Capacity-Achieving Distributions of Gaussian Multiple Access Channel with Peak Constraints

by

Babak Mamandipoor

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2012

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Characterizing probability distribution function for the input of a communication channel that achieves the maximum possible data rate, is one of the most fundamental problems in the field of information theory. In his ground-breaking paper, Shannon showed that the capacity of a point-to-point additive white Gaussian noise channel under an average power constraint at the input, is achieved by Gaussian distribution. Although imposing a limitation on the peak of the channel input is also very important in modelling the communication system more accurately, it has gained much less attention in the past few decades. A rather unexpected result of Smith indicated that the capacity achieving distribution for an AWGN channel under peak constraint at the input is unique and discrete, possessing a finite number of mass points.

In this thesis, we study multiple access channel under peak constraints at the inputs of the channel. By extending Smith's argument to our multi-terminal problem we show that any point on the boundary of the capacity region of the channel is only achieved by discrete distributions with a finite number of mass points. Although we do not claim uniqueness of the capacity-achieving distributions, however, we show that only discrete distributions with a finite number of mass points can achieve points on the boundary of the capacity region.

First we deal with the problem of maximizing the sum-rate of a two user Gaussian MAC with peak constraints. It is shown that generating the code-books of both users according to discrete distributions with a finite number of mass points achieves the largest sum-rate in the network. After that we generalize our proof to maximize the weighted sum-rate of the channel and show that the same properties hold for the optimum input distributions. This completes the proof that the capacity region of a two-user Gaussian MAC is achieved by discrete input distributions with a finite number of mass points.

Acknowledgements

First and foremost, I would like to express my endless gratitude to my supervisor, Professor Amir K. Khandani for his insightful thinking, support and unbounded enthusiasm. I have gained a great amount of experience and knowledge working with him. His useful comments and constructive advice has benefited me a lot. Through my work with professor Khandani I have also had the opportunity to collaborate with Dr. Kamyar Moshksar. I am very grateful for the opportunity I have had working with Kamyar.

I would like to extend my deepest gratitude to my parents for their support and help throughout the years, for their kind words, love and patience. Also special thanks to all my friends in Coding and Signal Transmission (CST) Lab., especially Akbar Ghasemi for many interesting discussions that we have had on this research.

to

my lovely parents

Table of Contents

List of Figures

Chapter 1

Introduction

1.1 Information Theory

A mathematical theory for the study of communication systems was first developed by Claude E. Shannon in 1948 [1]. In [1], Shannon defined information in mathematical terms and discussed limitations of data transmission rate over different communication channels. He Described the notion of *Channel Capacity* as the maximum transmission rate over the channel such that reliable communication is secured. Over the last half century a large portion of works in the field of Information Theory has been dedicated to finding the capacity of communication channels and implementing Coding strategies to achieve rates close to this capacity. Notions like *Entropy* of a random variable $\boldsymbol{x}$, denoted by $h(\boldsymbol{x})$, and the *Mutual Information* between two random variables, shown by $I(\boldsymbol{x};\boldsymbol{y})$, are of substantial importance in the development of the theory.

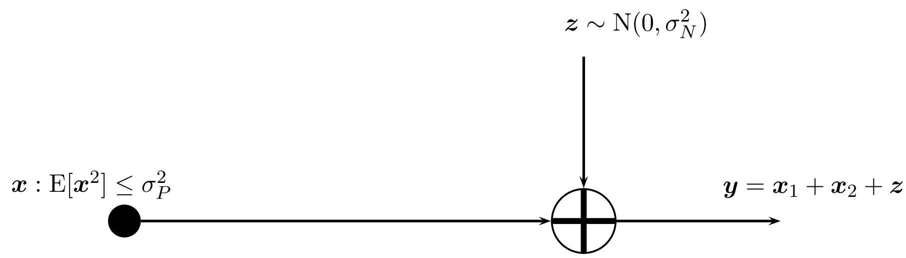

Figure 1.1: A point-to-point additive white Gaussian noise channel with average power constraint at the transmitter

1.2 Channel Capacity

Shannon modeled the channel by a probability transition matrix $p(y|x)$, and showed that its capacity is given by

$$C = \sup_{p_{\boldsymbol{x}}(x)} I(\boldsymbol{x}; \boldsymbol{y}), \tag{1.1}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the input and the output of the channel, respectively. It turns out that finding a closed-form formula for many of the communication channels is extremely complicated, but between those ones with a well-known capacity formula, Additive White Gaussian Noise (AWGN) channel is of great importance. Shannon showed that if the output of the channel is of the form $\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}$, where $\boldsymbol{z}$ is the additive noise with a Gaussian Probability Distribution Function (PDF), subject to an average power constraint at the input of the channel, is

$$C = \frac{1}{2} \log \left(1 + \frac{\sigma_P^2}{\sigma_N^2}\right), \tag{1.2}$$

where σ_P^2 and σ_N^2 denote the variance of $\boldsymbol{x}$ and $\boldsymbol{z}$, respectively. The capacity of this channel is achieved by Gaussian distribution with variance σ_P^2 at the input.

1.3 Literature Review

1.3.1 Scalar Gaussian Channel with Peak Constraint

In his seminal thesis, J. G. Smith studies a point to point Additive White Gaussian Noise channel under peak power constraint, as well as average power constraint [2, 3]. The main observation in [2], is that with only a "peak" power constraint at the input, the capacity achieving distribution is *unique* and *discrete* with a *finite* number of mass points. The same characteristics hold for the optimum distribution if we have both average and peak power constraints at the transmitter. We remark that characterizing the exact number and location of the mass points must be done numerically. This is a remarkable observation in the sense that, for a fixed peak power constraint, increasing the number of mass points of the input distribution does not necessarily result in an increase in the amount of mutual information between the input and the output of the channel. We will go more into the details of [2] in chapter 2, since there is a strong correlation betweens our proof for Theorem 1 in chapter 4, and Smith's approach in [2].

1.3.2 Capacity-Achieving Discrete Distributions

Discrete input distributions are capacity achieving in various scenarios. [4] shows that for some additive noise channels with peak constraint, where the noise has a piecewise-constant probability density, only discrete input distributions can achieve the capacity.

It is shown in [5] that under only an average power constraint at the transmitter in a point-to-point scenario, if the additive noise is "heavy-tailed" compared to Gaussian noise, the corresponding capacity-achieving distribution must be of bounded support. While, for a "light-tailed" additive noise compared to Gaussian noise, the capacity-achieving distribution must be of unbounded support[1]. Moreover, for heavy-tailed noises, if the marginal entropy function[2] admits an analytic extension to the complex plane, the optimum input distribution has a finite number of mass points.

Reference [6] provides an insightful methodology in dealing with similar problems considered in [2, 3]. In fact, [6] presents a general class of additive noise channels for which a discrete input with a finite number of mass points achieves the capacity under a peak constraint at the transmitter.

In [7], the author considers a point-to-point additive white Gaussian noise channel under small peak constraints. It is shown that if the constraint on the peak is small enough[3], a binary and equiprobable distribution with the mass points at the two ends of the input peak interval achieves the capacity of the channel.

The authors In [8] extend the results in [2, 3] to a quadrature additive Gaussian channel. It is shown that under average and peak power constraints, an input distribution with discrete amplitude, possessing a finite number of mass points, with a uniformly distributed and statistically independent phase, achieves the capacity of the channel. Geometrical representation of the optimal distribution is a finite number of concentric circles. The same properties hold for the capacity-achieving input distribution of an AWGN channel with Tikhonov phase error [9]. Capacity of an AWGN channel with block-independent carrier phase rotation is studied in [10], and it's been shown that an input distribution which is discrete in amplitude with infinite number of mass points, achieves the capacity.

In another framework, the authors in [11, 12] demonstrate optimality of discrete input distributions in channels with quantized output. Two different kinds of quantizers are studied in [12]; "saturation quantizer" in which the overflows are mapped to the closest

[1]For definition of "heavy-tailed" and "light-tailed" distributions, see [5].

[2]Marginal entropy function is defined by Smith in [2]; for details see Section().

[3]If the channel input is constrained to be in $[-A, A]$, then $A \leq 1.05$ is shown to be "small enough".

quantization level, and "modulo quantizer" where the overflows are mapped to the nearest quantization level after subtracting some multiple of the modulo period from the channel output[4]. It is shown that regardless of the noise distribution, the capacity achieving input distribution for a system with N-level modulo quantizer at the output, is discrete with exactly N mass points. Although these mass points may be at different locations for various noise distributions, but they are always uniformly distributed, i.e. mass points are equiprobable and equidistant. It is also shown that for Gaussian additive noise channel, the capacity of modulo quantization is a lower bound for the capacity of saturation quantization, and the gap between these two is reduced by increasing the number of quantization levels.

Conditionally Gaussian channels with peak and multiple cost constraint at the transmitter are studied in [13] where necessary and sufficient conditions for optimality of discrete input distributions are provided. [13] also provides a comprehensive list of research papers devoted to channels with peak and or power constraints.

In [14], the authors give a comprehensive survey on fading channels. In many cases, these channels were shown to have discrete capacity-achieving input distributions [15]-[21]. Reference [15] explores a memoryless point-to-point channel with fast Rayleigh fading, where fading is independent from one symbol to another. Assuming that the channel state information is unknown to both transmitter and receiver, it is shown that under only a constraint on average power at the transmitter the capacity-achieving input distribution is discrete. Moreover, authors in [21] show that Gaussian input distribution generates bounds on the mutual information of the channel.

It is shown in [19, 20] that the optimum input distribution of the discrete-time noncoherent AWGN channel, under only average power constraint, is discrete with finite num-

[4]For a detailed definition, see[12], Section (2).

ber of mass points.

For low power regime, it is shown that under general assumptions on point-to-point additive noise channels, the capacity-achieving input distributions are binary [23]. Moreover it has been shown that even for the channels with continuous optimum input distributions, simple discrete input constellations can get very close to the capacity of the channel.

1.4 Thesis Organization

This thesis consists of 6 chapters. Chapter 2 reviews scalar additive Gaussian channel under peak constraint at the transmitter and summarizes Smith's approach in characterizing the optimum input distribution.

Chapter 3 presents multiple access channel models with average and peak power constraints at the transmitters and introduces the capacity region of the channel under general assumptions. Our problem is formulated at the last section of chapter 3.

Chapter 4 studies the optimum input distributions corresponding to the maximum achievable sum-rate of a two user multiple access channel with peak constraints. We show that discrete distributions with a finite number of mass points can achieve the maximum sum-rate. Although we do not claim uniqueness of the answers but we show that discrete distributions are the only sum-capacity achieving distributions in this scenario.

Chapter 5 generalizes our results in chapter 4 for the optimum input distributions corresponding to the maximum weighted sum-rate in the same multiple access channel. This in turn proves that any point on the boundary of the capacity region is only achieved by discrete distributions with a finite number of mass points.

Chapter 6 summarizes our discussions in the previous chapters and states some possible future work directions.

1.5 Notation:

The set of natural, real and complex numbers are denoted by $\mathbb{N}$, $\mathbb{R}$ and $\mathbb{C}$, respectively. Other sets are shown by calligraphic letters such as $\mathcal{X}$. The real and imaginary parts of $s \in \mathbb{C}$ are denoted by $\mathrm{Re}(s)$ and $\mathrm{Im}(s)$, respectively. The imaginary number $\sqrt{-1}$ is shown by j. We denote the constant $\frac{1}{\sqrt{2\pi}}$ by c. Random variables are shown in bold such as $\boldsymbol{x}$ with realization x. The Probability Density Function (PDF), the Cumulative Distribution Function (CDF) and the characteristic function[5] of $\boldsymbol{x}$ are shown by $p_{\boldsymbol{x}}(\cdot)$, $F_{\boldsymbol{x}}(\cdot)$ and $\Phi_{\boldsymbol{x}}(\cdot)$, respectively. The law of a distribution function F on $\mathbb{R}$ is shown by μ_F, i.e, $\mu_F(\cdot)$ is the unique probability measure on $\mathbb{R}$ such that $F(x) = \mu_F((-\infty, x])$. A point $x \in \mathbb{R}$ is called a point of increase for a random variable $\boldsymbol{x}$ if for any open set $\mathcal{O}$ containing x, we have $\mu_{\boldsymbol{x}}(\mathcal{O}) > 0$. The support of a random variable $\boldsymbol{x}$ is the union of all points of increase for $\boldsymbol{x}$. The differential entropy of a continuous random variable $\boldsymbol{x}$ is shown by $h(\boldsymbol{x})$, and $I(\boldsymbol{x};\boldsymbol{y})$ denotes the mutual information between random variables $\boldsymbol{x}$ and $\boldsymbol{y}$. A normal random variable with mean m and variance σ^2 is denoted by $\mathrm{N}(m, \sigma^2)$. For two functions $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$, we say $f(x) = g(x)$ for almost all x if $f(x) = g(x)$ except possibly on a set of Lebesgue measure zero. The indicator function of any $\mathcal{A} \subset \mathbb{R}$ is shown by $\mathbf{1}_{\mathcal{A}}(\cdot)$. If $(x_n)_{n\in\mathbb{N}}$ is a sequence in a topological space, we write $\lim_n x_n$ as a short hand for $\lim_{n\to\infty} x_n$. In a metric space $\mathcal{X}$, a point $x_* \in \mathcal{X}$ is called a point of accumulation for a set $\mathcal{A} \subset \mathcal{X}$ if there exists a sequence $(x_n)_{n\in\mathbb{N}} \subset \mathcal{A}$ such that $\lim_n x_n = x_*$. The largest open set contained in a set $\mathcal{A}$ in a topological space is called the interior of $\mathcal{A}$ and is denoted by $\mathrm{int}(\mathcal{A})$. The smallest closed set containing a set $\mathcal{A}$ in a topological space is called the closure of $\mathcal{A}$ and is denoted by $\mathrm{cl}(\mathcal{A})$. The convex-hull of a subset $\mathcal{A}$ of a vector space $\mathcal{U}$ is denoted by $\mathrm{conv}(\mathcal{A}) = \{ax + (1-a)y : x, y \in \mathcal{A}, a \in [0,1]\}$.

[5]Characteristic function of $\boldsymbol{x}$ is the Fourier transform of $p_{\boldsymbol{x}}(\cdot)$.

Chapter 2

Peak-Constrained Scalar AWGN Channel

2.1 Capacity-Achieving Input Distributions

As we mentioned in the previous chapter, point-to-point additive white Gaussian noise channel under peak constraint at the transmitter, was first studied by Smith [2, 3]. In this chapter we try to briefly state the important claims, observations and steps of proof in [2]. The channel model is as following:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{z}, \tag{2.1}$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ denote the input and output of the channel, respectively, and the random variable $\boldsymbol{z} \sim \mathrm{N}(0, 1)$[1] represents the additive Gaussian noise. The information capacity of

[1] It is shown in [2] that normalizing noise variance to be equal to 1, does not affect our results.

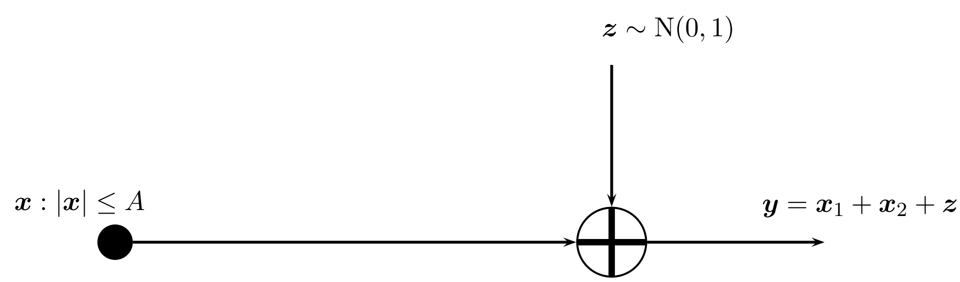

Figure 2.1: A point-to-point additive white Gaussian noise channel with peak power constraint at the transmitter

this channel is defined by

$$C \triangleq \sup_{\boldsymbol{x}:|\boldsymbol{x}|\leq A} I(\boldsymbol{x};\boldsymbol{y}). \tag{2.2}$$

Since $I(\boldsymbol{x};\boldsymbol{y}) = h(\boldsymbol{y}) - h(\boldsymbol{z})$, and $h(\boldsymbol{z})$ is a constant, one can rewrite (2.2) as

$$C = \sup_{\boldsymbol{x}:|\boldsymbol{x}|\leq A} h(\boldsymbol{y}). \tag{2.3}$$

So we need to find the supremum of the output differential entropy over all $F_{\boldsymbol{x}}(\cdot) \in \mathcal{F}_A$, where $\mathcal{F}_A$ denotes the set of all possible probability distribution functions with all points of increase in $[-A, A]$. By definition of differential entropy we have

$$h(\boldsymbol{y}) = -\int_{-\infty}^{\infty} p_{\boldsymbol{y}}(y) \log p_{\boldsymbol{y}}(y) dy. \tag{2.4}$$

Since $\boldsymbol{x}$ and $\boldsymbol{z}$ are independent random variables, we can write the output PDF as

$$p_{\boldsymbol{y}}(y) = \int_{-A}^{A} p_{\boldsymbol{z}}(y-x) dF_{\boldsymbol{x}}(x), \tag{2.5}$$

therefore by (2.4),

$$h(\boldsymbol{y}) = -\int_{-\infty}^{\infty} \int_{-A}^{A} p_{\boldsymbol{z}}(y-x) \log p_{\boldsymbol{y}}(y) dF_{\boldsymbol{x}}(x) dy. \tag{2.6}$$

Another notation for $h(\boldsymbol{y})$ that we use is $h(F_{\boldsymbol{x}})$ when $\boldsymbol{x} \sim F_{\boldsymbol{x}}(\cdot) \in \mathcal{F}_A$. It is shown in [2] that we are allowed to change the order of integration in (2.6),

$$h(\boldsymbol{y}) = \int_{-A}^{A} h(x; F_x) dF_{\boldsymbol{x}}(x), \tag{2.7}$$

where,

$$h(x; F_x) = -\int_{-\infty}^{\infty} p_{\boldsymbol{z}}(y - x) \log p_{\boldsymbol{y}}(y) dy. \tag{2.8}$$

$h(x; F_x)$ is called *marginal entropy function* and plays a crucial role in finding the necessary and sufficient conditions for the optimum distribution in (2.3). With these preliminaries in mind, we go to the next step which is applying an important optimization technique named "Karush-Kuhn-Tucker" (KKT) Theorem, to our problem. Here we state the KKT Theorem for completeness. We start by defining the concept of *directional derivative*.

Definition 1 *Let $f : \mathcal{V} \to \mathbb{R}$ be a function on a vector space $\mathcal{V}$. The directional derivative of f at $a \in \mathcal{V}$ along a direction $d \in \mathcal{V}$ is defined by*

$$\mathrm{D}_a(f; d) \triangleq \lim_{t \downarrow 0} \frac{1}{t}(f(a + td) - f(a)), \tag{2.9}$$

if this limit exists.

KKT Optimization Theorem: Let $f : \kappa \to \mathbb{R}$ be a continuous, weakly differentiable and strictly concave function from a convex and compact topological space κ, to $\mathbb{R}$. Then a unique $\omega_0 \in \kappa$ is the answer to the optimization problem

$$\sup_{\omega \in \kappa} f(\omega), \tag{2.10}$$

and the necessary and sufficient condition for ω_0 is $\mathrm{D}_{\omega_0}(f; \omega - \omega_0) \leq 0$ for all $\omega \in \kappa$. It is shown [3] that $h(\boldsymbol{y})$ and $\mathcal{F}_A$, satisfy all the conditions of the KKT Theorem. Therefore, by invoking this theorem to the optimization problem in (2.3), we have $F_0 \in \mathcal{F}_A$ to be the unique answer to (2.3), if and only if

$$\int_{-A}^{A} h(x; F_0) dF_{\boldsymbol{x}}(x) \leq h(F_0) \qquad \forall F_{\boldsymbol{x}}(\cdot) \in \mathcal{F}_A. \tag{2.11}$$

Let us denote the set of all points of increase of F_0 as $\mathcal{L}$. One can easily show that (2.11) is equivalent to the following:

$$h(x; F_0) \leq h(F_0) \qquad \forall x \in [-A, A], \tag{2.12}$$

$$h(x; F_0) = h(F_0) \qquad \forall x \in \mathcal{L}. \tag{2.13}$$

In the brilliant part of his work, Smith shows that the continuation of the marginal entropy function to the complex plane, denoted by $h(s; F_0)$ and $s \in \mathbb{C}$, is *analytic*[2] on the entire complex plane. Using a fundamental theorem in Complex Analysis, he proves finiteness of the number of the points of increase of F_0. We state this theorem here for completeness.

Identity Theorem in Complex Analysis: If two functions, analytic in some region of the complex plane, agree on an infinite set of points in that region and the set of points has a limit point in that region, then the functions are equal in that region.

Suppose F_0 has infinitely many points of increase. Base on Bolzano-Weierstrass Theorem [34], these points has to have a limit point in the interval $[-A, A]$. Invoking Identity Theorem of Complex Analysis, it is shown [2, 3] that this assumption leads to a contradiction. Therefore, the optimum input distribution for scalar AWGN channel with peak constraint has to be unique and discrete with a finite number of mass points.

[2]An analytic function is a function that is locally given by a convergent power series [38].

Chapter 3

Channel Models

3.1 Multiple Access Channel

In this chapter we want to give a brief introduction to the Multiple Access Channel (MAC) and the important notion of *Capacity Region*. MAC is referred to the channel where multiple transmitters are sending data to a single receiver simultaneously and over a "shared" noisy channel. The receiver is supposed to decode the data of all the users with arbitrary small probability of error. An error occurs when the receiver makes a wrong decision about the transmitted message of at least one of the users.

Consider a two user[1] MAC as shown in figure (3.1). User i is transmitting data with rate $R_i \geq 0$, for $i = 1, 2$, i.e. the codebook for i^{th} user contains 2^{nR_i} codewords of length n. A rate tuple (R_1, R_2) is said to be *achievable* if there exists codebooks for both users such that by making the length of the codewords large enough we can get arbitrary small probability of error at the receiver [25]. The following notions are important in defining the capacity region of MAC.

[1] All the definitions in this chapter can be extended to n-user multiple access channel.

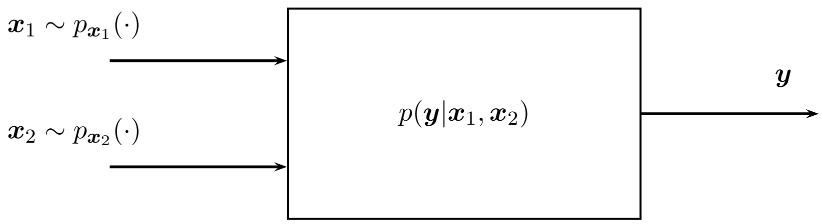

Figure 3.1: A two-user multiple access channel

Time Sharing: if (R_1, R_2) and $(R_1^{'}, R_2^{'})$ are achievable rate tuples for a two user multiple access channel, then $(\lambda R_1 + (1-\lambda)R_1^{'}, \lambda R_2 + (1-\lambda)R_2^{'})$ is also achievable for all $\lambda \in [0, 1]$.

Definition 2 *For some fixed input distribution $(\boldsymbol{x}_1, \boldsymbol{x}_2) \sim p_{\boldsymbol{x}_1}(\cdot)p_{\boldsymbol{x}_2}(\cdot)$, in a two user multiple access channel (figure (3.1)), the capacity region is characterized by*

$$\begin{aligned} 0 \leq R_1 \leq& I(\boldsymbol{x}_1; \boldsymbol{y}|\boldsymbol{x}_2) \\ 0 \leq R_2 \leq& I(\boldsymbol{x}_2; \boldsymbol{y}|\boldsymbol{x}_1). \\ R_1 + R_2 \leq& I(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{y}) \end{aligned} \tag{3.1}$$

This region outlines a pentagon shown in figure (3.2). Therefore, all the rate tuples lying in this pentagon are achievable for those particular fixed input distributions.

Using time-sharing and definition (2), one can say that the convex hull of the union of the capacity regions for all possible input distributions $(\boldsymbol{x}_1, \boldsymbol{x}_2) \sim p_{\boldsymbol{x}_1}(\cdot)p_{\boldsymbol{x}_2}(\cdot)$, is also achievable for multiple access channel [25]. It can also be shown that this region gives an upper bound on the achievable rate tuples, i.e. for any rate tuple at the boundary of the region, one cannot increase a user's rate without sacrificing another user's rate.

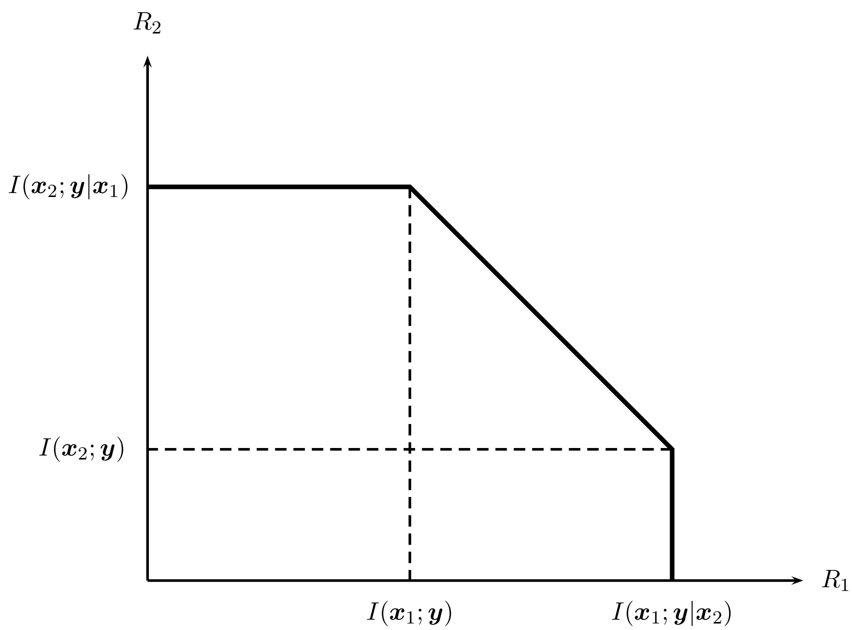

Figure 3.2: The capacity region of a two user MAC for some fixed input distribution $(\boldsymbol{x}_1, \boldsymbol{x}_2) \sim p_{\boldsymbol{x}_1}(\cdot)p_{\boldsymbol{x}_2}(\cdot)$.

Single-letter characterization of the *Capacity Region* of the multiple access channel was first derived in [26, 27]. This region characterizes the fundamental limits on the achievable data rates in the channel.

Capacity Region: The convex hull of the union of all the achievable rates (R_1, R_2), characterized by (3.1), for all of the possible product probability distributions $p_{\boldsymbol{x}_1}(\cdot)p_{\boldsymbol{x}_2}(\cdot)$ is the capacity region of the multiple access channel.

Capacity region of the multiple access channel has been characterized for different scenarios. References [28]-[31] introduce some interesting works for the multiple access channel under different assumptions.

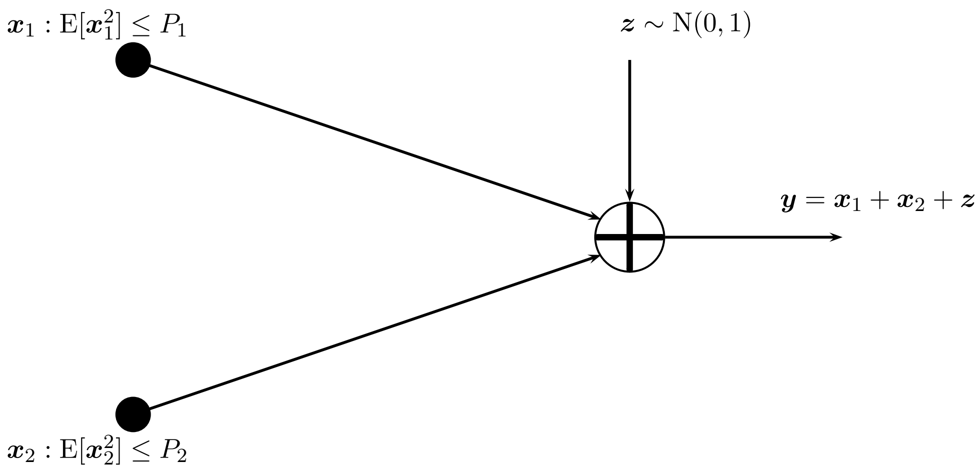

Figure 3.3: A two-user Gaussian multiple access channel with average power constraints

3.2 Gaussian Multiple Access Channel with Average Power Constraints

Consider a Multiple Access Channel with additive Gaussian noise, as shown in figure (3.3). Under average power constraints at the transmitters, the capacity region is a single pentagon characterized by

$$\begin{aligned} 0 \leq R_1 &\leq \frac{1}{2}\log(1+P_1) \\ 0 \leq R_2 &\leq \frac{1}{2}\log(1+P_2). \\ R_1 + R_2 &\leq \frac{1}{2}\log(1+P_1+P_2) \end{aligned} \tag{3.2}$$

This was first shown in [32, 33]. The reason for the simple structure of the capacity region is that, all of the constraints in (3.1) are maximized simultaneously by choosing $p_{\boldsymbol{x}_1}(\cdot)$ and $p_{\boldsymbol{x}_2}(\cdot)$ to be Gaussian distributions with variances P_1 and P_2, respectively. Therefore, the

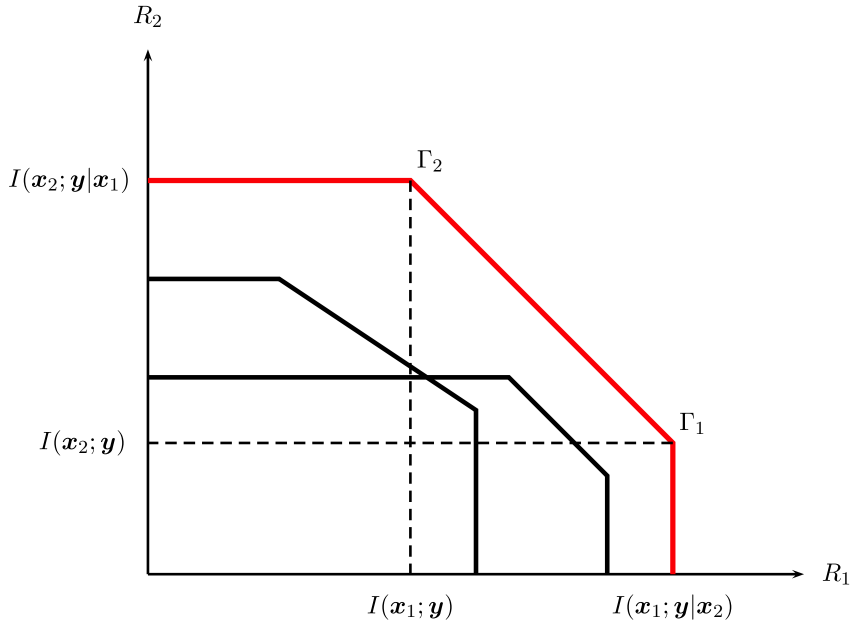

Figure 3.4: Red pentagon: Border of the capacity region of a two-user Gaussian multiple access channel with average power constraints; Black pentagons: examples of capacity regions corresponding to non-Gaussian input distributions

capacity region of this channel is of the form in figure (3.4). We know that if we add two independent random variables with Gaussian probability distributions, the result will also be Gaussian with a variance equal to the summation of the variances of the two primary random variables. Having this in mind, we can have a nice interpretation of the rate tuples at the corner points of the capacity region.

Consider the two-user scenario shown in figure (3.3). Suppose that the first user starts transmitting data with its maximum possible data rate while the second user is silent. In this case the first user can transmit data with rate $R_1 = I(\boldsymbol{x}_1;\boldsymbol{y}|\boldsymbol{x}_2)$. Now if we look at the point-to-point channel from the second user to the output of the channel when the signal from the first user is considered as noise, we will see a Gaussian channel with the noise variance equal to $1 + P_1$. Therefore the second user cannot send data with rates more that $R_2 = I(\boldsymbol{x}_2; \boldsymbol{y})$. This rate tuple (R_1, R_2) represents the corner point of the capacity region shown by Γ_1 in figure (3.4). The same interpretation holds for the point Γ_2, and by time-sharing, we can achieve any point on the line between Γ_1 and Γ_2.

Reference [22] proposes a code-division multiple-access scheme called rate-splitting multiple accessing (RSMA) for M-user Gaussian MAC. RSMA splits the M original sources into at most $2M - 1$ virtual sources and shows that the effort of finding the codes for the M users of MAC, is that of at most $2M - 1$ independent scalar Gaussian channels.

3.3 Problem Statement: Gaussian Multiple Access Channel with Peak Constraints

In this thesis we consider the problem of characterizing optimum input distributions for Gaussian multiple access channel with peak constraints at the transmitters (figure (3.5)), that achieve data rates at the boundary of the capacity region. Although the form of the capacity region is unknown, but we can show that only discrete input distributions with finite number of mass points can achieve rates at the boundary of the region. In the next chapter we prove those characteristics for input distributions that achieve the maximum sum-rate in the network. Later on in chapter 5, we extend our results to all weighted sum-rates which is equivalent to all of the points on the boundary of the capacity region.

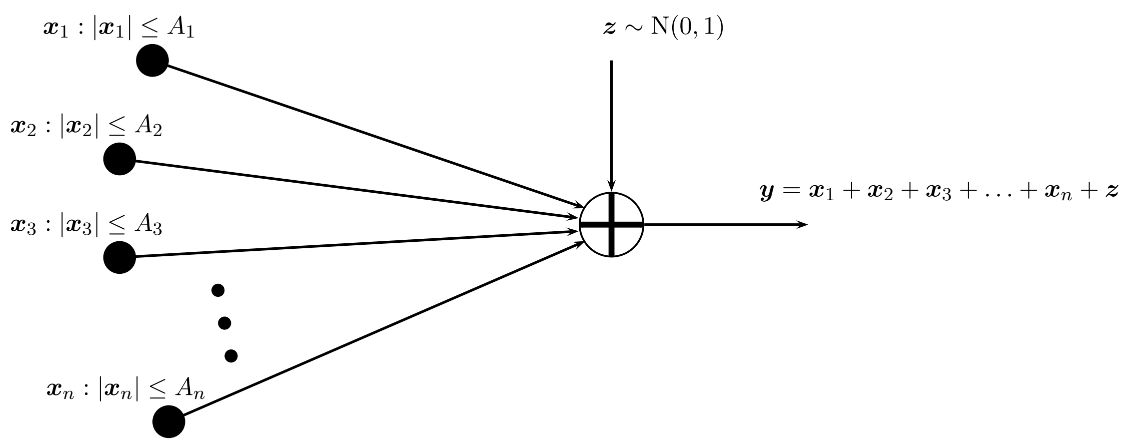

Figure 3.5: n-user Gaussian multiple access channel with peak constraints

Chapter 4

Sum-Capacity of Gaussian MAC with Peak Constraints

In this chapter, we consider a two-user Gaussian MAC under peak constraints at both transmitters[1] as shown in Fig. 4.1. Let $\boldsymbol{x}_i$ denote the signal transmitted by the i^{th} user on a certain transmission slot for $i = 1, 2$. The signal received at the common receiver is given by

$$\boldsymbol{y} = \boldsymbol{x}_1 + \boldsymbol{x}_2 + \boldsymbol{z}, \tag{4.1}$$

where $\boldsymbol{z} \sim \mathrm{N}(0, 1)$ is the additive noise at the common receiver. Note that $\boldsymbol{x}_1$, $\boldsymbol{x}_2$ and $\boldsymbol{z}$ are independent random variables. The transmitted signal by the i^{th} user is required to lie within the interval $[-A_i, A_i]$ for $i = 1, 2$ where $A_1, A_2 > 0$. The main part of the chapter is devoted to look for an optimal choice $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ such that the sum rate in the network is maximized, i.e.,

$$\boldsymbol{x}_1^*, \boldsymbol{x}_2^* = \arg \max_{\boldsymbol{x}_i : |\boldsymbol{x}_i| \leq A_i,\, i=1,2} I(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{y}). \tag{4.2}$$

[1]Our results can be easily extended to a Gaussian MAC with any finite number of users.

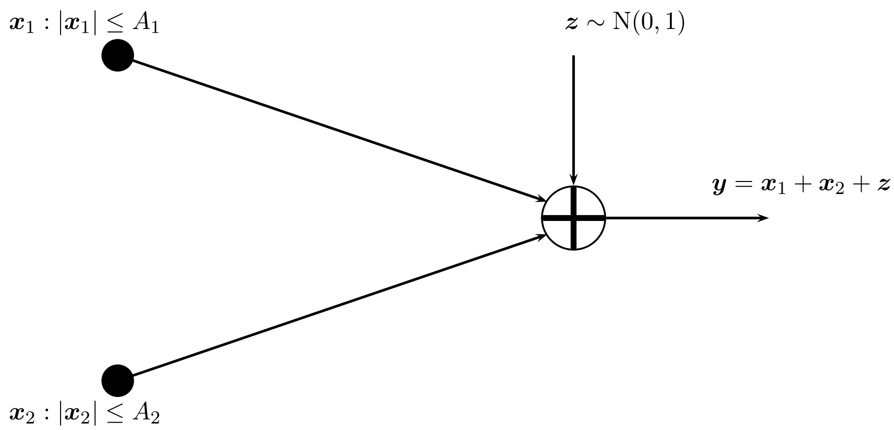

Figure 4.1: A two-user Gaussian MAC under peak constraints

Note that there might be more than one choice of $\boldsymbol{x}_1^*$ and $\boldsymbol{x}_2^*$ that satisfy (4.2). Since $I(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{y}) = h(\boldsymbol{y}) - h(\boldsymbol{z})$, one can alternatively write (4.2) as

$$\boldsymbol{x}_1^*, \boldsymbol{x}_2^* = \arg \max_{\boldsymbol{x}_i : |\boldsymbol{x}_i| \leq A_i,\, i=1,2} h(\boldsymbol{y}). \tag{4.3}$$

The main contribution here is that selecting $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ to be discrete random variables with a finite number of mass points is an answer to the optimization problem in (4.3). Although, we do not claim uniqueness for the answer to (4.3), however, we show that any answer to (4.3) must be discrete with a finite number of mass points. Let us fix $\boldsymbol{x}_1 = \boldsymbol{x}_1^*$. Note that the distribution of $\boldsymbol{x}_1^*$ is unknown at this point. Define $\widetilde{\boldsymbol{x}}_2$ by

$$\widetilde{\boldsymbol{x}}_2 \triangleq \arg \max_{\boldsymbol{x}_2 : |\boldsymbol{x}_2| \leq A_2} h(\boldsymbol{x}_1^* + \boldsymbol{x}_2 + \boldsymbol{z}). \tag{4.4}$$

Therefore,

$$h(\boldsymbol{x}_1^* + \boldsymbol{x}_2^* + \boldsymbol{z}) \leq h(\boldsymbol{x}_1^* + \widetilde{\boldsymbol{x}}_2 + \boldsymbol{z}). \tag{4.5}$$

According to (4.3),

$$h(\boldsymbol{x}_1^* + \boldsymbol{x}_2^* + \boldsymbol{z}) \geq h(\boldsymbol{x}_1^* + \widetilde{\boldsymbol{x}}_2 + \boldsymbol{z}). \tag{4.6}$$

Comparing (4.5) and (4.6), $h(\boldsymbol{x}_1^* + \boldsymbol{x}_2^* + \boldsymbol{z}) = h(\boldsymbol{x}_1^* + \widetilde{\boldsymbol{x}}_2 + \boldsymbol{z})$. However, it is shown in Theorem 1 that the answer to (4.4) is a *unique* discrete random variable $\widetilde{\boldsymbol{x}}_2$ with a finite number of mass points. Hence, $\boldsymbol{x}_2^* = \widetilde{\boldsymbol{x}}_2$ i.e. they have to have the same probability distribution functions. This shows any $\boldsymbol{x}_2^*$ satisfying (4.3) must be discrete with a finite number of mass points. A similar argument can be applied to verify the same property for $\boldsymbol{x}_1^*$.

Theorem 1 *Let $\boldsymbol{u}$ be a random variable with support $[-A, A]$ for some $A > 0$ and $\boldsymbol{z} \sim \mathrm{N}(0,1)$. For any $B > 0$, a unique and discrete random variable $\boldsymbol{x}$ with a finite number of mass points in $[-B, B]$ is the answer to the optimization problem $\sup_{\boldsymbol{x}:|\boldsymbol{x}|\leq B} I(\boldsymbol{x}; \boldsymbol{x}+\boldsymbol{u}+\boldsymbol{z})$.*

4.1 Proof of Theorem 1

In order to prove our result, we extend the approach taken in [2, 3]. The answer to our optimization problem in Theorem 1 is in fact the capacity of a point-to-point scalar additive white noise channel $\boldsymbol{v} = \boldsymbol{x} + \boldsymbol{u} + \boldsymbol{z}$ under the constraint $|\boldsymbol{x}| \leq B$ where $\boldsymbol{x}$ and $\boldsymbol{v}$ represent the input and output of the channel, respectively. Let us denote the set of CDFs for random variables whose points of increase lie in $[-B, B]$ by $\mathcal{F}_B$. Also, $p_{\boldsymbol{v}}(\cdot; F)$, $\Phi_{\boldsymbol{v}}(\cdot; F)$ and $h(\boldsymbol{v}; F)$ denote the PDF of $\boldsymbol{v}$, the characteristic function of $\boldsymbol{v}$ and the differential entropy of $\boldsymbol{v}$, respectively, when $\boldsymbol{x}$ is generated according to $F_{\boldsymbol{x}} = F$ for some $F \in \mathcal{F}_B$. We first show that $I(\boldsymbol{x}; \boldsymbol{x} + \boldsymbol{u} + \boldsymbol{z})$ achieves its supremum over $\mathcal{F}_B$. Note that $\arg\max_{\boldsymbol{x}:|\boldsymbol{x}|\leq B} I(\boldsymbol{x}; \boldsymbol{x} + \boldsymbol{u} + \boldsymbol{z}) = \arg\max_{F\in\mathcal{F}_B} h(\boldsymbol{v}; F)$. As such, we focus on the optimization problem $\sup_{F\in\mathcal{F}_B} h(\boldsymbol{v}; F)$. Our first observation is the following:

Lemma 1 $h(\boldsymbol{v}; \cdot) : \mathcal{F}_B \to \mathbb{R}$ *achieves its supremum over $\mathcal{F}_B$.*

Proof: For this purpose, it is enough to equip $\mathcal{F}_B$ with a topology under which $\mathcal{F}_B$ is compact and $h(\boldsymbol{v};\cdot) : \mathcal{F}_B \to \mathbb{R}$ is continuous. To proceed, we need the following definitions [35]:

Definition 3 *For any $F_1, F_2 \in \mathcal{F}_B$, the so-called Lévy metric is defined by*

$$d_L(F_1, F_2) \triangleq \inf \{\epsilon > 0 : F_1(x-\epsilon) - \epsilon \leq F_2(x) \leq F_1(x+\epsilon) + \epsilon, \textit{for any } x \in \mathbb{R}\} . \quad (4.7)$$

Definition 4 *A sequence $(F_n)_{n\in\mathbb{N}}$ of CDFs in $\mathcal{F}_B$ is said to converge weakly to $F_* \in \mathcal{F}_B$ if for all continuous functions $f : [-B, B] \to \mathbb{R}$, we have $\lim_n \int_{-B}^{B} f \mathrm{d}F_n = \int_{-B}^{B} f \mathrm{d}F_*$.*

We need the following technical result [35][2].

Helly-Bray Theorem: The Lévy metric $d_L : \mathcal{F}_B \times \mathcal{F}_B \to [0, \infty)$ is a metric on $\mathcal{F}_B$ that metrizes weak convergence of sequences in $\mathcal{F}_B$, i.e., $\lim_n \int_{-B}^{B} f \mathrm{d}F_n = \int_{-B}^{B} f \mathrm{d}F_*$ for any continuous function $f : [-B, B] \to \mathbb{R}$ if and only if $\lim_n d_L(F_n, F_*) = 0$ for any sequence $(F_n)_{n\in\mathbb{N}}$ and F_* in $\mathcal{F}_B$.
According to Lemma 8.10 in [36], the metric space $(\mathcal{F}_B, d_L)$ is sequentially compact[3]and hence, it is compact[4]. We remark that compactness of $\mathcal{F}_B$ under the Lévy metric is also shown in [3] based on a different approach.

It remains to verify continuity of $h(\boldsymbol{v};\cdot) : \mathcal{F}_B \to \mathbb{R}$. Defining the *total additive noise* by

$$\boldsymbol{w} \triangleq \boldsymbol{u} + \boldsymbol{z}, \quad (4.8)$$

[2]See Theorem 11.3.3 and problem 8 of section 11.3 in [35].

[3]A metric space $(\mathcal{X}, d)$ is called sequentially compact if any sequence $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{X}$ has a convergent subsequence.

[4]See Theorem 28.2 in [39] where it is shown that a metric space is compact if and only if it is sequentially compact.

we have

$$p_{\boldsymbol{v}}(v;F) = \int_{-B}^{B} p_{\boldsymbol{w}}(v-x)\mathrm{d}F(x). \tag{4.9}$$

Lemma 2 $p_{\boldsymbol{w}} : \mathbb{R} \to \mathbb{R}$ *is continuous.*

Proof: Note that

$$p_{\boldsymbol{w}}(w) = \int_{-A}^{A} p_{\boldsymbol{z}}(w-u)\mathrm{d}F_{\boldsymbol{u}}(u). \tag{4.10}$$

Let $(w_n)_{n\in\mathbb{N}} \subset \mathbb{R}$ and $\lim_n w_n = w_*$ for some $w_* \in \mathbb{R}$. Since $p_{\boldsymbol{z}} : \mathbb{R} \to \mathbb{R}$ is continuous, we have $\lim_n p_{\boldsymbol{z}}(w_n - u) = p_{\boldsymbol{z}}(w_* - u)$ for all $u \in [-A, A]$. Also, $p_{\boldsymbol{z}}(w_n - u) \leq c$ for all $u \in [-A, A]$. Noting that $\int_{-A}^{A} c \, \mathrm{d}F_{\boldsymbol{u}}(u) = c < \infty$, one can invoke the Lebesgue Dominated Convergence Theorem (LDCT) to deduce $\lim_n p_{\boldsymbol{w}}(w_n) = p_{\boldsymbol{w}}(w_*)$. This completes the proof. Here we state the Lebesgue Dominated Convergence Theorem (LDCT) for completeness [37].

Lebesgue Dominated Convergence Theorem (LDCT): Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of integrable functions which converges almost everywhere to a real valued measurable function f_*. If there exists an integrable function g such that $|f_n| \leq g$ for all n, then f_* is integrable and

$$\lim_n \int f_n \mathrm{d}\mu = \int f_* \mathrm{d}\mu. \tag{4.11}$$

Lemma 3 $p_{\boldsymbol{v}} : \mathbb{R} \times \mathcal{F}_B \to \mathbb{R}$ *is separately continuous in both arguments.*

Proof: Let us fix $F \in \mathcal{F}_B$. Using expression (4.9), proof of continuity for $p_{\boldsymbol{v}}(\cdot; F) : \mathbb{R} \to \mathbb{R}$ follows the same lines[5] as in the proof of continuity for $p_{\boldsymbol{w}} : \mathbb{R} \to \mathbb{R}$ offered in Lemma 2.

[5]Note that by Lemma 4, $p_{\boldsymbol{w}} : \mathbb{R} \to \mathbb{R}$ is continuous and by (4.10), $p_{\boldsymbol{w}}(w) \leq c$ for all $w \in \mathbb{R}$.

Next, let us fix $v \in \mathbb{R}$ and show that $p_{\boldsymbol{v}}(v; \cdot) : \mathcal{F}_B \to \mathbb{R}$ is continuous. Let $(F_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{F}_B$ such that $\lim_n d_L(F_n, F_*) = 0$ for $F_* \in \mathcal{F}_B$. Using continuity of $p_{\boldsymbol{w}} : \mathbb{R} \to \mathbb{R}$ (Lemma 2) and invoking Helly-Bray Theorem, $\lim_n \int_{-B}^{B} p_{\boldsymbol{w}}(v-x)\mathrm{d}F_n(x) = \int_{-B}^{B} p_{\boldsymbol{w}}(v-x)\mathrm{d}F_*(x)$, or equivalently, $\lim_n p_{\boldsymbol{v}}(v, F_n) = p_{\boldsymbol{v}}(v, F_*)$. This completes the proof.

We are ready to show that for any sequence $(F_n)_{n\in\mathbb{N}}$ and F_* in $\mathcal{F}_B$ such that $\lim_n d_L(F_n, F_*) = 0$, then $\lim_n h(\boldsymbol{v}; F_n) = h(\boldsymbol{v}; F_*)$. Based on Lemma 3, $\lim_n p_{\boldsymbol{v}}(v; F_n) = p_{\boldsymbol{v}}(v; F_*)$ for any $v \in \mathbb{R}$. Since $f : [0, \infty) \to \mathbb{R}$ defined by $f(\zeta) = \zeta \log \zeta$ for $\zeta > 0$ and $f(0) = 0$ is continuous, then

$$\lim_n p_{\boldsymbol{v}}(v; F_n) \log p_{\boldsymbol{v}}(v; F_n) = p_{\boldsymbol{v}}(v; F) \log p_{\boldsymbol{v}}(v; F). \tag{4.12}$$

If we can show there exists a function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $|p_{\boldsymbol{v}}(v; F_n) \log p_{\boldsymbol{v}}(v; F_n)| \le \varphi(v)$ for any $v \in \mathbb{R}$ and $\int_{\mathbb{R}} \varphi(v)\mathrm{d}v < \infty$, we can invoke LDCT to conclude $\lim_n h(\boldsymbol{v}; F_n) = h(\boldsymbol{v}; F_*)$. As such, the next step is to construct such a function φ. By (4.10), $p_{\boldsymbol{w}}(w) = \int_{-A}^{A} ce^{-\frac{1}{2}(w-u)^2}\mathrm{d}F_{\boldsymbol{u}}(u)$. This yields

$$\theta_A(w) \le p_{\boldsymbol{w}}(w) \le \Theta_A(w), \tag{4.13}$$

where

$$\theta_A(w) \triangleq \min_{u\in[-A,A]} ce^{-\frac{1}{2}(w-u)^2} = \begin{cases} ce^{-\frac{1}{2}(w+A)^2} & w \ge 0 \\ ce^{-\frac{1}{2}(w-A)^2} & w < 0 \end{cases} \tag{4.14}$$

and

$$\Theta_A(w) \triangleq \max_{u\in[-A,A]} ce^{-\frac{1}{2}(w-u)^2} = \begin{cases} ce^{-\frac{1}{2}(w-A)^2} & w \ge A \\ c & -A < w < A \\ ce^{-\frac{1}{2}(w+A)^2} & w \le -A \end{cases}. \tag{4.15}$$

Also, using (4.13) in (4.9),

$$\min_{x\in[-B,B]} \theta_A(v-x) \le p_{\boldsymbol{v}}(v; F_n) \le \max_{x\in[-B,B]} \Theta_A(v-x). \tag{4.16}$$

It is easy to see that

$$\min_{x\in[-B,B]} \theta_A(v-x) = \theta_{A+B}(v) \tag{4.17}$$

and

$$\max_{x\in[-B,B]} \Theta_A(v-x) = \Theta_{A+B}(v). \tag{4.18}$$

Hence,

$$\theta_{A+B}(v) \le p_{\boldsymbol{v}}(v; F_n) \le \Theta_{A+B}(v). \tag{4.19}$$

Remark 1 *Let $f : \mathbb{R} \to (0, \infty)$ be a bounded function, i.e., $0 < f(v) \le b < \infty$ for any $v \in \mathbb{R}$ and some $b > 0$. For any $v \in \mathbb{R}$,*

$$|\log f(v)| + \log f(v) = \begin{cases} 0 & \log f(v) \le 0 \\ 2\log f(v) & \log f(v) > 0 \end{cases} \tag{4.20}$$

and

$$\log f(v) \le \log b. \tag{4.21}$$

Combining (4.20) and (4.21),

$$|\log f(v)| \le -\log f(v) + 2|\log b|. \tag{4.22}$$

We use (4.22) in several instances throughout the proof.

By (4.16) and Remark 1,

$$\begin{aligned} |-p_{\boldsymbol{v}}(v; F_n) \log p_{\boldsymbol{v}}(v; F_n)| &\le \Theta_{A+B}(v)|\log p_{\boldsymbol{v}}(v; F_n)| \\ &\le \Theta_{A+B}(v)\left(-\log p_{\boldsymbol{v}}(v; F_n) + 2|\log c|\right) \\ &\le \Theta_{A+B}(v)\left(-\log \theta_{A+B}(v) + 2|\log c|\right). \end{aligned} \tag{4.23}$$

As $\Theta_{A+B}(v)$ decays like e^{-v^2} and $\log\theta_{A+B}(v)$ only grows like v^2, it is easy to see that

$$\varphi(v) \triangleq \Theta_{A+B}(v)\left(-\log\theta_{A+B}(v) + 2|\log c|\right) \tag{4.24}$$

is an integrable function.

$$\int_{\mathbb{R}} \varphi(v)\mathrm{d}v = -\int_{\mathbb{R}} \Theta_{A+B}(v)\log\theta_{A+B}(v)\mathrm{d}v + 2|\log c|\int_{\mathbb{R}} \Theta_{A+B}(v)\mathrm{d}v. \tag{4.25}$$

On one hand,

$$\begin{aligned}
\int_{\mathbb{R}} \Theta_{A+B}(v)\mathrm{d}v &= c\int_{-\infty}^{-A-B} e^{-\frac{1}{2}(v+A+B)^2}\mathrm{d}v + c\int_{A+B}^{\infty} e^{-\frac{1}{2}(v-A-B)^2}\mathrm{d}v + c\int_{-A-B}^{A+B}\mathrm{d}v \\
&\leq c\int_{-\infty}^{\infty} e^{-\frac{1}{2}(v+A+B)^2}\mathrm{d}v + c\int_{\infty}^{\infty} e^{-\frac{1}{2}(v-A-B)^2}\mathrm{d}v + 2c(A+B) \\
&= 2c\int_{-\infty}^{\infty} e^{-\frac{v^2}{2}}\mathrm{d}v + 2c(A+B) \\
&= 2c\left(\sqrt{2\pi} + A + B\right) < \infty.
\end{aligned} \tag{4.26}$$

On the other hand,

$$
\begin{aligned}
&\int_{\mathbb{R}} \Theta_{A+B}(v)\log\theta_{A+B}(v)\mathrm{d}v = \log(c)\int_{\mathbb{R}}\Theta_{A+B}(v)\mathrm{d}v - \frac{1}{2}\int_0^\infty \Theta_{A+B}(v)(v+A+B)^2\mathrm{d}v \\
&-\frac{1}{2}\int_{-\infty}^0 \Theta_{A+B}(v)(v-A-B)^2\mathrm{d}v \\
= \;& \log(c)\int_{\mathbb{R}}\Theta_{A+B}(v)\mathrm{d}v - \frac{1}{2}\int_0^{A+B}\Theta_{A+B}(v)(v+A+B)^2\mathrm{d}v \\
&-\frac{1}{2}\int_{-A-B}^0 \Theta_{A+B}(v)(v-A-B)^2\mathrm{d}v \\
&-\frac{1}{2}\int_{A+B}^\infty \Theta_{A+B}(v)(v+A+B)^2\mathrm{d}v - \frac{1}{2}\int_{-\infty}^{-A-B}\Theta_{A+B}(v)(v-A-B)^2\mathrm{d}v \\
= \;& \log(c)\int_{\mathbb{R}}\Theta_{A+B}(v)\mathrm{d}v - \frac{c}{2}\int_0^{A+B}(v+A+B)^2\mathrm{d}v - \frac{c}{2}\int_{-A-B}^0 (v-A-B)^2\mathrm{d}v \\
&-\frac{c}{2}\int_{A+B}^\infty (v+A+B)^2 e^{-\frac{1}{2}(v-A-B)^2}\mathrm{d}v - \frac{c}{2}\int_{-\infty}^{-A-B}(v-A-B)^2 e^{-\frac{1}{2}(v+A+B)^2}\mathrm{d}v \\
= \;& \log(c)\int_{\mathbb{R}}\Theta_{A+B}(v)\mathrm{d}v - \frac{7c(A+B)^3}{3} \\
&-\frac{c}{2}\int_{A+B}^\infty (v+A+B)^2 e^{-\frac{1}{2}(v-A-B)^2}\mathrm{d}v - \frac{c}{2}\int_{-\infty}^{-A-B}(v-A-B)^2 e^{-\frac{1}{2}(v+A+B)^2}\mathrm{d}v.
\end{aligned}
\tag{4.27}
$$

This yields

$$
\begin{aligned}
&\left|\int_{\mathbb{R}} \Theta_{A+B}(v)\log\theta_{A+B}(v)\mathrm{d}v\right| \le |\log c|\int_{\mathbb{R}}\Theta_{A+B}(v)\mathrm{d}v + \frac{7c(A+B)^3}{3} \\
&+\frac{c}{2}\int_{A+B}^{\infty}(v+A+B)^2 e^{-\frac{1}{2}(v-A-B)^2}\mathrm{d}v + \frac{c}{2}\int_{-\infty}^{-A-B}(v-A-B)^2 e^{-\frac{1}{2}(v+A+B)^2}\mathrm{d}v \\
\overset{(a)}{\le}\;& 2c\left(\sqrt{2\pi}+A+B\right)|\log c| + \frac{7c(A+B)^3}{3} \\
&+\frac{c}{2}\int_{-\infty}^{\infty}(v+A+B)^2 e^{-\frac{1}{2}(v-A-B)^2}\mathrm{d}v + \frac{c}{2}\int_{-\infty}^{\infty}(v-A-B)^2 e^{-\frac{1}{2}(v+A+B)^2}\mathrm{d}v \\
=\;& 2c\left(\sqrt{2\pi}+A+B\right)|\log c| + \frac{7c(A+B)^3}{3} \\
&+\frac{c}{2}\int_{-\infty}^{\infty}(v+2(A+B))^2 e^{-\frac{v^2}{2}}\mathrm{d}v + \frac{c}{2}\int_{-\infty}^{\infty}(v-2(A+B))^2 e^{-\frac{v^2}{2}}\mathrm{d}v \\
=\;& 2c\left(\sqrt{2\pi}+A+B\right)|\log c| + \frac{7c(A+B)^3}{3} + c\int_{-\infty}^{\infty}\left(v^2+4(A+B)^2\right)e^{-\frac{v^2}{2}}\mathrm{d}v \\
=\;& 2c\left(\sqrt{2\pi}+A+B\right)|\log c| + \frac{7c(A+B)^3}{3} + \sqrt{2\pi}c(1+4(A+B)^2) < \infty, \qquad (4.28)
\end{aligned}
$$

where (a) follows by (4.26). By (4.25), (4.26) and (4.28), we conclude that φ is an integrable function. This completes the proof of continuity for $h(\boldsymbol{v};\cdot):\mathcal{F}_B\to\mathbb{R}$.

The next Lemma guarantees that $h(\boldsymbol{v};\cdot):\mathcal{F}_B\to\mathbb{R}$ attains its global maximum for a unique distribution in $\mathcal{F}_B$ and it admits no local maxima.

Lemma 4 $h(\boldsymbol{v};\cdot):\mathcal{F}_B\to\mathbb{R}$ *is strictly concave.*

Proof: It is evident that $\mathcal{F}_B$ is a convex set. It is also well-known [24] that $h(\boldsymbol{v};\cdot):\mathcal{F}_B\to\mathbb{R}$ is concave, i.e., for any $F_1,F_2\in\mathcal{F}_B$ and $\lambda\in[0,1]$,

$$h(\boldsymbol{v};\lambda F_1+(1-\lambda)F_2)\geq\lambda h(\boldsymbol{v};F_1)+(1-\lambda)h(\boldsymbol{v};F_2), \tag{4.29}$$

with equality if and only if $p_{\boldsymbol{v}}(v,F_1)=p_{\boldsymbol{v}}(v,F_2)$ for almost all $v\in\mathbb{R}$. Assuming $p_{\boldsymbol{v}}(v,F_1)=p_{\boldsymbol{v}}(v,F_2)$ for almost all $v\in\mathbb{R}$ and noting that $p_{\boldsymbol{v}}(v,F)\leq c$ for any $v\in\mathbb{R}$ and $F\in\mathcal{F}_B$, the Fourier transforms of $p_{\boldsymbol{v}}(\cdot,F_1)$ and $p_{\boldsymbol{v}}(\cdot,F_2)$ are identical, i.e., $\Phi_{\boldsymbol{v}}(\omega;F_1)=\Phi_{\boldsymbol{v}}(\omega;F_2)$ for all $\omega\in\mathbb{R}$. This yields $\Phi_{\boldsymbol{w}}(\omega)(\Phi_1(\omega)-\Phi_2(\omega))=0$ for all $\omega\in\mathbb{R}$ where $\Phi_i(\cdot)$ is the Fourier transform of the probability law on $\mathbb{R}$ induced by $F_i(\cdot)$, i.e., $\Phi_i(\omega)=\int_{\mathbb{R}}e^{jwx}\mathrm{d}F_i(x)$ for $i=1,2$. The following Lemma shows that zeros of $\Phi_{\boldsymbol{w}}(\cdot)$ as isolated.

Lemma 5 $\Phi_{\boldsymbol{w}}:\mathbb{R}\to\mathbb{C}$ *can be zero in at most a countably infinite set of isolated points.*

Proof: Let us consider the moment generating function $M_{\boldsymbol{w}}:\mathbb{C}\to\mathbb{C}$ of $\boldsymbol{w}$ defined by[6]

$$M_{\boldsymbol{w}}(s)\triangleq\int_{\mathbb{R}}e^{sw}\mathrm{d}F_{\boldsymbol{w}}(w),\ \ s\in\mathbb{C}. \tag{4.30}$$

Note that $M_{\boldsymbol{w}}(s)$ is defined for any $s\in\mathbb{C}$, i.e., $|M_{\boldsymbol{w}}(s)|<\infty$. In fact,

$$\begin{aligned}|M_{\boldsymbol{w}}(s)| &\leq \int_{\mathbb{R}}|e^{sw}|\mathrm{d}F_{\boldsymbol{w}}(w)\\ &= \int_{\mathbb{R}}e^{\mathrm{Re}(s)w}p_{\boldsymbol{w}}(w)\mathrm{d}w\\ &\leq \int_{\mathbb{R}}e^{\mathrm{Re}(s)w}\Theta_A(w)\mathrm{d}w. \end{aligned}\tag{4.31}$$

[6]Note that $\Phi_{\boldsymbol{w}}(\omega)=M_{\boldsymbol{w}}(j\omega)$ for all $\omega\in\mathbb{R}$.

It is straightforward to see that the right side in (4.31) is finite. Next, we show that $M_{\boldsymbol{w}}(\cdot)$ is an analytic function everywhere on $\mathbb{C}$. To show this we need Morera's Theorem [40] which is mentioned here for completeness:

Morera's Theorem: Let f be a continuous, complex valued function on a connected open set $\mathcal{D}$ in the complex plane. If $\oint_{\Delta} f \mathrm{d}s = 0$ for every triangular path Δ in $\mathcal{D}$, then f is analytic on $\mathcal{D}$.

Let us first show that $M_{\boldsymbol{w}}(\cdot)$ is a continuous function. Assume $(s_n)_{n\in\mathbb{N}}$ is a sequence in $\mathbb{C}$ such that $\lim_n s_n = s_* \in \mathbb{C}$. Therefore, there exists $b > 0$ such that $|s_n| \leq b$ for any $n \in \mathbb{N}$. Moreover, for any $w \in \mathbb{R}$, $|e^{s_n w}| = e^{\mathrm{Re}(s_n) w} \leq e^{b|w|}$. But,

$$\begin{aligned} \int_{\mathbb{R}} e^{b|w|} \mathrm{d}F_{\boldsymbol{w}}(w) &= \int_{\mathbb{R}} e^{b|w|} p_{\boldsymbol{w}}(x) \mathrm{d}w \\ &\leq \int_{\mathbb{R}} e^{b|w|} \Theta_A(w) \mathrm{d}w \\ &< \infty. \end{aligned} \tag{4.32}$$

The last step in (4.32) follows similar lines as in (4.24)-(4.28) where we showed integrability of φ. Hence, we can use LDCT to write $\lim_n M_{\boldsymbol{w}}(s_n) = M_{\boldsymbol{w}}(s_*)$, i.e., $M_{\boldsymbol{w}} : \mathbb{C} \to \mathbb{C}$ is continuous.

Let Δ be a triangular path in $\mathbb{C}$. Note that

$$\left| \oint_{\Delta} M_{\boldsymbol{w}}(s) \mathrm{d}s \right| \leq \mathrm{len}(\Delta) \max_{s\in\Delta} M_{\boldsymbol{w}}(s) < \infty, \tag{4.33}$$

where $\mathrm{len}(\Delta)$ is the length of Δ and we have used the fact that the continuous function $M_{\boldsymbol{w}}(\cdot)$ attains its maximum on Δ as Δ is a closed and bounded set of points, or equivalently, a compact set in $\mathbb{C}$. Then (4.33) allows us to invoke Tonelli-Fubini Theorem [35] that justifies exchanging the order of integration, i.e.,

$$\begin{aligned}\oint_{\Delta} M_{\boldsymbol{w}}(s)\mathrm{d}s &= \oint_{\Delta}\int_{\mathbb{R}} e^{sw}\mathrm{d}F_{\boldsymbol{w}}(w)\mathrm{d}s \\ &= \int_{\mathbb{R}}\oint_{\Delta} e^{sw}\mathrm{d}s\mathrm{d}F_{\boldsymbol{w}}(w) \\ &= 0, \end{aligned} \tag{4.34}$$

where the last step follows by the fact that $f:\mathbb{C}\to\mathbb{C}$ defined by $f(s)=e^{sw}$ is analytic on $\mathbb{C}$ for any $w\in\mathbb{R}$, and therefore, by Cauchy's integral Theorem, $\oint_{\Delta} e^{sw}\mathrm{d}s=0$. Hence, by Morera's Theorem, $M_{\boldsymbol{w}}(\cdot)$ is analytic everywhere on $\mathbb{C}$. It is well-known [40] that zeros of an analytic function are isolated. Since $\Phi_{\boldsymbol{w}}(\omega)=M_{\boldsymbol{w}}(j\omega)$, we conclude that $\Phi_{\boldsymbol{w}}(\cdot)$ can be zero for at most a countably infinite set of isolated points. Based on lemma 5, if we denote the isolated zeros (if any) of $\Phi_{\boldsymbol{w}}(\cdot)$ by $\omega_1,\omega_2,\omega_3,\cdots$, then $\Phi_1(\omega)=\Phi_2(\omega)$ for any $\omega\in\mathbb{R}$ except possibly for $\omega\in\{\omega_1,\omega_2,\omega_3,\cdots\}$. Fixing $n\geq 1$, let $(\omega_{n,m})_{m\in\mathbb{N}}$ be a sequence of rational numbers that tends to ω_n from above and $\{\omega_{n,m}: m\in\mathbb{N}\}\cap\{\omega_1,\omega_2,\omega_3,\cdots\}=\emptyset$. Then $\Phi_1(\omega_{n,m})=\Phi_2(\omega_{n,m})$ for any $m\in\mathbb{N}$. Using the fact that the characteristic function of any random variable is (uniformly) continuous[7] on $\mathbb{R}$, we get $\Phi_1(\omega_n)=\lim_m\Phi_1(\omega_{n,m})=\lim_m\Phi_2(\omega_{n,m})=\Phi_2(\omega_n)$ for any $n\in\mathbb{N}$. Therefore, $\Phi_1(\omega)=\Phi_2(\omega)$ for all $\omega\in\mathbb{R}$. This in turn yields $F_1=F_2$.

We are ready to present necessary and sufficient conditions for the unique distribution F_0 that satisfies $h(\boldsymbol{v};F_0)=\sup_{F\in\mathcal{F}_B}h(\boldsymbol{v};F)$. The following Lemma is essential for developing the rest of the proof.

Lemma 6 *Let $f:\mathcal{C}\to\mathbb{R}$ be a concave function where $\mathcal{C}$ is a convex set. For any $a\in\mathrm{int}(\mathcal{C})$ and for any $a'\in\mathcal{C}$, the directional derivative of f at a along $a'-a$ exists and is finite. Moreover, a is a point of global maximum for f if and only if $\mathrm{D}_a(f;a'-a)\leq 0$ for any $a'\in\mathcal{C}$.*

[7] See Lemma 9.3 in [36].

Proof: See Propositions 2.1.1, 2.1.2 and 3.1.2 in [34].

One can write $h(\boldsymbol{v}; F)$ as

$$h(\boldsymbol{v}; F) = \int_{-B}^{B} k(x; F)\, \mathrm{d}F(x), \quad F \in \mathcal{F}_B, \tag{4.35}$$

where $k : [-B, B] \times \mathcal{F}_B \to \mathbb{R}$ is given by

$$k(x; F) \triangleq -\int_{\mathbb{R}} p_{\boldsymbol{w}}(v - x) \log p_{\boldsymbol{v}}(v; F) \mathrm{d}v. \tag{4.36}$$

Note that existence and finiteness of $\mathrm{D}_{F_0}(h(\boldsymbol{v}; \cdot); F - F_0)$ is guaranteed by Lemma 6. Calculating the directional derivative $\mathrm{D}_{F_0}(h(\boldsymbol{v}; \cdot); F - F_0)$ is a straightforward task and we refer the reader to [3] for details[8]. In fact,

$$\mathrm{D}_{F_0}(h(\boldsymbol{v}; \cdot); F - F_0) = \int_{-B}^{B} k(x; F_0)\, \mathrm{d}F(x) - h(\boldsymbol{v}; F_0). \tag{4.37}$$

Using Lemma 6, the optimum distribution F_0 satisfies $\mathrm{D}_{F_0}(h(\boldsymbol{v}; \cdot); F - F_0) \leq 0$. As such, the necessary and sufficient condition for optimality of F_0 is given by

$$\int_{-B}^{B} k(x; F_0)\, \mathrm{d}F(x) \leq h(\boldsymbol{v}; F_0), \quad F \in \mathcal{F}_B. \tag{4.38}$$

Taking $F(x) = \mathbf{1}_{[y,\infty)}(x)$ in (4.38), it is seen that

$$k(y; F_0) \leq h(\boldsymbol{v}; F_0), \;\; y \in [-B, B]. \tag{4.39}$$

Let $\mathcal{I} \subset [-B, B]$ be the set of points of increase for F_0 and $x_0 \in \mathcal{I}$, i.e., $\mu_{F_0}(\mathcal{O}) > 0$ for any open set $\mathcal{O}$ containing x_0. Following the steps in [13], we show $k(x_0; F_0) = h(\boldsymbol{v}; F_0)$. Assume on the contrary that $k(x_0; F_0) - h(\boldsymbol{v}; F_0) = -\epsilon$ for some $\epsilon > 0$. It is verified in the appendix [9] that $k(\cdot; F) : \mathbb{R} \to \mathbb{R}$ is a continuous function for any $F \in \mathcal{F}_B$. Then there is

[8]See the Lemma that appears on page 29 of [3].

[9]In fact, appendix verifies a much stronger statement about $k(\cdot; F)$ where it is shown that continuation of $k(\cdot; F)$ to the whole complex plane is analytic everywhere on $\mathbb{C}$.

an open set $\mathcal{O}$ containing x_0 such that $k(x;F_0)-h(\boldsymbol{v};F_0)\leq -\frac{\epsilon}{2}$ for any $x\in\mathcal{O}$. Integrating both sides with respect to $\mu_{F_0}(\cdot)$, we get

$$\int_{\mathcal{O}}(k(x;F_0)-h(\boldsymbol{v};F_0))\mathrm{d}F_0(x)\leq -\frac{\epsilon}{2}\mu_{F_0}(\mathcal{O}). \tag{4.40}$$

On the other hand,

$$\begin{aligned}\int_{\mathcal{O}}(k(x;F_0)-h(\boldsymbol{v};F_0))\mathrm{d}F_0(x) &\overset{(a)}{\geq} \int_{\mathbb{R}}(k(x;F_0)-h(\boldsymbol{v};F_0))\mathrm{d}F_0(x)\\ &\overset{(b)}{=} \int_{\mathbb{R}}k(x;F_0)\mathrm{d}F_0(x)-h(\boldsymbol{v};F_0)\\ &\overset{(c)}{=} 0, \end{aligned} \tag{4.41}$$

where (a) follows by (4.38), (b) is due to the fact that $\int_{\mathbb{R}}\mathrm{d}F_0(x)=1$ and (c) is a consequence of (4.35). Combining (4.40) and (4.41), we get $\mu_{F_0}(\mathcal{O})\leq 0$ which is a contradiction. Hence,

$$k(x_0;F_0)=h(\boldsymbol{v};F_0),\quad x_0\in\mathcal{I}. \tag{4.42}$$

It is shown in the appendix that continuation of $k(\cdot;F)$ to the whole complex plain is well-defined and in fact, $k(\cdot;F)$ is analytic on the whole complex plain. It is well-known [40] that zeros of an analytic function are isolated points on its domain[10]. Therefore, the equation $k(s;F_0)=h(\boldsymbol{v};F_0)$ can have at most a countably infinite set of isolated solutions denoted by $\mathcal{S}$. We can write $\mathcal{I}\subset\mathcal{S}\cap[-B,B]$. If $\mathcal{I}$ is not finite, i.e., if it is countably infinite, it must have a point of accumulation x_* by Bolzano-Weierstrass Theorem [34]. Hence, there is a sequence of points $(x_n)_{n\in\mathbb{N}}$ in $\mathcal{I}$ such that $\lim_n x_n=x_*$. Since $k(x_n;F_0)=h(\boldsymbol{v};F_0)$ for all $n\in\mathbb{N}$ and $k(\cdot;F_0)$ is continuous, we get $k(x_*;F_0)=\lim_n k(x_n;F_0)=h(\boldsymbol{v};F_0)$. As such, $x_*\in\mathcal{I}$ which is a contradiction as all points in $\mathcal{I}$ are isolated. Therefore, $\mathcal{I}$ must be a *finite* set of isolated points. This completes the proof of Theorem 1.

[10] A set $\mathcal{S}\subset\mathbb{C}$ is said to be a set of isolated points if for any $s\in\mathcal{S}$, there is an open set $\mathcal{O}\subset\mathbb{C}$ containing s such that $\mathcal{S}\cap\mathcal{O}=\{s\}$.

Chapter 5

Maximizing the weighted sum rate under peak constraints

This chapter is devoted to study the weighed sum rate of a two-user Gaussian MAC. Let us denote the capacity region of a Gaussian MAC with peak constraints shown in Fig. 4.1 by $\mathcal{R}_{\mathrm{MAC}}(A_1, A_2)$. In fact, $\mathcal{R}_{\mathrm{MAC}}(A_1, A_2)$ is the set of all tuples (R_1, R_2) where R_i is the transmission rate of the i^{th} user such that the receiver can decode the messages of both users with arbitrarily small probability of error. By [24],

$$\mathcal{R}_{\mathrm{MAC}}(A_1, A_2) = \mathrm{cl}\left(\mathrm{conv}\left(\bigcup_{\boldsymbol{x}_i : |\boldsymbol{x}_i| \leq A_i, i=1,2} \mathcal{R}_{\mathrm{MAC}}(A_1, A_2; \boldsymbol{x}_1, \boldsymbol{x}_2)\right)\right), \tag{5.1}$$

where $\mathcal{R}_{\mathrm{MAC}}(A_1, A_2; \boldsymbol{x}_1, \boldsymbol{x}_2)$ is the set of all tuples (R_1, R_2) that satisfy

$$\begin{aligned} 0 \leq R_1 &\leq I(\boldsymbol{x}_1; \boldsymbol{y}|\boldsymbol{x}_2) \\ 0 \leq R_2 &\leq I(\boldsymbol{x}_2; \boldsymbol{y}|\boldsymbol{x}_1) \quad . \\ R_1 + R_2 &\leq I(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{y}) \end{aligned} \tag{5.2}$$

Any point on the boundary of $\mathcal{R}_{\mathrm{MAC}}(A_1, A_2)$ corresponds to a solution for $\arg\sup_{(R_1,R_2)\in\mathcal{R}_{\mathrm{MAC}}(A_1,A_2)} R_1 + mR_2$ for some $m > 0$. Using the structure of $\mathcal{R}_{\mathrm{MAC}}(A_1, A_2)$, one may alternatively describe any point on the boundary of $\mathcal{R}_{\text{MAC}}(A_1, A_2)$ by solving for $\arg\sup_{\boldsymbol{x}_i:|\boldsymbol{x}_i|\leq A_i,\, i=1,2} \sup_{(R_1,R_2)\in\mathcal{R}_{\text{MAC}}(A_1,A_2;\boldsymbol{x}_1,\boldsymbol{x}_2)} R_1 + mR_2$ for some $m > 0$. This leads to maximizing $I(\boldsymbol{x}_1;\boldsymbol{y}|\boldsymbol{x}_2) + mI(\boldsymbol{x}_2;\boldsymbol{y})$ and $I(\boldsymbol{x}_1;\boldsymbol{y}) + mI(\boldsymbol{x}_2;\boldsymbol{y}|\boldsymbol{x}_1)$ for $0 < m < 1$ and $m > 1$, respectively over the set of all $\boldsymbol{x}_i$ such that $|\boldsymbol{x}_i| \leq A_i$ for $i = 1, 2$. We show that the answer to these optimization problems are discrete random variables $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ with a finite number of mass points in $[-A_1, A_1]$ and $[-A_2, A_2]$, respectively. In what follows, we focus on the case $0 < m < 1$. The case $m > 1$ is treated similarly.

Let

$$\begin{aligned} \boldsymbol{x}_1^*, \boldsymbol{x}_2^* &\triangleq \arg \max_{\boldsymbol{x}_i:|\boldsymbol{x}_i|\leq A_i,\, i=1,2} I(\boldsymbol{x}_1;\boldsymbol{y}|\boldsymbol{x}_2) + mI(\boldsymbol{x}_2;\boldsymbol{y}) \\ &= \arg \max_{\boldsymbol{x}_i:|\boldsymbol{x}_i|\leq A_i,\, i=1,2} I(\boldsymbol{x}_1;\boldsymbol{x}_1+\boldsymbol{z}) + mI(\boldsymbol{x}_2;\boldsymbol{x}_2+\boldsymbol{x}_1+\boldsymbol{z}). \end{aligned} \tag{5.3}$$

By Theorem 1, the term $I(\boldsymbol{x}_2;\boldsymbol{x}_2+\boldsymbol{x}_1+\boldsymbol{z})$ is maximized for a discrete $\boldsymbol{x}_2$ with a finite number of mass points in $[-A_2, A_2]$ for any choice of $\boldsymbol{x}_1$ where $|\boldsymbol{x}_1| \leq A_1$. Therefore, $\boldsymbol{x}_2^*$ is a discrete random variable with a finite number of mass points in $[-A_2, A_2]$. Then

$$\begin{aligned} \boldsymbol{x}_1^* &= \arg \max_{\boldsymbol{x}_1:|\boldsymbol{x}_1|\leq A_1} I(\boldsymbol{x}_1;\boldsymbol{x}_1+\boldsymbol{z}) + mI(\boldsymbol{x}_2^*;\boldsymbol{x}_2^*+\boldsymbol{x}_1+\boldsymbol{z}) \\ &= \arg \max_{\boldsymbol{x}_1:|\boldsymbol{x}_1|\leq A_1} (1-m)h(\boldsymbol{x}_1+\boldsymbol{z}) + mh(\boldsymbol{x}_1+\boldsymbol{x}_2^*+\boldsymbol{z}). \end{aligned} \tag{5.4}$$

Let us define

$$\boldsymbol{w}_1 \triangleq \boldsymbol{z},\ \ \boldsymbol{w}_2 \triangleq \boldsymbol{x}_2^* + \boldsymbol{z} \tag{5.5}$$

and

$$\boldsymbol{v}_i \triangleq \boldsymbol{x}_1 + \boldsymbol{w}_i,\ \ i = 1, 2. \tag{5.6}$$

Following our previous notation, we are looking for $F_{\boldsymbol{x}_1} = F_0 \in \mathcal{F}_{A_1}$ given by

$$F_0 = \arg \max_{F\in\mathcal{F}_{A_1}} (1-m)h(\boldsymbol{v}_1;F) + mh(\boldsymbol{v}_2;F). \tag{5.7}$$

Mimicking the same lines of proof for Lemma 1 and Lemma 4, we can show that $(1 - m)h(\boldsymbol{v}_1; \cdot) + mh(\boldsymbol{v}_2; \cdot) : \mathcal{F}_{A_1} \to \mathbb{R}$ achieves its supremum over $\mathcal{F}_{A_1}$ and is strictly concave. Moreover, let $k_i : [-A_1, A_1] \times \mathcal{F}_{A_1} \to \mathbb{R}$ be given by

$$k_i(x; F) \triangleq -\int_{\mathbb{R}} p_{\boldsymbol{w}_i}(v - x) \log p_{\boldsymbol{v}_i}(v; F) \mathrm{d}v, \;\; i = 1, 2, \tag{5.8}$$

for any $x \in [-A_1, A_1]$ and $F \in \mathcal{F}_{A_1}$. Then one can write $h(\boldsymbol{v}_i; F)$ as

$$h(\boldsymbol{v}_i; F) = \int_{-A_1}^{A_1} k_i(x; F) \mathrm{d}F(x), \;\; i = 1, 2, \tag{5.9}$$

for any $F \in \mathcal{F}_{A_1}$. Invoking similar arguments as the ones in the appendix where we show $k(\cdot; F) : \mathbb{C} \to \mathbb{C}$ given in (4.36) is analytic on the whole complex plane, we can show that continuation of $k_i(\cdot; F)$ to the whole complex plane is everywhere analytic for any $F \in \mathcal{F}_{A_1}$ and $i = 1, 2$. Verifying that any point of increase x_0 for F_0 satisfies

$$(1 - m)k_1(x; F_0) + mk_2(x; F_0) = (1 - m)h(\boldsymbol{v}_1; F_0) + mh(\boldsymbol{v}_2; F_0), \tag{5.10}$$

using the same lines of reasoning that appear after (4.42), one can show that the points of increase for F_0 consist of only a finite number of isolated points in $[-A_1, A_1]$. This completes the proof for the fact that any point on the boundary of the region $\mathcal{R}_{\mathrm{MAC}}(A_1, A_2)$ can be achieved only by input distributions $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ whose CDFs have only a finite number of points of increase in the intervals $[-A_1, A_1]$ and $[-A_2, A_2]$, respectively.

Chapter 6

Conclusion

6.1 Summary

In this thesis we have studied Gaussian multiple access channels under peak constraints at the transmitters. We have characterized the optimum input distributions corresponding to the points at the boundary of the capacity region to be discrete with a finite number of points of increase. First we prove our claim for maximizing sum-capacity of the channel then we generalize the idea to the weighted sum-rate. This concludes that only discrete distributions with finite number of mass points achieve points at the boundary of the capacity region but does not claim uniqueness of those distributions.

Central part of the thesis is the arguments presented in chapter 4 where we break down the multiple access channel problem into a point-to-point problem with two additive noises, one is the additive Gaussian noise, and the other is an amplitude-limited noise with an arbitrary probability distribution function. In Theorem 1 we claimed that a unique and discrete distribution with a finite number of mass points achieves the capacity of this simpler channel. We adopted the standard methodology of Smith [2] to prove our claim in Theorem1.

6.2 Future works

We have characterized the optimum input distributions to achieve points on the boundary of the capacity region of a multiple access channel with peak constraints at the transmitters. One possible extension to this work would be characterizing the general form of the capacity region of this channel and in particular answering to the question that is there just a single point on the boundary of the capacity region that corresponds to the maximum sum-rate or there is a collection of infinitely many points that achieve the sum-capacity, i.e. the boundary of the capacity region contains a straight line that corresponds to the maximum sum-rate?

Other interesting question on this channel would be on the optimality of different schemes such as Time Division Multiple Access (TDMA) or Frequency Division Multiple Access (FDMA) to achieve points on the boundary of the capacity region. It has been shown for Gaussian MAC with average power constraints, that both schemes of TDMA and FDMA can achieve exactly one point on the boundary of the capacity region corresponding to maximum sum-rate. Proving optimality (at least at one point on the capacity region) or non-optimality of TDMA and FDMA in our MAC with peak constraints would be an interesting contribution to the topic.

Appendix

Fixing $F \in \mathcal{F}_B$, let us define $\rho : \mathbb{C} \to \mathbb{C}$ by

$$\rho(s) \triangleq \int_{\mathbb{R}} p_{\boldsymbol{w}}(v-s) \log p_{\boldsymbol{v}}(v;F) \mathrm{d}v. \tag{6.1}$$

Note that $k(x;F) = \rho(x)$ for $x \in \mathbb{R}$. We show that $\rho(\cdot)$ is analytic everywhere on $\mathbb{C}$. Let us start by verifying that $|\rho(s)| < \infty$ for any $s \in \mathbb{C}$. By (4.10),

$$|p_{\boldsymbol{w}}(v-s)| \le \int_{-A}^{A} |p_{\boldsymbol{z}}(v-s-u)| \mathrm{d}F_{\boldsymbol{u}}(u). \tag{6.2}$$

Since

$$|p_{\boldsymbol{z}}(v-s-u)| = ce^{-\frac{1}{2}\mathrm{Re}((v-s-u)^2)} = ce^{\frac{1}{2}(\mathrm{Im}(s))^2} e^{-\frac{1}{2}(v-\mathrm{Re}(s)-u)^2}, \tag{6.3}$$

we get

$$\begin{aligned} |p_{\boldsymbol{w}}(v-s)| &\le ce^{\frac{1}{2}(\mathrm{Im}(s))^2} \max_{u\in[-A,A]} e^{-\frac{1}{2}(v-\mathrm{Re}(s)-u)^2} \\ &= e^{\frac{1}{2}(\mathrm{Im}(s))^2} \Theta_A(v-\mathrm{Re}(s)), \end{aligned} \tag{6.4}$$

where we have used the definition of $\Theta_A(\cdot)$ in (4.15). Using (6.4) and recalling the inequality $|\log p_{\boldsymbol{v}}(v)| \le -\log \theta_{A+B}(v) + 2|\log c|$ from (4.23),

$$\begin{aligned} |\rho(s)| &\le \int_{\mathbb{R}} |p_{\boldsymbol{w}}(v-s)| |\log p_{\boldsymbol{v}}(v;F)| \mathrm{d}v \\ &\le e^{\frac{1}{2}(\mathrm{Im}(s))^2} \int_{\mathbb{R}} \Theta_A(v-\mathrm{Re}(s)) \left(-\log \theta_{A+B}(v) + 2|\log c|\right) \mathrm{d}v. \end{aligned} \tag{6.5}$$

Following a similar argument that we used to prove integrability of $\varphi(\cdot)$ defined in (4.24), it can be seen that the right side in (6.5) is finite. Hence, $|\rho(s)| < \infty$ for any $s \in \mathbb{C}$. To show that $\rho(\cdot)$ is an analytic function on $\mathbb{C}$, we invoke Morera's Theorem. For this purpose, we need the following Lemmas:

Lemma 7 *$\rho(\cdot) : \mathbb{C} \to \mathbb{C}$ is everywhere continuous.*

Proof: Let $(s_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{C}$ such that $\lim_n s_n = s_*$ for some $s_* \in \mathbb{C}$. Our goal is to show $\lim_n \rho(s_n) = \rho(s_*)$. We proceed as follows:

1) We show $\lim_n p_{\boldsymbol{w}}(v - s_n) = p_{\boldsymbol{w}}(v - s_*)$ for any $v \in \mathbb{R}$. To see this, note that $\lim_n p_{\boldsymbol{z}}(v - s_n - u) = p_{\boldsymbol{z}}(v - s_* - u)$ for any $u, v \in \mathbb{R}$ as $p_{\boldsymbol{z}}(\cdot)$ is continuous (in fact analytic) everywhere on $\mathbb{C}$. Moreover, since $(s_n)_{n\in\mathbb{N}}$ converges, there is a $b > 0$ such that $|s_n| \leq b$ for any $n \in \mathbb{N}$. By (6.3),

$$\begin{aligned}
|p_{\boldsymbol{z}}(v - s_n - u)| &\leq c \max_{s\in\mathbb{C}:|s|\leq b} e^{\frac{1}{2}(\mathrm{Im}(s))^2} e^{-\frac{1}{2}(v-\mathrm{Re}(s)-u)^2} \\
&\leq c e^{\frac{b^2}{2}} \max_{s\in\mathbb{C}:\mathrm{Re}(s)\in[-b,b]} e^{-\frac{1}{2}(v-\mathrm{Re}(s)-u)^2} \\
&= e^{\frac{b^2}{2}} \Theta_b(v-u), \;\; u, v \in \mathbb{R}, \;\; n \in \mathbb{N}. \qquad (6.6)
\end{aligned}$$

Since

$$\begin{aligned}
\int_{-A}^{A} \Theta_b(v-u) \mathrm{d}F_{\boldsymbol{u}}(u) &\leq \max_{u\in[-A,A]} \Theta_b(v-u) \\
&= \Theta_{b+A}(v) < \infty, \;\;\; v \in \mathbb{R}, \qquad (6.7)
\end{aligned}$$

we can apply LDCT to (4.10) in order to conclude $\lim_n p_{\boldsymbol{w}}(v - s_n) = p_{\boldsymbol{w}}(v - s_*)$.

2) We show that there exists a function $\phi : \mathbb{R} \to \mathbb{R}$ such that $|p_{\boldsymbol{w}}(v - s_n) \log p_{\boldsymbol{v}}(v; F)| \leq \phi(v)$ for any $v \in \mathbb{R}$ and $\int_{\mathbb{R}} \phi(v) \mathrm{d}v < \infty$. Using this fact and the previous step, we conclude

$\lim_n \rho(s_n) = \rho(s_*)$ by applying LDCT again. By (6.2), (6.6) and (6.7) and recalling the inequality $|\log p_{\boldsymbol{v}}(v)| \leq -\log \theta_{A+B}(v) + 2|\log c|$ from (4.23),

$$|p_{\boldsymbol{w}}(v - s_n) \log p_{\boldsymbol{v}}(v; F)| \leq e^{\frac{b^2}{2}} (-\log \theta_{A+B}(v) + 2|\log c|)\, \Theta_{b+A}(v), \ v \in \mathbb{R}. \quad (6.8)$$

Let us define

$$\phi(v) \triangleq e^{\frac{b^2}{2}} (-\log \theta_{A+B}(v) + 2|\log c|)\, \Theta_{b+A}(v), \ v \in \mathbb{R}. \quad (6.9)$$

Following our argument in (4.24)-(4.28) to prove integrability of $\varphi(\cdot)$, it can be seen that $\int_{\mathbb{R}} \phi(v) \mathrm{d}v$ is finite. This completes the proof of continuity for $\rho(\cdot)$.

Lemma 8 *For any $v \in \mathbb{R}$, define $f_v : \mathbb{C} \to \mathbb{C}$ by $f_v(s) = p_{\boldsymbol{w}}(v - s)$. Then f_v is analytic everywhere.*

Proof: By (6.4), f_v is well-defined, i.e., $|f_v(s)| < \infty$ for any $s \in \mathbb{C}$. Moreover, $f_v : \mathbb{C} \to \mathbb{C}$ is everywhere continuous by Lemma 7. Next, let Δ be an arbitrary triangular path in $\mathbb{C}$. Then we can find $b' > 0$ such that $|s| \leq b'$ for any $s \in \Delta$. Since f_v is continuous on the compact set $\{s \in \mathbb{C} : |s| \leq b'\}$, we get $\sup_{s\in\Delta} |f_v(s)| < \infty$. Therefore,

$$\begin{aligned} \left|\oint_{\Delta}\int_{-A}^{A} p_{\boldsymbol{z}}(v-s-u)\mathrm{d}F_{\boldsymbol{u}}(u)\mathrm{d}s\right| &= \left|\oint_{\Delta} f_v(s)\mathrm{d}s\right| \\ &\leq \mathrm{len}(\Delta)\sup_{s\in\Delta}|f_v(s)| \\ &< \infty. \end{aligned} \tag{6.10}$$

Then by Tonelli-Fubini Theorem,

$$\begin{aligned} \oint_{\Delta} f_v(s)\mathrm{d}s &= \oint_{\Delta}\int_{-A}^{A} p_{\boldsymbol{z}}(v-s-u)\mathrm{d}F_{\boldsymbol{u}}(u)\mathrm{d}s \\ &= \int_{-A}^{A}\oint_{\Delta} p_{\boldsymbol{z}}(v-s-u)\mathrm{d}s\mathrm{d}F_{\boldsymbol{u}}(u) \\ &= 0, \end{aligned} \tag{6.11}$$

where the last step is by Cauchy's integral Theorem, as $p_{\boldsymbol{z}}(\cdot)$ is analytic everywhere. Since $\oint_{\Delta} f_v(s)\mathrm{d}s = 0$ for any triangular path Δ and $f_v(\cdot)$ is continuous everywhere, by Morera's Theorem, we conclude that $f_v(\cdot)$ is analytic.

Next, let Δ be an arbitrary triangular path in $\mathbb{C}$. Then there exists $b'' > 0$ such that for any $s \in \Delta$, we have $|s| \leq b''$. By Lemma 7, $\rho(\cdot)$ is everywhere continuous. Since the set $\{s \in \mathbb{C} : |s| \leq b''\}$ is compact, we conclude that $\sup_{s\in\Delta} |\rho(s)| < \infty$. Therefore,

$$\begin{aligned} \left|\oint_{\Delta}\int_{\mathbb{R}} p_{\boldsymbol{w}}(v-s)\log p_{\boldsymbol{v}}(v;F)\mathrm{d}v\mathrm{d}s\right| &= \left|\oint_{\Delta}\rho(s)\mathrm{d}s\right| \\ &\leq \mathrm{len}(\Delta)\sup_{s\in\Delta}|\rho(s)| \\ &< \infty. \end{aligned} \tag{6.12}$$

Hence, by Tonelli-Fubini Theorem [35],

$$\begin{aligned}\oint_{\Delta} \rho(s)\mathrm{d}s &= \oint_{\Delta} \int_{\mathbb{R}} p_{\boldsymbol{w}}(v-s) \log p_{\boldsymbol{v}}(v;F)\mathrm{d}v\mathrm{d}s \\ &= \int_{\mathbb{R}} \oint_{\Delta} p_{\boldsymbol{w}}(v-s)\mathrm{d}s \log p_{\boldsymbol{v}}(v;F)\mathrm{d}v \\ &= 0, \end{aligned} \tag{6.13}$$

where the last step follows by Lemma 8 and Cauchy's integral Theorem. Therefore, using Morera's Theorem, $\rho(\cdot)$ is everywhere analytic.

Bibliography

[1] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.,* vol. 37, pp. 623-656 (pt. II), July/Oct. 1948.

[2] J. G. Smith, "The Information Capacity of Amplitude and Variance-Constrained Scalar Gaussian Channels", *Inform. Contr.*, vol. 18, pp. 203-219, 1971.

[3] J. G. Smith,"On the information capacity of peak and average power constrained Gaussian channels", *Ph.D. dissertation.*, Univ. of California, Berkeley, CA, Dec. 1969.

[4] W. Oettli, "Capacity-achieving input distributions for some amplitude-limited channels with additive noise," *IEEE Tran. Inf. Theory*, vol. 20, no. 3, May 1974.

[5] A. Das, "Capacity-achieving distributions for non-Gaussian additive noise channels," *Int. Symp. Inf. Theory, ISIT2000*, Sorrento, Italy, June 2000.

[6] A. Tchamkerten, "On the Discreteness of Capacity-Achieving Distributions", *IEEE Trans. on Inf. Theory*, vol. 50, no. 11, Nov. 2004.

[7] M. Raginsky, "On the Information Capacity of Gaussian Channels Under Small Peak Power Constraints," *46th Annual Allerton Conference*, Illinois, USA, 2008.

[8] S. Shamai (Shitz) and I. Bar-David, “The capacity of average and peak-power limited quadrature Gaussian channel”, *IEEE Trans. Inf. Theory*, vol. 41, pp. 1060-1071, July 1995.

[9] P. Hou, B. J. Belzer, and T. R. Fischer, “On the capacity of the partially coherent additive white Gaussian noise channel,” *Int. Symp. on Inf. Theory*, Yokohama, Japan, June 2003.

[10] R. Nuriyev and A. Anastasopoulos, “Capacity characterization for the non-coherent block-independent AWGN channel,” *Int. Symp. on Inf. Theory*, Yokohama, Japan, June 2003.

[11] J. Singh,O. Dabeer and U. Madhow, ”On the limits of communication with low-precision analog-to-digital conversion at the receiver,” *IEEE Trans. on Commun.*, vol.57, no.12, pp.3629-3639, Dec. 2009. March 29, 1929.

[12] Y. Wu, L.M. Davis and R. Calderbank, ”On the capacity of the discrete-time channel with uniform output quantization,” *Int. Symp. Inf. Theory, ISIT2009*, Seoul, South Korea, June 2009.

[13] T. H. Chan, S. Hranilovic and F. R. Kschischang, “Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs,” *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073-2088, June 2005.

[14] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communication aspects,” *IEEE Tran. Inf. Theory*, vol. 44, no. 6, pp. 2619-2692, 1998.

[15] I. C. A. Faycal, M. D. Trott, and S. Shamai, “The capacity of discrete-time memoryless Rayleigh-fading channels”, *IEEE Trans. Inf. Theory*, vol. 47, pp. 1290-1301, May 2001.

[16] R. Palanki, "On the capacity achieving distributions of some fading channels," 40^{th} *Allerton Conference on Communication, Control, and Computing*, 2002.

[17] M. Gursoy, H. Poor, S. Verdu, "The non-coherent Rician fading channel - part I: Structure of capacity achieving input," *IEEE Tran. Wireless Communication*, vol. 4, no. 5, pp. 2193-2206, Sept. 2005.

[18] M. Gursoy, H. Poor, S. Verdu, "The non-coherent Rician fading channel - part II: Spectral efficiency in the low-power regime," *IEEE Tran. Wireless Communication*, vol. 4, no. 5, pp. 2207-2221, Sept. 2005.

[19] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time non-coherent additive white Gaussian noise channel," in *IEEE Int. Symp. on Inf. Theory*, 2002, p. 165.

[20] M. Katz and S. Shamai, "On the capacity-achieving distribution of the discrete-time non-coherent and partially coherent AWGN channels," in *IEEE Tran. Inf. Theory*, vol. 50, no. 10, pp. 2257-2270, Oct. 2004.

[21] R. Chen, B. Hajek, and R. Koetter, "On fixed input distributions for non-coherent communication over high SNR Rayleigh fading channels," *IEEE Tran. Inf. Theory*, vol. 50, no. 12, 2004.

[22] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the Gaussian multiple-access channel," *IEEE Tran. Inf. Theory*, vol. 42, no. 2, March 1996.

[23] J. Huang, S. P. Meyn, "Characterization and computation of optimal distributions for channel coding," *IEEE Tran. Inf. Theory*, vol. 51, no. 7, July 2005.

[24] T. M. Cover and J. A. Thomas, "Elements of information theory", *Wiley Series in Telecommunications and Signal Processing*, 2006.

[25] A. El Gamal, Y. Kim, "Network Information Theory," *Cambridge University Press*, 2011.

[26] R. Ahlswede, "Multi-way Communication Channels," *in proc. 2nd Int. Symp. Inf. Theory*, Budapest, Hungary: Hungarian Academy of Science, 1973, pp. 23-52.

[27] H. Liao, "Multiple Access Channels," *Ph.D. dissertation*, Dept. Elec. Eng., Univ. of Hawai, 1972.

[28] S. Verdu, "The capacity region of the symbol-asynchronous Gaussian multiple access channel," *IEEE Tran. Inf. Theory*, vol. 35, no. 4, July 1989.

[29] S. Verdu, "Multiple-access channels with memory with and without frame synchronism," *IEEE Tran. Inf. Theory*, vol. 35, no. 3, May 1989.

[30] R. S. Cheng and S. Verdu, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Tran. Inf. Theory*, vol. 39, no. 3, May 1993.

[31] D. N. C. Tse, and S. V. Hanly, "Multiaccess fading channels-Part I: Polymatroid structure, optimal resource allocation, and throughput capacities," *IEEE Tran. Inf. Theory*, vol. 44, no. 7, November 1998.

[32] T. Cover, "Some advances in broadcast channels," *Advances in Communication Systems*, vol. 4. New York: Academic, 1975, pp. 229-260.

[33] A. D. Wyner, "Recent results in the Shannon theory," *IEEE Trans. Inf. Theory*, vol. IT-20, Jan. 1974.

[34] J. M. Borwein and A. S. Lewis, "Convex analysis and nonlinear optimization", *CMS Books in Mathematics*, 2010.

[35] R. M. Dudley, "Real analysis and probability", *Cambridge University Press*, 2002.

[36] L. B. Koralov and Y. G. Sinai, "Theory of probability and random processes", *Universitext, Springer-Verlag*, 2007.

[37] R. G. Bartle, "The elements of integration and Lebesgue measure," *John Wiley and Sons, Inc.*, p:44, 1995.

[38] A. D. Aleksandrov, A. N. Kolmogorov, and M. A. Lavrent'ev, "Mathematics, Its Content, Methods and Meaning". *Dover Publications Inc.*, 1999.

[39] J. R. Munkres, "Topology", *Prentice Hall Inc.*, 2000.

[40] J. B. Conway, "Functions of One Complex Variable I", *Graduate Texts in Mathematics*, 1973.