R)$, is derived in [2]. More detailed analysis of the function $E_0(\rho, \mathbf{P}(x))$, $E_r(R)$, and proof of the $E_x(R)$, [2] can be examined.

REFERENCES

[1] A. Ingber and M. Feder, "Finite blocklength coding for channels with side information at the receiver," in *2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel*, pp. 000798–000802, IEEE, 2010.

[2] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Transactions on Information Theory*, vol. 11, pp. 3–18, Jan. 1965.

[3] T. Ericson, "A gaussian channel with slow fading (corresp.)," *IEEE Transactions on Information Theory*, vol. 16, no. 3, pp. 353–355, 1970.

[4] W. K. M. Ahmed and P. J. McLane, "The information theoretic reliability function for multipath fading channels with diversity," in *Proceedings of ICUPC - 5th International Conference on Universal Personal Communications*, vol. 2, pp. 886–890 vol.2, 1996.

[5] W. K. M. Ahmed and P. J. McLane, "Achievable performance over fading channels with antenna diversity," in *WCNC. 1999 IEEE Wireless Communications and Networking Conference (Cat. No.99TH8466)*, vol. 1, pp. 25–29 vol.1, 1999.

[6] W. K. M. Ahmed and P. J. McLane, "Random coding error exponents for two-dimensional flat fading channels with complete channel state information," *IEEE Transactions on Information Theory*, vol. 45, no. 4, pp. 1338–1346, 1999.

[7] W. K. M. Ahmed and P. J. McLane, "Information theoretic considerations for coded modulation over fading channels," *IEEE Transactions on Communications*, vol. 48, no. 12, pp. 1970–1974, 2000.

[8] M. Z. I. Sarkar, H. Shin, and M. Z. Win, "Random coding exponent for mimo channels," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, pp. 559–563, 2008.

[9] H. Shin and M. Z. Win, "Gallager's exponent for mimo channels: a reliability-

rate tradeoff," *IEEE Transactions on Communications*, vol. 57, pp. 972–985, April 2009.

[10] E. Malkamaki and H. Leib, "Coded diversity on block-fading channels," *IEEE Transactions on Information Theory*, vol. 45, pp. 771–781, March 1999.

[11] J. S. Richters, *Communication over fading dispersive channels*. Technical report 464, Research Laboratory of Electronics at Massachusetts Institute of Technology, Cambridge, MA, 1967. (`http://hdl.handle.net/1721.1/4279`).

[12] H. Shin, *Capacity and error exponents for multiple-input multipleoutput wireless channels*. PhD thesis, Ph. D. dissertation, Seoul National University, Seoul, Korea, 2004.

[13] E. Biglieri, G. Caire, and G. Taricco, "Coding for the block-fading channel: optimum and suboptimum power-allocation schemes," in *1998 Information Theory Workshop (Cat. No.98EX131)*, pp. 96–97, 1998.

[14] J. Xue, M. Z. I. Sarkar, and T. Ratnarajah, "Error exponents for nakagami-m fading channels," *Problems of Information Transmission*, vol. 50, pp. 144–170, April 2014.

[15] J. Xue, T. Ratnarajah, and C. Zhong, "Error exponents for multi-keyhole mimo channels," *Problems of Information Transmission*, vol. 51, pp. 1–19, January 2015.

[16] J. Xue, M. Z. I. Sarkar, C. Zhong, and T. Ratnarajah, "Error exponents for orthogonal stbc in generalized-k fading mimo channels," in *2012 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1925–1929, 2012.

[17] J. Xue, M. Z. I. Sarkar, T. Ratnarajah, and C. Zhong, "Error exponents for rayleigh fading product mimo channels," in *2012 IEEE International Symposium on Information Theory Proceedings*, pp. 2166–2170, 2012.

[18] G. Alfano, C. Chiasserini, A. Nordio, and S. Zhou, "Information-theoretic characterization of mimo systems with multiple rayleigh scattering," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5312–5325, 2018.

[19] G. Alfano, C. . Chiasserini, A. Nordio, and S. Zhou, “A unifying analysis of error exponents for mimo channels with application to multiple-scattering,” in *2015 International Symposium on Wireless Communication Systems (ISWCS)*, pp. 321–325, 2015.

[20] J. Zhang, M. Matthaiou, G. K. Karagiannidis, Z. Tan, and H. Wang, “Gallager's error exponent analysis of stbc systems over η-μ fading channels,” in *2013 IEEE International Conference on Communications (ICC)*, pp. 5829–5834, 2013.

[21] A. Karadimitrakis, A. L. Moustakas, and R. Couillet, “Gallager bound for mimo channels: Large- n asymptotics,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1323–1330, 2018.

[22] B. Nakiboğlu, “The Augustin Capacity and Center,” *Problems of Information Transmission*, vol. 55, pp. 299–342, October 2019. (arXiv:1803.07937 [cs.IT]).

[23] U. Augustin, *Noisy Channels*. Habilitation thesis, Universität Erlangen-Nürnberg, 1978.

[24] H.-C. Cheng and B. Nakiboğlu, “On the existence of the augustin mean,” in *2021 IEEE Information Theory Workshop (ITW)*, pp. 1–6, 2021.

[25] H.-C. Cheng, M. H. Hsieh, and M. Tomamichel, “Quantum sphere-packing bounds with polynomial prefactors,” *IEEE Transactions on Information Theory*, vol. 65, pp. 2872–2898, May 2019. (arXiv:1704.05703 [quant-ph]).

[26] B. Nakiboğlu, “The Rényi Capacity and Center,” *IEEE Transactions on Information Theory*, vol. 65, pp. 841–860, Feb 2019. (arXiv:1608.02424 [cs.IT]).

[27] C. E. Shannon, “A mathematical theory of communication,” *Bell System Technical Journal, The*, vol. 27, pp. 379–423 and 623–656, July and October 1948.

[28] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I,” *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.

[29] B. Nakiboğlu, “The Sphere Packing Bound for Memoryless Channels,” *Problems of Information Transmission*, vol. 56, pp. 201–244, July 2020. (arXiv:1804.06372 [cs.IT]).

[30] E. Arıkan and E. Telatar, “BEC and BSC are E_0 extremal.” unpublished technical note.

[31] A. Guillen i Fabregas, I. Land, and A. Martinez, “Extremes of error exponents,” *IEEE Transactions on Information Theory*, vol. 59, no. 4, pp. 2201–2207, 2013.

[32] M. Alsan, “Extremality for Gallager’s reliability function E_0,” *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4277–4292, 2015.