ACHIEVABLE CODING RATES FOR AWGN AND BLOCK FADING CHANNELS IN THE FINITE BLOCKLENGTH REGIME

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

BY

MEHMET VURAL

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

SEPTEMBER 2010

Approval of the thesis:

ACHIEVABLE CODING RATES FOR AWGN AND BLOCK FADING CHANNELS IN THE FINITE BLOCKLENGTH REGIME

submitted by **MEHMET VURAL** 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. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. İsmet Erkmen
Head of Department, **Electrical and Electronics Engineering** ________________

Assoc. Prof. Dr. Ali Özgür Yılmaz
Supervisor, **Electrical and Electronics Engineering Dept., METU** ________________

Examining Committee Members:

Prof. Dr. Yalçın Tanık
Electrical and Electronics Engineering Dept., METU ________________

Assoc. Prof. Dr. Ali Özgür Yılmaz
Electrical and Electronics Engineering Dept., METU ________________

Prof. Dr. Mete Severcan
Electrical and Electronics Engineering Dept., METU ________________

Prof. Dr. T. Engin Tuncer
Electrical and Electronics Engineering Dept., METU ________________

Ülkü Oktel, M.S.
ASELSAN ________________

Date: ________________

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, Last Name: MEHMET VURAL

Signature :

ABSTRACT

ACHIEVABLE CODING RATES FOR AWGN AND BLOCK FADING CHANNELS IN THE FINITE BLOCKLENGTH REGIME

Vural, Mehmet
M.S., Department of Electrical and Electronics Engineering
Supervisor : Assoc. Prof. Dr. Ali Özgür Yılmaz

September 2010, 65 pages

In practice, a communication system works with finite blocklength codes because of the delay constraints and the information-theoretic bounds which are proposed for finite blocklength systems can be exploited to determine the performance of a designed system. In this thesis, achievable rates for given average error probabilities are considered for finite blocklength systems. Although classical bounds can be used to upper bound the error probability, these bounds require the optimization of auxiliary variables. In this work, a bound which is called the dependence testing (DT) bound that is free of any auxiliary variables is exploited. The DT bound is evaluated by introducing a normal approximation to the information density. Simulations carried out both for the Gaussian and discrete input alphabets show the proposed approximation enables very good prediction of the achievable rates. The proposed approximation is also used to calculate the average error probability for block fading channels. Simulations performed for Rayleigh block fading channels demonstrate that the total blocklength of the system in addition to the number of fading blocks should be accounted for especially when the number of fading blocks is large. A power allocation problem in block fading channels when the channel state information is available to the transmitting side is investigated in the final part of this work. The DT bound is optimized for a given channel state vector by allocating different power levels to each fading block by exploiting short-term power allocation. A simple power allocation algorithm is proposed which comes out with very similar results compared with the analytically computed values.

Keywords: Finite blocklength, Block fading channels, DT bound, information density, power allocation

ÖZ

SONLU BLOK UZUNLUĞU DURUMUNDA AWGN VE BLOK SÖNÜMLEMELİ KANALLAR İÇİN ERİŞİLEBİLİR KODLAMA HIZLARI

Vural, Mehmet
Yüksek Lisans, Elektrik ve Elektronik Mühendisliği Bölümü
Tez Yöneticisi : Doç. Dr. Ali Özgür Yılmaz

Eylül 2010, 65 sayfa

Gerçek hayattaki bir iletişim sistemi gecikme limitleri yüzünden sonlu uzunluktaki kodlarla çalışmak durumundadır ve sonlu blok uzunluklu sistemler için önerilen bilgi teorisindeki sınırlar, tasarlanan sistemin performansını belirlemede kullanılabilir. Bu tezde, sonlu blok uzunluklu sistemlerde verilen bir ortalama hata olasılığı için erişilebilir kodlama hızlarına odaklanılmıştır. Hata olasılığını üstten sınırlamak için klasik yöntemler kullanılabilirse de bu yöntemler yardımcı değişkenlerin optimizasyonunu gerektirir. Bu çalışmada bu tür yardımcı değişkenlere bağlı olmayan bağlılık testi (BT) sınırından faydalanılmaktadır. BT sınırı, bilgi yoğunluğuna Gauss yaklaştırması kullanılarak hesaplanmaktadır. Gauss ve sonlu girdi alfabeleriyle yapılan benzetimler önerilen yaklaştırma ile erişilebilir kodlama hızının tahmininin çok iyi olduğunu göstermiştir. Ayrıca önerilen yaklaştırma blok sönümlemeli kanalların ortalama hata olasılığını hesaplamak için de kullanılmaktadır. Rayleigh blok sönümlemeli kanallar için yapılan benzetimler sönümlenen blokların sayısına ek olarak sistemin toplam blok uzunluğunun da özellikle sönümlenen blok sayısı uzadıkça dikkate alınması gerektiğini göstermektedir. Bu çalışmanın son kısmında gönderen tarafta kanal durum bilgisi bulunduğunda güç tahsisi problemi ele alınmıştır. Kısa süreli güç tahsis yöntemi kullanılarak verilen bir kanal durum vektörü için her sönümlenen bloğa farklı güç seviyeleri dağıtılarak BT sınırı optimize edilmektedir. Her bloğa dağıtılacak güç seviyesini bulan basit bir güç tahsis algoritması önerilmiş ve bu algoritmanın verdiği sonuçların analitik olarak elde edilen değerlerle uyuştuğu gözlenmiştir.

Anahtar Kelimeler: Sonlu blok uzunluğu, Blok sönümlemeli kanallar, BT sınırı, bilgi yoğunluğu, güç paylaştırma

To my parents

ACKNOWLEDGMENTS

I would like to express my gratitude to my supervisor Assoc. Prof. Dr. Ali Özgür Yılmaz for his excellent guidance throughout this work. His infinite support, constructive criticism and patience on my study has always motivated me. Without his support, this thesis would not have been completed.

I would like to thank ASELSAN Inc. for allowing and supporting me to spend time on this thesis. I would also like to thank all my colleagues for the good times we shared.

I thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) for their financial support during my graduate study.

I would also like to thank my flatmates Anıl and Gencer and my friend Ayhan for their encouragement and support to complete this work.

I am grateful to my parents for their endless love and support. I also present my special thanks to my sisters Yasemin and Gülsemin who have always encouraged and supported me in my studies.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

Introduction

In wireless communications there has been great interest to examine the performance of finite blocklength systems after Shannon's work on channel capacity in [1]. The channel capacity that is defined and calculated in [1] is achieved only when the length of the codeword is increased into infinity. When the codeword length is a finite number, some other performance criteria arise in wireless communication problems. One of those criteria is to find a lower bound on the size of the codebook for a given blocklength and error probability. This type of bounds are defined as achievability bounds since the existence of at least one code is guaranteed such that the given error probability is satisfied for the given length of codewords.

Achievability bounds tell us what the blocklength must be to approach the channel capacity while a desired coding rate is satisfied. Equally, for a given blocklength, one can determine how much the coding rate is away from the channel capacity. Since it is a lower bound, an achievability bound becomes tighter if it promises a larger size of the codebook. In this work, we investigate the gap between the channel capacity and some achievability bounds to compare their strengths.

In the early years of the information theory studies, three main achievability bounds are proposed. In 1954, Feinstein proposed a lower bound on the maximal probability of error in [2]. In 1957, Shannon's achievability bound for average probability of error is published in [3] which was a slightly strengthened version of Feinstein's bound but it was different from Feinstein's bound in the sense that 'maximal' is replaced with 'average' probability of error. The last of the main bounds for finite blocklength systems is given by Gallager in [4] which leads us to the well-known Gallager's error exponent.

Although these bounds give one a hint to calculate how far a system's operation is away from the channel capacity when a finite blocklength is considered for a certain error probability, optimization of auxiliary constants is required in order to obtain tight bounds. Recently, new bounds on the channel coding rate were introduced in [6] for channels with additive noise. These bounds turned out to be tighter than the classical bounds and they require no selection of auxiliary constants. However, closed form solutions cannot be obtained when one attempts to calculate these bounds.

In wireless channels, it is not only the additive noise that disrupts the transmitted signal but also the random variation of the channel coefficients introduced by constructive or destructive addition of signal components over the communication medium randomly [9]. This phenomenon is called fading. The block fading channel model introduced in [7] is a useful model to analyze many of today's communication systems. There have been many studies to inspect the performance of communication systems over block fading channels and the main performance indicator has been chosen as the outage probability [8]. However, when calculating the outage probability the length of the codeword is usually assumed to be infinite and it is assumed there is a finite number of fading blocks over the codeword. We believe that it could also be useful to inspect the effect of blocklength and the number of fading blocks over the achievability bounds for block fading channels. In [18], the blocklength and the number of fading blocks is considered by computing Gallager's error exponent [4] for block fading channels, however due to the complexity of the optimization of the auxiliary variable some simplifications are proposed to calculate the error probability bound.

In this thesis, we will deal with one of the bounds given in [6] which is called the dependence testing (DT) bound. We propose a computation method to the DT bound by applying a normal approximation to information density. First, we consider the case for the additive white Gaussian noise (AWGN) channel. By using similar arguments for the AWGN channel, we also calculate the DT bound for block fading channels. Our approximation in computing the DT bound generates very close results in relation with Monte Carlo simulations. Approximations similar to the one proposed in this thesis are performed to obtain Feinstein's bound [2] in [19] for AWGN channels, and a normal approximation to the number of errors in a binary symmetric channel to calculate Wolfowitz's bound [20] is proposed in [21].

When the channel state information is available to the transmitting side in fading channels, it is possible that the transmitter adapts the transmitted power according to the channel state in time, frequency or space. By applying a power allocation scheme to a communication system, it is possible to enhance the performance. We also optimize and compute the DT bound when channel state information at the transmitter (CSIT) is available and utilized.

1.1 Outline of the Thesis

In Chapter 2, we present our system model used throughout the thesis. We define the ergodic channel capacity for infinite blocklength systems and the delay-constrained capacity when a finite blocklength system is considered. We introduce the outage probability for block fading channels and investigate how it arises in a communication problem by using error probability expressions. We also present the major existing bounds for finite blocklength systems proposed by Feinstein, Shannon, and Gallager.

In Chapter 3, the DT bound is investigated for finite blocklength channels. Since information density arises as a random variable when calculating the DT bound, we investigate if it can be modeled as a Gaussian random variable by using the central limit theorem [27]. We provide closed form expressions for the DT bound by using the normal approximation for real and complex AWGN channels with Gaussian input alphabet. We show that the proposed approximation characterizes the DT bound tightly for blocklengths as small as 100. We also exploit the Gaussian approximation to calculate the DT bound for block fading channels and present closed form expressions for a given channel coefficient vector. Thus we obtain a method to estimate the performance of a communication system by taking into account the length of the codeword block. We also use the normal approximation to calculate the DT bound for constrained input alphabets.

In Chapter 4, we extend the approach given in Chapter 3 to the case where CSIT is available. We present the existing power allocation schemes in the literature for fading channels. Then we present how the DT bound is optimized when CSIT is available. Since the expressions for the allocated power for a given fading coefficient vector become very complex, we present an algorithm which is called the maximum marginal allocation algorithm given in [32] to compute the power to be allocated to each fading block in the runtime of a communication system. By using this algorithm we analyze the effect of number of fading blocks and the blocklength on the DT bound for equal power and adaptive power schemes.

Chapter 5 points out the key points in this thesis and outlines the future work.

CHAPTER 2

Background Information

2.1 Introduction

In this chapter, we will present the channel model that is used in our work. In addition, we will recall general results about the Shannon capacity of a channel and the outage probability concept for finite blocklength channels. Since Shannon established the convergence of the coding rate to the channel capacity as the blocklength is increased to infinity, many researchers studied the penalty that forms with finite blocklengths. The major error probability bounds for finite blocklength channels will be stated and their proofs will be outlined in this chapter. These bounds can be extremely useful in finding the highest rate that can be achieved when operating with a given blocklength and error probability. Outline of this chapter is as follows: In Section 2.2, the block fading channel model presented by Ozarow in [7] is given and in Section 2.3 the result of the channel coding theorem and the outage capacity for finite blocklength channels are presented. The bounds for the finite blocklength are introduced in Section 2.4.

2.2 Finite Blocklength Channels

The block fading channel introduced in [7] which models slowly varying fading proves useful and simplifies the design and analysis of many of the today's communication systems. Slow frequency hopping schemes encountered in GSM and EDGE systems as well as OFDM systems are well-modeled using the block-fading channel. In [7], a transmitted codeword is divided into K blocks of duration T and each individual block is sent through a faded narrowband multipath channel in a TDMA system. The channel is assumed to stay constant over one block but may change from one block to another. By the use of this model, it is proven that for the same outage probability, higher rates can be achieved by increasing the number of blocks. It is also emphasized in [7] that the same model can be applied to CDMA and FDMA systems.

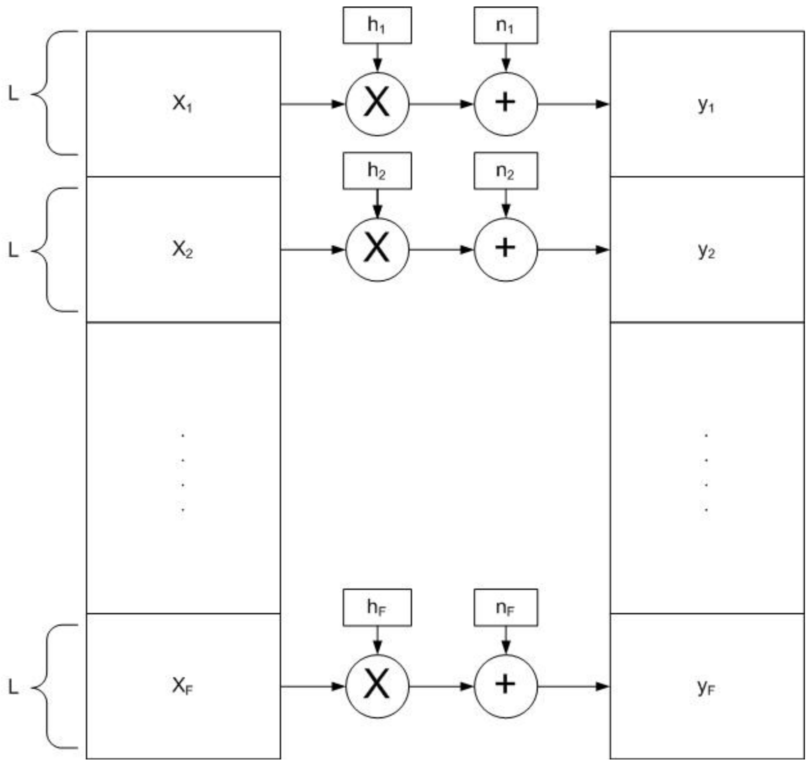

Figure 2.1: Block fading channel model

The block fading channel model is defined as follows: Consider the transmission scheme over a narrowband AWGN fading channel with F blocks of L channel uses each. For every block $f = 1, \dots, F$, the fading coefficient of the channel is given as h_f. This model is shown in Figure 2.1. Thus the channel model can be written in the vector form as

$$\mathbf{y}_f = h_f \mathbf{x}_f + \mathbf{n}_f, \tag{2.1}$$

where $\mathbf{y}_f$ is the received signal, $\mathbf{x}_f$ is the transmitted codeword in block f, h_f is the channel coefficient for block f and $\mathbf{n}_f$ is the noise vector with independent, identically distributed Gaussian entries with zero mean and unit variance $N(0, 1)$. Vectors $\mathbf{y}_f$, $\mathbf{x}_f$ and $\mathbf{n}_f$ are of length L.

In our work, the block fading channel coefficients are considered as realizations of a random variable whose magnitude is Rayleigh distributed which is a commonly used model in multipath channels. The pdf and cdf of Rayleigh distribution [9] is given as

$$f(x,\sigma) = \frac{x}{\sigma^2} exp(-\frac{x^2}{2\sigma^2}) \tag{2.2}$$

and

$$F(x,\sigma) = 1 - exp(-\frac{x^2}{2\sigma^2}) \tag{2.3}$$

respectively for an average power value of σ^2.

2.3 The Channel Capacity and Outage Probability

In this section, we recall the fundamental results about the channel capacity which is the largest rate at which information can be reliably transferred from one point to another. The channel coding theorem for stationary memoryless channels is proved by Shannon in [1] and can be expressed as

$$C = \max_{p_X(\cdot)} I(X;Y), \tag{2.4}$$

where $p_X(\cdot)$ is the probability distribution of the input X to the channel and Y is the channel output. The mutual information $I(X;Y)$ between the input and output of the channel [10] is defined as

$$I(X;Y) = \sum_{x,y} p(x,y) log \frac{p(x,y)}{p(x)p(y)}. \tag{2.5}$$

Shannon's channel coding theorem proved that any coding rate below the channel capacity is achievable with asymptotically small error probability. Conversely, any coding rate larger than the channel capacity must have an error probability bounded away from zero.

In [10], it is proven that the capacity is achieved by Gaussian distributed input alphabets for AWGN channels and the channel capacity is calculated as

$$C = \frac{1}{2} log\left(1 + \frac{P}{N}\right), \tag{2.6}$$

where P is the variance of the input alphabet and N is the variance of the additive noise. The variance is regarded as power in this context.

In the fading case the ergodic channel capacity is given in [9] as

$$C = E_h\left[\frac{1}{2} log(1 + |h|^2 \frac{P}{N})\right] \tag{2.7}$$

and it is the calculated by taking the expected value of the AWGN channel capacity with respect to the channel fading statistics.

The channel capacity is achieved when the blocklength of the coding scheme grows to infinity. For that reason, the capacity in (2.7) is referred to as the 'delay-unconstrained' [11] or the ergodic channel capacity. In order to achieve the capacity for the fading case, the received codeword should be affected by all realizations of the channel states. For the finite blocklength case where the channel relies on particular realizations over a finite number of independent fading coefficients, the channel is non-ergodic and defined as information unstable [12]. This 'delay-constrained' case corresponds to real-time transmission over slowly fading channels and it is more relevant in wireless system applications. In this type of channels, the Shannon (ergodic) capacity for Rayleigh fading statistics is calculated as zero [13]. In fact, there may be a non-zero probability that no matter how small the value of the actual transmitted rate is, the channel is not able to support any nonzero data rate for some channel states, which is called the 'outage event'.

Data transmission is carried out by a codebook having M codewords so that the rate of transmission is $R = (logM)/FL$ bits/channel use. The distribution of the input source is given as p_X. For a given channel state $\mathbf{H} = \mathbf{h}$, the average codeword error probability which corresponds to the frame error rate can be bounded as in [8]

$$\overline{P_{error|\mathbf{H}=\mathbf{h}}} \leq 2^{-FL(E_0(\rho, p_{\mathbf{X}}, \mathbf{H}=\mathbf{h}) - \rho R)}, \tag{2.8}$$

where

$$E_0(\rho, p_{\mathbf{X}}, \mathbf{H} = \mathbf{h}) = -\frac{1}{FL} log \int \cdots \int \left(\int \cdots \int p_{\mathbf{X}}(\mathbf{x}) p_{\mathbf{Y}|\mathbf{X},\mathbf{H}}(\mathbf{y}|\mathbf{x}, \mathbf{h})^{1/(1+\rho)} d\mathbf{x} \right)^{1+\rho} d\mathbf{y} \tag{2.9}$$

and ρ is in [0,1]. If (2.9) is maximized over ρ, the average error probability can be rewritten as

$$\overline{P_{error|\mathbf{H}=\mathbf{h}}} \leq \begin{cases} 1, I_F < R \\ 2^{-FLE_r(R, p_{\mathbf{X}}, \mathbf{H}=\mathbf{h})}, I_F \geq R \end{cases} \tag{2.10}$$

where

$$E_r(R, p_{\mathbf{X}}, \mathbf{H} = \mathbf{h}) = \max_{0 \leq \rho \leq 1} E_0(\rho, p_{\mathbf{X}}, \mathbf{H} = \mathbf{h}) - \rho R \tag{2.11}$$

and I_F is the instantenous mutual information for the given channel state which can be expressed as

$$I_F = \frac{1}{FL} \sum_{f=1}^{F} \int \int p_{\mathbf{Y},\mathbf{X}|\mathbf{H}_f}(\mathbf{y_f}; \mathbf{x_f}|h_f) \cdot log \frac{p_{\mathbf{Y}|\mathbf{X},\mathbf{H}}(\mathbf{y_f}|\mathbf{x_f}, h_f)}{p_{\mathbf{Y}|\mathbf{H}}(\mathbf{y_f}|h_f)} d\mathbf{y_f} d\mathbf{x_f}. \tag{2.12}$$

By using (2.10), we can express the average error probability over the fading statistics as

$$\overline{P_{error}} = E_{\mathbf{H}}\overline{P_{error|\mathbf{H}=\mathbf{h}}} \leq P_{out}(R, p_{\mathbf{X}}) + \int_{\mathbf{h}:I_F \geq R} 2^{-FLE_r(R,p_{\mathbf{X}},\mathbf{H}=\mathbf{h})} dF_{\mathbf{h}}(\mathbf{h}), \tag{2.13}$$

where the outage probability is defined as

$$P_{out}(R, p_{\mathbf{X}}) = \text{Prob}(I_F < R). \tag{2.14}$$

With increasing L to infinity, the second term in (2.13) converges to zero. On the contrary, the first term is an irreducible one which is independent of L. Thus arbitrarily small error probabilities are not achievable in general for finite block length channels. In order to distinguish the capacity of an F-block channel from the channel capacity in the delay-unlimited case, the former one is defined as the delay-limited capacity in [17]. The delay-limited capacity for a block fading AWGN channel can be expressed as

$$C_{\text{delay}} = \lim_{\epsilon \to 0} sup(R : P_{out}(R, p_{\mathbf{X}}) \leq \epsilon). \tag{2.15}$$

In most of the studies dealing with the finite block fading channel, the outage probability is taken as the main parameter to be minimized and the second term in (2.13) is considered as zero which means that L is taken to be infinite. Outage probability is considered as the appropriate performance limit indicator [7], [14].

2.4 On the bounds for the finite blocklength channel

This section will be dedicated to the major existing bounds for the finite blocklength channels. We will discuss the Feinstein's achievability bound for a given maximal probability of error [2], Shannon's average probability of error bound [3] and also Gallager's random coding bound [4] which leads us to the random coding error exponent. These bounds calculate an average or maximal coding error probability for a given blocklength and coding rate. Equivalently they give us a lower bound on the size of a code that can be guaranteed to exist with a given blocklength and error probability.

We first define the fundamental problem in communications as shown in Figure 2.2. Let us consider an input alphabet A, an output alphabet B and a probability function $P_{\mathbf{Y}|\mathbf{X}} : A \to B$ which can be called the 'channel'. A subset of the space A^N is called a codebook $\{c_1, \cdots, c_M\}$ of length M with the coding rate

$$R = \frac{logM}{N}, \tag{2.16}$$

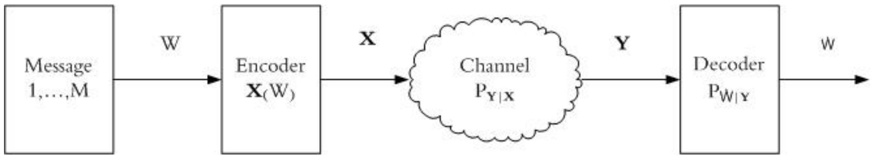

Figure 2.2: Channel model for a communication problem

where N is the blocklength of the codeword and M is the number of messages that can be sent with this codebook. The encoder is a function from the message set into the codebook $\mathbf{X}(W)$ and the decoder is a transformation $P_{\hat{W}|\mathbf{Y}} : B^N \rightarrow \{1, \cdots, M\}$. If $\hat{W} = W$, then the decoding operation works correctly and the message is transmitted without error. If the input messages to the encoder are of equal probability, then the average probability of error of the decoder can be expressed as

$$1 - \frac{1}{M} \sum_{m=1}^{M} P_{\hat{W}|\mathbf{X}}(m|c_m), \tag{2.17}$$

where $P_{\hat{W}|\mathbf{X}}(m|c_m)$ is the probability of correct decoding of the m^{th} message. If the decoder together with the codebook satisfies an average probability of error less than ϵ, then the pair is called an (M, ϵ) code with average probability of error. If the decoder satisfies a probability of correct reception $P_{\hat{W}|\mathbf{X}}(m|c_m) \geq 1 - \epsilon$ for $\forall m \in 1, \cdots, m$, then this codebook and the decoder are called an (M, ϵ) code with maximal probability of error. If the codeword length is N, then this code is further called an (N, M, ϵ) code. The fundamental rate limit in this communication scenario can be expressed as

$$M^*(N, \epsilon) = max\left(M : \exists (N, M, \epsilon) - \text{code}\right). \tag{2.18}$$

Feinstein's lemma for the lower bound of maximal probability of error is expressed as follows.

Theorem 2.4.1 (Feinstein's theorem [2]) *Given M and any $a > 0$, for any input distribution $p_{\mathbf{X}}$, there exists an (N, M, ϵ) code with maximal probability of error for the transition probability $P_{\mathbf{Y}|\mathbf{X}}$ such that*

$$\epsilon \leq Me^{-a} + Pr\left\{i(\mathbf{x}; \mathbf{y}) \leq a\right\}, \tag{2.19}$$

where the information density is defined as

$$i(\mathbf{x}, \mathbf{y}) = log \frac{p_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x})}{p_{\mathbf{Y}}(\mathbf{y})}. \tag{2.20}$$

The proof is provided in Appendix A.

If we choose $a = logM + N\gamma$ with $\gamma > 0$, the Feinstein's bound can be rewritten as

$$\epsilon \leq e^{-\gamma N} + Pr\left\{\frac{1}{N}i(\mathbf{x};\mathbf{y}) \leq \frac{1}{N}logM + \gamma\right\}, \tag{2.21}$$

Note that the decoder is constructed such that the probability of error is satisfied for every codeword in the codebook so that the error bound that is proved to hold is the maximal probability of error. Thus, Feinstein's theorem implies that for any input distribution $p(\mathbf{x})$, there exists a code with rate R, for any $\gamma > 0$ and $N > 0$

$$\epsilon \leq e^{-\gamma N} + Pr\left\{\frac{1}{N}i(\mathbf{x};\mathbf{y}) \leq R + \gamma\right\}, \tag{2.22}$$

in the maximal probability of error sense.

The following theorem establishes Shannon's achievability bound [3] for the average probability of error.

Theorem 2.4.2 (Shannon's theorem [3]) *Given M and any $\beta > 0$, for any input distribution $p_{\mathbf{X}}$, there exists an (N, M, ϵ) code with average probability of error for the transition probability $p_{\mathbf{Y}|\mathbf{X}}$ such that*

$$\epsilon \leq \frac{M-1}{\beta} + Pr\left\{i(\mathbf{x};\mathbf{y}) \leq log\beta\right\}, \tag{2.23}$$

where the information density is defined in (2.20).

The proof is outlined in Appendix B.
When we replace $\beta = e^a$ in order to compare the bounds of Feinstein and Shannon, we can conclude that Theorem 2.4.1 is a slightly weakened version of Theorem 2.4.2 where $M - 1$ is replaced by M. However, Theorem 2.4.2 is a weakened version of Theorem 2.4.1 in the sense that maximal is replaced with the average probability of error.

Before giving Gallager's probability of error bound, we present the following lemma from [22].

Lemma 2.4.3 *Let $P(A_1), \cdots, P(A_m)$ be the probabilities of a set of events $A_1, \cdots, A_m$ and*

$$P(\bigcup_m A_m) \tag{2.24}$$

be the probability of their union. For any ρ*,* $0 < \rho \leq 1$*, we have*

$$P(\bigcup_m A_m) \leq \left[\sum_{m=1}^{M} P(A_m)\right]^{\rho}. \tag{2.25}$$

Proof. We can write that

$$P(\bigcup_m A_m) \leq \begin{cases} \sum_{m=1}^{M} P(A_m) \\ 1. \end{cases} \tag{2.26}$$

The first bound is the simple union bound on the probabilities and the second one is because of the probability cannot be greater than one. If the former bound in (2.26) is satisfied then we further increase $\sum P(A_m)$ by raising it to the power by ρ and (2.25) follows. If the latter bound in (2.26) is satisfied, then $[\sum P(A_m)]^{\rho} \geq 1$ and again (2.25) follows. ∎

Theorem 2.4.4 (Gallager's theorem [4]) *Given* M *and any* $0 \leq \rho \leq 1$*, for any input distribution* $p_{\mathbf{X}}$*, there exists an* (N, M, ϵ) *code with average probability of error for the transition probability* $p_{\mathbf{Y}|\mathbf{X}}$ *such that*

$$\epsilon \leq (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} p(\mathbf{x}) p(\mathbf{y}|\mathbf{x})^{1/(1+\rho)}\right]^{(1+\rho)}. \tag{2.27}$$

The proof of the theorem is given in Appendix C.

Theorem 2.4.4 leads us to the Gallager's error exponent for discrete memoryless channels [22]. For discrete memoryless channels

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{N} p(y_i|x_i) \tag{2.28}$$

and also if the codewords are independently generated, i.e.,

$$p(\mathbf{x}) = \prod_{i=1}^{N} p(x_i). \tag{2.29}$$

Substituting (2.28) and (2.29) into (2.27), we get

$$\epsilon \leq (M-1)^{\rho} \sum_{\mathbf{y}} \left[\sum_{\mathbf{x}} \prod_{i=1}^{N} p(x_i) p(y_i|x_i)^{1/(1+\rho)}\right]^{(1+\rho)}. \tag{2.30}$$

If we restate the bracketed term in (2.30) as

$$\epsilon \leq (M-1)^{\rho} \sum_{\mathbf{y}} \left[\prod_{i=1}^{N} \sum_{\mathbf{x}} p(x_i) p(y_i|x_i)^{1/(1+\rho)}\right]^{(1+\rho)}, \tag{2.31}$$

by using the arithmetic rule for interchanging products of sums. Finally, by interchanging the power and multiplication operation in the bracketed term and applying the same rule to the outer summation we get

$$\epsilon \leq (M-1)^{\rho} \prod_{i=1}^{N} \sum_{y_i} \left[\sum_{x_i} p(x_i) p(y_i|x_i)^{1/(1+\rho)} \right]^{(1+\rho)}. \tag{2.32}$$

If we upper bound $(M-1)$ by 2^{NR}, (2.32) can be expressed as

$$\epsilon \leq 2^{-N\left[-log\left\{\sum_{y_i}[\sum_{x_i} p(x_i)p(y_i|x_i)^{1/(1+\rho)}]^{(1+\rho)}\right\}-\rho R\right]}. \tag{2.33}$$

Since (2.33) is the average for the ensemble of codewords, there must exist at least one codebook that satisfies the upper bound. If we define

$$E_0(\rho, p) = -log\left\{ \sum_{y_i} \left[\sum_{x_i} p(x_i) p(y_i|x_i)^{1/(1+\rho)} \right]^{(1+\rho)} \right\}, \tag{2.34}$$

we obtain

$$\epsilon \leq 2^{-N[E_0(\rho,p)-\rho R]}. \tag{2.35}$$

Since p and ρ are arbitrary in (2.34) and (2.35), we obtain the tightest bound by choosing p and ρ to maximize $E_0(\rho, p) - \rho R$ which leads to the random coding exponent defined as

$$E_r(R) = \max_{\rho} \max_{p} \left[E_0(\rho, p) - \rho R \right]. \tag{2.36}$$

All theorems stated above requires a selection of auxiliary constants. We need to optimize those constants to obtain tight bounds which may result in complex operations. In Chapter 3, we will deal with some other bounds on the achievable rate with a certain average error probability that do not need auxiliary constants.

CHAPTER 3

Dependence Testing Bound and a Gaussian approximation

3.1 Introduction

The dependence testing (DT) bounds proposed in [5], [6] provides an average error probability bound for a given blocklength and coding rate. The DT bound is calculated by using Feinstein's decoder in which information density is exploited in the decoding operation. One of the advantages of the DT bound compared to the bounds introduced in Section 2.4 is that it does not require any optimization of auxiliary variables. In addition to the DT bound, another achievability bound which is called the random coding union (RCU) bound is introduced in [6]. The RCU bound is stronger than the DT bound, however the computation complexity of the RCU bound is much higher. Also, the authors in [6] propose $\kappa - \beta$ bound for input constrained case where the input symbols of the codewords are chosen from a subset of the input alphabet such that a given constraint is satisfied and optimization of auxiliary variables is performed. In this work, we do not consider the limitations on the elements of the codeword and in order to focus on bounds with no auxiliary variables, we use the dependence testing (DT) bound for AWGN channels. We also inspect channel coding rates with certain error probability performance for discrete input alphabets and fading channels. When one attempts to calculate the DT bound for these type of channels, closed form solutions cannot be obtained due to the distribution of the information density. Thus, we introduce a normal approximation to the information density to calculate the DT bound.

As discussed in [23], Strassen proposed a maximal rate expression for a given blocklength and average error probability by using a normal approximation in the error probability expressions. However, we use a normal approximation only to the information density which arises as the random variable in the error expressions for the DT bound. Thus, instead of using an approximation in order to calculate the error probability bound as in [23], the DT bound can be calculated by our proposed method.

We also calculate the DT bound for block fading channels introduced in Section 2.2. We show for block fading channels that, in addition to the number of fading blocks, the blocklength also plays an important role in the error probability. As mentioned in Section 2.3, in outage capacity evaluations for the block fading channels, it is assumed that the blocklength is increased to infinity and the term that does not converge to zero in this limiting case is calculated. Thus only the number of fading blocks are accounted for. In our work, the effect of the total blocklength is also considered and the error probability results are given as a function of the blocklength for fading channels.

3.2 Overview of the DT bound

In this section, we present a summary of the DT bounding technique proposed in [5]. We will introduce a normal approximation to calculate that bound and draw our results based on this approximation.

Theorem 3.2.1 (Dependence Testing Bound [5]) *For any distribution p_X on the input alphabet A, there exists a code with M codewords and average probability of error not exceeding*

$$\epsilon \leq E\left[exp(-\left|i(\mathbf{x},\mathbf{y}) - log\frac{M-1}{2}\right|^{+})\right] \tag{3.1}$$

where

$$|a|^{+} = \begin{cases} a, a \geq 0 \\ 0, a < 0 \end{cases} . \tag{3.2}$$

Proof. Consider the following identity for arbitrary $z \geq 0$ and $\gamma > 0$:

$$exp\left\{-|log\frac{z}{\gamma}|^{+}\right\} = \mathbf{1}_{\{z\leq\gamma\}} + \frac{\gamma}{z}\mathbf{1}_{\{z>\gamma\}}, \tag{3.3}$$

where

$$\mathbf{1}_{\{s\}} = \begin{cases} 1, \text{s is true} \\ 0, \text{s is false} \end{cases} . \tag{3.4}$$

If we let $z = \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})p(\mathbf{y})}$, then (3.3) can be expressed as

$$exp\left\{-|i(\mathbf{x},\mathbf{y}) - log\gamma|^{+}\right\} = \mathbf{1}_{\{i(\mathbf{x},\mathbf{y})\leq log\gamma\}} + \frac{\gamma p(\mathbf{x})p(\mathbf{y})}{p(\mathbf{x},\mathbf{y})}\mathbf{1}_{\left\{\frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{x})p(\mathbf{y})}>\gamma\right\}}. \tag{3.5}$$

When the average of the right-hand side of (3.5) is taken with respect to $p(\mathbf{x},\mathbf{y})$, one obtains

$$Pr\left\{i(\mathbf{x},\mathbf{y}) \leq log\gamma\right\} + \gamma\sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x})p(\mathbf{y})\mathbf{1}_{\{i(\mathbf{x},\mathbf{y})>log\gamma\}}. \tag{3.6}$$

If we write $p(\mathbf{y})$ as $p(\overline{\mathbf{y}})$ and define the joint probability $p_{XY\overline{Y}}(\mathbf{x},\mathbf{y},\overline{\mathbf{y}}) = p_X(\mathbf{x})p_{Y|X}(\mathbf{y}|\mathbf{x})p_Y(\overline{\mathbf{y}})$ which means $\overline{\mathbf{y}}$ has the same distribution as $\mathbf{y}$ but independent of $\mathbf{x}$, we can express (3.6) as

$$Pr\left\{i(\mathbf{x},\mathbf{y}) \leq log\gamma\right\} + \gamma\sum_{\mathbf{x}}\sum_{\overline{\mathbf{y}}} \frac{p(\mathbf{x})p(\mathbf{y}|\mathbf{x})p(\overline{\mathbf{y}})}{p(\mathbf{y}|\mathbf{x})}\mathbf{1}_{\{i(\mathbf{x},\overline{\mathbf{y}})>log\gamma\}}, \tag{3.7}$$

which is equal to

$$Pr\left\{i(\mathbf{x},\mathbf{y}) \leq log\gamma\right\} + \gamma\sum_{\mathbf{x}}\sum_{\overline{\mathbf{y}}} p(\mathbf{x},\overline{\mathbf{y}})\mathbf{1}_{\{i(\mathbf{x},\overline{\mathbf{y}})>log\gamma\}}. \tag{3.8}$$

Thus we can write the average of both sides of (3.5) with respect to $p(\mathbf{x},\mathbf{y})$ as

$$E\left[exp\left\{-[i(\mathbf{x},\mathbf{y}) - log\gamma]^{+}\right\}\right] = Pr\left\{i(\mathbf{x},\mathbf{y}) \leq log\gamma\right\} + \gamma Pr\left\{i\left(\mathbf{x},\overline{\mathbf{y}} > log\gamma\right)\right\}. \tag{3.9}$$

Replacing $\gamma = log\frac{M-1}{2}$, we need to prove that

$$\epsilon \leq Pr\left\{i(\mathbf{x},\mathbf{y}) \leq log\frac{M-1}{2}\right\} + \frac{M-1}{2}Pr\left\{i(\mathbf{x},\overline{\mathbf{y}}) > log\frac{M-1}{2})\right\}. \tag{3.10}$$

Let $\{Z_{\mathbf{x}}\}_{\mathbf{x}\in A} : B \mapsto \{0,1\}$ be the function whose input space is the output alphabet B defined as

$$Z_{\mathbf{x}}(\mathbf{y}) = \mathbf{1}_{\left\{i(\mathbf{x},\mathbf{y})>log\frac{M-1}{2}\right\}}. \tag{3.11}$$

For a given codebook $c_1,\cdots,c_M$, the decoding rule is given as follows: The receiver runs M likelihood ratio binary hypothesis tests between $P_{\mathbf{Y}|\mathbf{c_j}}(\mathbf{y})$ and $P_{\mathbf{Y}}$ and computes the values $Z_{\mathbf{c_j}}(\mathbf{y})$ by evaluating $i(\mathbf{c_j},\mathbf{y}) = log\frac{P_{\mathbf{Y}|\mathbf{c_j}}(\mathbf{y})}{P_{\mathbf{Y}}(\mathbf{y})}$ for the received channel output $\mathbf{y}$ and decides that $\mathbf{c_j}$ is sent by the transmitter where j is the lowest index for which $Z_{\mathbf{c_j}} = 1$, or the decoder returns an error if there is no such index for $1,\cdots,M$. We can write the error probability given that the j^{th} codeword is sent as

$$\epsilon(\mathbf{c_j}) = Pr\left[\left\{Z_{\mathbf{c_j}}(\mathbf{y}) = 0\right\}\bigcup_{i<j}\left\{Z_{\mathbf{c_i}}(\mathbf{y}) = 1\right\}|\mathbf{x} = \mathbf{c_j}\right]. \tag{3.12}$$

If we use the union bound for the probability expressions,

$$Pr\left[\left\{Z_{\mathbf{c_j}}(\mathbf{y}) = 0\right\}\bigcup_{i<j}\left\{Z_{\mathbf{c_i}}(\mathbf{y}) = 1\right\}|\mathbf{x} = \mathbf{c_j}\right] \leq \quad Pr\left[i(\mathbf{c_j},\mathbf{y}) \leq log\tfrac{M-1}{2}|\mathbf{x} = \mathbf{c_j}\right] + \sum_{i<j} Pr\left[i(\mathbf{c_i},\mathbf{y}) > log\frac{M-1}{2}|\mathbf{x} = \mathbf{c_j}\right]. \tag{3.13}$$

We use the random coding argument and average (3.13) over codebooks that are generated as independent random variables with distribution $p(\mathbf{x})$ and rewrite the bound as

$$Pr\left[i(\mathbf{x},\mathbf{y}) \leq log\frac{M-1}{2}\right] + (j-1)Pr\left[i(\mathbf{x},\overline{\mathbf{y}}) > log\frac{M-1}{2}\right]. \tag{3.14}$$

Note that $\overline{\mathbf{y}}$ has the same distribution as $\mathbf{y}$ but independent of the transmitted codeword $\mathbf{x}$. We calculated the average probability of error for the j^{th} codeword. Assuming the probability of transmitting any codeword $c_1, \cdots, c_M$ is $1/M$, the average error probability becomes

$$\frac{1}{M}\sum_{j=1}^{M} Pr\left[i(\mathbf{x},\mathbf{y}) \leq log\frac{M-1}{2}\right] + \frac{1}{M}\sum_{j=1}^{M}(j-1)Pr\left[i(\mathbf{x},\overline{\mathbf{y}}) > log\frac{M-1}{2}\right] \tag{3.15}$$

by averaging over all the codewords since

$$\epsilon = \frac{1}{M}\sum_{j=1}^{M}\epsilon(\mathbf{c_j}). \tag{3.16}$$

(3.15) leads to the following expression for the average over all the codewords

$$\epsilon = Pr\left[i(\mathbf{x},\mathbf{y}) \leq log\frac{M-1}{2}\right] + \frac{(M-1)}{2}Pr\left[i(\mathbf{x},\overline{\mathbf{y}}) > log\frac{M-1}{2}\right]. \tag{3.17}$$

Thus, we obtained the expression given in (3.10). Since (3.17) is the average for the random codewords, there must exist at least one codebook that satisfies the upper bound which completes the proof. ■

Note that the bound given in Theorem 3.2.1 does not require any selection of auxiliary constants unlike the bounds given in Section 2.4. In Figure 3.1, the DT bound is compared with Shannon's bound given in 2.4.2 and Gallager's bound stated in Theorem 2.4.4 for the average error probability 10^{-3} and the average SNR (signal to noise ratio) is taken to be 0 dB. In this work, average SNR is defined as the ratio of the average signal power (E_s) to the average noise power (N_0). Feinstein's bound gives almost the same results with Shannon's but satisfying maximal error probability constraint. Since these bounds provide lower bounds for rates, we observe that the DT bound is tighter than the other two bounds. While computing the Gallager's and Shannon's bound the auxiliary constants are optimized by Monte Carlo simulations.

In Section 3.3, we will demonstrate how to calculate the DT bound by using a normal approximation.

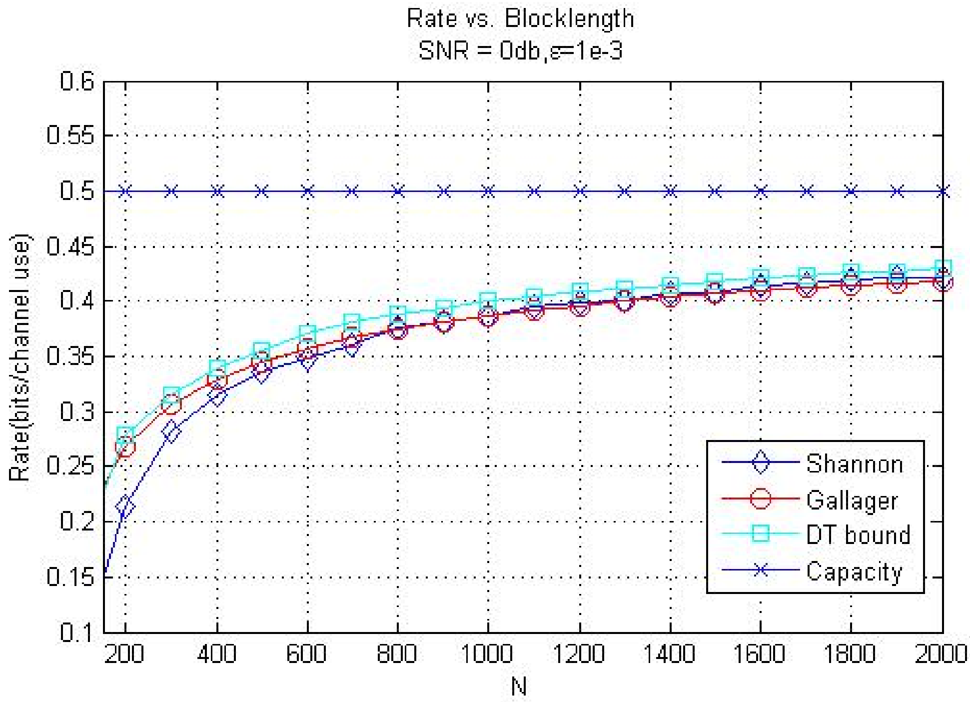

Figure 3.1: Rate vs. Blocklength for Gallager, Shannon and DT bounds

3.3 A Normal Approximation to the DT bound

In order to calculate the average probability of error for a given coding rate by the use of the DT bound presented in Section 3.2, we use the total probability theorem and extract the right handside of (3.1) as

$$\begin{aligned} &1 \cdot Pr\left(i(\mathbf{x},\mathbf{y}) \leq log\frac{M-1}{2}\right) + \\ &E\left[exp\left\{-\left(i(\mathbf{x},\mathbf{y}) - log\frac{M-1}{2}\right)\right\} | i(\mathbf{x},\mathbf{y}) > log\frac{M-1}{2}\right] Pr\left(i(\mathbf{x},\mathbf{y}) > log\frac{M-1}{2}\right). \end{aligned} \tag{3.18}$$

Thus we need the following to calculate the DT bound:

(1) $Pr\left(i(\mathbf{x},\mathbf{y}) \leq log\frac{M-1}{2}\right)$,

(2) $E\left[exp\left\{-\left(i(\mathbf{x},\mathbf{y}) - log\frac{M-1}{2}\right)\right\} | i(\mathbf{x},\mathbf{y}) > log\frac{M-1}{2}\right]$.

In this section, we obtain a normal approximation of the information density in the real and complex additive white Gaussian channel cases. The approximation is stated for the real case as follows:

Theorem 3.3.1 (Gaussian approximation-the real case) *If the information density of the real AWGN channel is approximated as a Gaussian random variable, then its mean is*

$$\mu = \frac{N}{2} log_2(1 + P) \tag{3.19}$$

and its variance is given as

$$\sigma^2 = \left(\frac{log_2 e}{2}\right)^2 \left[4N\frac{P}{1 + P}\right] \tag{3.20}$$

where N is the blocklength, P is the variance of the input x and the noise variance is normalized to 1.

Proof. We will calculate the mean and variance of the information density $i(\mathbf{x}, \mathbf{y})$ for the real AWGN channel represented in the vector form as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{3.21}$$

where **x** is real Gaussian input to the channel, **n** is the real Gaussian noise with elements of zero mean and unit variance, and **y** is the channel output. The variance of zero mean x_i's is P for all i which implies that y_i's are Gaussian random variables with zero mean and variance $1 + P$ for all i. The numerator of information density in (2.20) is the pdf of a jointly Gaussian random vector

$$\mathbf{y}|\mathbf{x} \sim \mathcal{N}(\mathbf{x}, \mathbf{I}_N) \tag{3.22}$$

and the denominator is also the pdf of a jointly Gaussian random vector

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_N(1 + P)). \tag{3.23}$$

In the expressions above, $\mathbf{I}_N$ is the identity matrix of dimension N, $\mathcal{N}(\mathbf{m_z}, \Sigma)$ is the pdf of a real normal random vector **z** of length N with mean vector $\mathbf{m_z}$ and covariance matrix $\mathbf{\Sigma}$ and given as

$$\mathcal{N}(\mathbf{m_z}, \mathbf{\Sigma}) = \frac{1}{(2\pi)^{\frac{N}{2}} det(\mathbf{\Sigma})^{\frac{1}{2}}} exp\left(\frac{-(\mathbf{z} - \mathbf{m_z})^T \mathbf{\Sigma}^{-1} (\mathbf{z} - \mathbf{m_z})}{2}\right), \tag{3.24}$$

where ^{T} is the transpose operation. Then (3.22) becomes

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}} exp\left(-\frac{1}{2}(\mathbf{y} - \mathbf{x})^T (\mathbf{y} - \mathbf{x})\right) \tag{3.25}$$

and (3.23) becomes

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{N}{2}}(P+1)^{\frac{N}{2}}} exp\left(-\frac{1}{2(P+1)}\mathbf{y}^T\mathbf{y}\right). \tag{3.26}$$

By inserting (3.25) and (3.26) to the information density expression,

$$i(\mathbf{x}, \mathbf{y}) = \frac{log_2(e)}{2}\left[\frac{\mathbf{y}^T\mathbf{y}}{P+1} - (\mathbf{y}-\mathbf{x})^T(\mathbf{y}-\mathbf{x})\right] + log_2\left((P+1)^{N/2}\right), \tag{3.27}$$

which is equal to

$$\frac{N}{2}log_2(P+1) + \frac{log_2(e)}{2}\sum_{i=1}^{N}\frac{{y_i}^2}{P+1} - {n_i}^2. \tag{3.28}$$

We apply normal approximation to the summation term in (3.28) for large N. The mean can be calculated as follows:

$$E\left[\sum_{i=1}^{N}\frac{{y_i}^2}{P+1} - {n_i}^2\right] = N\left(\frac{E\left[y_i^2\right]}{P+1} - E\left[n_i^2\right]\right), \tag{3.29}$$

for any i. Since both y_i and n_i are zero mean,

$$E\left[n_i^2\right] = 1, \tag{3.30}$$

$$E\left[y_i^2\right] = (1+P). \tag{3.31}$$

Thus the mean of the summation term becomes

$$N\left(\frac{P+1}{P+1} - 1\right) = 0. \tag{3.32}$$

In order to calculate the variance

$$E\left[\left(\sum_{i=1}^{N}\frac{y_i^2}{P+1} - n_i^2\right)^2\right] = NE\left[\left\{\frac{y_i^2}{P+1} - n_i^2\right\}^2\right], \tag{3.33}$$

for any i since y_i and n_i are independent for $i \neq j$. (3.33) is then equal to

$$N\left(E\left[\frac{y_i^4}{(P+1)^2}\right] - \frac{2}{P+1}E\left[y_i^2 n_i^2\right] + E\left[n_i^4\right]\right). \tag{3.34}$$

The expected values of the different random variables in (3.34) are given as:

$$E\left[y_i^4\right] = 3(P+1)^2, \tag{3.35}$$

$$E\left[y_i^2 n_i^2\right] = P+3, \tag{3.36}$$

and

$$E\left[|n_i|^4\right] = 3. \tag{3.37}$$

In (3.35) and (3.37), we used the 4^{th} order moment expression for a real Gaussian random variable [26]. To show the equality given in (3.36), we rewrite $y_i^2 n_i^2$ as

$$E\left[y_i^2 n_i^2\right] = E\left[(x_i + n_i)^2 n_i^2\right], \tag{3.38}$$

which is equal to

$$E\left[x_i^2 n_i^2\right] + E\left[n_i^4\right] + E\left[2x_i n_i^3\right]. \tag{3.39}$$

Again by using the 4^{th} order moment for n_i and the independence of x_i and n_i, we obtain the expression given in (3.36). By using (3.35),(3.36) and (3.37), (3.34) becomes

$$6N - 2N\frac{P+3}{P+1} = \frac{4NP}{P+1}. \tag{3.40}$$

Thus, we obtain the expressions given in (3.19) and (3.20) which completes the proof. ■

In Figure 3.2, we show the histogram of the information density, where the x-axis is the equally spaced bins between the minimum and maximum values of the information density and the y-axis represents the number of values that fall into each bin, and the normal density by using the mean and variance values given in (3.19) and (3.20) for $N = 300$ and SNR = 0 dB. We conclude that the information density is well-approximated by a normal random variable.

We use Theorem 3.3.1 to calculate the DT bound.
Define $\alpha = i(\mathbf{x},\mathbf{y})$ - $\log\frac{M-1}{2}$. Then, the mean and variance of α can be written as

$$\mu_\alpha = \frac{N}{2}log_2(1+P) - log\frac{M-1}{2} \tag{3.41}$$

and

$$\sigma_\alpha^2 = \left(\frac{log_2 e}{2}\right)^2 \left[4N\frac{P}{1+P}\right] \tag{3.42}$$

by using Theorem 3.3.1. In order to find the average probability of error given in (3.1), we need to calculate $Pr(\alpha \leq 0)$ and $E\left[exp(-\alpha)|\alpha > 0\right]$. Then, we can express the DT bound as

$$Pr\{\alpha \leq 0\} + Pr\{\alpha > 0\} E\left[exp(-\alpha)|\alpha > 0\right]. \tag{3.43}$$

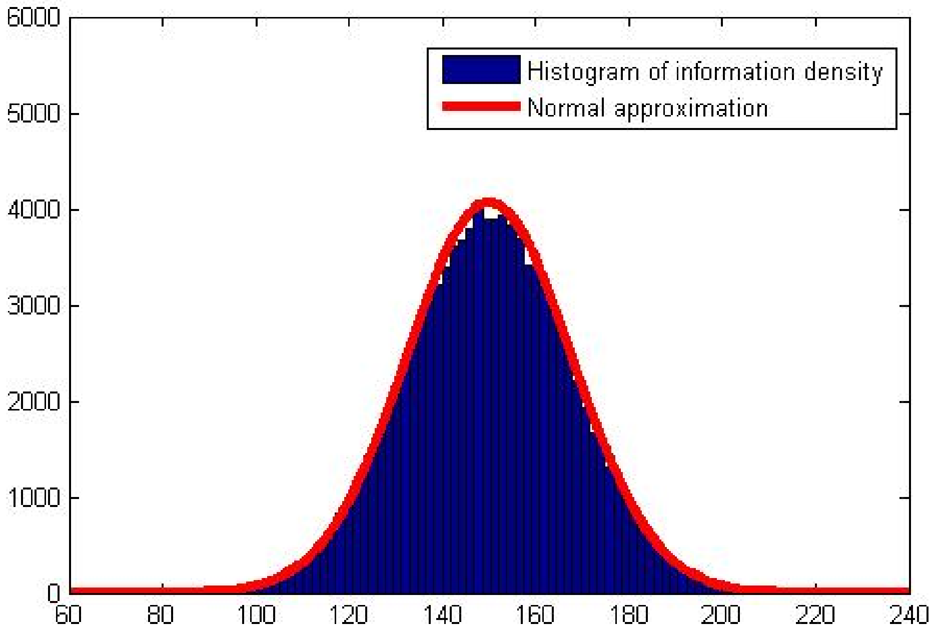

Figure 3.2: Normal approximation to the information density (real case)

The former one can be calculated as follows: If we define

$$\eta_1 = -\frac{\mu_\alpha}{\sigma_\alpha} \tag{3.44}$$

then this probability can be rewritten as

$$Pr(\alpha \leq \eta_1 \sigma_\alpha + \mu_\alpha), \tag{3.45}$$

so that the probability in (3.45) can be expressed as

$$Pr\left(\frac{\alpha - \mu_\alpha}{\sigma_\alpha} < \eta_1\right) = erfc\left(-\frac{\eta_1}{\sqrt{2}}\right)/2 \tag{3.46}$$

in terms of the complementary error function

$$erfc(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty exp(-t^2)\,dt. \tag{3.47}$$

Using definitions directly, the other term is written as

$$E\left[exp(-\alpha)|\alpha > 0\right] = \int_0^\infty exp(-\alpha) p_\alpha(\alpha|\alpha > 0)\,d\alpha. \tag{3.48}$$

If $K = \frac{1}{Pr(\alpha>0)}$ is defined and since α is considered to be a Gaussian random variable (3.48) can be rewritten as

$$K \int_0^\infty exp(-\alpha) \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} exp\left(-\frac{(\alpha - \mu_\alpha)^2}{2\sigma_\alpha^2}\right) d\alpha \tag{3.49}$$

$$= K\int_0^\infty \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} exp\left(-\left(\frac{(\alpha-\mu_\alpha)^2}{2\sigma_\alpha^2}+\alpha\right)\right) d\alpha. \tag{3.50}$$

Completing the squares of the exponential term, one obtains the following:

$$= Kexp\left(\frac{\sigma_\alpha^2}{2}-\mu_\alpha\right)\int_0^\infty \frac{1}{\sqrt{2\pi\sigma_\alpha^2}} exp\left(-\frac{(\alpha-(\mu-\sigma^2))^2}{2\sigma_\alpha^2}\right) d\alpha, \tag{3.51}$$

which is equal to

$$= \frac{K}{2} exp\left(\frac{\sigma_\alpha^2}{2}-\mu_\alpha\right) erfc\left(\frac{\eta_2}{\sqrt{2}}\right), \tag{3.52}$$

where

$$\eta_2 = -\frac{\mu_\alpha - \sigma_\alpha^2}{\sigma_\alpha}. \tag{3.53}$$

Thus, the average probability of error is bounded as

$$\epsilon \leq \left[erfc\left(-\frac{\eta_1}{\sqrt{2}}\right)/2\right] + \left[1-\left(erfc(-\frac{\eta_1}{\sqrt{2}})/2\right)\right] Kexp\left(\frac{\sigma_\alpha^2}{2}-\mu_\alpha\right)\left[erfc\left(\frac{\eta_2}{\sqrt{2}}/2\right)\right] \tag{3.54}$$

for the AWGN channel.

By using the definition of K, we can simplify (3.54) as

$$\epsilon \leq \left(erfc(-\frac{\eta_1}{\sqrt{2}})/2\right) + exp\left(\frac{\sigma_\alpha^2}{2}-\mu_\alpha\right)\left(erfc(\frac{\eta_2}{\sqrt{2}}/2)\right). \tag{3.55}$$

We can calculate the DT bound easily by using (3.55). The calculation for the complex AWGN case is similar and is stated in Theorem 3.3.2.

Theorem 3.3.2 (Gaussian approximation-the complex case) *If the information density of the complex AWGN channel is approximated as a Gaussian random variable, then its mean is*

$$\mu = Nlog_2(1+P) \tag{3.56}$$

and its variance is given as

$$\sigma^2 = (log_2 e)^2\left[2N\frac{P}{1+P}\right] \tag{3.57}$$

where N is the blocklength and P is the variance of the input x.

Proof. We will calculate the mean and variance of the information density $i(\mathbf{x},\mathbf{y})$ for the complex AWGN channel given as

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \tag{3.58}$$

where $\mathbf{x}$ is complex Gaussian input to the channel, $\mathbf{n}$ is the circularly symmetric complex Gaussian noise with elements of zero mean and unit variance, and $\mathbf{y}$ is the complex channel output. The variance of zero mean x_i's is P for all i which implies that y_i's are complex Gaussian with zero mean and variance $1 + P$ for all i. The numerator of information density in (2.20) is the pdf of a jointly Gaussian complex random vector

$$\mathbf{y}|\mathbf{x} \sim \mathcal{CN}(\mathbf{x}, I_N) \tag{3.59}$$

and the denominator is also the pdf of a jointly Gaussian random vector

$$\mathbf{y} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N(1 + P)). \tag{3.60}$$

In (3.59) and (3.60), $\mathcal{CN}(\mathbf{m_z}, \Sigma)$ is the pdf of a complex normal random vector $\mathbf{z}$ of length N with mean vector $\mathbf{m_z}$ and covariance matrix $\mathbf{\Sigma}$ and given as

$$p(\mathbf{z}) = \frac{1}{(2\pi)^N det(\mathbf{\Sigma})} exp\left(\frac{-(\mathbf{z} - \mathbf{m_z})^H \mathbf{\Sigma}^{-1}(\mathbf{z} - \mathbf{m_z})}{2}\right), \tag{3.61}$$

where the superscript ^{H} is the Hermitian operation. The numerator of the information density can be expressed as

$$p(\mathbf{y}|\mathbf{x}) = \frac{2}{(2\pi)^N} exp\left(-(\mathbf{y} - \mathbf{x})^H(\mathbf{y} - \mathbf{x})\right) \tag{3.62}$$

and (3.60) becomes

$$p(\mathbf{y}) = \frac{2}{(2\pi)^N (P + 1)^N} exp\left(-\frac{1}{P + 1}\mathbf{y}^H\mathbf{y}\right). \tag{3.63}$$

Rewriting the information density expression by using (3.62) and (3.63),

$$i(\mathbf{x}, \mathbf{y}) = log_2(e)\left[\frac{\mathbf{y}^H\mathbf{y}}{P + 1} - (\mathbf{y} - \mathbf{x})^H(\mathbf{y} - \mathbf{x})\right] + log_2\left((P + 1)^N\right), \tag{3.64}$$

which is equal to

$$Nlog_2(P + 1) + log_2(e)\sum_{i=1}^{N}\frac{|y_i|^2}{P + 1} - |n_i|^2. \tag{3.65}$$

As in the real case, we apply normal approximation to the summation term in (3.65) by using the central limit theorem [27]. The mean can be calculated as follows:

$$E\left[\sum_{i=1}^{N}\frac{|y_i|^2}{P + 1} - |n_i|^2\right] = N\left(\frac{E\left[|y_i|^2\right]}{P + 1} - E\left[|n_i|^2\right]\right), \tag{3.66}$$

for any i. Since both y_i and n_i are zero mean,

$$E\left[|n_i|^2\right] = 1, \tag{3.67}$$

$$E\left[|y_i|^2\right] = (1 + P). \tag{3.68}$$

Thus the mean of the summation term becomes

$$N\left(\frac{P+1}{P+1} - 1\right) = 0. \tag{3.69}$$

In order to calculate the variance

$$E\left[\left(\sum_{i=1}^{N} \frac{|y_i|^2}{P+1} - |n_i|^2\right)^2\right] = NE\left[\left\{\frac{|y_i|^2}{P+1} - |n_i|^2\right\}^2\right], \tag{3.70}$$

for any i since the y_i and n_i are independent for $i \neq j$. (3.70) is then equal to

$$N\left(E\left[\frac{|y_i|^4}{(P+1)^2}\right] - \frac{2}{P+1}E\left[|y_i|^2|n_i|^2\right] + E\left[|n_i^4|\right]\right). \tag{3.71}$$

By using the same argument in the proof of Theorem 3.3.1, the expected values of the different random variables arising in (3.71) can be expressed as:

$$E\left[|y_i|^4\right] = 2(P + 1)^2, \tag{3.72}$$

$$E\left[|y_i|^2|n_i|^2\right] = P + 2, \tag{3.73}$$

and

$$E\left[|n_i|^4\right] = 2. \tag{3.74}$$

By using (3.72),(3.73) and (3.74), (3.71) becomes

$$4N - 2N\frac{P+2}{P+1}. \tag{3.75}$$

Thus, we obtain the equations given in (3.56) and (3.57) which completes the proof. ■

In Figure 3.3, we demonstrate how well the normal approximation of the information density is, for the complex AWGN channel with $N = 300$ and SNR = 0 dB where the axes are the same as in Figure 3.2. We generated complex normal random variables for the channel input and the additive noise and plotted the histogram of the information density. When the histogram is compared with the pdf of a Gaussian random variable with mean and variance given in Theorem 3.3.2, one conclude that the information density can be well approximated as a Gaussian random variable for the complex case.

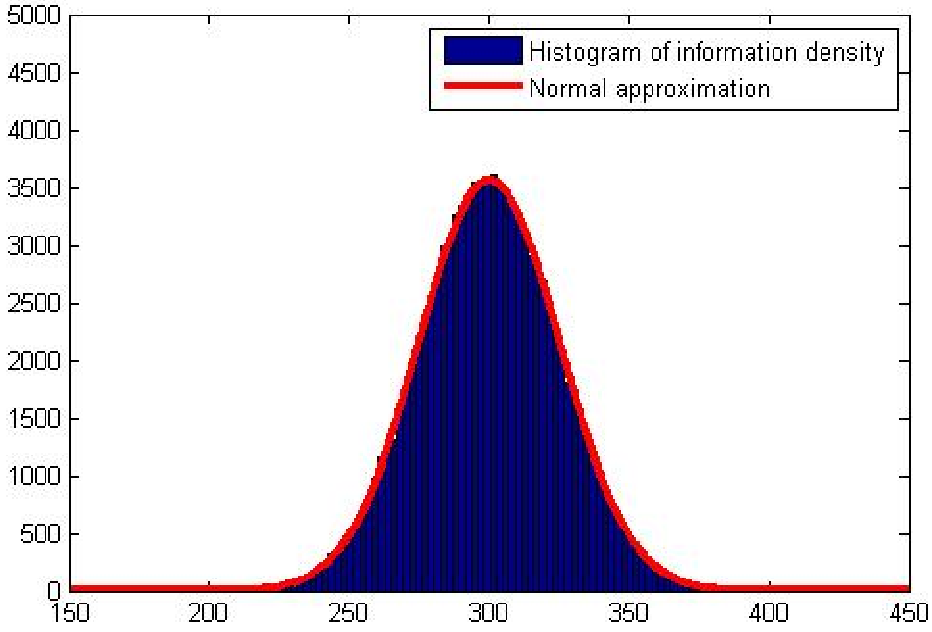

Figure 3.3: Normal approximation to the information density (complex case)

In order to calculate the DT bound for the complex AWGN channel, we redefine the mean and variance of α as

$$\mu_\alpha = Nlog_2(1+P) - log\frac{M-1}{2} \tag{3.76}$$

and

$$\sigma_\alpha^2 = (log_2 e)^2\left[2N\frac{P}{1+P}\right] \tag{3.77}$$

Then, the same arguments given for the real case can be applied.

In Figure 3.4, the maximum achieved coding rate as a function of the blocklength N for the real AWGN channel and the Gaussian input alphabet is displayed for the proposed normal approximation scheme by using Theorem 3.3.1 and (3.55) . The average probability of error is taken as 10^{-3} and the average SNR is taken to be 6 dB. The DT bound is also calculated by 250,000 Monte Carlo simulations for the same system. Shannon capacity for the same channel is shown in the figure. In Figure 3.5, the DT bound is calculated for the complex AWGN channel with the same parameters. In Figures 3.6, 3.7, the DT bound is calculated for real and complex cases respectively where SNR is 0 dB and the error probability is taken as 10^{-3}.

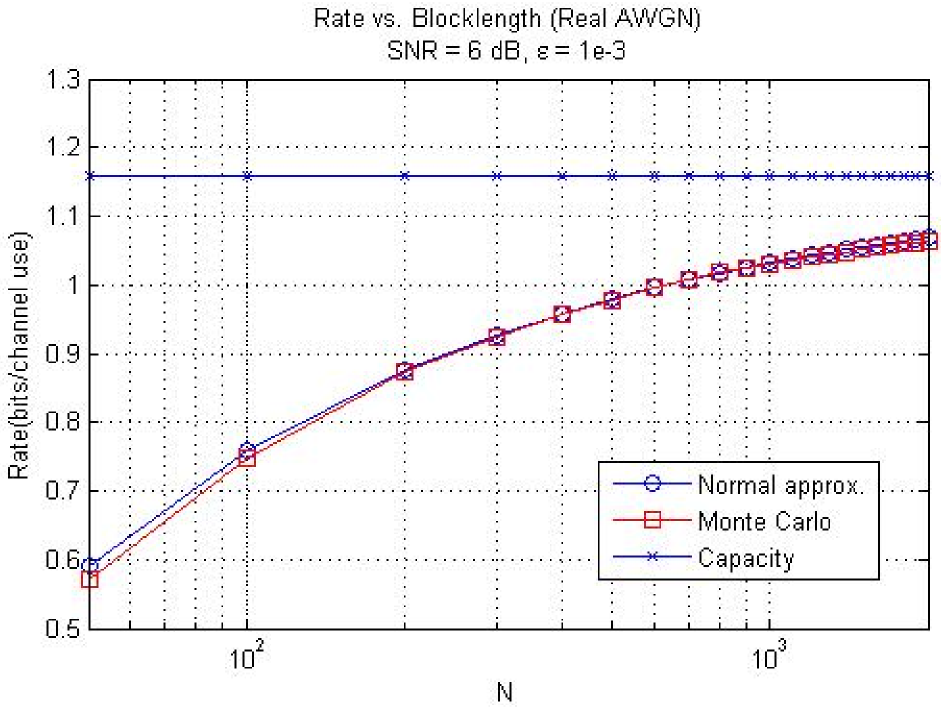

Figure 3.4: Rate vs. Blocklength for real Gaussian alphabet, SNR = 6 dB, $\epsilon = 10^{-3}$

It can be seen that as the blocklength N increases, maximum coding rate approaches the channel capacity and our proposed normal approximation model provide results very similar to those of the Monte Carlo simulations. These results show that the Gaussian approximation model works well for computing the finite blocklength channel capacity.

3.4 The DT bound for Block Fading Channels

We will apply the normal approximation presented in Section 3.3 to the block fading channel model which is introduced in Section 2.2. We can represent the block fading channel as

$$\mathbf{y} = \mathbf{hx} + \mathbf{n}, \tag{3.78}$$

where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_N \end{bmatrix}, \mathbf{n} = \begin{bmatrix} n_1 \\ \vdots \\ n_N \end{bmatrix} \tag{3.79}$$

and $\mathbf{h}$ is the diagonal matrix whose diagonal entries are the complex valued $h_1, \cdots, h_N$ which are not necessarily different. If the number of fading blocks is F, then there exist N/F inde-

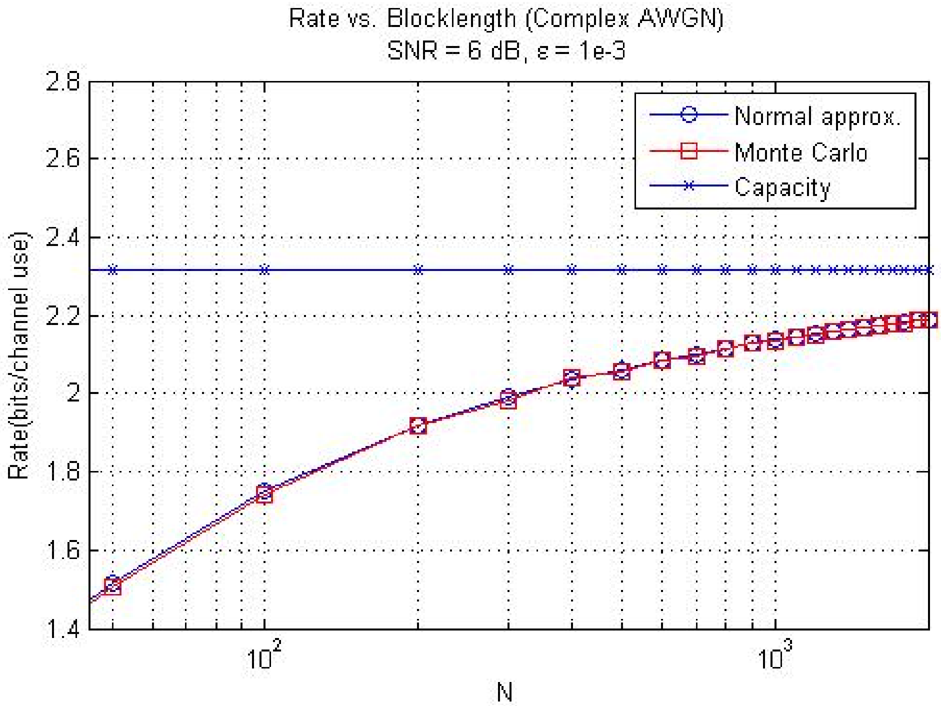

Figure 3.5: Rate vs. Blocklength for complex Gaussian alphabet, SNR = 6 dB, $\epsilon = 10^{-3}$

pendent realizations of the fading coefficients.

Theorem 3.4.1 (Gaussian approximation-block fading channels) *If the information density $i(\mathbf{x}, \mathbf{y}|\mathbf{h})$ of the fading channel for a given fading coefficient matrix $\mathbf{h}$ is approximated as a Gaussian random variable, then its mean is*

$$\mu = \sum_{i=1}^{N} log_2 \left(1 + |h_i|^2 P\right) \tag{3.80}$$

and its variance is given as

$$\sigma^2 = (log_2 e)^2 \left[4N - 2 \sum_{i=1}^{N} \frac{2 + |h_i|^2 P}{1 + |h_i|^2 P} \right] \tag{3.81}$$

where N is the blocklength, P is the variance of the input x with complex Gaussian alphabet and the variance of the complex normal noise $\mathbf{n}$ is normalized to 1.

Proof. The information density for given fading coefficients can be expressed as

$$i(\mathbf{x}, \mathbf{y}|\mathbf{h}) = log_2 \frac{p(\mathbf{y}|\mathbf{x}, \mathbf{h})}{p(\mathbf{y}|\mathbf{h})}. \tag{3.82}$$

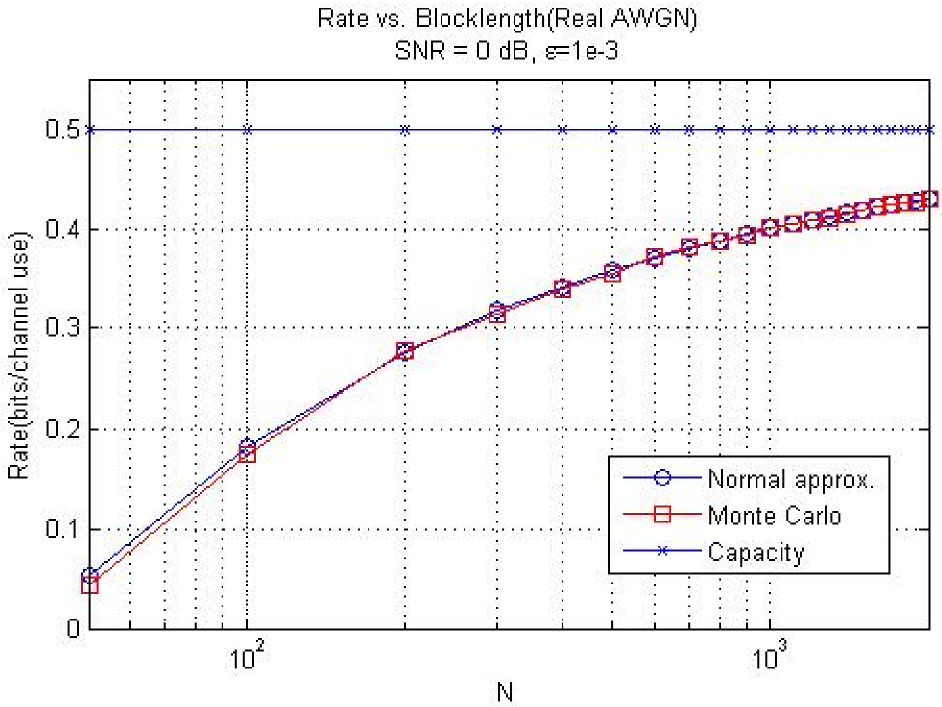

Figure 3.6: Rate vs. Blocklength for real Gaussian alphabet, SNR = 0 dB, $\epsilon = 10^{-3}$

The numerator in (3.82) is the pdf of a jointly Gaussian complex random vector which can be represented as

$$\mathbf{y}|\mathbf{x},\mathbf{h} \sim \mathcal{CN}(\mathbf{hx}, \mathbf{I}_N) \tag{3.83}$$

and the denominator in (3.82) is again the pdf of a jointly Gaussian complex random vector with zero mean

$$\mathbf{y}|\mathbf{h} \sim \mathcal{CN}(\mathbf{0}, P\mathbf{h_{abs}} + \mathbf{I}_N), \tag{3.84}$$

where $\mathbf{h_{abs}}$ is the diagonal matrix whose diagonal entries are $|h_1|^2, \cdots, |h_N|^2$.
(3.83) and (3.84) can be expressed as

$$p(\mathbf{y}|\mathbf{x},\mathbf{h}) = \frac{2}{(2\pi)^N} exp(-(\mathbf{y}-\mathbf{hx})^H(\mathbf{y}-\mathbf{hx})) \tag{3.85}$$

and

$$p(\mathbf{y}|\mathbf{h}) = \frac{2}{(2\pi)^N \prod_{i=1}^{N}\left(|h_i|^2P+1\right)} exp\left(-\mathbf{y}^H diag\left\{\frac{1}{|h_1|^2P+1}, \cdots, \frac{1}{|h_N|^2P+1}\right\}\mathbf{y}\right), \tag{3.86}$$

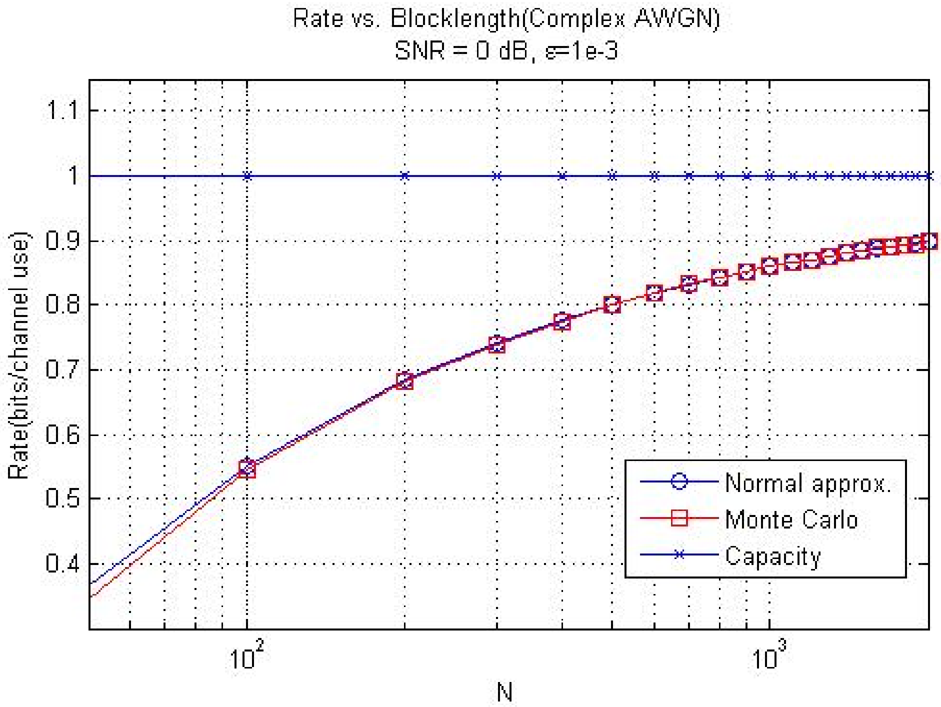

Figure 3.7: Rate vs. Blocklength for complex Gaussian alphabet, SNR = 0 dB, $\epsilon = 10^{-3}$

where $diag\{a_1, \cdots, a_N\}$ is an $N \times N$ diagonal matrix with diagonal entries $a_1, \cdots, a_N$. By inserting (3.85) and (3.86) into the information density equations and after several straightforward steps, (3.82) can be written as

$$i(\mathbf{x}, \mathbf{y}|\mathbf{h}) = \left(\sum_{i=1}^{N} log_2(1 + |h_i|^2 P)\right) + log_2 e \left[\sum_{i=1}^{N} \frac{|y_i|^2}{1 + |h_i|^2 P} - |n_i|^2\right]. \tag{3.87}$$

We apply normal approximation to the summation term on the right in (3.87) again by using the central limit theorem.

The mean is computed as in:

$$E\left[\sum_{i=1}^{N} \frac{|y_i|^2}{|h_i|^2 P + 1} - |n_i|^2\right], \tag{3.88}$$

which is equal to

$$E\left[\sum_{i=1}^{N} \frac{|h_i x_i + n_i|^2}{|h_i|^2 P + 1} - |n_i|^2\right]. \tag{3.89}$$

We rewrite (3.89) as

$$\sum_{i=1}^{N} \left(\frac{|h_i x_i + n_i|^2 P + 1}{|h_i|^2 P + 1} - 1\right), \tag{3.90}$$

which becomes zero after simplifications.

In order to calculate the variance, we have to calculate

$$E\left[\left(\sum_{i=1}^{N}\frac{|h_i x_i + n_i|^2}{|h_i|^2 P + 1} - |n_i|^2\right)^2\right]. \tag{3.91}$$

(3.91) can be expressed as

$$\sum_{i=1}^{N} E\left[\frac{|h_i x_i + n_i|^2}{|h_i|^2 P + 1} - |n_i|^2\right]^2 \tag{3.92}$$

since y_i and n_j are independent for $i \neq j$ given the fading coefficients h_i's. If we rewrite (3.92) as

$$\sum_{i=1}^{N} E\left[\frac{|y_i|^4}{(|h_i|^2 P + 1)^2}\right] + E\left[|n_i|^4\right] - \frac{2}{|h_i|^2 P + 1} E\left[|y_i|^2 |n_i|^2\right] \tag{3.93}$$

and using (3.72),(3.73) and (3.74), (3.93) is equal to

$$4N - 2\sum_{i=1}^{N}\frac{2 + |h_i|^2 P}{1 + |h_i|^2 P}. \tag{3.94}$$

Then, the mean and variance of the information density is derived as in (3.80) and (3.81). ■

Thus for a given fading coefficient matrix, we can use Theorem 3.4.1 in order to calculate the DT bound for block fading channels. If we consider that there are only F probably different fading coefficients we may take the mean and variance of $\alpha = i(\mathbf{x}, \mathbf{y}|\mathbf{h}) - log\frac{M-1}{2}$ to calculate the DT bound as

$$\mu_\alpha = N\frac{1}{F}\sum_{f=1}^{F}\left[log_2(1 + |h_f|^2 P)\right] - log\frac{M-1}{2} \tag{3.95}$$

and

$$\sigma_\alpha^2 = (log_2 e)^2 \left[4N - 2N\frac{1}{F}\sum_{f=1}^{F} 1 + \frac{1}{1 + |h_f|^2 P}\right]. \tag{3.96}$$

We take the expected value of the error probability given in (3.55) over fading coefficients which gives the average error probability for block fading channels. Note that the blocklength is taken into account when calculating the error probability unlike the case in outage probability calculations given in (2.13) where the blocklength is considered to be infinite.

In Figure 3.8, the average error probability of the Rayleigh fading channel presented in Section 2.2 for different average SNR values is demonstrated for a fixed code rate $R = 0.6$ with varying number of fading blocks. In Figure 3.9, the same is displayed for complex AWGN channel. The blocklength N is taken to be 100, 200, 400 and the number of fading blocks F is taken as 4,10, 20. For the fading case, 250,000 different fading coefficient vectors are generated and the mean of the resulting error bound is plotted. Since the diversity order increases as the number of fading blocks increases, the average probability of error decreases. Also the difference between different blocklength values increases as the number of fading blocks is increased.

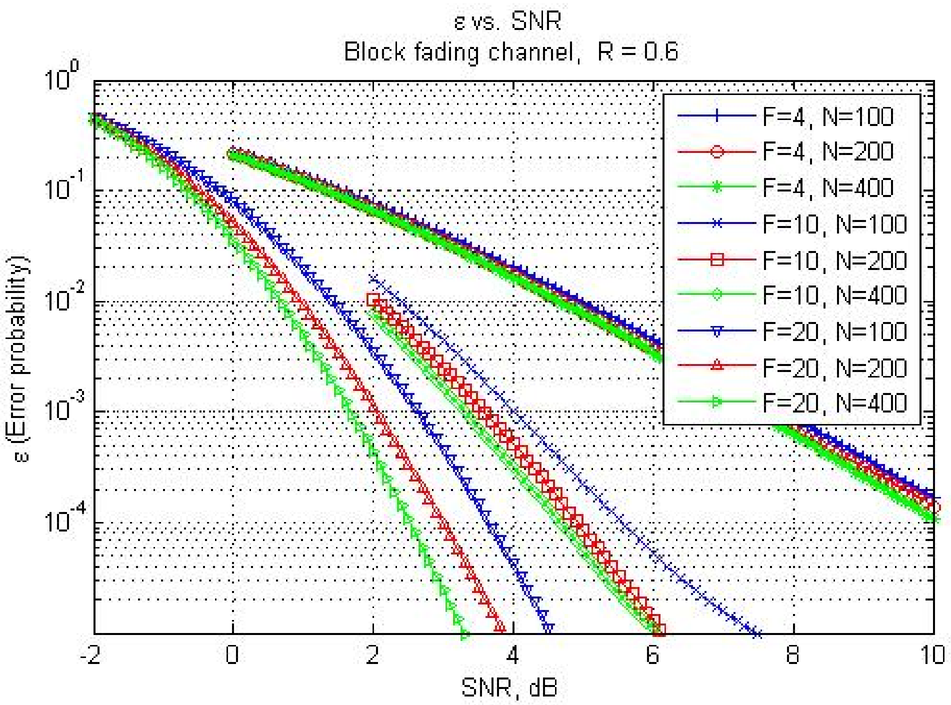

Figure 3.8: Error vs. SNR for Block Fading Channel

We see that the effect of the blocklength is still an important parameter as the number of fading blocks is increased. The SNR gain for 20 fading blocks is almost 1 dB when N is incremented from 100 to 400. The SNR gain (which is 1.8 dB) is greater in the AWGN channel than the fading channels which implies that the effect of blocklength decreases in the fading case. It can be concluded from these results that in the fading case the blocklength still plays an important role when calculating the code rate for a given average error probability.

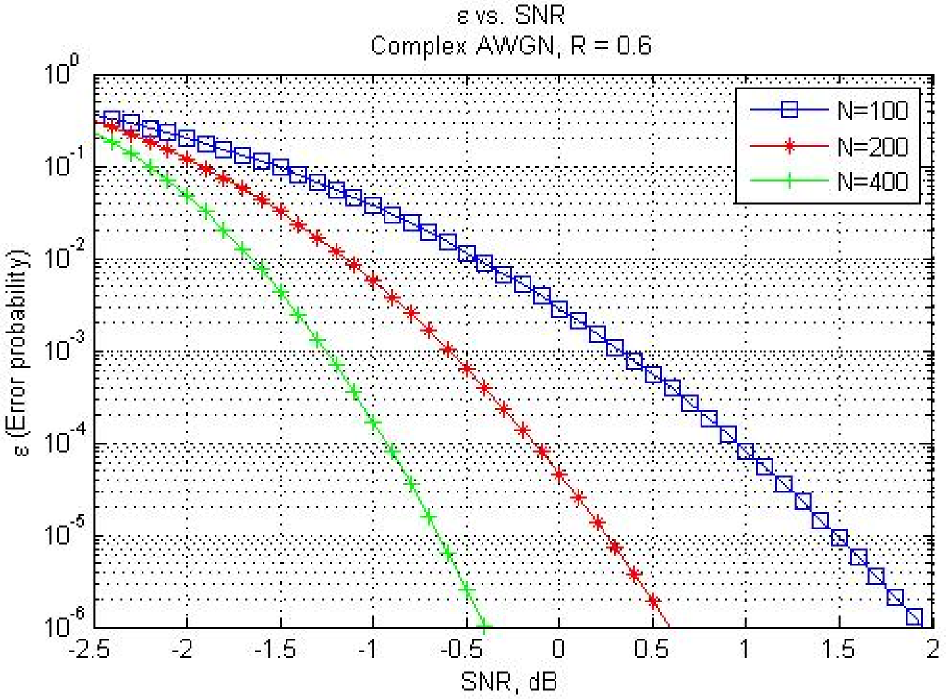

Figure 3.9: Error vs. SNR for Complex AWGN Channel

3.5 The DT bound for Discrete Alphabets

In previous sections of Chapter 3, we analyzed the Dependence Testing bound for Gaussian input alphabets. In this section, we will present the results on the channels with constrained input alphabets. Unfortunately, we cannot calculate the mean and variance of the information density for the constrained input case directly. Thus, we used Monte Carlo simulations to calculate the mean and the variance [28] by calculating the information density

$$i(\mathbf{x}, \mathbf{y}) = Nlog_2|S| - \sum_{i=1}^{N} log_2 \sum_{x_i' \in S} exp\left(\frac{-\left[|x_i - x_i' + n_i|^2 - |n_i|^2 \right]}{N_0} \right) \quad (3.97)$$

where S is the constrained alphabet with $E(|x_i|^2 = 1)$ and N_0 is the noise variance. After calculating μ_α and σ_α^2 by using (3.97) with Monte Carlo simulations, the same arguments introduced in Section 3.3 can be used to calculate the DT bound.

In Figure 3.10, the maximum achievable code rate for the QPSK input alphabet as a function of the blocklength is depicted with SNR = 6 dB and $\epsilon = 10^{-3}$ and in Figure 3.11, the DT bound is calculated for the 8PSK input alphabet with the same average SNR and error probability. In order to compare the validity of our normal approximation, the DT bound is also calculated with 100,000 Monte Carlo simulations. The channel capacity for the given input alphabet [28] is also plotted. We conclude that our normal approximation to compute the DT bound for QPSK inputs gets close to the Monte Carlo results for blocklength $N > 500$, however for 8PSK input alphabet and for $N > 200$, the results with Gaussian approximation and Monte Carlo simulations are almost the same. Thus, the DT bound can also be calculated easily for constrained input alphabets by making use of the normal approximation. If we compare Figures 3.10 and 3.11 with Figure 3.5, it can be concluded that the DT bound for constrained input channels approaches the channel capacity faster than for Gaussian input channels. If we compare the blocklength values where the code rate attains 0.9 of the Shannon capacity, we see that N is equal to 600 for complex Gaussian alphabet, whereas $N \approx 200$ for QPSK and 300 for 8PSK input alphabet. The convergence turns out to be faster in QPSK than 8PSK.

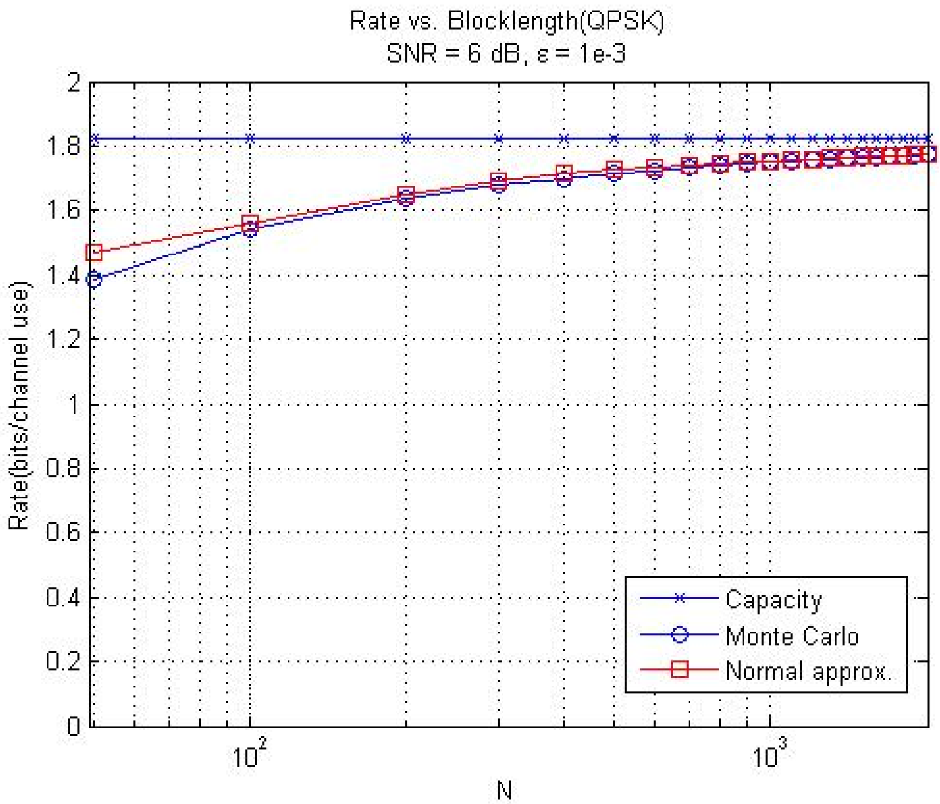

Figure 3.10: Rate vs. Blocklength for QPSK alphabet, SNR = 6 dB, $\epsilon = 10^{-3}$

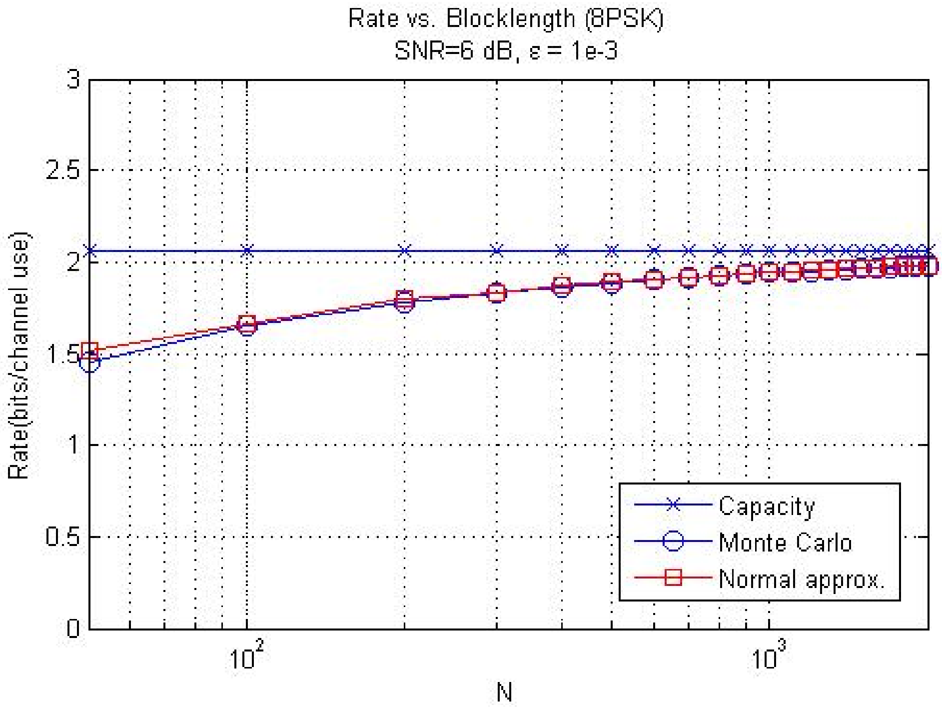

Figure 3.11: Rate vs. Blocklength for 8PSK alphabet, SNR = 6 dB, $\epsilon = 10^{-3}$

3.6 Conclusion

We developed a normal approximation to the Dependence Testing bound in this chapter. Although there had been previous bounds for the finite blocklength channels in the literature, those ones required the selection and optimization of auxiliary constants to provide tight bounds for the error probability. The DT bound provides a bound which does not include any auxiliary variables but an analytical computation cannot be obtained. With the normal approximation approach we propose, analytical solutions can be obtained. We have proposed a Gaussian approximation to the information density in order to calculate the DT bound both for AWGN and block fading channels. The proposed approximation provides us results very close to the DT bound. We also used the normal approximation and computed the DT bound for constrained input alphabets.

CHAPTER 4

Power Allocation with CSIT

4.1 Introduction

In fading channels, it is possible to change the instantaneous power of the transmitted signal so that the channel capacity is increased or the error probability is decreased, if the channel state information is available to the transmitter. In Section 2.3, we discussed the channel capacity where channel state information is not available to the transmitter and the transmitted power is constant over the whole transmission. Also the outage probability calculations are performed without considering any channel state information to the transmitter. In Section 2.4 and in Chapter 3, the bounds for the finite blocklength channels are developed with no power adaptation at the transmitter side. In this chapter, we will consider power allocation strategies when CSIT is available to the transmitter. The channel capacity with CSIT will be introduced and the outage probability minimization for a given code rate will be discussed. In addition, the existing power allocation strategies in the literature to minimize the error bounds for finite blocklength channels will be discussed. The DT bound will be calculated for the case where power allocation is allowed for block fading channels, by using the normal approximation proposed in Chapter 3.

4.2 Channel Capacity with CSIT

In the case that the transmitter has the CSIT, the transmitter may vary the power and the coding rate of the transmitted codeword by taking into account the channel state [9]. The transmitter may adapt its transmission strategy by using the channel state information. Since the transmission power is limited, there is usually an average power constraint

$$\int_0^\infty P(h)p(h)dh \leq \overline{P} \tag{4.1}$$

when evaluating the channel capacity, where $P(h)$ is the allocated power for a given fading state h, $p(h)$ is the pdf of the channel states and $\overline{P}$ is the average power that is transmitted. By making use of the CSIT, Shannon capacity is calculated in [15]. In the fading case and with the power optimally distributed over time we can modify (2.7) and define the fading channel capacity with average power constraint as

$$C = \max_{P(h):\int P(h)p(h)dh\leq\overline{P}} \int_0^\infty log_2\left(1 + \frac{|h|^2 P(h)}{N}\right)p(h)dh. \tag{4.2}$$

The main idea behind the proof presented in [15] is a time diversity system with multiplexed input and demultiplexed output. The range of fading states are first quantized to a finite set $\left\{h_j : 1 \leq j \leq K\right\}$. Then for each j, an encoding/decoding architecture for an AWGN channel with SNR $\gamma_j = \frac{P|h_j|^2}{N}$ is designed. The input x_j for the j^{th} encoder has average power $P(h_j)$ and encoding rate $R_j = C_j$, where C_j is the capacity of an AWGN channel with received SNR γ_j. The fading coefficients are associated with the corresponding encoder/decoder pair and the average rate on the channel is just the sum of rates associated with each of the h_j channels weighted by $p(h_j)$. This produces the average capacity formula expressed in (4.2) when the number of states is increased to the infinity.

We need to find the optimal power allocation strategy in order to calculate the channel capacity with power allocation. Let's form the Lagrangian operator for the allocated power as

$$J(P(h)) = \int_0^\infty log_2\left(1 + \frac{|h|^2 P(h)}{N}\right)p(h)dh - \lambda \int P(h)p(h)dh. \tag{4.3}$$

If we differantiate (4.3) with respect to the $P(h)$ and using Kuhn-Tucker conditions [15], we can express the power to be allocated as

$$P(h) = \left[\frac{ln2}{\lambda} - \frac{N}{|h|^2}\right]^+ . \tag{4.4}$$

Defining $\gamma_0 = \frac{\lambda}{ln2}$ and $\gamma = \frac{|h|^2}{N}$, which is the instantaneous SNR for a given fading coefficient h and unity allocated power, (4.4) can be expressed as

$$P(h) = \left[\frac{1}{\gamma_0} - \frac{1}{\gamma}\right]^+ . \tag{4.5}$$

The threshold γ_0 in (4.5) must be found to satify the average power constraint given in (4.1) and it can be written as

$$\int_{\gamma_0}^\infty \left[\frac{1}{\gamma_0} - \frac{1}{\gamma}\right]p(h)dh = \overline{P}. \tag{4.6}$$

The capacity is achieved with a time-varying data rate where the rate is related to the instantaneous SNR value. Note that as the fading coefficient value increases the power allocated for that channel state is incremented after a threshold value. This strategy is referred to as water-filling. The water-filling strategy refers to the fact that the line $1/\gamma$ is the bottom of a bowl and the water is poured to the bowl until the water level is reached to a constant level $1/\gamma_0$. As the channel quality decreases, less power is allocated and the coding rate is decreased. When the channel conditions are in a good state, we take advantage of it and send high data rates by allocating more power.

Let's investigate the benefits of the power allocation scheme in the fading case. In a fading channel, if we do not have CSI at the transmitting side and thus do not implement any power allocation then the capacity formula given in (2.7) applies. By using Jensen's inequality we can bound the ergodic channel capacity in the fading case as

$$E_h\left[log(1+|h|^2\frac{P}{N})\right] \leq log\left(1+E\left[|h|^2\right]\frac{P}{N}\right) = log\left(1+\overline{SNR}\right), \tag{4.7}$$

where $E\left[|h|^2\right]\frac{P}{N}$ is the average SNR which is denoted as $\overline{SNR}$. Thus we can conclude that the Shannon capacity of a fading channel without CSIT is less than or equal to the Shannon capacity of an AWGN channel with the same average SNR. In Figure 4.1, the capacity when the transmitter has the channel state information and applies the water-filling formula given in (4.5) to the transmitted power is compared with the AWGN channel capacity and the capacity of the fading channel without CSIT. The Rayleigh fading channel model is used in obtaining the curves. Note that the fading channel will occasionally have high instantaneous SNR values even in the low average SNR region. We can see that the fading channel capacity is larger than that of the AWGN channel at low SNR values when the transmitter has the channel state information. This is because, in the fading case, the transmitter saves its power when the channel has a small fading gain and utilizes it at high valued channel gain states. Also we see that the difference between the capacity with and without CSIT becomes negligible at high SNR. This is due to the fact that there is no reason to preserve power for high SNR values since the channel gain is high almost all the time and thus power allocation is not needed.

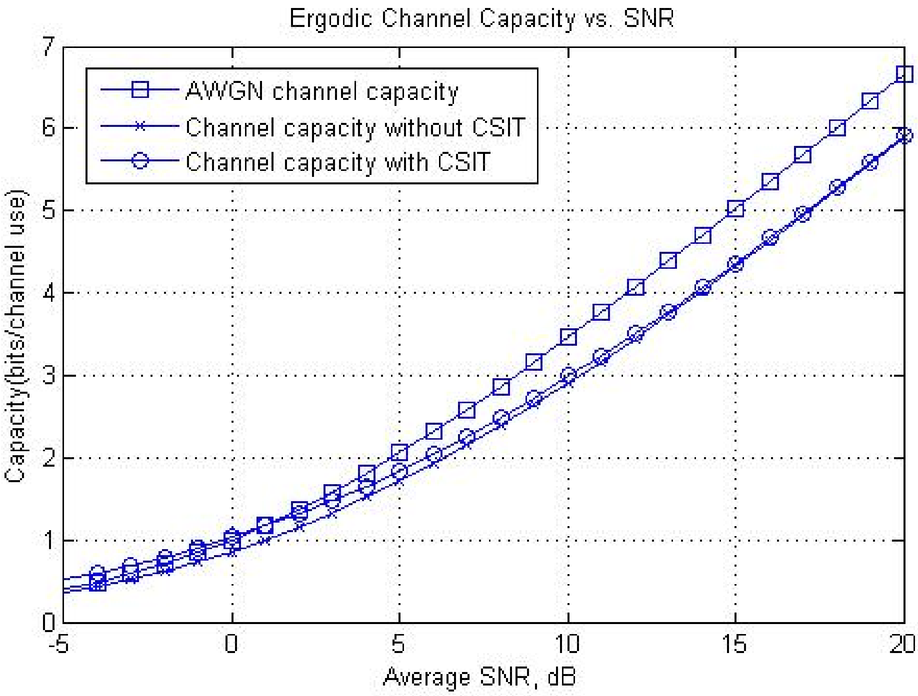

Figure 4.1: Ergodic Channel Capacity in Rayleigh fading

4.3 Power Control for Outage Probability Minimization

In Section 2.3, we discussed the outage probability concept which is taken as the main parameter to be investigated in most of the studies regarding the finite blocklength channels. The power allocation strategies can be applied to minimize the outage probability for the finite blocklength case if the channel states are known by the transmitter before sending the codeword. The power constraint can be applied in two ways [16]. One is the long-term power contraint, which can be expressed as

$$\begin{cases} \text{Minimize}\,(P_{out}(R, P_{\mathbf{h}})) \\ \text{Subject to}\, E\left[\frac{1}{F}\sum_{i=1}^{F} P_{\mathbf{h}}(i)\right] \leq \overline{P} \end{cases}, \tag{4.8}$$

where $P_{\mathbf{h}}(i)$ is the power allocated to the i^{th} fading block with fading coefficients $\mathbf{h}$ of length F. In this case, the transmitter does not need to consume all the power for every codeword block. Only the mean of the power that is allocated should satisfy the average power constraint. However, this may bring in peak power limitations which are studied in [29]. The second way is the short-term power constraint which can be stated as

$$\begin{cases} \text{Minimize}\,(P_{out}(R, P_{\mathbf{h}})) \\ \text{Subject to}\ \frac{1}{F}\sum_{i=1}^{F} P_{\mathbf{h}}(i) \leq \overline{P}, \text{with probability 1} \end{cases} . \tag{4.9}$$

In the short-term limitation, the transmitter should use all the available power for every codeword block and it is only allowed to make power allocation between F fading blocks. The solution to the short-term constraint is given in [16] as

$$P_{\mathbf{h}} = \begin{cases} P^{st}(\mathbf{h}), \text{if}\ \mathbf{h} \notin U(R, P_h) \\ g(\mathbf{h}), \text{if}\ \mathbf{h} \in U(R, P_h) \end{cases} . \tag{4.10}$$

In (4.10), $g(\mathbf{h})$ is an arbitrary function such that $\sum_{f=1}^{F} g(h_f) \leq P$ and $P^{st}(\mathbf{h})$ is the solution of the maximization problem which is expressed as

$$\begin{cases} \text{Maximize}\, I_F(\mathbf{h}, P_h) \\ \text{Subject to}\ P_{\mathbf{h}} \leq \overline{P}, \text{with probability 1} \end{cases} , \tag{4.11}$$

where $I_F(\mathbf{h}, P_{\mathbf{h}})$ is the instantenous mutual information given in (2.12) and can be rewritten for the power allocation scheme and Gaussian alphabets as

$$I_F(\mathbf{h}, P_{\mathbf{h}}) = \frac{1}{F}\sum_{f=1}^{F} log_2(1 + |h_f|^2 P_{\mathbf{h}}(f)), \tag{4.12}$$

with unit variance additive noise and independent Gaussian inputs $\mathcal{CN}(0, P_{\mathbf{h}}(f))$. The region $U(R, P_{\mathbf{h}})$ is the outage region defined as

$$U(R, P_{\mathbf{h}}) = \{\mathbf{h} : I_F(\mathbf{h}, P_{\mathbf{h}}) < R\}\,. \tag{4.13}$$

(4.11) is solved in [16] by using the Lagrangian method and Kuhn-Tucker conditions and the solution is expressed as follows.

Define the subset $Q \subset \Re_+^F$ such that $Q = \left\{|\mathbf{h}|^2 \in \Re_+^F : |h_1|^2 \geq \cdots \geq |h_f|^2\right\}$ and assume that $|\mathbf{h}|^2 \in Q$. $P^{st}(h_f)$ is given as

$$P^{st}(h_f) = \left[\lambda^{st}(\mu, \mathbf{h}) - \frac{1}{|h_f|^2}\right]^+, \tag{4.14}$$

where

$$\lambda^{st} = \frac{1}{\mu}\sum_{l=0}^{\mu-1}\frac{1}{|h_l|^2} + \frac{F}{\mu}\overline{P} \tag{4.15}$$

and μ is such that $\frac{1}{|h_f|^2} \leq \lambda^{st}$, for $f < \mu$ and $\frac{1}{|h_f|^2} > \lambda^{st}$, for $f \geq \mu$.
Since every fading gain vector $\left\{|h_1|^2, \cdots, |h_F|^2\right\}$ can be permuted such that it is in Q, the power allocation rule that is expressed above can be used to minimize the outage probability for short-term power allocation. When the fading coefficients are in the set defined in (4.13),i.e., the outage event occurred, the most sensible choice is to set $g(\mathbf{h}) = 0$, which means turning off the transmission.
The authors in [16] also presented a solution to the long term power allocation problem stated in (4.8). The outage probability minimization problem is solved by using the dual of (4.9)

$$\begin{cases} \text{Minimize}\, P_{\mathbf{h}} \\ \text{Subject to}\, I_F(\mathbf{h}, P_{\mathbf{h}}) \geq R \end{cases} . \tag{4.16}$$

If we again assume that the fading gain vector $\left\{|h_1|^2, \cdots, |h_F|^2\right\}$ is in Q, the f^{th} component of $P^{lt}(\mathbf{h})$ which is the solution to (4.16) is given by

$$P^{lt}_f(h) = \left[\lambda^{lt}(\mu, \mathbf{h}) - \frac{1}{|h_f|^2}\right]^+, \tag{4.17}$$

where

$$\lambda^{lt}(\mu, \mathbf{h}) = \left(\frac{e^{2FR}}{\prod_{l=0}^{\mu-1} |h_l|^2}\right)^{1/\mu} \tag{4.18}$$

and μ is such that $\frac{1}{|h_f|^2} \leq \lambda^{lt}$, for $f < \mu$ and $\frac{1}{|h_f|^2} > \lambda^{lt}$, for $f \geq \mu$.
If we define the set $R(s)$ as

$$R(s) = \left\{ |\mathbf{h}|^2 \in \Re^F_+ : \sum_{f=1}^{F} P^{lt}_f(h) < s \right\} \tag{4.19}$$

and

$$s^* = sup\left\{s : P(s) < \overline{P}\right\}, \tag{4.20}$$

where

$$P(s) = \int_{R_s} \sum P^{lt}(\mathbf{h}) p(\mathbf{h}) d\mathbf{h}, \tag{4.21}$$

the long term power allocation rule is defined in [16] as

$$P_{\mathbf{h}} = \begin{cases} P^{lt}(\mathbf{h}), \text{if}\, \mathbf{h} \in R(s^*) \\ 0, \text{if}\, \mathbf{h} \notin R(s^*) \end{cases} . \tag{4.22}$$

The expression in (4.22) suggests to turn off the transmission if the fading coefficient vector is below the threshold s^* and otherwise the power is allocated according to (4.17) that requires the minimum power to avoid an outage event as expressed in (4.16).

When the outage probabilities are plotted for short-term and long-term power constraints it occurs that the short-term power allocation does not yield significant power gains when compared to the constant power allocation. However, it is concluded in [16] that the power gain obtained by applying the long-term constraint is very large for Rayleigh block fading channel model(more than 10 dB for $R = 0.5$, $P_{out} = 10^{-3}$ and $F = 2$). It is also stated that even arbitrarily small error probabilities are possible at finite SNR depending on the rate and number of fading blocks.

4.4 On the bounds for block fading channels with CSIT

In this section, we will deal with the bounds presented in Section 2.4 especially with Gallager's error exponent, when the channel state information is available to the transmitter. There is not much work in the literature up to our knowledge where any power allocation strategy to minimize the error bounds for finite blocklength channels is applied and we reproduce the results of [30], [31]. In [30], the authors presented a suboptimum power allocation scheme for block fading channels in order to maximize the error exponent and thus minimizing the error probability. The auxiliary constant is also changed in the maximization operation. In [31], the random coding error exponent is optimized for the time-independent flat fading channel where the auxiliary variable ρ is not considered as a parameter to be optimized and it is taken as a constant in the power expressions.

When we consider the block fading AWGN channels and assume that the channel state is known to the transmitter, the error exponent expressed in (2.34) is given in [30] as

$$E_0(\rho, h, P_{\mathbf{h}}) = \frac{1}{F}\sum_{f=1}^{F}\left\{-log\left[\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} p(x)p(y|x, h_f, P_{h_f})^{1/(1+\rho)}dx\right)^{(1+\rho)} dy\right]\right\} \tag{4.23}$$

for a given input distribution $p(x)$, fading coefficients $\mathbf{h}$ of length F and allocated power vector $P_{\mathbf{h}}$ dependent on the channel state. The average error probability is bounded by

$$\overline{\epsilon} \leq E\left[2^{-N(E_0(\rho,\mathbf{h},P_{\mathbf{h}})-\rho R)}\right]. \tag{4.24}$$

Note that the auxiliary constant ρ interacts with both $\mathbf{h}$ and $P_{\mathbf{h}}$ and thus if the optimization of ρ is done before averaging over the fading coefficients $\mathbf{h}$, the bound becomes tighter. In [30], two input distributions is considered. We will concentrate on the Gaussian distribution in order to compare the results in [30] with our results for the DT bound which will be presented in the coming sections. With normal input distribution the error exponent can be written as

$$E_0(\rho, \mathbf{h}, P_{\mathbf{h}}) = \frac{1}{2F}\rho\sum_{f=1}^{F} log\left(1 + \frac{|h_f|^2 P_{\mathbf{h}}(f)}{1+\rho}\right). \tag{4.25}$$

When the long term power allocation is considered the power constraint can be expressed as

$$\frac{1}{F}\sum_{f=1}^{F} E\left[P_{\mathbf{h}}(f)\right] = \overline{P}, \tag{4.26}$$

where $\overline{P}$ is the maximum available power in average. It is suggested in [30] to use a power allocation scheme which is called the selective channel inversion in order to increase the error exponent expressed in (4.25). This power allocation strategy can be described by

$$P_{\mathbf{h}}(f) = \begin{cases} \frac{K}{|h_f|^2}, \text{if } |h_f|^2 \text{ is in the (F-}\mu\text{) greatest gains in } |h|^2 \\ 0, \text{ otherwise} \end{cases}, \tag{4.27}$$

where K is chosen such that the average power constraint given in (4.26) is satisfied. It is only allowed to allocate power to the $F - \mu$ fading blocks and the SNR of those blocks become equal after power allocation, since channel inversion is applied in the power allocation rule. By using (4.27), the error exponent can be rewritten as

$$E_0(\rho, \mathbf{h}, P_{\mathbf{h}}) = \frac{F - \mu}{2F} \rho log \left(1 + \frac{K}{1 + \rho}\right). \tag{4.28}$$

When the random coding exponent defined in (2.36) by using (4.28) is plotted the authors conclude in [30] that approximately ten-fold increase of the error exponent bound may be seen at low rates with respect to constant power allocation.

When we consider the time-independent flat fading channels as in Section 4.2, the error exponent maximization is presented in [31]. In this case the error exponent is expressed as

$$E_r(p_{X|h}(x|h), R) = \max_{0 \le \rho \le 1} \{E_0(p_{X|h}(x|h), \rho) - \rho R\} \tag{4.29}$$

and

$$E_0(p_{X|h}(x|h), \rho) = -log \int_h p(h) \int_y \left[\int_x p(x|h) p(y|x, h)^{\frac{1}{1+\rho}} dx \right]^{1+\rho} dydh. \tag{4.30}$$

If the input distribution is Gaussian, (4.30) is expressed in [31] as

$$E_0(p_{X|h}(x|h), \rho) = -logE_h \left\{ \left(1 + \frac{|h|^2 P(h)}{1 + \rho}\right)^{-\rho} \right\} \tag{4.31}$$

and the optimum power allocation to maximize (4.31) is given as

$$P(h) = \left[\left(\frac{\alpha |h|^{2\rho}}{(1 + \rho)^{\rho}} \right)^{-\frac{1}{1+\rho}} - \frac{1 + \rho}{|h|^2} \right]^+, \tag{4.32}$$

where α is chosen such that the power constraint given in (4.1) is satisfied. If we rearrange (4.31) by using (4.32)

$$E_0(\rho) = -log \int_{a_0}^{\infty} p(a) \left(\frac{a^2}{\alpha(1 + \rho)} \right)^{-\rho/(1+\rho)} da, \tag{4.33}$$

where $a = |h|$, $a_0 = \sqrt{\alpha(1 + \rho)}$ and α is such that

$$\int_{a_0}^{\infty} p(a) \left[\left(\frac{\alpha a^{2\rho}}{(1 + \rho)^{\rho}} \right)^{-1/(1+\rho)} - \frac{1 + \rho}{a^2} \right] da = \overline{P}. \tag{4.34}$$

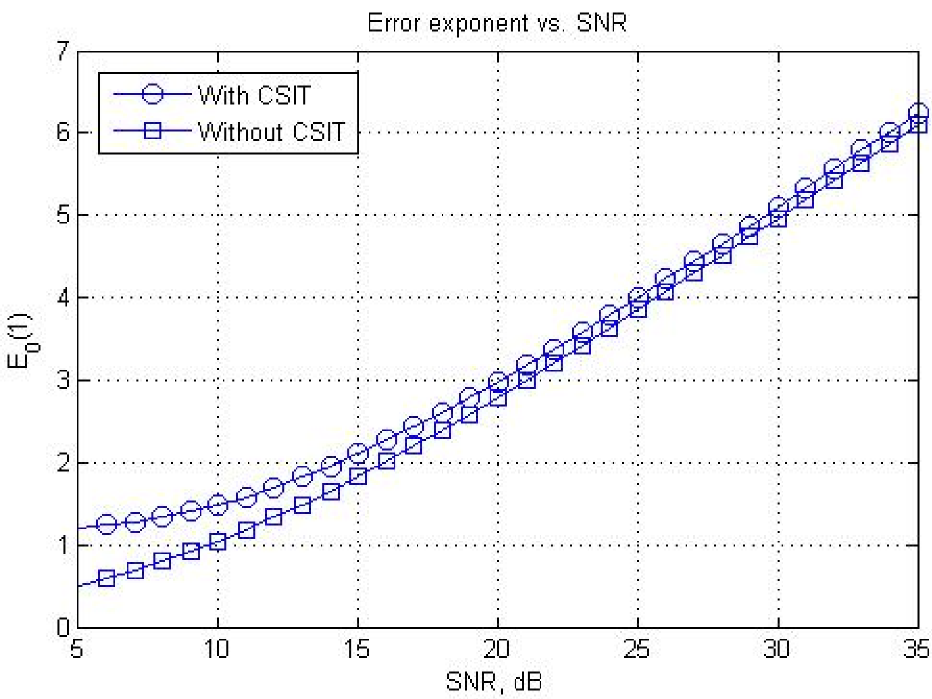

Figure 4.2: Error exponent vs. SNR with and without CSIT

If the error exponent is plotted by using (4.33) and (4.34) and assuming $\rho = 1$, the difference between the power allocation scheme and constant power allocation can be seen in Figure 4.2. We can conclude that even in high SNR, there is still an advantage of using the described power allocation scheme to increase the error exponent and thus decrease the average error probability. In Section 4.5, we will construct a power allocation scheme to minimize the DT bound in block fading channels.

4.5 DT bound optimization with CSIT

In Section 3.4, we have presented a normal approximation to calculate the DT bound for fading channels with the assumption that the channel state is not available to the transmitter and the transmitted codewords are sent with constant power. If the transmit side has the CSI, the power can be adjusted so that the DT bound given in Theorem 3.2.1 is minimized in the fading case. We will apply a short-term power allocation scheme in this section and present a simple algorithm to calculate the power to be allocated to each fading block.

We will use the Gaussian approximation developed in Chapter 3 and since we want to minimize the DT bound by using short-term power allocation, we can represent the function to be minimized by using (3.55) as

$$\min\left\{\left(erfc(-\frac{\eta_1}{\sqrt{2}})/2\right) + exp\left(\frac{\sigma_\alpha^2}{2} - \mu_\alpha\right)\left(erfc(\frac{\eta_2}{\sqrt{2}}/2)\right)\right\}, \tag{4.35}$$

where

$$\mu_\alpha = N\frac{1}{F}\sum_{f=1}^{F} log_2\left(1 + |h_f|^2 P_{\mathbf{h}}(f)\right) - log\frac{M-1}{2} \tag{4.36}$$

and

$$\sigma_\alpha^2 = (log_2 e)^2\, N\left[4 - \frac{2}{F}\sum_{f=1}^{F} 1 + \frac{1}{1 + |h_f|^2 P_{\mathbf{h}}(f)}\right] \tag{4.37}$$

and the following were defined in Chapter 3.

$$\eta_1 = -\frac{\mu_\alpha}{\sigma_\alpha} \tag{4.38}$$

$$\eta_2 = -\frac{\mu_\alpha - \sigma_\alpha^2}{\sigma_\alpha} \tag{4.39}$$

If we extract the complementary error functions in (4.35) in their integral form, we obtain that

$$\min\left\{\frac{1}{\sqrt{\pi}}\int_{\mu_\alpha/(\sigma_\alpha\sqrt{2})}^{\infty} exp(-t^2)dt + \frac{1}{\sqrt{\pi}}exp(\frac{\sigma_\alpha^2}{2} - \mu_\alpha)\int_{\frac{\sigma_\alpha - \frac{\mu_\alpha}{\sigma_\alpha}}{\sqrt{2}}}^{\infty} exp(-t^2)dt\right\} \tag{4.40}$$

subject to

$$\sum_{f=1}^{F} P_{\mathbf{h}}(f) = F \cdot P, \tag{4.41}$$

where $P_{\mathbf{h}}(f)$ is the allocated power to block $f \in \{1, \cdots, F\}$.

The Lagrangian multiplier is constructed as follows

$$J = \frac{1}{\sqrt{\pi}} \int_{\mu_\alpha/(\sigma_\alpha \sqrt{2})}^{\infty} exp(-t^2)dt + \frac{1}{\sqrt{\pi}} exp\left(\frac{\sigma_\alpha^2}{2} - \mu_\alpha\right) \int_{\frac{\sigma_\alpha - \frac{\mu_\alpha}{\sigma_\alpha}}{\sqrt{2}}}^{\infty} exp(-t^2)dt$$
$$- \lambda \left[\sum_f (P_{\mathbf{h}}(f) - FP) \right]. \tag{4.42}$$

In order to minimize the DT bound, we should satisfy that

$$\frac{\partial J}{\partial P_{\mathbf{h}}(f)} = 0, \tag{4.43}$$

for any f. Let's extract the lower bounds of the integral operations in (4.42) and define

$$g_1 = \frac{\sqrt{N}\left(\frac{1}{F}\sum_f log_2\left(1 + |h_f|^2 P_{\mathbf{h}}(f)\right) - \frac{1}{N} log_2\left(\frac{M-1}{2}\right)\right)}{\sqrt{2} log_2 e \sqrt{4 - \frac{2}{F}\sum_f (1 + \frac{1}{1 + |h_f|^2 P_{\mathbf{h}}(f)})}} \tag{4.44}$$

and

$$g_2 = \frac{\sqrt{N}\left(4\,(log_2 e)^2 - \frac{2(log_2 e)^2}{F} \sum_f \frac{2 + |h_f|^2 P_{\mathbf{h}}(f)}{1 + |h_f|^2 P_{\mathbf{h}}(f)}\right)}{log_2 e \sqrt{4 - \frac{2}{F}\sum(1 + \frac{1}{1+|h_f|^2 P_{\mathbf{h}}(f)})}}$$
$$- \frac{\sqrt{N}\left(\sum \frac{log_2(1+|h_f|^2 P_{\mathbf{h}}(f))}{F} - log_2(\frac{M-1}{2})^{\frac{1}{N}}\right)}{log_2 e \sqrt{4 - \frac{2}{F}\sum(1 + \frac{1}{1+|h_f|^2 P_{\mathbf{h}}(f)})}}. \tag{4.45}$$

Then we can rewrite (4.42) as

$$J = \frac{1}{\sqrt{\pi}} \int_{g_1}^{\infty} exp(-t^2)dt + \frac{1}{\sqrt{\pi}} exp\left(\frac{\sigma_\alpha^2}{2} - \mu_\alpha\right) \int_{g_2}^{\infty} exp(-t^2)dt$$
$$- \lambda \left[\sum_f (P_{\mathbf{h}}(f) - FP) \right]. \tag{4.46}$$

By using the Leibniz integral rule we can express (4.43) as

$$\frac{\partial J}{\partial P_{\mathbf{h}}(f)} = -exp(-(g_1)^2)\frac{\partial g_1}{\partial P_{\mathbf{h}}(f)} + \frac{\partial\left(exp(\frac{\sigma_\alpha^2}{2} - \mu_\alpha)\right)}{\partial P_{\mathbf{h}}(f)}\left(erfc(\frac{\eta_2}{\sqrt{2}}/2)\right)$$
$$-exp\left(\frac{\sigma_\alpha^2}{2} - \mu_\alpha\right) exp\left(-(g_1)^2\right) \frac{\partial g_2}{\partial P_{\mathbf{h}}(f)} - \lambda. \tag{4.47}$$

The partial derivatives arising in (4.47) are calculated easily. Unfortunately, we cannot obtain an analytical solution to the problem given in (4.43) and thus $P_{\mathbf{h}}(f)$ cannot be expressed in an exact form for a given fading coefficient vector $\mathbf{h}$. However, we can solve the system of $(F + 1)$ equations the last one of which is given in (4.41) numerically in MATLAB.

In order to be able to calculate the power to be allocated in a real system where excessive numerical calculations are not possible, we will use the algorithm described in [32] which is defined as the maximum marginal allocation approach. The algorithm provides an easy way to calculate the allocated power $P_{\mathbf{h}}(f)$ which minimizes the DT bound. The algorithm is defined as follows. We first discretize the range of the power to be allocated. Since (4.41) should hold, we let $P_{\mathbf{h}}(f)$ in the set $\{0, \Delta, 2\Delta, \cdots, D\Delta\}$ where $D\Delta = FP$. We use FP units of power by allocating Δ units of power in each step of the algorithm. We define $\epsilon(P_{\mathbf{h}}(1), P_{\mathbf{h}}(2), \cdots, P_{\mathbf{h}}(F))$ as the error probability calculated by assigning $P_{\mathbf{h}}(f)$ units of power to block f and using (3.55). In the first step, we set $P_{\mathbf{h}}(j) = \Delta$ if $\epsilon(0, 0, \cdots, P_{\mathbf{h}}(j) = \Delta, \cdots, 0) < \epsilon(0, 0, \cdots, P_{\mathbf{h}}(k) = \Delta, \cdots, 0)$ for $k \neq j$. In the second step, recalling the fact that we allocated Δ units of power to the j^{th} block, the power allocation scheme occurs such that we take the minimum element of the set

$$(\epsilon(P_{\mathbf{h}}(1) = \Delta, 0, \cdots, P_{\mathbf{h}}(j) = \Delta, \cdots, 0), \cdots, \epsilon(0, \cdots, P_{\mathbf{h}}(j) = 2\Delta, \cdots, 0),$$
$$\cdots, \epsilon(0, \cdots, P_{\mathbf{h}}(j) = \Delta, \cdots, P_{\mathbf{h}}(F) = \Delta)). \quad (4.48)$$

We continue in the same way until we reach the D^{th} step. Thus, the power allocation algorithm takes place in D steps. In Table 4.1, we show the allocated power results for some randomly generated fading coefficients where $F = 2$ and $SNR = 0\,\mathrm{dB}$.

Table 4.1: Comparison of the algorithmic result and the numerical solution to (4.43).

$\|h\|^2$		Algorithmic		Numerical	
$\|h_1\|^2$	$\|h_2\|^2$	$P_{\mathbf{h}}(1)$	$P_{\mathbf{h}}(2)$	$P_{\mathbf{h}}(1)$	$P_{\mathbf{h}}(2)$
1.2374	1.7332	0.82	1.18	0.838	1.162
0.944	1.4937	0.7	1.3	0.723	1.277
0.8494	2.1437	0.34	1.66	0.377	1.623
0.3431	1.5031	0	2	0	2

We have used (4.47) in order to find the theoretical result in Table 4.1 and solved the optimum power allocation problem numerically in MATLAB. We have taken $D = 100$ in our simulations when using the maximum marginal allocation algorithm. The results show that the presented algorithm give us similar results with the analytic solution and it can be used in order to calculate the optimum power to be allocated to each block.

We have evaluated the results by generating 100,000 random fading coefficient vectors and using the proposed algorithm to find the allocated power to each block. In Figure 4.3, the equal power allocation and optimum power allocation results for DT bound is plotted for $F = 4$ and $N = 100$. The input alphabet is taken as complex Gaussian for the channel coding rate $R = 0.6$. The power allocation scheme provide an advantage of approximately 2 dB power gain when the average error probability is taken as 10^{-3}. In Figure 4.4 and in Figure 4.5, the results for $N = 400$ and $N = 1000$ are presented respectively. For the blocklength $N = 400$, the power saving is approximately 1.5 dB and for $N = 1000$ we have about 1 dB SNR gain when power allocation is used. In Figure 4.6, the results for $F = 10$ and $N = 100$ are presented. In this case, the power gain resulted in approximately 1.5 dB for $\epsilon = 10^{-3}$. In Figure 4.7, the results for $F = 10$ and $N = 400$ are plotted. The power saving by power allocation is decreased to 1 dB for this case.

We conclude that when the number of fading blocks is increased the advantage of the power allocation is decreased. We also conclude that as the blocklength is increased for a given number of fading blocks, the advantage gained by power allocation is decreased. As discussed in Section 2.3, in outage probability calculations blocklength is considered to be infinite. The results for DT bound with power allocation coincides with the results in [16], as presented in Section 4.3, where the short term power allocation resulted in no advantage for outage probability minimization.

The results presented in this section and in Section 4.3 do not comply with the power allocation results for ergodic capacity given in Section 4.2 where it is concluded that power allocation does not increase the maximum coding rate substantially. Thus, one should pay attention to the results developed for finite blocklength channels when designing a system in real life, since they can much differ from the infinite blocklength results.

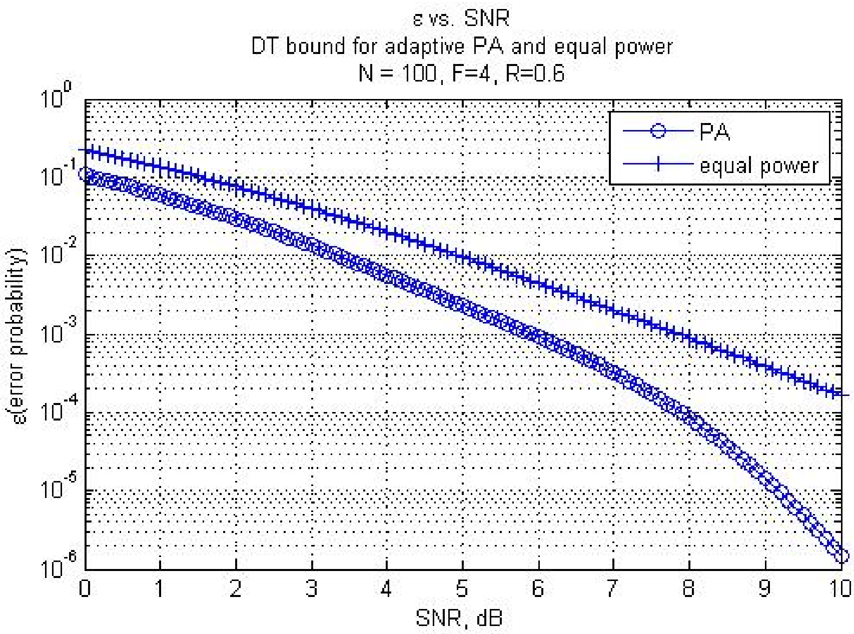

Figure 4.3: Comparison of DT bound for equal and adaptive power allocation, $F = 4$, $N = 100$

4.6 Conclusion

We have discussed the water-filling scheme to find the channel capacity when the blocklength is taken as infinite. In addition, we have mentioned the proposed power allocation strategies to minimize the outage probability or error probability for finite blocklength channels. Then, we presented the short-term power allocation scheme in order to minimize the error probability calculated by using the DT bound. Since, the expressions turned out to be quite complex, we also introduced an algorithm called maximal marginal allocation in order to calculate the power to be allocated for a given fading coefficient vector. It turned out that this algorithm gives very similar results with the ones calculated by using the theoretical derivation. There is 1-2 dB gain in performance with short term power allocation in the scenarios investigated.

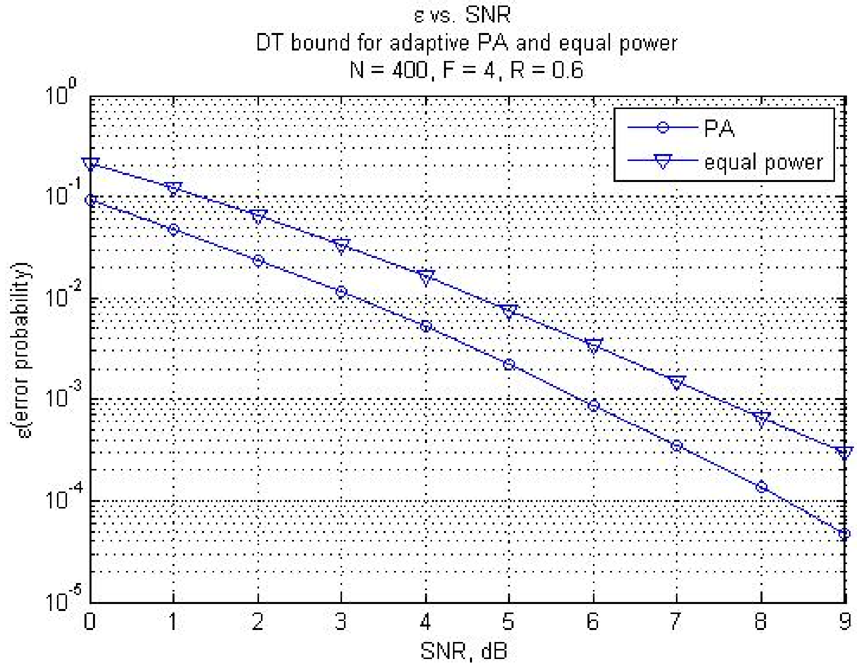

Figure 4.4: Comparison of DT bound for equal and adaptive power allocation, $F = 4, N = 400$

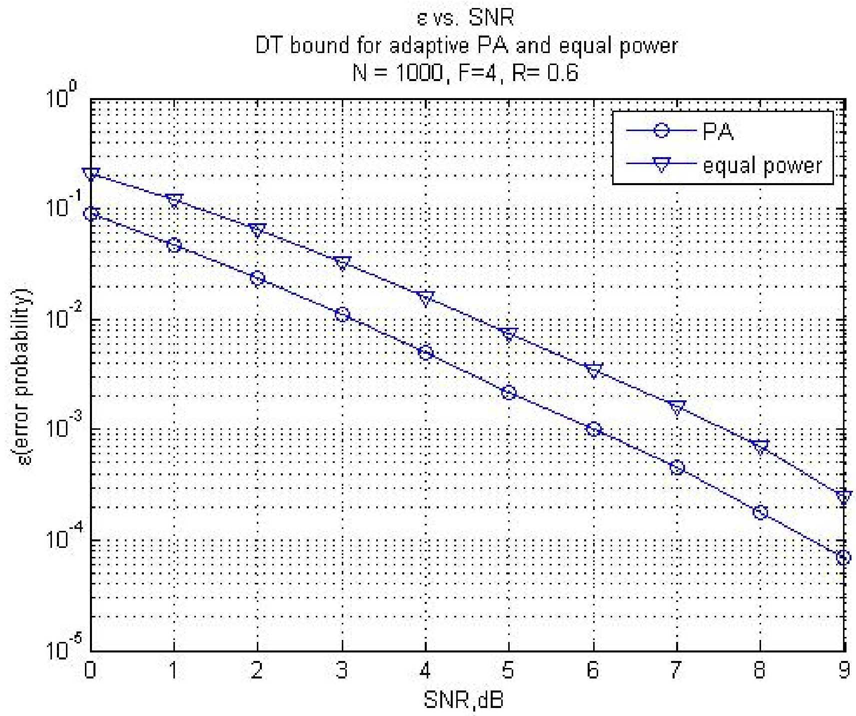

Figure 4.5: Comparison of DT bound for equal and adaptive power allocation, $F = 4, N = 1000$

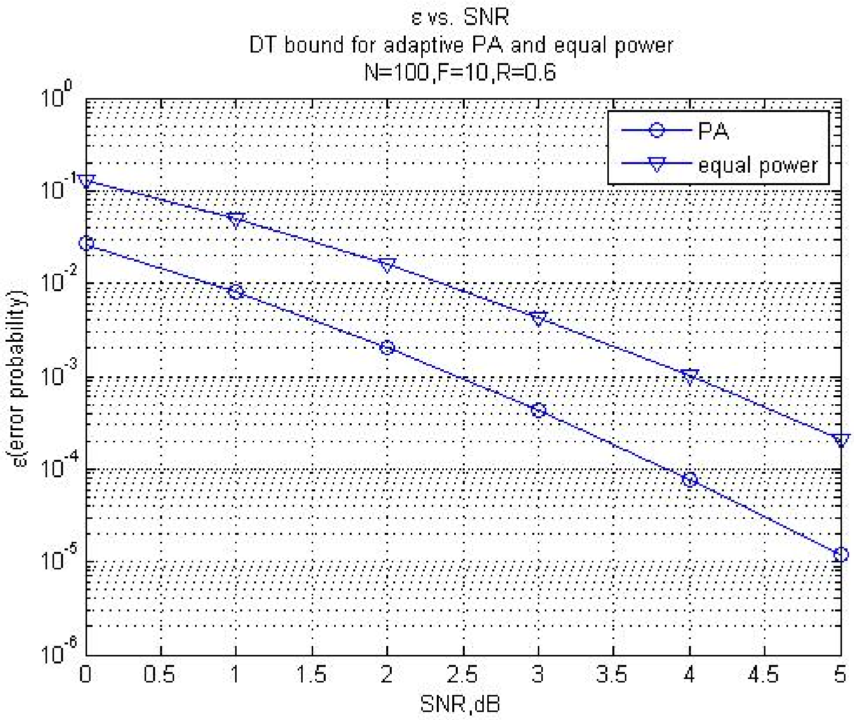

Figure 4.6: Comparison of DT bound for equal and adaptive power allocation, $F = 10$,$N = 100$

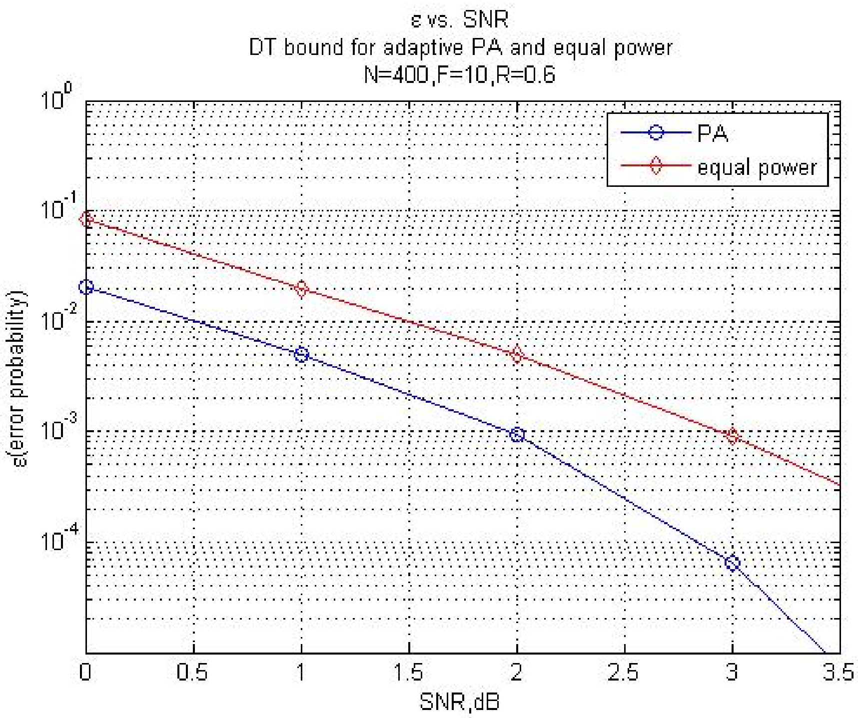

Figure 4.7: Comparison of DT bound for equal and adaptive power allocation, $F = 10, N = 400$

CHAPTER 5

Conclusions and Future Work

In this work, we have dealt with information theoretic error bounds for finite blocklength systems with a given coding rate and proposed a method to approximately calculate the DT bound for different input alphabets and different channel scenarios. We made use of the normal approximation to the information density while calculating the DT bound.

In Chapter 2, we addressed our system model and investigated the well-known information theoretic results when the coding blocklength is both infinite and finite. Ergodic channel capacity when the blocklength is infinite and the delay-limited capacity for finite blocklength channels are defined and how the outage probability term used for block fading channels emerge in our system model has been shown. In addition, we discussed the major existing achievability bounds for finite blocklength channels.

In Chapter 3, we proposed a method to calculate the DT bound. After introducing the DT bound, it was compared with the previously existing bounds and we concluded that this bound is tighter than the ones given in Chapter 2. In addition, no auxiliary variables are needed to be optimized while calculating the DT bound. However, no closed-form expressions can be obtained when one attempts to calculate it. Thus, a normal approximation is proposed to calculate the DT bound. We have developed our approximation for Gaussian input alphabets and both for AWGN and block fading channels. When the results of the proposed approximation is compared with the exact results, it turned out that our approximation characterizes the bound for blocklengths as small as 100. We have investigated the effect of the blocklength and the number of fading blocks for block fading channels. The average error probability became smaller as the number of fading blocks is increased since the diversity order of the system is increased. We concluded that blocklength is an important parameter for large values of the number of fading blocks. In addition, we used the same method in our Monte Carlo simulations for constrained alphabets. For constrained alphabets the achievable rate for a given error probability and blocklength is closer to the corresponding constrained capacity than those for Gaussian inputs to the channel capacity. Thus, in a real communication system, we can approach to the information theoretic limits at even smaller blocklengths.

The proposed method is used to find the short term power allocation strategy for block fading channels to minimize the error expression in the DT bound in Chapter 4 when CSIT is available. First, the existing power allocation strategies in the literature are investigated and the comparisons with the equal power case are made through. Then, the expressions to calculate the optimal power to be allocated to each block for the DT bound are presented. Since the expressions turned out to be very complex to calculate, we also made use of an algorithm called the maximum marginal allocation approach in order to calculate the allocated power in a real-time system for a given fading coefficient vector. This approach gave very similar results with the numerical solution as long as the number of steps to calculate the power is taken large enough. After presenting the simulations for different number of fading blocks and different blocklengths, we have concluded that the power allocation results for the ergodic channel capacity may much differ from the ones for finite blocklength systems. In addition, the power savings for the DT bound decreases as the blocklength is increased which coincides with the results for outage probability where the blocklength is assumed to be infinite and the short term power allocation yielded nearly zero power gain.

In summary, one can use the proposed method to calculate the DT bound for AWGN channels without the need of an optimization of auxiliary variables as in the major existing bounds. For block fading channels the effect of blocklength is taken into account when an error bound is selected as the performance indicator which acts as an advantage when compared with the outage probability.

As a future work to our thesis, one can investigate the effect of long term power adaptation for the optimization of DT bound. This can again lead to complex expressions to calculate the optimal power and an algorithm should again be proposed for long term power allocation. Since the outage probability optimization with long term power constraint provides one with high values of power saving when compared with the short term optimization, it would be worthwhile to compare the results for the DT bound.

In addition, the proposed method can be used to calculate the DT bound for Multi-Input Multi-Output (MIMO) systems. Thus, one can obtain the effect of the number of antennas for a MIMO system with finite blocklength channels.

Future studies may also include the case where the channel is known at the receiver imperfectly. In our thesis, we have considered the case where the channel estimation at the transmitting side is perfect. When the channel state is estimated both from the training symbols and the information symbols, the effect of blocklength may become more important.

REFERENCES

[1] C. E. Shannon. 1948, pp.379-423. *A Mathematical Theory of Communication*. Bell Sys. Tech. Journal.

[2] A. Feinstein. 1954, pp.2-22. *A new basic theorem of information theory*. IRE Trans. Inform. Theory.

[3] C. E. Shannon. 1957, pp.6-25. *Certain results in coding theory for noisy channels*. Inform. Control.

[4] R. G. Gallager. 1965, pp.3-18. *A simple derivation of the coding theorem and some applications*. IEEE Trans. Inform. Theory.

[5] Y. Polyanskiy, H.V. Poor, S. Verdu. 2008, July, pp.1763-1767. *New channel coding achievability bounds*. IEEE Int. Symp. Inform. Theory.

[6] Y. Polyanskiy, H.V. Poor, S. Verdu. 2010, May, pp.2307-2359. *Channel Coding Rate in the Finite Blocklength Regime*. IEEE Trans. Inform. Theory.

[7] L. H. Ozarow, S. Shamai, A.D. Wyner. 1994, May, pp.359-378. *Information theoretic considerations in cellular mobile radio*. IEEE Trans. Vehicular Technology.

[8] R. Knopp, P. A. Humblet. 2000, January, pp.189-205. *On coding for block fading channels*. IEEE Trans. Inform. Theory.

[9] A. Goldsmith. 2005. *Wireless Communications*. Cambridge University Press.

[10] T. Cover, J. Thomas. 2006. *Elements of Information Theory*. Wiley.

[11] G. Caire, G. Taricco, E. Biglieri. 1999, July, pp.1468-1489. *Optimum Power Control over Fading Channels*. IEEE Trans. Inform. Theory.

[12] S. Verdu,T. S. Han. 1994, July, pp.1147-1157. *A general formula for channel capacity*. IEEE Trans. Inform. Theory.

[13] E. Biglieri,J. Proakis, S. Shamai. 1998, October, pp.2619-2692. *Fading channels: information-theoretic and communications aspects*. IEEE Trans. Inform. Theory.

[14] G. Kaplan, S. Shamai. 1995, pp.192-205. *Error probabilities for the block-fading Gaussian channel*. Arch. Elektronik Ubertragungstech.

[15] A. Goldsmith, P. P. Varaiya. 1997, November, pp.1986-1992. *Capacity of fading channels with channel side information*. IEEE Trans. Inform. Theory.

[16] G. Caire, G. Taricco, E. Biglieri. 1999, July, pp.1468-1489. *Optimum power control over fading channels*. IEEE Trans. Inform. Theory.

[17] S. V. Hanly, D. Tse. 1998, November, pp.2816-2831. *Multiaccess fading channels-Part II: Delay-Limited Capacities*. IEEE Trans. Inform. Theory.

[18] E. Malkamaki, H. Leib. 1999, March, pp.771-781. *Coded diversity on Block Fading Channels*. IEEE Trans. Inform. Theory.

[19] D. Buckingham, M. C. Valenti. 2008, March, pp.390-395. *The information-outage probability of finite-length codes over AWGN channels*. 42nd Annual Conference on Inf. Sciences and Sys.

[20] J. Wolfowitz. 1978. *Coding theorems of information theory*. Springer-Verlag.

[21] D. Baron, M. A. Khojastepour, R.G. Baraniuk. 2004, November, pp.1096-1100. *How quickly can we approach channel capacity?*. Asimolar Conference on Signals, Systems and Computers

[22] R. G. Gallager. 1968. *Information Theory and Reliable Communication*. Wiley.

[23] Y. Polyanskiy, H.V. Poor, S. Verdu. 2009, June, pp.2204-2208. *Dispersion of Gaussian channels*. IEEE Int. Symp. Inform. Theory.

[24] V. Strassen. 1962, pp.689-723. *Asymptotic estimations in Shannon's information theory*. Trans. Third Prague Conf. Information Theory.

[25] H. L. Van Trees. 2001. *Detection, Estimation and Modulation Theory, Part I*. Wiley.

[26] R. G. Gallager. 1996. *Discrete Stocahistic Processes*. Kluwer Boston.

[27] S. M. Ross. 2003. *Introduction to Probability Models*. Elsevier.

[28] G. Ungerboeck. 1982, January, pp.55-67. *Channel coding with multilevel/phase signals*. IEEE Trans. Inform. Theory.

[29] K. D. Nguyen, A. G. Fabregas, L. K. Rasmussen. 2009, May, pp.2514-2523. *Power Allocation for Block-Fading Channels with Arbitrary Input Constellations*. IEEE Trans. Wireless Communications.

[30] E. Biglieri, G. Caire, G. Taricco. 1998, June, pp.96-97. *Coding for the block-fading channel: optimum and suboptimum power-allocation schemes*. IEEE ITW.

[31] W. K. M. Ahmed, P. J. McLane. 1999, February, pp.49-51. *On the error exponent for memoryless flat fading channels with channel-state-information feedback*. IEEE Comm. Letters.

[32] S. Kay. 2009, July, pp.1153-1166. *Waveform Design for Multistatic Radar Detection*. IEEE Trans. Aerospace and Electronic Systems.

APPENDIX A

Proof of Theorem 2.4.1

Define the set $G = \{(\mathbf{x}, \mathbf{y}) : i(\mathbf{x}; \mathbf{y}) > a\}$. Set

$$\epsilon = Me^{-a} + Pr\{i(\mathbf{x}; \mathbf{y}) \leq a\} = Me^{-a} + P(G^C) \tag{A.1}$$

and assume $\epsilon < 1$. Thus $P(G^C) \leq \epsilon < 1$ and therefore

$$Pr\{i(\mathbf{x}; \mathbf{y}) > a\} = P(G) > 1 - \epsilon > 0. \tag{A.2}$$

Define $G_{\mathbf{x}} = \{\mathbf{y} : (\mathbf{x}, \mathbf{y}) \in G\}$ and

$$R = \{\mathbf{x} : P(G_{\mathbf{x}}|\mathbf{x}) > 1 - \epsilon\}. \tag{A.3}$$

We choose our first codeword $\mathbf{x_1}$ of the codebook in the set R and for the decoder we decide that $\mathbf{x_1}$ is sent if $\mathbf{y} \in F_1 = G_{\mathbf{x}_1}$. If possible, for the next codeword we choose $\mathbf{x_2} \in R$ such that $P(G_{\mathbf{x}_2} - F_1|\mathbf{x}_2) > 1 - \epsilon$ and let $F_2 = G_{\mathbf{x}_2} - F_1$. We continue in the same way until all the codewords are constructed or all the points in R are exhausted. When we are given the pairs $\left\{\mathbf{x_j}, F_j\right\}$ for $j = 1, \cdots (i-1)$, we should find $\mathbf{x_i} \in R$ such that

$$P(G_{\mathbf{x_i}} - \bigcup_{j<i} F_j|\mathbf{x_i}) > 1 - \epsilon \tag{A.4}$$

and let $F_i = G_{\mathbf{x_i}} - \bigcup_{j<i} F_j$. If the points in R satifying the above conditions are finished before we collected M points, then denote the final point's index as n. For every $i \leq n$, the following inequality is satisfied:

$$P(F_i^c|\mathbf{x_i}) \leq \epsilon. \tag{A.5}$$

Thus we found a decoding rule at the receiver that the bound in (2.19) is satisfied if and only if we can prove that n is not strictly less than M. For the proof let us assume that $n < M$ and let us define $F = \bigcup_{i=1}^{n} F_i$ and express $P(G)$ as

$$P(G) = P\left(G \cap (A \times F)\right) + P\left(G \cap (A \times F^c)\right). \tag{A.6}$$

We can bound the first term in (A.6) as

$$P(G \cap (A \times F)) \leq P(A \times F) = P(F) = \sum_{i=1}^{n} P(F_i). \tag{A.7}$$

The last equality follows from the fact that F_i's are disjoint.

Define f($\mathbf{x}$,$\mathbf{y}$) as

$$f(\mathbf{x}, \mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})}, \tag{A.8}$$

which means that $i(\mathbf{x}; \mathbf{y}) = log(f(\mathbf{x}, \mathbf{y}))$. The following inequalities

$$P(F_i) = \sum_{\mathbf{y} \in F_i} p(\mathbf{y}) \leq \sum_{\mathbf{y} \in G_{\mathbf{x_i}}} p(\mathbf{y}) \leq \sum_{\mathbf{y} \in G_{\mathbf{x_i}}} \frac{f(\mathbf{x_i}, \mathbf{y})}{e^a} p(\mathbf{y}) \leq e^{-a} \sum_{\mathbf{y}} p(\mathbf{y}|\mathbf{x_i}) = e^{-a}, \tag{A.9}$$

lead to

$$P(G \cap (A \times F)) \leq ne^{-a}. \tag{A.10}$$

We now bound the second term in (A.6) as

$$\begin{aligned} P(G \cap (A \times F^c)) &= \sum_{\mathbf{x}} P(G \cap (A \times F^c))|\mathbf{x})p(\mathbf{x}) \\ &= \sum_{\mathbf{x}} P(G_{\mathbf{x}} \cap F^c)|\mathbf{x})p(\mathbf{x}) = \sum_{\mathbf{x}} P(G_{\mathbf{x}} - \bigcup_{i=1}^{n} F_i)|\mathbf{x})p(\mathbf{x}) \end{aligned} \tag{A.11}$$

If we define the set

$$B = \left\{ \mathbf{x} : P(G_{\mathbf{x}} - \bigcup_{i=1}^{n} F_i|\mathbf{x}) > 1 - \epsilon \right\}, \tag{A.12}$$

it must be true that $P(B) = 0$, otherwise there must be one extra point x_{n+1} satisfying

$$P(G_{\mathbf{x_{n+1}}} - \bigcup_{i=1}^{n+1} F_i|\mathbf{x_n}) > 1 - \epsilon. \tag{A.13}$$

Thus we bound the second term as

$$P(G \cap (A \times F^c)) \leq 1 - \epsilon \tag{A.14}$$

and we get that

$$P(G) \leq ne^{-a} + 1 - \epsilon. \tag{A.15}$$

We defined ϵ in (A.1) and we can write $P(G)$ as

$$P(G) = 1 - \epsilon + Me^{-a}. \tag{A.16}$$

The last two equations gives a contradiction and hence $n \leq M$. This result completes the proof.

APPENDIX B

Proof of Theorem 2.4.2

For a given M and β, consider the pairs $(\mathbf{x}, \mathbf{y})$ of input and output words and define the set T that consists of those pairs for which $i(\mathbf{x}, \mathbf{y}) > log\beta$. Then the probability that the $(\mathbf{x}, \mathbf{y})$ pair will belong to the set T is $1 - Pr(i(\mathbf{x}, \mathbf{y}) \leq log\beta)$. We consider the ensemble of codes obtained as follows. The integers $1, 2, \cdots, M = 2^{NR}$ are associated independently with B different possible input words $\mathbf{x_1}, \mathbf{x_2}, \cdots, \mathbf{x_B}$ with probabilities $p(\mathbf{x_1}), \cdots, p(\mathbf{x_B})$. Thus an ensemble of codes are produced each of which uses M (or less) input words since each codeword is constructed independently. If there are B different input words, there will be B^M different codes in this ensemble. The constructed codes have different probabilities. As to say, the code in which all integers $1, \cdots, M$ are mapped to a single input word $\mathbf{x_1}$ has probability $p(\mathbf{x_1})^M$. A code in which d_k of the integers are mapped into $\mathbf{x_k}$ has probability $\prod_k p(\mathbf{x_k})^{d_k}$. In order to obtain the bound given in (2.23) we will deal with the average probability of error for this ensemble of codes. We assume that the use of each codeword is equiprobable with probability $1/M$. The decoding rule is defined as follows. A received vector $\mathbf{y}$ is decoded to the integer m with the largest probability conditional on the received $\mathbf{y}$, which actually corresponds to MAP (maximal a posteriori) decoding. After defining the decoding rule, we are ready to calculate the average probability of error for this ensemble of codes. Consider a particular message which is the integer '1'. It will be mapped to $\mathbf{x}$ with probability $p(\mathbf{x})$. When it is mapped to $\mathbf{x}$ and a word $\mathbf{y}$ is received, an error will occur if there are one or more integers mapped into the set $S_{\mathbf{y}}(\mathbf{x})$ of input words which have an a posteriori probability of higher than or equal to that of $\mathbf{x}$ in the code in question. Let us define the probability of all these input words as

$$Q_{\mathbf{y}}(\mathbf{x}) = \sum_{\mathbf{x}' \in S_{\mathbf{y}}(\mathbf{x})} p(\mathbf{x}'). \tag{B.1}$$

Thus (B.1) is the probability associated with all words more probable or as probable as $\mathbf{x}$ conditioned on the received word $\mathbf{y}$. The fraction of codes in which the second message '2' is not in $S_{\mathbf{y}}(\mathbf{x})$ is thus equal to $1 - Q_{\mathbf{y}}(\mathbf{x})$. The fraction of codes in which $S_{\mathbf{y}}(\mathbf{x})$ is free of all other integers is $\left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1}$. The same argument can be applied to any other integer $2, \cdots, M$. Thus the probability of error in this ensemble where the message is mapped into input codeword $\mathbf{x}$ and received as $\mathbf{y}$ can be expressed as

$$p(\mathbf{x}, \mathbf{y}) \left[1 - \left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1}\right]. \tag{B.2}$$

The average probability of error can then be written as

$$\epsilon = \sum_{\mathbf{x},\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \left[1 - \left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1}\right]. \tag{B.3}$$

In order to bound the expression in (B.3) we break the sum into two parts. The first is a sum over $(\mathbf{x}, \mathbf{y})$ where the pair belongs to the set T and the second over the complementary set T^C.

$$\epsilon = \sum_{T^C} p(\mathbf{x}, \mathbf{y}) \left[1 - \left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1}\right] + \sum_{T} p(\mathbf{x}, \mathbf{y}) \left[1 - \left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1}\right] \tag{B.4}$$

We replace $\left[1 - \left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1}\right]$ by 1 in the first summation in (B.4) increasing the summation. The first term becomes $\sum_{T^C} p(\mathbf{x}, \mathbf{y})$ which is equal to $Pr(i(\mathbf{x}, \mathbf{y}) \leq log\beta)$. For the second sum we use a basic inequality which is $\left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1} \geq 1 - (M-1)Q_{\mathbf{y}}(\mathbf{x})$. Then the second sum is increased by replacing $\left[1 - \left(1 - Q_{\mathbf{y}}(\mathbf{x})\right)^{M-1}\right]$ by $(M-1)Q_{\mathbf{y}}(\mathbf{x})$. Thus the bound can be expressed as

$$\epsilon \leq Pr(i(\mathbf{x}, \mathbf{y}) \leq log\beta) + (M-1) \sum_{T} p(\mathbf{x}, \mathbf{y}) Q_{\mathbf{y}}(\mathbf{x}). \tag{B.5}$$

We must show that for $(\mathbf{x}, \mathbf{y})$ in T, $Q_{\mathbf{y}}(\mathbf{x}) \leq \frac{1}{\beta}$. With $(\mathbf{x}, \mathbf{y})$ in T

$$log \frac{p(\mathbf{y}|\mathbf{x})}{p(\mathbf{y})} > log\beta, \tag{B.6}$$

and

$$p(\mathbf{y}|\mathbf{x}) > p(\mathbf{y})\beta. \tag{B.7}$$

If $\mathbf{x}'$ is in $S_{\mathbf{y}}(\mathbf{x})$,

$$p(\mathbf{y}|\mathbf{x}') \geq p(\mathbf{y}|\mathbf{x}) > p(\mathbf{y})\beta \tag{B.8}$$

which results in

$$p(\mathbf{x}', \mathbf{y}) > p(\mathbf{x}')p(\mathbf{y})\beta \tag{B.9}$$

and

$$p(\mathbf{x}'|\mathbf{y}) > p(\mathbf{x}')\beta. \tag{B.10}$$

Summing each side of (B.10) over $\mathbf{x}' \in S_{\mathbf{y}}(\mathbf{x})$ gives

$$1 \geq \sum_{\mathbf{x}' \in S_{\mathbf{y}}(\mathbf{x})} p(\mathbf{x}'|\mathbf{y}) > \beta Q_{\mathbf{y}}(\mathbf{x}). \tag{B.11}$$

The left inequality in (B.11) holds since the sum of a set of disjoint probabilities cannot be greater than one. Hence we get

$$Q_{\mathbf{y}}(\mathbf{x}) < \frac{1}{\beta}. \tag{B.12}$$

By using (B.12) in (B.5) we obtain

$$\epsilon \leq Pr(i(\mathbf{x}, \mathbf{y}) \leq log\beta) + \frac{M-1}{\beta} \sum_{T} p(\mathbf{x}, \mathbf{y}) \tag{B.13}$$

and

$$\epsilon \leq Pr(i(\mathbf{x}, \mathbf{y}) \leq log\beta) + \frac{M-1}{\beta}, \tag{B.14}$$

by using the fact that the sum of probabilities of disjoint sets cannot exceed one. Since the average probability of error over the chosen ensemble of codes is satisfied there must be at least one code that satisfies the bound which concludes the proof.

APPENDIX C

Proof of Theorem 2.4.4

A maximum-likelihood decoder operates at the receiver so that the decoder maps the observation $\mathbf{y}$ into the integer m if

$$p(\mathbf{y}|\mathbf{x_m}) > p(\mathbf{y}|\mathbf{x_{m'}}), \forall m' \neq m, 1 \leq m' \leq M. \tag{C.1}$$

By using the random coding argument, the average probability of error can be written as

$$\epsilon = \sum_{\mathbf{x_m}} \sum_{\mathbf{y}} p(\mathbf{x_m}) p(\mathbf{y}|\mathbf{x_m}) p\left\{\text{error}|m, \mathbf{x_m}, \mathbf{y}\right\}, \tag{C.2}$$

where $p\{\text{error}|m, \mathbf{x_m}, \mathbf{y}\}$ is the probability of decoding error conditioned on the message m entering the encoder, reception of $\mathbf{y}$ at the receiver and also the selection of $\mathbf{x_m}$ as the m^{th} codeword.

For given $m, \mathbf{x}, \mathbf{y}$, define the incorrect decoding event $A_{m'}$ for every $m' \neq m$ as the event that the codeword m' is selected since $p(\mathbf{y}|\mathbf{x_{m'}}) \geq p(\mathbf{y}|\mathbf{x_m})$. Then we have

$$p\left\{error|m, \mathbf{x_m}, \mathbf{y}\right\} \leq P(\bigcup_{m \neq m'} A_{m'}) \leq \left[\sum_{m \neq m'} P(A_{m'})\right]^{\rho}, \text{for any } 0 < \rho \leq 1. \tag{C.3}$$

The first inequality in (C.3) is due to the fact that the decoder does not necessarily make an error if $p(\mathbf{y}|\mathbf{x_m}) = p(\mathbf{y}|\mathbf{x_{m'}})$ for some m' and the second inequality comes from Lemma 2.4.3. If we use the definition of $A_{m'}$ we have

$$P(A_{m'}) = \sum_{\mathbf{x_{m'}}: p(\mathbf{y}|\mathbf{x_{m'}}) \geq p(\mathbf{y}|\mathbf{x_m})} p(\mathbf{x_{m'}}) \tag{C.4}$$

We can bound (C.4) by multiplying each term by $[p(\mathbf{y}|\mathbf{x_{m'}})/p(\mathbf{y}|\mathbf{x})]^s$, for any $s > 0$ and if we further bound by summing over all $\mathbf{x'_m}$, we obtain

$$P(A_{m'}) \leq \sum_{\mathbf{x_{m'}}} p(\mathbf{x_{m'}}) \frac{p(\mathbf{y}|\mathbf{x_{m'}})^s}{p(\mathbf{y}|\mathbf{x_m})^s}, s > 0. \tag{C.5}$$

The upper bound in (C.5) is satisfied for every m' so we can rewrite (C.3) as

$$p\{\text{error}|m, \mathbf{x_m}, \mathbf{y}\} \leq \left[(M-1)\sum_{\mathbf{x}} p(\mathbf{x})\frac{p(\mathbf{y}|\mathbf{x_{m'}})^s}{p(\mathbf{y}|\mathbf{x_m})^s}\right]^\rho . \tag{C.6}$$

If we substitute (C.6) into (C.2), we get

$$\epsilon \leq (M-1)^\rho \sum_{\mathbf{y}} \left[\sum_{\mathbf{x_m}} p(\mathbf{x_m})p(\mathbf{y}|\mathbf{x_m})^{1-s\rho}\right]\left[\sum_{\mathbf{x_k}} p(\mathbf{x_k})p(\mathbf{y}|\mathbf{x_k})^s\right]^\rho \tag{C.7}$$

If we choose $s = 1/(1+\rho)$, then

$$\epsilon \leq (M-1)^\rho \sum_{\mathbf{y}} \left[\sum_{\mathbf{x_m}} p(\mathbf{x_m})p(\mathbf{y}|\mathbf{x_m})^{\frac{1}{1+\rho}}\right]\left[\sum_{\mathbf{x_k}} p(\mathbf{x_k})p(\mathbf{y}|\mathbf{x_k})^{\frac{1}{1+\rho}}\right]^\rho \tag{C.8}$$

Since $\mathbf{x_m}$ is a dummy variable for the summation the bound in (2.27) is finally obtained.