

2008

## Information-outage analysis of finite-length codes

David Scott Buckingham  
*West Virginia University*

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

---

### Recommended Citation

Buckingham, David Scott, "Information-outage analysis of finite-length codes" (2008). *Graduate Theses, Dissertations, and Problem Reports*. 1923.  
<https://researchrepository.wvu.edu/etd/1923>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact [researchrepository@mail.wvu.edu](mailto:researchrepository@mail.wvu.edu).

# Information-Outage Analysis of Finite-Length Codes

by

David Scott Buckingham

Thesis submitted to the  
College of Engineering and Mineral Resources  
at West Virginia University  
in partial fulfillment of the requirements  
for the degree of

Master of Science  
in  
Electrical Engineering

Daryl S. Reynolds, Ph.D.  
Natalia A. Schmid, Ph.D.  
Matthew C. Valenti, Ph.D., Chair

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia  
2008

Keywords: Information-outage probability, mutual information, finite-length, error rates

Copyright 2008 David Scott Buckingham

## Abstract

### Information-Outage Analysis of Finite-Length Codes

by

David Scott Buckingham  
Master of Science in Electrical Engineering

West Virginia University

Matthew C. Valenti, Ph.D., Chair

The performance of random error control codes approaches the Shannon capacity limit as the code length goes to infinity. When the code length is finite, then the code will be unable to achieve arbitrarily low error probability and a nonzero codeword error rate is inevitable. Information-theoretic bounds on codeword error rate may be found as a function of length through traditional methods such as sphere packing. Alternatively, the behavior of finite-length codes can be characterized in terms of an information-outage probability. The information-outage probability is the probability that the mutual-information rate, which is a random variable, is less than the code rate.

In this thesis, a Gaussian approximation is proposed that accurately models the information-outage probability for codewords of moderately short length. The information-outage probability is related to several previously derived bounds, including Shannon's sphere-packing and random coding bounds, as well as a bound on maximal error probability known as Feinstein's lemma. It is shown that the information-outage probability is a useful predictor of achievable error rate.

# Acknowledgements

I would first like to thank my committee chair and advisor, Dr. Matthew C. Valenti, for giving me the opportunity to work with him and his students. This thesis would not be possible without his constant guidance and support.

I would also like to thank Dr. Daryl S. Reynolds and Dr. Natalia A. Schmid for being on my committee. I have been fortunate to have had the opportunity to take courses with all of my committee members, and their teachings have been essential to my understanding of the subject.

Next, I would also like to thank the students in the wireless communications research lab with whom I've had the pleasure of working alongside. In particular, I would like to thank my colleague John Mazzie, who has been a great help to me over the last few years.

Finally, I would like to express my gratitude to my family. First to my wife, Sarah, for keeping me going through all the stresses and deadlines, while also dealing with her own. I would be remiss if I did not also express my thanks to my mother and father. Their support seemingly has no limit, and has been much appreciated throughout my life.

# Contents

<b>Acknowledgements</b>	<b>iii</b>
<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>Notation</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Thesis Outline . . . . .	2
<b>2 Digital Modulation</b>	<b>5</b>
2.1 System Model . . . . .	5
2.2 Modulation . . . . .	8
2.3 Demodulation . . . . .	11
2.3.1 Demodulator . . . . .	11
2.3.2 Detector . . . . .	13
2.3.3 Computing Bit LLRs . . . . .	15
2.4 Chapter Summary . . . . .	17
<b>3 Mutual Information</b>	<b>18</b>
3.1 Review of Mutual Information . . . . .	19
3.2 Modulation-Constrained Input . . . . .	20
3.3 Unconstrained Input . . . . .	23
3.3.1 AWGN Channel . . . . .	23
3.3.2 Rayleigh Fading Channel . . . . .	28
3.4 Chapter Summary . . . . .	34
<b>4 Information-Outage Probability</b>	<b>36</b>
4.1 Block Fading Channels . . . . .	37
4.2 Finite-Blocklength Codes . . . . .	37
4.2.1 AWGN Channel . . . . .	38
4.2.2 Rayleigh Fading Channel . . . . .	42
4.3 Coded Performance . . . . .	46

4.4 Chapter Summary . . . . .	48
<b>5 Alternative Bounds</b>	<b>50</b>
5.1 Feinstein's Lemma . . . . .	50
5.2 Sphere-Packing Bound . . . . .	52
5.3 Random Coding Bound . . . . .	54
5.4 Chapter Summary . . . . .	57
<b>6 Conclusion</b>	<b>58</b>
6.1 Summary . . . . .	58
6.2 Future Research . . . . .	60
<b>References</b>	<b>64</b>

# List of Figures

2.1	Simple system model for a digital communications system. . . . .	6
2.2	Basic symbol-level channel model. . . . .	7
2.3	Constellation diagram for 16-QAM. . . . .	10
3.1	The exact Bessel-K pdf and its Gaussian approximation for $n = \{1, 5, 10\}$ and $\mathcal{E}_s/N_0 = 3/4$ . . . . .	26
3.2	A histogram of $10^7$ realizations of $Z_{10}$ for $\mathcal{E}_s/N_0 = 3/4$ along with the exact and approximate pdfs. . . . .	27
3.3	The exact pdf of the mutual-information rate in (3.44) (from $10^7$ samples of mutual-information rate) and its Gaussian approximation for $n = \{1, 5, 10\}$ and $\mathcal{E}_s/N_0 = 3/4$ . . . . .	33
4.1	CDF of the exact and approximate mutual-information variables for $n = 20$ and $\mathcal{E}_s/N_0 = 3/4$ . . . . .	40
4.2	Information outage probability for AWGN vs. SNR for various blocklengths at binary rate $R_2 = 2/3$ . . . . .	41
4.3	Minimum SNR ( $E_b/N_o$ ) in dB required to achieve $P_o = 10^{-4}$ at binary rate $R_2 = 2/3$ as a function of blocklength, $n$ . . . . .	43
4.4	CDF of the exact and approximate mutual-information variables in Rayleigh fading for $n = 20$ and $\mathcal{E}_s/N_0 = 3/4$ . . . . .	44
4.5	Information-outage probability for Rayleigh fading vs. SNR for various blocklengths at binary rate $R_2 = 2/3$ . . . . .	45
4.6	Frame-error rate (FER) of the turbo code from the LTE standard and the information-outage probability of the same blocklength over AWGN channel. $k$ is the number of bits per block and the number of symbols per block is $(3k + 12)/2$ . . . . .	47
4.7	Frame-error rate (FER) of the turbo code from the LTE standard and the information-outage probability of the same blocklength over Rayleigh fading channel. $k$ is the number of bits per block and the number of symbols per block is $(3k + 12)/2$ . . . . .	48
5.1	LTE Frame-error rate (FER) and IOP curves from Fig. 4.6 with the maximal error rate determined from Feinstein's lemma superimposed. . . . .	51
5.2	LTE Frame-error rate (FER) and IOP curves from Fig. 4.6 with sphere-packing bound superimposed. . . . .	53

5.3	LTE Frame-error rate (FER) and IOP curves from Fig. 4.6 with random coding bound superimposed. . . . .	55
5.4	Minimum SNR ( $E_b/N_o$ ) in dB required to achieve $P_o = 10^{-4}$ at binary rate $R_2 = 2/3$ as a function of blocklength, $n$ . . . . .	56
6.1	Information-outage probability of unconstrained Gaussian input and modulation-constrained input for AWGN channel. Frame error rates of turbo code included for reference with outage curves. . . . .	61
6.2	Information-outage probability of unconstrained Gaussian input and modulation-constrained input for Rayleigh fading channel. Frame error rates of turbo code included for reference with outage curves. . . . .	62



# List of Tables

4.1	Minimum SNR ( $E_b/N_o$ ) in dB Required for Various Outage Probabilities and Blocklengths - AWGN Channel (Binary Rate = $2/3$ ) . . . . .	42
4.2	Minimum SNR ( $E_b/N_o$ ) in dB Required for Various Outage Probabilities and Blocklengths - Rayleigh Fading Channel (Binary Rate = $2/3$ ). Differences in parentheses indicate the approximation required higher SNR . . . . .	46

# Notation

We use the following notation and symbols throughout this thesis.

$(\cdot)^H$	: Complex conjugate transpose
$(\cdot)^*$	: Complex conjugate
$E[\cdot]$	: Expectation operator
$F_X(x)$	: Cumulative Distribution Function (CDF) of a random variable $X$
$p_X(x)$	: Probability density function (pdf) of a random variable $X$
$f(\cdot)$	: Likelihood function (proportional to probability)
$\Lambda$	: Log-likelihood function
$\lambda$	: Log-likelihood ratio
$\ \cdot\ $	: Euclidian norm
$\Re\{\cdot\}$	: Real part of the argument
$\Im\{\cdot\}$	: Imaginary part of the argument
$ \cdot $	: Cardinality of a set
$\mathcal{N}$	: Gaussian Distribution
$\mathcal{N}_c$	: Circular Complex Gaussian Distribution

Bold upper case letters denote matrices and bold lower case letters denote vectors.

# Chapter 1

## Introduction

### 1.1 Introduction

In 1948 Shannon defined the channel capacity as the maximum rate at which data could be transmitted over a channel with arbitrarily low probability of error [1]. The capacity can be found by finding the *average mutual information* between channel input and output, maximized over all input distributions. The average mutual information is simply the expectation of the *mutual information* random variable. However, the Shannon capacity does not exist for nonergodic channels, such as slow fading channels or channels with finite-length inputs. For such systems, the mutual information is a random variable whose value will change from one channel realization to the next.

The probability that the mutual information is less than the code rate  $R$  is called the *information-outage probability* (IOP) [2, 3]. The information-outage probability is the cumulative distribution function (CDF) of the mutual information evaluated at  $R$ . Information outage probability provides some insight into the frame-error rate (FER) that may be achieved by practical codes operating in nonergodic settings.

Past work [3, 4, 5] on information-outage probability has concentrated on the block-fading channel. The mutual information per block is taken to be  $\log(1 + SNR)$  while the SNR is allowed to vary from block to block. The overall mutual information is then found by averaging across the blocks. In effect, it is assumed that the channel is able to achieve the Shannon capacity for the duration of each block and the randomness in the mutual

information is due to the randomness of the SNR. However, there is a paradox in that a block must be sufficiently small to span just one coherence interval, yet sufficiently long that the Shannon capacity is achieved.

Even in the additive white Gaussian noise (AWGN) channel, the mutual information between channel input and output is a random variable if the blocklength is finite. An information-outage probability can be defined for the AWGN channel with finite blocklengths just as it can be defined for the block-fading channel. This outage probability has been called the *information spectrum* [6, 7] in the recent literature. In [8, 9], the mutual information is shown to be drawn from a Bessel-K distribution which describes the average of  $n$  independent and identically distributed (i.i.d.) Laplacian random variables, where  $n$  is the number of symbols per codeword.

We revisited this distribution in [10], where a Gaussian approximation to it was proposed. In this thesis, we continue the work started in [10], expanding the focus to include the fading channel in addition to AWGN.

## 1.2 Thesis Outline

Before we delve into the information-theoretic material concerning the mutual-information rate, we first take a step back to look at a digital communications system as a whole in chapter 2. A brief description of the individual components of a communications system is provided. For more detailed explanations concerning digital communications systems, the reader is encouraged to refer to [11, 12, 13].

In chapter 3 we review key concepts from information theory, including mutual information, average mutual information and the mutual-information rate. The mutual-information rate is key to our application of information-outage probability and is used throughout the thesis. The mutual-information rate is first derived for a channel with a modulation-constraint imposed on the input distribution. We then look at the mutual-information rate for unconstrained Gaussian (capacity-achieving) input. This was done previously in [9], where a mean-shifted Laplacian random variable was found to describe the mutual information for the AWGN channel. It was shown that the mutual-information rate was equivalent

to the average of  $n$  samples of the mutual information if the inputs to the channel are independent and identically distributed (i.i.d). The average of  $n$  i.i.d. Laplacian random variables is a Bessel-K distribution. Therefore, a mean-shifted Bessel-K distribution was found to represent the mutual-information rate for AWGN.

Recalling the Central Limit Theorem [14], we know that the sum of  $n$  i.i.d. random variables with finite mean and variance will approach a Gaussian distribution as  $n \rightarrow \infty$ . Therefore, we introduce a Gaussian distribution in [10] to approximate the mutual information, and by extension the mutual-information rate. We also introduce a Gaussian approximation for the fading channel, for which an exact distribution for the mutual-information rate is not defined.

In chapter 4, the concept of information-outage probability is applied to the case of finite-length codes. Because the codewords are of finite-length, then the mutual-information rate is a random variable. Similar to the block-fading approach to information-outage, we use the term *instantaneous capacity*. If we let the instantaneous capacity be an individual sample of the mutual-information rate, then in theory, this gives us the instantaneous rate at which the codeword can be communicated with arbitrarily low probability of error. For a fixed code rate, the codeword may or may not be able to be transmitted across the channel with low probability of error, depending on the sampled mutual-information rate. Therefore, the event of an *information-outage* is the event that the instantaneous capacity (or mutual-information rate) is less than the code rate. The probability that a random variable,  $X$ , takes on a value less than or equal to an argument,  $x$  is [14]

$$F_X(x) = P[X \leq x],$$

where  $F_X(x)$ , is the CDF of the random variable,  $X$ . Therefore, the probability of an information-outage is simply the CDF of the mutual-information rate, evaluated at the code rate. Given the CDF of the mutual-information rate, we can then compute the information-outage probability for a code with specified length and rate.

We will see that for the AWGN channel, the CDF of the mutual-information rate is somewhat cumbersome and difficult to calculate for even moderately low blocklength. However, the computation of the information-outage probability using the Gaussian approximation

requires only the use of the Gaussian Q-function. For the fading channel, a distribution for the mutual-information rate cannot be readily found in closed form. Therefore the resulting information-outage probabilities must be computed using simulation, unless an exact distribution is derived. Again, we can apply our Gaussian approximation to the mutual-information rate for a fading channel. We then show that the approximations are a good fit for blocklengths as low as  $n = 120$ .

Having found an easily computable method for approximating the information-outage probability, we compare the IOP with the frame-error rate (FER) that may be achieved with actual codes (in this case, the turbo code from the UMTS-LTE standard). A relationship is observed between the IOP and the simulated FER; the two curves have a similar slope with a fixed horizontal offset.

While these results might suggest that the information-outage probability is a useful quantity for predicting the performance of good finite-length codes, we will compare it to other previously derived bounds in chapter 5. The bounds that we compare against are: Feinstein's Lemma, the sphere-packing bound, and the random coding bound. However, we will see that not only is the computation of the bounds non-trivial, but the information-outage probability is often a better predictor of performance.

Finally in chapter 6, we review the findings of the thesis, and suggest ideas for future research.

# Chapter 2

## Digital Modulation

Many of today's modern communications systems utilize digital modulation techniques in order to reliably transmit and receive information. Digital communications offers several advantages to traditional analog techniques. Because of the inherent flexibility of digital data, a digital signal can represent a discrete data source or a digitized analog source [15]. Digital communications also allows the use of techniques such as *source coding*, or compression, and *forward error correction (FEC) coding*, which help the receiver to recover the transmitted signal in a noisy environment.

In this chapter, we will review the concept of digital modulation. In section 2.1 we will present a simple system model that describes a typical digital communications system. After briefly discussing the components of the model, the process of modulation will be addressed in section 2.2. This essentially involves selecting a signal to transmit based on the bits or symbols that are input to the modulator. In section 2.3, the demodulation process will be explained. During demodulation, the receiver attempts to select the most likely transmitted symbol given the received signal, which has been corrupted by the channel. Finally a summary of the chapter will be given in section 2.4.

### 2.1 System Model

To accurately predict the performance of a digital communications system, it is important to develop an appropriate system model. Shown in Fig. 2.1 is a simple system model used

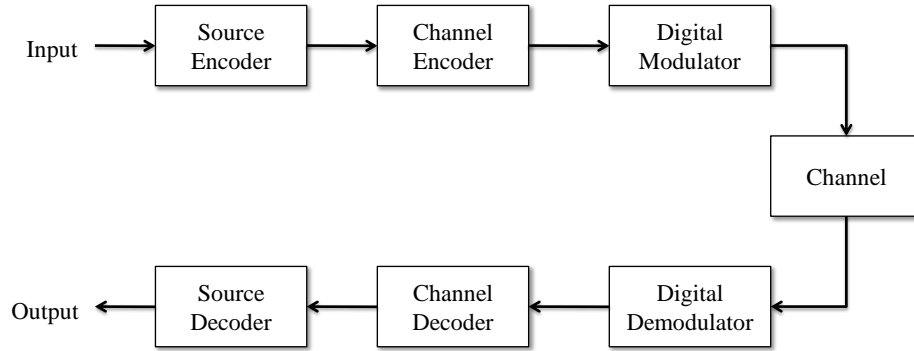


Figure 2.1: Simple system model for a digital communications system.

commonly for digital communications systems.

Let the input to the system,  $\mathbf{d}$ , be a length- $l$  vector of data bits. This vector is optionally passed through a *source encoder*. The purpose of the source encoder is to remove uncontrolled redundancy, thereby reducing the number of bits required to represent the original data. This is done using an appropriate source coding algorithm, such as Huffman coding [16] or the Lempel-Ziv algorithm. The source coding process is also known as *data compression*, and can be lossy or lossless, depending on the algorithm used. The output of the source encoder is a length- $k$  vector,  $\mathbf{u}$ , of information bits, more commonly referred to as the message. The message is then passed into a *channel encoder*.

The purpose of the channel encoder is to introduce controlled redundancy to the message that the receiver can use to overcome the adverse effects of the channel. There are a number of methods available to accomplish this. In fact, the subject of channel coding has been heavily studied in the past half-century. As a result there are many different types of channel codes, ranging from codes as simple as repeating a bit  $n$  times (i.e. repetition codes), to more complex coding schemes, such as Turbo codes and Low-Density Parity-Check (LDPC) codes. The latter have been shown to approach the Shannon limit [17].

The output of the channel encoder is a length- $n$  vector of code symbols  $\mathbf{c}$ , also called the *codeword*. Note that the output of the encoder is a vector of symbols, which are not restricted to only bits. Not all channel codes are binary. One of the more common classes of nonbinary codes is the Reed-Solomon codes [18], which are used to encode data on many



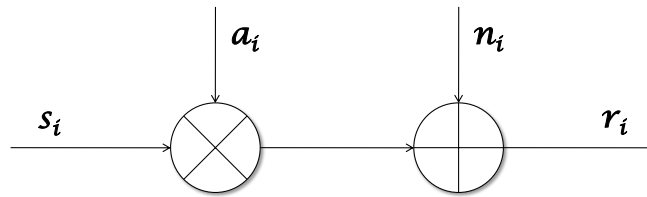


Figure 2.2: Basic symbol-level channel model.

of today's physical digital media. One of the advantages of using nonbinary codes are their inherent ability to handle burst errors. Each code has advantages and disadvantages, with a tradeoff commonly between performance and complexity. Selecting the best channel code for a given system is an art, and many factors must be taken into consideration. Once the message is encoded, the resulting codeword is then passed to the *digital modulator*.

The digital modulator is used to map bits or groups of bits to modulated signals/symbols that will be transmitted across the channel. It is important to better define what we call a *signal* and a *symbol*. Let a *signal* be the modulated waveform defined in the continuous-time domain, and a *symbol* be its discrete-time representation. For the purposes of this thesis, it is sufficient to analyze the system in discrete-time, and therefore we will use symbols rather than signals. The modulation process is discussed in greater detail later in section 2.2. The output of the digital modulator is a length- $N$  vector,  $\mathbf{s}$ , of modulated symbols. The symbols are selected from  $M$  available symbols, depending on the modulation.  $M$  is also known as the *order* of the modulation. Each symbol is in itself a  $K$ -dimensional vector, with  $K$  being the *dimensionality* of the modulation. As we will see later, the dimensionality is also dependent on the type of modulation being used. In this thesis, we limit our attention to one or two dimensional modulation, in which case the symbol may be conveniently represented as a real or complex scalar, respectively. These symbols are transmitted over the *channel*.

A basic channel model is shown in Fig. 2.2. The figure shows the effects of the channel at the symbol level. The channel takes an input symbol,  $s_i$ , which is first multiplied by a channel gain,  $a_i$ . In the case of an AWGN channel, it is a constant value (which we select to be unity). Alternatively, it could represent a single sample of a flat-fading process. In

general the fading could be correlated. However, in this thesis we will assume i.i.d. fading with each symbol scaled by a distinct sample. This is a valid assumption when the fading is fast or when an interleaver is used between the encoder and modulator. Once the input symbol is multiplied by the channel gain, a noise sample,  $n_i$  is added to the result. The noise samples are taken from an additive white Gaussian noise (AWGN) process. Each noise sample,  $n_i$ , is a  $K$ -dimensional Gaussian vector written compactly as

$$n_i \sim \mathcal{N}\left(\mathbf{0}, \frac{N_o}{2} I_K\right) \quad (2.1)$$

In other words, the noise is zero-mean Gaussian with variance  $N_o/2$  in each dimension. The resulting symbol,  $r_i$ , can be expressed as

$$r_i = a_i s_i + n_i. \quad (2.2)$$

The *digital demodulator* receives the channel output vector,  $\mathbf{r}$ . For each received symbol,  $r_i$ , the demodulator computes the likelihood of each of the  $M$  possible symbols. These *symbol likelihoods* are then used to compute the likelihood of each of the  $\log_2(M)$  bits mapped to the symbol. The demodulation process is defined more clearly in section 2.3. The resulting likelihoods are passed to the *channel decoder*, which attempts to decode data by finding the most likely codeword,  $\hat{\mathbf{c}}$ . Depending on the type of channel coding, this process may be iterative. Once the most likely codeword is found, the resulting most likely message,  $\hat{\mathbf{u}}$ , is the message which generates the codeword,  $\hat{\mathbf{c}}$ . If source coding was used, then the appropriate algorithm is used to expand the message,  $\hat{\mathbf{u}}$ , into its uncompressed form,  $\hat{\mathbf{d}}$ . If the system is performing ideally, then  $\hat{\mathbf{d}}$ , will be identical to  $\mathbf{d}$ , with high probability.

## 2.2 Modulation

Let us now better define the process of digital modulation. The modulator selects from  $M$  available signal waveforms to transmit. In general,  $M$  is a power of 2, and can be as small as 2, in which case the modulation is *binary*. Binary modulation maps each individual bit to one of two possible *signal waveforms*,  $s_0(t)$  or  $s_1(t)$ . Typically a bit value of 0 is mapped to signal  $s_0(t)$ , and a value of 1 is mapped to signal  $s_1(t)$ . In general, signal waveforms will

also be known as *modulated symbols*.  $M$ -ary modulation maps a group of  $b$  bits to one of the  $M$  available signal waveforms, where  $M = 2^b$ .

Each modulated signal can be represented as a weighted linear combination of orthonormal functions, known as the *basis functions*. The cardinality,  $K$ , of the set of basis functions is also called the dimensionality of the modulation. Let

$$K = |\Phi| \quad (2.3)$$

where

$$\Phi = \{\phi_0(t), \phi_1(t), \dots, \phi_{K-1}(t)\} \quad (2.4)$$

$K$  is modulation-specific, and can vary from  $K = 1$ , in the case of Binary Phase-Shift Keying (BPSK) or Pulse-Amplitude-Modulation (PAM), to  $K = M$ , in the case of orthogonal Frequency-Shift Keying (FSK). Many common  $M$ -ary modulations are defined with dimensionality  $K = 2$ . For modulations with dimensionality  $K \leq 2$ , it is convenient to display the set of modulated signals in the form of a *constellation diagram*. An example of one such diagram is shown in Fig. 2.3 for Quadrature Amplitude Modulation (QAM) with  $M = 16$ . Notice that each of the axes corresponds to the projection of the signal onto one of the basis functions.

The constellation *mapping* determines which symbol that each combination of  $b$  bits represents. The mapping is important in terms of the bit-error performance of the system. For example, consider that if a symbol is received incorrectly, it will most likely be misinterpreted as a nearby symbol (in terms of the signal constellation). Depending on the mapping, the representative bits of each symbol could differ in 1 or up to  $b$  bit positions. So a single symbol error could represent a single bit error or up to  $b$  bit errors. The optimal mapping is modulation-dependent, and work has been done to determine not only optimal mappings for specific modulations, but also algorithms that can be used to recursively determine the optimal mapping for any signal constellation [19, 20].

Let  $s_{m,k}$  be the coefficients or weights of the basis functions,  $\phi_k(t)$  for  $k$  from 0 to  $K - 1$ , required to create the modulated signal,  $s_m(t)$ ,  $0 \leq m \leq M - 1$ . Each constellation point,

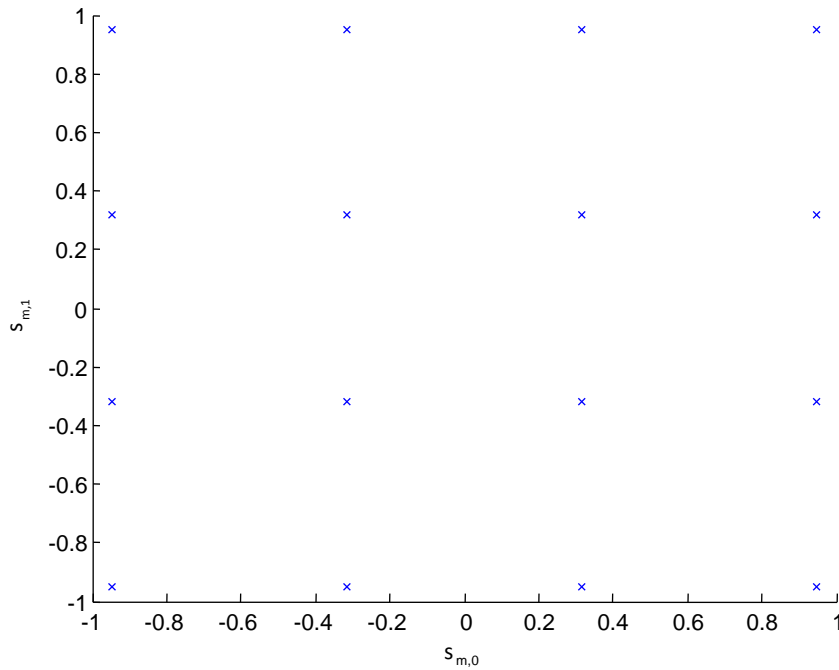


Figure 2.3: Constellation diagram for 16-QAM.

$s_m(t)$  is then

$$s_m(t) = \sum_{k=0}^{K-1} s_{m,k} \phi_k(t). \quad (2.5)$$

As discussed previously, it is sufficient to view the system as discrete in time. Therefore we can use the coefficients  $s_{m,k}$ , in an equivalent discrete-time representation, which we call a symbol. For 2-D modulations, it is convenient to represent each constellation point as a complex number. If we let  $\phi_1(t)$  represent the real axis, and likewise,  $\phi_2(t)$  represent the imaginary axis, it follows that

$$s_m = s_{m,0} + j s_{m,1} \quad (2.6)$$

There are several motivations for using complex numbers to represent modulated symbols. For example, in software a vector of symbols can be represented as a single complex vector, rather than two vectors representing the real and complex parts. Let

$$\mathcal{S} = \{s_0, \dots, s_{M-1}\} \quad (2.7)$$

be the set of all  $M$  possible modulated symbols. For each transmitted symbol, the modulator selects  $b = \log_2(M)$  bits and maps them to the appropriate symbol from the set,  $\mathcal{S}$ . This symbol is then transmitted, where it undergoes the effects of the channel.

## 2.3 Demodulation

Now we will briefly go over the process of demodulation. The purpose of the demodulator is to process the received signal and determine the symbol/bit likelihoods. Once this is done, a hard-decision may be made by selecting the symbol from the set,  $\mathcal{S}$ , which has the largest likelihood. Once the most likely symbol is selected, then the appropriate bit (binary) or bits ( $M$ -ary) can be selected from the mapping used. A receiver can be said to have a *front end* and a *back end*. The front end is known as a *demodulator*, and the back end is known as a *detector*.

In this discussion we consider only the case of *coherent detection*. In the case of coherent detection, the receiver is synchronized with the phase of the transmitter. If the basis functions (with correlation demodulation) or the filters (with matched-filter demodulation) are synchronized in phase with the received signal, then the SNR at the output of the demodulator is maximized.

### 2.3.1 Demodulator

The front end, or demodulator, transforms the received signal waveform,  $r(t)$ , into a  $K$ -dimensional vector  $\mathbf{r}$ , where  $K$  is the dimensionality of the modulation, defined in (2.3). Each value,  $r_i$ , in the vector,  $\mathbf{r}$ , is the correlation of  $r(t)$  and the  $i$ -th basis function. Two types of demodulators are the *correlation demodulator* and the *matched-filter demodulator*.

A correlation demodulator first passes the received signal,  $r(t)$ , through a bank of  $K$  correlators. In the  $i$ -th correlator, the received signal is first multiplied by the  $i$ -th basis function, then integrated over the period of the basis function, and finally sampled once at the end of the period. Essentially, each correlator finds the time-domain inner product of

the received signal and the  $i$ -th basis function:

$$\begin{aligned} r_i &= \int_0^{T_s} r(t)\phi_i(t)dt \\ &= \langle r(t), \phi_i(t) \rangle \end{aligned} \quad (2.8)$$

Let the transmitted signal,  $s(t)$ , be the  $j$ -th signal of the signal set. Again we can represent  $s(t)$  by a summation of weighted basis functions as in (2.5). For the AWGN case, we have

$$r(t) = s(t) + n(t). \quad (2.9)$$

Using the correlation demodulator, the  $i$ -th value in the received vector  $\mathbf{r}$  is found to be

$$\begin{aligned} r_i &= \int_0^{T_s} r(t)\phi_i(t)dt \\ &= \int_0^{T_s} \left[ \sum_{k=0}^{K-1} s_{j,k}\phi_k(t) + n(t) \right] \phi_i(t)dt \\ &= \sum_{k=0}^{K-1} s_{j,k} \int_0^{T_s} \phi_k(t)\phi_i(t)dt + \int_0^{T_s} n(t)\phi_i(t)dt \\ &= s_{j,i} + n_i, \end{aligned} \quad (2.10)$$

where

$$n_i = \int_0^{T_s} n(t)\phi_i(t)dt \quad (2.11)$$

is the correlation of the noise with the  $i$ -th basis function. The received signal  $r(t)$ , is thus transformed to the vector  $\mathbf{r}$ , where

$$\mathbf{r} = \mathbf{s}_j + \mathbf{n}, \quad (2.12)$$

where  $\mathbf{n} = [n_0, \dots, n_{K-1}]$  is a vector containing the correlations of the noise with the  $K$  basis functions.

A matched-filter demodulator first passes the received signal,  $r(t)$ , through  $K$  matched-filters, with impulse responses given by

$$h_i(t) = \phi_i(T_s - t), \quad 0 \leq t \leq T_s, i \in [0 : K - 1] \quad (2.13)$$

where  $T_s$  is the duration of the  $i$ -th basis function,  $\phi_i(t)$ . The result is then sampled at the end of the epoch. The output of a filter is the convolution of its input with its impulse response. Therefore the output of the  $i$ -th matched-filter is

$$\begin{aligned}
 r_i &= r(t) * h_i(t)|_{t=T_s} \\
 &= \int_{-\infty}^{\infty} r(\lambda)h_i(\lambda)d\lambda |_{t=T_s} \\
 &= \int_{-\infty}^{\infty} r(\lambda)h_i(T_s - \lambda)d\lambda \\
 &= \int_0^{T_s} r(\lambda)\phi_i(\lambda)d\lambda \\
 &= \langle r(t), \phi_i(t) \rangle
 \end{aligned} \tag{2.14}$$

We see that the output of the matched-filter, shown in (2.14) is equivalent to the output of the correlation demodulator, shown in (2.8).

### 2.3.2 Detector

The back end, or detector, picks the most likely transmitted symbol,  $\hat{\mathbf{s}}$ , given the received vector,  $\mathbf{r}$ . Two types of detectors are the *MAP detector* and the *ML detector*. The detectors are similar, with the only exception being that the MAP detector can handle input distributions that are not uniform.

A *maximum a posteriori* (MAP) detector determines which signal was most likely sent by picking the estimated symbol,  $\hat{\mathbf{s}}$ , that is most probable given the received symbol,  $\mathbf{r}$ . The MAP rule for selecting the estimated symbol is

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}_i \in \mathcal{S}} p(\mathbf{s}_i | \mathbf{r}). \tag{2.15}$$

The probabilities of symbol  $\mathbf{s}_i$  given received vector,  $\mathbf{r}$ , are not necessarily known. However, if we consider (2.12), and we instead condition  $\mathbf{r}$  on  $\mathbf{s}_i$ , the resulting distribution is conditionally Gaussian and given by

$$\mathbf{r} | \mathbf{s}_i \sim \mathcal{N} \left( \mathbf{s}_i, \frac{N_o}{2} I_K \right) \tag{2.16}$$

The MAP rule in (2.15) can be written in terms of the desired probabilities by using Bayes' Theorem, which states that

$$p(\mathbf{s}_i|\mathbf{r}) = \frac{p(\mathbf{r}|\mathbf{s}_i)p(\mathbf{s}_i)}{p(\mathbf{r})} \quad (2.17)$$

The MAP rule can now be stated as

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}_i \in \mathcal{S}} \frac{p(\mathbf{r}|\mathbf{s}_i)p(\mathbf{s}_i)}{p(\mathbf{r})}. \quad (2.18)$$

Because we are attempting to maximize (2.18) over each possible symbol, we can drop the pdf of the received vector,  $\mathbf{r}$ , since it is constant over all input symbols. Our final expression for the MAP rules is then

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}_i \in \mathcal{S}} p(\mathbf{r}|\mathbf{s}_i)p(\mathbf{s}_i). \quad (2.19)$$

A *maximum likelihood* (ML) detector uses the same rule in (2.19), but simplifies it by assuming that the input distribution is uniform. Therefore the probability of each symbol is equally likely and it can be removed from the expression. The ML rule is thus

$$\hat{\mathbf{s}} = \arg \max_{\mathbf{s}_i \in \mathcal{S}} p(\mathbf{r}|\mathbf{s}_i). \quad (2.20)$$

We know this distribution, as defined in (2.16). The pdf of the distribution is actually a joint conditional pdf of the marginal pdf's in each dimension. Because the basis functions that represent each dimension are orthonormal, and the noise is white, the joint pdf is simply the product of the marginals. The resulting joint conditional pdf is given by

$$p(\mathbf{r}|\mathbf{s}_i) = \frac{1}{(\pi N_o)^{K/2}} \exp \left\{ -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}_i\|^2 \right\} \quad (2.21)$$

The goal of the ML detector is to maximize the expression in (2.21) over all possible symbols. We can simplify the expression in (2.21) by first removing the constant in front of the exponential. The resulting expression,

$$f(\mathbf{r}|\mathbf{s}_i) = \exp \left\{ -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}_i\|^2 \right\}, \quad (2.22)$$

is known as the *likelihood* function. If we consider that the logarithm is a monotonically increasing function, we can further reduce the expression by taking the log of the likelihood



function,  $f(\mathbf{r}|\mathbf{s}_i)$ . The resulting *log-likelihood* of symbol  $\mathbf{s}_i$ , denoted by  $\Lambda_i$ , is then

$$\begin{aligned}\Lambda_i &= \log f(\mathbf{r}|\mathbf{s}_i) \\ &= -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}_i\|^2.\end{aligned}\tag{2.23}$$

With these manipulations, we can write the ML rule as

$$\begin{aligned}\hat{\mathbf{s}} &= \arg \max_{\mathbf{s}_i \in \mathcal{S}} f(\mathbf{r}|\mathbf{s}_i) \\ &= \arg \max_{\mathbf{s}_i \in \mathcal{S}} \log f(\mathbf{r}|\mathbf{s}_i) \\ &= \arg \max_{\mathbf{s}_i \in \mathcal{S}} -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}_i\|^2.\end{aligned}\tag{2.24}$$

To further simplify the expression we again remove the constant. We can also remove the negation by finding the symbol which minimizes the resulting expression. The final expression can be reduced to

$$\begin{aligned}\hat{\mathbf{s}} &= \arg \max_{\mathbf{s}_i \in \mathcal{S}} -\|\mathbf{r} - \mathbf{s}_i\| \\ &= \arg \min_{\mathbf{s}_i \in \mathcal{S}} \|\mathbf{r} - \mathbf{s}_i\|\end{aligned}\tag{2.25}$$

We see in (2.25), that the ML rule is equivalent to picking the symbol with the minimum Euclidian distance to the received vector,  $\mathbf{r}$ .

### 2.3.3 Computing Bit LLRs

Bit interleaved coded modulation (BICM) [21], introduces an interleaver between the channel encoder and the modulator. This interleaving of bits has several applications, one such being that of combating the effects of fading [22]. If BICM is used, then the symbol log-likelihoods are no longer sufficient. The corresponding bit log-likelihoods must also be calculated before the de-interleaving process on the receiver side.

Because there are only two possible values for a single bit, it is convenient to represent the bit log-likelihoods as a single bit *log-likelihood ratio (LLR)*.

Let  $\lambda_n$  be the bit LLR of the  $n^{\text{th}}$  code bit of a given group of  $b = \log_2(M)$  bits.

$$\lambda_n = \log \frac{p(c_n = 1|\mathbf{r})}{p(c_n = 0|\mathbf{r})}\tag{2.26}$$

where  $p(c_n = 1|\mathbf{r})$  is the probability that the bit is a 1 given the received vector  $\mathbf{r}$ . From (2.26) we can see that if the bit is more likely to be 1, the resulting LLR will be positive.

We can represent the probability of the  $n^{\text{th}}$  code bit in terms of the probabilities of all symbols whose bitwise mappings contain the same value in the  $n^{\text{th}}$  bit position. The LLR expression in (2.26) can thus be represented as

$$\begin{aligned}\lambda_n &= \log \frac{\sum_{\mathbf{s}_i \in \mathcal{S}_n^{(1)}} p(\mathbf{s}_i|\mathbf{r})}{\sum_{\mathbf{s}_i \in \mathcal{S}_n^{(0)}} p(\mathbf{s}_i|\mathbf{r})} \\ &= \log \frac{\sum_{\mathbf{s}_i \in \mathcal{S}_n^{(1)}} p(\mathbf{r}|\mathbf{s}_i) p[s_i]}{\sum_{\mathbf{s}_i \in \mathcal{S}_n^{(0)}} p(\mathbf{r}|\mathbf{s}_i) p[s_i]} \\ &= \log \left[ \sum_{\mathbf{s}_i \in \mathcal{S}_n^{(1)}} p(\mathbf{r}|\mathbf{s}_i) \right] - \log \left[ \sum_{\mathbf{s}_i \in \mathcal{S}_n^{(0)}} p(\mathbf{r}|\mathbf{s}_i) \right],\end{aligned}\quad (2.27)$$

where  $\mathcal{S}_n^{(1)}$  represents the set of symbols whose  $n^{\text{th}}$  bit position is a 1. With equally likely signals, the  $p[x_i]$  term is a constant and can therefore be pulled outside of both summations, in effect canceling each other.

The objective is to turn the already calculated symbol likelihoods into bit LLRs. We have an expression for the symbol likelihoods in (2.24). We can express (2.27) in terms of the symbol likelihoods by recalling the relationship between the logarithm and its base,  $e$ .

$$\lambda_n = \log \left[ \sum_{\mathbf{s}_i \in \mathcal{S}_n^{(1)}} \exp \{ \log(p(\mathbf{r}|\mathbf{s}_i)) \} \right] - \log \left[ \sum_{\mathbf{s}_i \in \mathcal{S}_n^{(0)}} \exp \{ \log(p(\mathbf{r}|\mathbf{s}_i)) \} \right]. \quad (2.28)$$

In (2.28) we arrive at an expression containing a sum of exponentials. This is also known as a *Jacobian logarithm*. In [23], a recursive function named  $\max^*$  is presented to solve such a logarithm, and is defined as

$$\begin{aligned}\max^*(x, y) &= \log [e^x + e^y] \\ &= \max(x, y) + \log [1 + e^{-|y-x|}] \\ &= \max(x, y) + f_c(|y - x|),\end{aligned}\quad (2.29)$$

where  $f_c$  is a correction function which depends only on the magnitude of the difference of  $x$

and  $y$ . The LLR expression in (2.28) can thus be represented in terms of the  $\max *$  function

$$\begin{aligned}\lambda_n &= \max_{\mathbf{s}_i \in \mathcal{S}_n^{(1)}} * \log [p(\mathbf{r}|\mathbf{s}_i)] - \max_{\mathbf{s}_i \in \mathcal{S}_n^{(0)}} * \log [p(\mathbf{r}|\mathbf{s}_i)] \\ &= \max_{\mathbf{s}_i \in \mathcal{S}_n^{(1)}} * \left[ -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}_i\|^2 \right] - \max_{\mathbf{s}_i \in \mathcal{S}_n^{(0)}} * \left[ -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}_i\|^2 \right].\end{aligned}\quad (2.30)$$

The expression in (2.30) can then be used to determine the bit log-likelihood ratios before the de-interleaving process.

## 2.4 Chapter Summary

In this chapter, the process of digital modulation was discussed. A basic system model for a digital communications system was presented, along with a brief description of each component of the system. The modulation process involves mapping bits or symbols to a specific signal from a finite set of signals. Each signal is a weighted combination of basis functions specified by the specific modulation. As the signal is transmitted across the channel, it is corrupted by fading and additive noise. Once the signal is received, it is passed through a correlator or matched-filter demodulator, which gives a  $K$ -dimensional vector representation of the received signal. Each dimensional value represents the correlation of the received signal with the basis function specific to that dimension. This vector is then passed through a MAP or ML detector, which selects the most likely symbol sent, given the received vector.

# Chapter 3

## Mutual Information

Shannon defined the channel capacity as the maximum rate at which data could be reliably transmitted over a channel [1]. The expression for the capacity was found in terms of the signal-to-noise ratio (SNR) by using a quantity known as *mutual information*. The channel capacity was then found to be the mutual information between the channel input and output maximized over all input distributions. A Gaussian input distribution maximizes the mutual information, which leads to the expression for channel capacity.

In this chapter we will examine the mutual information between the input and output of several different channels. Section 3.1 contains an overview of mutual information as well as its related measures, the average mutual information and the mutual-information rate. In section 3.2 we will apply a modulation constraint on the channel input distribution and derive expressions for the modulation-constrained mutual-information rate. In section 3.3 we remove the constraint and instead use a Gaussian input distribution. The Gaussian distribution is used because it is capacity-achieving. Expressions for the mutual-information rate are found for both AWGN and Rayleigh fading channels and Gaussian approximations to the mutual-information rates will be introduced. Finally a brief review of the chapter's contents can be found in section 3.4.

### 3.1 Review of Mutual Information

The *mutual information* between the events  $\{X = x\}$  and  $\{Y = y\}$  is the function [24, 11]

$$i(x; y) = \log \frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} \quad (3.1)$$

where  $p_{X,Y}(x, y)$  is the joint pdf and  $p_X(x)$  and  $p_Y(y)$  are the marginal pdfs of the random variables  $X$  and  $Y$ , respectively. If the random variables  $X$  and  $Y$  are independent, then the joint pdf is equivalent to the product of the marginals. In this case the mutual information is

$$\begin{aligned} i(x; y) &= \log \frac{p_X(x)p_Y(y)}{p_X(x)p_Y(y)} \\ &= \log 1 \\ &= 0 \end{aligned}$$

However, more interesting cases involve random variables that have some correlation. The *average* mutual information is found by taking the expectation over the joint pdf  $p_{X,Y}(x, y)$

$$I(X; Y) = E[i(X; Y)]. \quad (3.2)$$

While some authors [25] use the term *mutual information* to refer to (3.2), we follow the convention of [24, 11] and reserve the term for (3.1).

When  $X$  and  $Y$  are random variables representing the input and output, respectively of a given channel, we can find a quantity known as the *ergodic channel capacity*[1]. The channel capacity is found by maximizing the average mutual information over the input distribution.

$$\mathcal{C} = \max_{p_X(x)} I(X; Y) \quad (3.3)$$

In [1], Shannon found that the distribution that maximized the average mutual information was Gaussian. For a  $K$ -dimensional vector channel, the channel capacity can be expressed as

$$\mathcal{C} = \frac{K}{2} \log \left( 1 + \frac{2\mathcal{E}_s}{KN_o} \right) \quad (3.4)$$

Later in this chapter we will examine the mutual information between a codeword which is sent over a channel, and the received vector. Each codeword can be represented by a

length- $n$  vector. For the vector channel, the average of the mutual information between an input sample vector  $\mathbf{x}$  and output sample vector  $\mathbf{y}$  is defined as the function

$$\frac{1}{n}i(\mathbf{x}; \mathbf{y}) = \frac{1}{n} \log \frac{p_{\mathbf{X}, \mathbf{Y}}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{X}}(\mathbf{x})p_{\mathbf{Y}}(\mathbf{y})}. \quad (3.5)$$

This function is called the (normalized) *mutual-information rate* in [7, 9]. In all cases explored in this thesis, the inputs to the channel,  $\mathbf{x}$ , are independent and identically distributed (i.i.d.), and there is no correlation between either fading or noise samples. For these reasons, the mutual-information rate in (3.5) can be rewritten in a more convenient expression

$$\frac{1}{n}i(\mathbf{x}; \mathbf{y}) = \frac{1}{n} \sum_{k=1}^n i(x_k; y_k), \quad (3.6)$$

with  $i(x, y)$  as defined in (3.1). To find the mutual-information rate between length- $n$  sample vectors  $\mathbf{x}$  and  $\mathbf{y}$ , we need only average  $n$  samples of the mutual information between  $x$  and  $y$ .

## 3.2 Modulation-Constrained Input

In a digital communications system, a modulation scheme is used to transmit signals across a channel. This modulation has a finite number of signals,  $M$ , with which to transmit. For simplicity, we assume that each signal has an equally likely probability of being selected. In other words, the probability that a specific symbol is transmitted is  $p(\mathbf{s}) = 1/M$ .

From (3.6), we know that the mutual-information rate is equivalent to the average of  $n$  samples of the mutual information. Keeping that in mind, we can find the mutual information for a single input symbol and output symbol, and apply the result to find the mutual-information rate.

With channel input symbol,  $\mathbf{s}$  and output symbol, we see from (2.2) that the expression for  $\mathbf{r}$  is

$$\mathbf{r} = \mathbf{s} + \mathbf{n}. \quad (3.7)$$

The mutual information between the input symbol and output symbol is

$$\begin{aligned} i(\mathbf{s}; \mathbf{r}) &= \log \frac{p(\mathbf{s}, \mathbf{r})}{p(\mathbf{s})p(\mathbf{r})} \\ &= \log \frac{p(\mathbf{s}|\mathbf{r})}{p(\mathbf{s})}. \end{aligned} \quad (3.8)$$

This simplification is a result of the definition of a conditioned pdf

$$p(\mathbf{s}|\mathbf{r}) = \frac{p(\mathbf{r}, \mathbf{s})}{p(\mathbf{r})}$$

Splitting the log and substituting the appropriate input distribution ( $p(\mathbf{s}) = 1/M$ ), we arrive at the following equation:

$$\begin{aligned} i(\mathbf{s}; \mathbf{r}) &= \log \frac{1}{p(\mathbf{s})} + \log p(\mathbf{s}|\mathbf{r}) \\ &= \log M + \log p(\mathbf{s}|\mathbf{r}) \end{aligned} \quad (3.9)$$

Note that the 2nd term of the resulting equation is the Maximum a Posteriori (MAP) decoding metric, shown in (2.15). Concentrating on this term, we will divide the probability by the sum of conditional probabilities over the entire signal set, or the equivalent of 1 (which does not numerically alter the equation). This will allow us to employ Bayes' rule to reverse the condition.

$$\begin{aligned} \log p(\mathbf{s}|\mathbf{r}) &= \log \frac{p(\mathbf{s}|\mathbf{r})}{\sum_{\mathbf{s}' \in \mathcal{S}} p(\mathbf{s}'|\mathbf{r})} \\ &= \log \frac{p(\mathbf{r}|\mathbf{s})}{\sum_{\mathbf{s}' \in \mathcal{S}} p(\mathbf{r}|\mathbf{s}')} \end{aligned} \quad (3.10)$$

The pdf of  $\mathbf{r}|\mathbf{s}$  was already listed in (2.21). Substituting this pdf:

$$\begin{aligned} \log \frac{p(\mathbf{r}|\mathbf{s})}{\sum_{\mathbf{s}' \in \mathcal{S}} p(\mathbf{r}|\mathbf{s}')} &= \log \frac{\exp \left\{ -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}\|^2 \right\}}{\sum_{\mathbf{s}' \in \mathcal{S}} \exp \left\{ -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}'\|^2 \right\}} \\ &= \log f(\mathbf{r}|\mathbf{s}) - \log \sum_{\mathbf{s}' \in \mathcal{S}} e^{\log f(\mathbf{r}|\mathbf{s}')} \end{aligned} \quad (3.11)$$

where from (2.23) we have

$$\log f(\mathbf{r}|\mathbf{s}) = -\frac{1}{N_o} \|\mathbf{r} - \mathbf{s}\|^2 \quad (3.12)$$

In (3.11), we again have a logarithm of a sum of exponentials. Using the  $\max *$  function, the initial probability can be replaced by these functions:

$$\log p(\mathbf{s}|\mathbf{r}) = \log f(\mathbf{r}|\mathbf{s}) - \max_{\mathbf{s}' \in \mathcal{S}} * \log f(\mathbf{r}|\mathbf{s}') \quad (3.13)$$

The modulation-constrained mutual information can now be expressed as

$$i(\mathbf{s}; \mathbf{r}) = \log M + \log f(\mathbf{r}|\mathbf{s}) - \max_{\mathbf{s}' \in \mathcal{S}} * \log f(\mathbf{r}|\mathbf{s}'). \quad (3.14)$$

The modulation constrained mutual-information rate is the average of  $n$  samples of the mutual information random variables given in (3.14)

$$i(\mathbf{s}; \mathbf{r}) = \log M + \frac{1}{n} \sum_{k=1}^n \log f(\mathbf{r}_k|\mathbf{s}) - \max_{\mathbf{s}' \in \mathcal{S}} * \log f(\mathbf{r}_k|\mathbf{s}'). \quad (3.15)$$

Because the mutual-information rate is defined over a codeword of finite blocklength,  $n$ , it is in itself a random variable. We can find the modulation constrained channel capacity by taking the expectation of the mutual information. There is no need to maximize over the input distribution, because it is determined by the specific choice of modulation. The expectation is

$$\begin{aligned} \mathcal{C} &= E[i(\mathbf{s}; \mathbf{r})] \\ &= \log M + E[\log p(\mathbf{s}|\mathbf{r})] \\ &= \log M + E_{p(\mathbf{s}|\mathbf{r})} \left[ \log f(\mathbf{r}|\mathbf{s}) - \max_{\mathbf{s}' \in \mathcal{S}} * \log f(\mathbf{r}|\mathbf{s}') \right] \\ &= \log M + \sum_{i=0}^{M-1} \int p(\mathbf{s}_i, \mathbf{r}) \left[ \log f(\mathbf{r}|\mathbf{s}) - \max_{\mathbf{s}' \in \mathcal{S}} * \log f(\mathbf{r}|\mathbf{s}') \right] d\mathbf{r} \end{aligned} \quad (3.16)$$

As shown in (3.16), the expectation can be expressed as a summation of integrals. The pdf can be replaced with a conditional pdf that is already known:

$$\begin{aligned} p(\mathbf{s}_i, \mathbf{r}) &= p(\mathbf{r}|\mathbf{s}_i)p(\mathbf{s}_i) \\ &= \frac{1}{M}p(\mathbf{r}|\mathbf{s}_i) \end{aligned}$$

The final equation for modulation constrained capacity is expressed as:

$$\mathcal{C} = \log M + \frac{1}{M} \sum_{i=0}^{M-1} \int p(\mathbf{r}|\mathbf{s}_i) \left[ \log f(\mathbf{r}|\mathbf{s}) - \max_{\mathbf{s}' \in \mathcal{S}'} * \log f(\mathbf{r}|\mathbf{s}') \right] d\mathbf{r} \quad (3.17)$$



Monte Carlo simulation is useful in computing the modulation constrained capacity shown in (3.17). Simulation techniques can also be applied to the modulation constrained mutual-information rate in (3.15). Because the mutual-information rate is random, it can be used to calculate information-outage probabilities by comparing the mutual-information rate samples to the desired rate.

### 3.3 Unconstrained Input

Shannon showed that a Gaussian input distribution maximizes the mutual information between a channel's input and output[1]. With a constraint imposed on the modulation, there are a finite alphabet of symbols available to transmit across the channel. A Gaussian input distribution can take on an infinite number of values, and therefore is not suitable for the modulation constrained case. With this in mind, we will now look at the unconstrained input case. First we will look at the AWGN channel case in section 3.3.1, then apply the problem to a Rayleigh fading channel in section 3.3.2.

#### 3.3.1 AWGN Channel

Using our general model for an AWGN channel we have

$$Y = X + N \quad (3.18)$$

where  $X \sim \mathcal{N}_c(0, \mathcal{E}_s)$  is the channel input and  $N \sim \mathcal{N}_c(0, N_0)$  is the noise. Again note that the input has been chosen to be Gaussian in the interest of achieving the AWGN capacity, and the output  $Y \sim \mathcal{N}_c(0, \mathcal{E}_s + N_0)$ .

Note that the distributions used are circular complex Gaussian. The probability density function (pdf) of a complex Gaussian  $Z \sim \mathcal{N}_c(m, \sigma^2)$  is given by

$$p_Z(z) = \frac{1}{\pi\sigma^2} \exp \left\{ \frac{-|z - m|^2}{\sigma^2} \right\}. \quad (3.19)$$

To find the mutual information between the channel input and output, we need the joint pdf of  $X$  and  $Y$ . However this pdf is not readily available, so we will use the definition of

conditional probability to find a different way to express the mutual information equation from (3.1) as

$$i(x; y) = \log \frac{p_{Y|X}(y|x)}{p_Y(y)}, \quad (3.20)$$

where  $p_{Y|X}(y|x)$  is the conditional pdf of  $Y$  given  $X$  which is  $\mathcal{N}_c(X, N_0)$ . Substituting the Gaussian pdf (3.19) for  $Y|X$  and  $Y$  into (3.20) yields

$$\begin{aligned} i(x; y) &= \log \left[ \frac{\mathcal{E}_s + N_o}{N_o} \exp \left\{ \frac{|y|^2}{\mathcal{E}_s + N_0} - \frac{|y - x|^2}{N_0} \right\} \right] \\ &= \log \left( 1 + \frac{\mathcal{E}_s}{N_o} \right) + \frac{|y|^2}{\mathcal{E}_s + N_0} - \frac{|y - x|^2}{N_0}. \end{aligned} \quad (3.21)$$

Remember that the mutual information given by the expression in (3.21) is a random variable. We can see from (3.21) that the mutual information random variable is the sum of a constant and a function of the random variables  $X$  and  $Y$ . We can confirm Shannon's results for the Gaussian channel by taking the expectation of the mutual information random variable.

$$\begin{aligned} I(X; Y) &= E[i(X; Y)] \\ &= \log \left( 1 + \frac{\mathcal{E}_s}{N_0} \right) + \frac{E[|Y|^2]}{\mathcal{E}_s + N_0} - \frac{E[|Y - X|^2]}{N_0}. \end{aligned} \quad (3.22)$$

Because  $E[|Y|^2] = \mathcal{E}_s + N_0$  and  $E[|Y - X|^2] = [N] = N_0$ , we are left with the original constant in (3.21)

$$I(X; Y) = \log \left( 1 + \frac{\mathcal{E}_s}{N_0} \right), \quad (3.23)$$

which is the ergodic capacity of the AWGN channel with two-dimensional signaling. This is confirmed from (3.4) by setting  $K = 2$ .

### Mutual-Information Rate

Capacity is related to the average mutual information only in the limit of infinite-length codewords. We are concerned with information-theoretic measures of performance for codewords of finite length. Let  $n$  be the length of the codeword and let the output  $\mathbf{Y} \in \mathbb{C}^n$  be

$$\mathbf{Y} = \mathbf{X} + \mathbf{N}, \quad (3.24)$$

where  $\mathbf{X}$  is a vector of independent and identically distributed (i.i.d.) variables  $X_i \sim \mathcal{N}_c(0, \mathcal{E}_s), 1 \leq i \leq n$ , and  $\mathbf{N}$  is white noise with i.i.d. elements  $N_i \sim \mathcal{N}_c(0, N_0), 1 \leq i \leq n$ .

For the vector channel, the mutual-information rate between an input sample vector  $\mathbf{x}$  and output sample vector  $\mathbf{y}$  is defined in (3.5). However, we saw for i.i.d. channel inputs and white noise, the mutual-information rate can be expressed in (3.6) as the average of  $n$  samples of the mutual information random variable in (3.21). As previously mentioned, when  $n$  is finite, the quantity  $i(\mathbf{x}; \mathbf{y})$  given by (3.6) is a random variable. Let  $Z_n$  be the mutual information between a length- $n$   $\mathbf{x}$  and length- $n$   $\mathbf{y}$ , which has form

$$Z_n = \log \left( 1 + \frac{\mathcal{E}_s}{N_0} \right) + W_n, \quad (3.25)$$

where

$$\begin{aligned} W_n &= \frac{1}{n} \sum_{k=1}^n \left( \frac{|Y_k|^2}{\mathcal{E}_s + N_0} - \frac{|Y_k - X_k|^2}{N_0} \right) \\ &= \frac{1}{n} \sum_{k=1}^n \left( \frac{|Y_k|^2}{\mathcal{E}_s + N_0} - \frac{|N_k|^2}{N_0} \right) \\ &= \frac{1}{n} \sum_{k=1}^n \left( |Y'_k|^2 - |N'_k|^2 \right), \end{aligned} \quad (3.26)$$

where  $Y'_k \sim \mathcal{N}_c(0, 1)$  and  $N'_k \sim \mathcal{N}_c(0, 1)$  are a pair of correlated complex Gaussian variables with correlation coefficient that depends on the ratio  $\mathcal{E}_s/N_0$ .

When  $n = 1$ ,  $W_n$  is a Laplacian random variable with zero mean and variance

$$\sigma_W^2 = \frac{2\mathcal{E}_s}{\mathcal{E}_s + N_0}. \quad (3.27)$$

See [8] and [9] for details of this derivation. An alternative but equivalent representation for the mutual-information rate shown in (3.25) is provided in [9] and given by

$$i(\mathbf{x}; \mathbf{y}) = \log \left( 1 + \frac{\mathcal{E}_s}{N_0} \right) + \frac{1}{n} \sqrt{\frac{\mathcal{E}_s/N_0}{\mathcal{E}_s/N_0 + 1}} \sum_{l=1}^n w_l, \quad (3.28)$$

where  $w_l, 1 \leq l, \leq n$  are i.i.d. Laplacian with zero mean and variance two.

For  $n > 1$ ,  $W_n$  is the average of  $n$  i.i.d. Laplacian random variables and is a Bessel-K random variable with pdf given by

$$f_{W_n}(w) = \frac{2^{1-n}}{\sqrt{\pi}\Gamma(n)\sigma_W} \left( \frac{\sqrt{2}|w|}{\sigma_W} \right)^{n-\frac{1}{2}} K_{n-\frac{1}{2}} \left( \frac{\sqrt{2}|w|}{\sigma_W} \right) \quad (3.29)$$

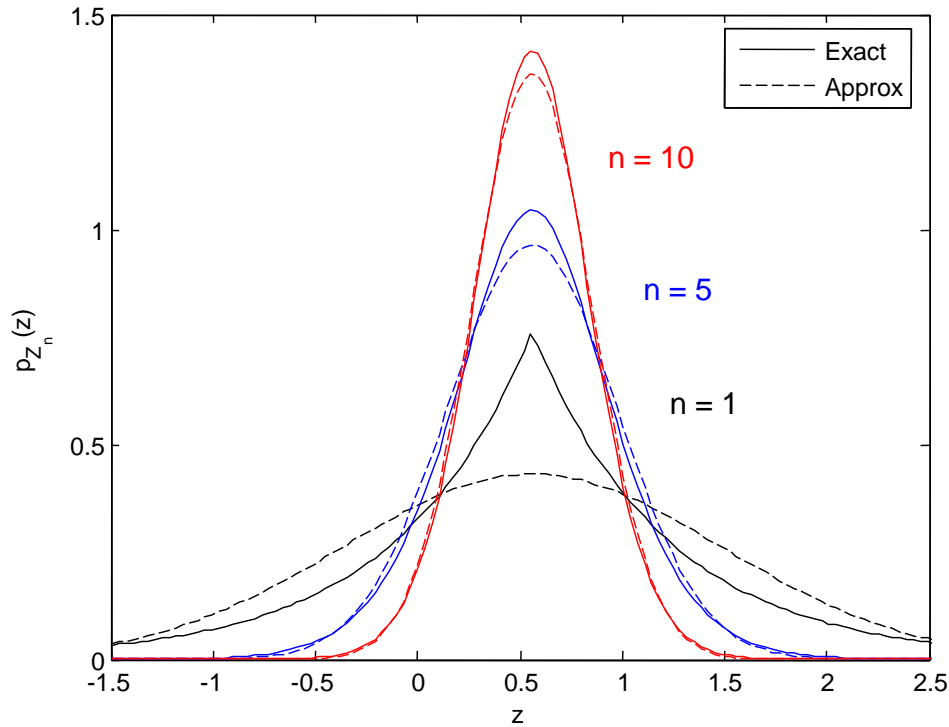


Figure 3.1: The exact Bessel-K pdf and its Gaussian approximation for  $n = \{1, 5, 10\}$  and  $\mathcal{E}_s/N_0 = 3/4$ .

where

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$$

is the gamma function and  $K_\nu(z)$  is the modified Bessel function of the second kind with index  $\nu$ . When  $n$  is an integer,  $K_{n-\frac{1}{2}}(z)$  can be expressed as

$$K_{n-\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2z}} e^{-z} \sum_{l=0}^{n-1} \frac{\Gamma(n+1)}{\Gamma(n-l)\Gamma(l+1)} (2z)^{-l}. \quad (3.30)$$

### Approximation to the Mutual-Information Rate

For large  $n$ , it is inconvenient to deal directly with the pdf of the Bessel-K random variable. However for large  $n$ , the Central Limit Theorem applies, and thus  $W_n$  may be approximated by a Gaussian variable. Let  $\tilde{W}_n$  represent the Gaussian approximation to  $W_n$ . Since  $\tilde{W}_n$  is the average of  $n$  i.i.d. variables with zero mean and variance  $\sigma_W^2$ ,  $\tilde{W}_n$  will be

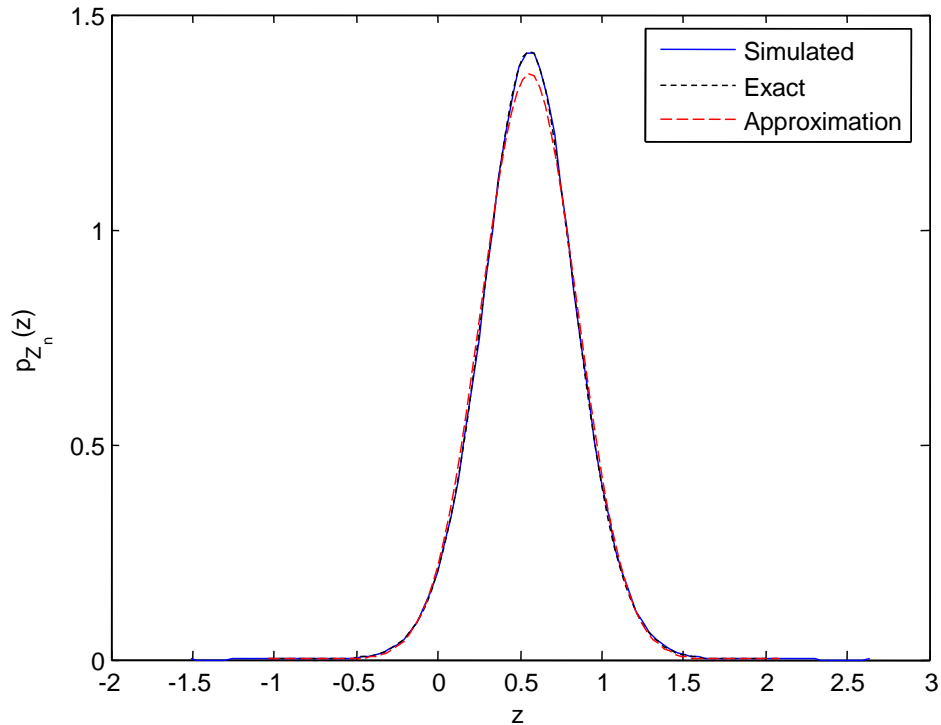


Figure 3.2: A histogram of  $10^7$  realizations of  $Z_{10}$  for  $\mathcal{E}_s/N_0 = 3/4$  along with the exact and approximate pdfs.

a scalar Gaussian with zero mean and variance  $\sigma_W^2/n$ . Since  $Z_n$  is merely  $W_n$  with shifted mean, it may be approximated with  $\tilde{Z}_n$ , where

$$\tilde{Z}_n \sim \mathcal{N}\left(\log\left(1 + \frac{\mathcal{E}_s}{N_0}\right), \frac{2\mathcal{E}_s}{n(\mathcal{E}_s + N_0)}\right). \quad (3.31)$$

Note that the variance of the Gaussian approximation in (3.31) is a function of the blocklength,  $n$ . As  $n$  approaches infinity, the variance approaches 0. At this point the Gaussian approximation is no longer random, but converges to the deterministic value of the channel capacity.

A comparison of the exact and approximate densities is shown in Fig. 3.1 for  $n = \{1, 5, 10\}$  at  $\mathcal{E}_s/N_0 = 3/4$ . The Laplacian density is apparent from the  $n = 1$  case. For  $n$  as small as 10, the exact and approximate densities are already taking on the same appearance. Further illustration of the approximation is shown in Fig. 3.2, which shows a normalized histogram curve produced by drawing  $10^7$  realizations of the mutual-information random variable using (3.21),  $n = 10$ , and  $\mathcal{E}_s/N_0 = 3/4$ . The exact and approximate densities are

superimposed on the plot to verify their correctness.

### 3.3.2 Rayleigh Fading Channel

The model for a Rayleigh fading channel is similar to that of the AWGN channel given in (3.18), only with the addition of a multiplicative fading coefficient. The model is given by

$$Y = aX + N \quad (3.32)$$

where  $X \sim \mathcal{N}_c(0, \mathcal{E}_s)$  is the channel input,  $a \sim \mathcal{N}_c(0, \sigma_a^2)$  is the fading coefficient, and  $N \sim \mathcal{N}_c(0, N_0)$  is the noise. We will assume that the fading coefficients are known to the receiver and are i.i.d. with variance  $\sigma_a^2 = 1$ . These fading coefficients are independent of the noise,  $N$  and the channel input,  $X$ .

We can maintain the complex Gaussian distribution for the channel output,  $Y$ , by conditioning on the event that the fading coefficient,  $a$ , takes on a specific value  $a$ . From [9], we see that because the fading is known to the decoder, the effective channel output is  $(y, a)$  and the mutual information can be expressed as

$$\begin{aligned} i(x; y, a) &= \log \frac{p_{y, a|x}(y, a|x)}{p_{y, a}(y, a)} \\ &= \log \frac{p_{y|x, a}(y|x, a)}{p_{y|a}(y|a)} \\ &= i(x; y|a). \end{aligned} \quad (3.33)$$

By conditioning on the fading coefficient, which is random, the conditioned channel output,  $Y|a$ , maintains its Gaussian distribution from the AWGN case and is now defined as

$$Y|a \sim \mathcal{N}_c(0, |a|^2 \mathcal{E}_s + N_0). \quad (3.34)$$

Substituting the Gaussian pdf expressions into the mutual information equation in (3.20), we now arrive at the following result

$$\begin{aligned} i(x; y|a) &= \log \left[ \frac{|a|^2 \mathcal{E}_s + N_0}{N_0} \exp \left\{ \frac{|y|^2}{|a|^2 \mathcal{E}_s + N_0} - \frac{|y - ax|^2}{N_0} \right\} \right] \\ &= \log \left( 1 + \frac{|a|^2 \mathcal{E}_s}{N_0} \right) + \frac{|y|^2}{|a|^2 \mathcal{E}_s + N_0} - \frac{|y - ax|^2}{N_0}. \end{aligned} \quad (3.35)$$

Similar to the AWGN case (3.23), we can find the ergodic channel capacity over a fading channel by taking the expectation of the mutual information in (3.35). However, we will use a more convenient form for the mutual information in (3.35), from [9]

$$i(x; y|a) = \log(1 + \lambda) + \sqrt{\frac{\lambda}{\lambda + 1}} w \quad (3.36)$$

where  $\lambda = |a|^2 \mathcal{E}_s / N_o$  is an exponential random variable representing the SNR. The mean of the exponential is known as the average SNR, and is equal to  $\mathcal{E}_s / N_o$ . Again,  $w$  is a Laplacian random variable with zero mean and variance two. The random variables  $\lambda$  and  $w$  are independent, and this independence can be utilized in the following expectation

$$\begin{aligned} I(X; Y|a) &= E[i(X; Y|a)] \\ &= E[\log(1 + \lambda)] + E\left[\sqrt{\frac{\lambda}{\lambda + 1}} w\right] \\ &= E[\log(1 + \lambda)] + E\left[\sqrt{\frac{\lambda}{\lambda + 1}}\right] E[w] \\ &= E[\log(1 + \lambda)]. \end{aligned} \quad (3.37)$$

The 2nd term in (3.37) was eliminated because the Laplacian random variable has zero mean. The pdf of an exponential random variable with mean  $E[X] = \Gamma$  takes the form

$$p_X(x) = \frac{1}{\Gamma} e^{-x/\Gamma}. \quad (3.38)$$

Substituting the pdf (3.38) into the expectation (3.37) gives

$$\begin{aligned} E[\log(1 + \lambda)] &= \frac{1}{\Gamma} \int_0^\infty \log(1 + \lambda) e^{-\lambda/\Gamma} d\lambda \\ &= \frac{e^{1/\Gamma}}{\Gamma} \int_1^\infty \log(1 + \lambda') e^{-\lambda'/\Gamma} d\lambda' \\ &= e^{1/\Gamma} E_1\left(\frac{1}{\Gamma}\right), \end{aligned} \quad (3.39)$$

where

$$E_1(z) = \int_z^\infty e^{-t} t^{-1} dt \quad (3.40)$$

is the *exponential integral function*. Again the mean of the exponential random variable associated with the SNR is  $\mathcal{E}_s / N_o$ . So the ergodic capacity over a Rayleigh fading channel

where the fading coefficients are known to the receiver is

$$\begin{aligned} \mathcal{C} &= I(X; Y | \mathbf{a}) \\ &= e^{N_o/\mathcal{E}_s} E_1 \left( \frac{N_o}{\mathcal{E}_s} \right). \end{aligned} \quad (3.41)$$

### Mutual-Information Rate

We would like to apply the results from the AWGN channel case in Section 3.3.1 to the fading channel in order to find an expression for the mutual-information rate. As in the AWGN case, we define the model for the fading channel over codewords of blocklength  $n$  as

$$\mathbf{Y} = \mathbf{a}\mathbf{X} + \mathbf{N}, \quad (3.42)$$

where  $\mathbf{X}$  is a vector of (i.i.d.) variables  $X_i \sim \mathcal{N}_c(0, \mathcal{E}_s), 1 \leq i \leq n$ ,  $\mathbf{a}$  is a vector of i.i.d. fading coefficients  $a_i \sim \mathcal{N}_c(0, 1), 1 \leq i \leq n$ , and  $\mathbf{N}$  is white noise with i.i.d. elements  $N_i \sim \mathcal{N}_c(0, N_0), 1 \leq i \leq n$ . Note the multiplication of input vector,  $\mathbf{X}$ , with the fading coefficient vector,  $\mathbf{a}$ , is an element-by-element operation.

A convenient expression for the mutual-information rate  $i(\mathbf{x}; \mathbf{y} | \mathbf{a})/n$ , conditioned on  $\{\mathbf{a} = \mathbf{a}\}$ , is provided in [9] as

$$\frac{1}{n} i(\mathbf{x}; \mathbf{y} | \mathbf{a}) = \log(1 + \lambda) + \frac{1}{n} \sqrt{\frac{\lambda}{\lambda + 1}} \sum_{l=1}^n w_l, \quad (3.43)$$

where  $\lambda = |\mathbf{a}|^2 \mathcal{E}_s / N_0$  is an exponential random variable representing the received SNR.  $\mathcal{E}_s / N_0$  is then the average SNR, or equivalently, the mean of the SNR random variable,  $\lambda$ . As previously,  $w_l, 1 \leq l, \leq n$  are i.i.d. Laplacian with zero mean and variance two. However, this expression (3.43), is given for static fading, which is defined as the case when the fading coefficient is constant over the entire codeword. For our model, the fading coefficient changes with each symbol. Therefore, we manipulate the expression in (3.43) to the form

$$\frac{1}{n} i(\mathbf{x}; \mathbf{y} | \mathbf{a}) = \frac{1}{n} \sum_{l=1}^n \left[ \log(1 + \lambda_l) + \sqrt{\frac{\lambda_l}{\lambda_l + 1}} w_l \right], \quad (3.44)$$

where we are now conditioning on the event,  $\{\mathbf{a} = \mathbf{a}\}$ , and  $\lambda_l = |a_l|^2 \mathcal{E}_s / N_0, 1 \leq l \leq n$  are exponential random variables.

Now that we have an expression for the mutual-information rate, we will derive a Gaussian approximation.



### Approximation to the Mutual-Information Rate

From the expression for the conditioned mutual-information rate in (3.44), it would be difficult to obtain an expression for the pdf. Similar to the AWGN case, we introduce a Gaussian approximation to the mutual-information rate in (3.44).

To find the parameters for the Gaussian approximation, we can find the mean and variance of the mutual information. We have already found the mean in (3.41), which is also the ergodic channel capacity. The variance can be found in a similar way by using the alternative expression for the mutual information given by (3.44),

$$\begin{aligned}
\sigma_i^2 &= E \left[ (i(x; y|a) - E[i(x; y|a)])^2 \right] \\
&= E \left[ \left( \log(1 + \lambda) + \sqrt{\frac{\lambda}{\lambda + 1}} w_1 - e^{N_o/\mathcal{E}_s} E_1 \left( \frac{N_o}{\mathcal{E}_s} \right) \right)^2 \right] \\
&= E \left[ \log^2(1 + \lambda) \right] + E \left[ \frac{\lambda}{\lambda + 1} w_1^2 \right] - e^{2N_o/\mathcal{E}_s} \left[ E_1 \left( \frac{N_o}{\mathcal{E}_s} \right) \right]^2
\end{aligned} \tag{3.45}$$

The expression for the variance in (3.45) can be found by taking advantage of the independence of the Laplacian random variable,  $w_1$ , with the exponential random variable,  $\lambda$ . Now we need to find expressions for the two expectations in (3.45).

- $E \left[ \log^2(1 + \lambda) \right]$

To find the expectation, we use the pdf for an exponential random variable (3.38) and perform the following integral

$$\begin{aligned}
E \left[ \log^2(1 + \lambda) \right] &= \int_0^\infty \log^2(1 + \lambda) \left( \frac{N_o}{\mathcal{E}_s} \right) e^{-\lambda N_o/\mathcal{E}_s} d\lambda \\
&= \int_1^\infty \log^2(\lambda') \left( \frac{N_o}{\mathcal{E}_s} \right) e^{-(\lambda'-1)N_o/\mathcal{E}_s} d\lambda' \\
&= \left( \frac{N_o}{\mathcal{E}_s} \right) e^{N_o/\mathcal{E}_s} \int_1^\infty \log^2(\lambda') e^{-\lambda' N_o/\mathcal{E}_s} d\lambda'
\end{aligned} \tag{3.46}$$

From [26], we have

$$\int_0^\infty \log^2(x) e^{-\mu x} dx = \frac{1}{\mu} \left[ \frac{\pi^2}{6} + (\mathbf{C} + \log(\mu))^2 \right], \tag{3.47}$$

where  $\mathbf{C}$ , is *Euler's constant*. The lower limit of the integral in (3.47) is less than the lower limit of the integral in (3.46), so we need to find the integral of the additional area. In Mathematica we find that

$$\int_0^1 \log^2(x) e^{-\mu x} dx = {}_2F_3([1, 1, 1], [2, 2, 2], -\mu) \quad (3.48)$$

where  ${}_pF_q$  is the *generalized Hypergeometric function*, shown in [26] in series form as

$${}_pF_q(\alpha_1, \alpha_2, \dots, \alpha_p; \beta_1, \beta_2, \dots, \beta_q; z) = \sum_{k=0}^{\infty} \frac{(\alpha_1)_k (\alpha_2)_k \dots (\alpha_p)_k}{(\beta_1)_k (\beta_2)_k \dots (\beta_q)_k} \frac{z^k}{k!}. \quad (3.49)$$

The final result for the expectation in (3.46) can be found by subtracting the expression in (3.48) from (3.47).

$$\begin{aligned} E[\log^2(1 + \lambda)] &= \left(\frac{N_o}{\mathcal{E}_s}\right) e^{N_o/\mathcal{E}_s} \int_1^{\infty} \log^2(\lambda') e^{-\lambda' N_o/\mathcal{E}_s} d\lambda' \\ &= \left(\frac{N_o}{\mathcal{E}_s}\right) e^{N_o/\mathcal{E}_s} \left[ \int_0^{\infty} \log^2(\lambda') e^{-\lambda' N_o/\mathcal{E}_s} d\lambda' - \int_0^1 \log^2(\lambda') e^{-\lambda' N_o/\mathcal{E}_s} d\lambda' \right] \\ &= e^{N_o/\mathcal{E}_s} \left[ \frac{\pi^2}{6} + \left( \mathbf{C} + \log\left(\frac{N_o}{\mathcal{E}_s}\right) \right)^2 \right] - \\ &\quad 2e^{N_o/\mathcal{E}_s} \left(\frac{N_o}{\mathcal{E}_s}\right) {}_3F_3\left([1, 1, 1], [2, 2, 2], -\frac{N_o}{\mathcal{E}_s}\right) \end{aligned} \quad (3.50)$$

- $E\left[\frac{\lambda}{\lambda+1} w_1^2\right]$

We can exploit the independence of the random variables  $\lambda$  and  $w_1$  to simplify the expectation

$$\begin{aligned} E\left[\frac{\lambda}{\lambda+1} w_1^2\right] &= E\left[\frac{\lambda}{\lambda+1}\right] E[w_1^2] \\ &= 2E\left[\frac{\lambda}{\lambda+1}\right] \end{aligned} \quad (3.51)$$

Because the Laplacian random variable,  $w_1$ , is zero mean, the second moment is equivalent to the variance, which is 2. The remaining expectation in (3.51) can be found by integrating over the exponential pdf (3.38)

$$E\left[\frac{\lambda}{\lambda+1}\right] = \int_0^{\infty} \frac{\lambda}{\lambda+1} \left(\frac{N_o}{\mathcal{E}_s}\right) e^{-\lambda N_o/\mathcal{E}_s} d\lambda. \quad (3.52)$$

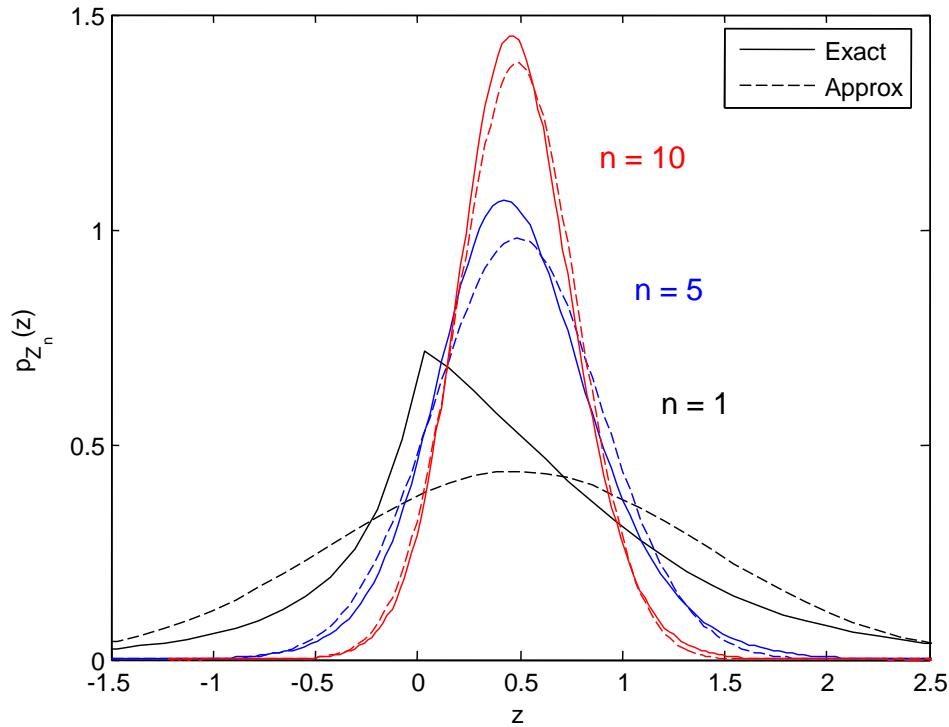


Figure 3.3: The exact pdf of the mutual-information rate in (3.44) (from  $10^7$  samples of mutual-information rate) and its Gaussian approximation for  $n = \{1, 5, 10\}$  and  $\mathcal{E}_s/N_0 = 3/4$ .

In [26], we can find an integral in the form

$$\int_0^{\infty} \frac{x}{x+1} e^{-x\mu} dx = \frac{1}{\mu} - e^{\mu} E_1(\mu). \quad (3.53)$$

Thus, our final expression for the expectation in (3.51) is

$$E \left[ \frac{\lambda}{\lambda+1} w_1^2 \right] = 2 - 2 \left( \frac{N_o}{\mathcal{E}_s} \right) e^{N_o/\mathcal{E}_s} E_1 \left( \frac{N_o}{\mathcal{E}_s} \right). \quad (3.54)$$

We now have our expression for the variance of our Gaussian approximation to the mutual-information rate.

$$\tilde{Z}_n \sim \mathcal{N} \left( e^{N_o/\mathcal{E}_s} E_1 \left( \frac{N_o}{\mathcal{E}_s} \right), \frac{\sigma_i^2}{n} \right). \quad (3.55)$$

where

$$\begin{aligned} \sigma_i^2 = & e^{N_o/\mathcal{E}_s} \left[ \frac{\pi^2}{6} + \left( \mathbf{C} + \log \left( \frac{N_o}{\mathcal{E}_s} \right) \right)^2 - 2 \left( \frac{N_o}{\mathcal{E}_s} \right) {}_3F_3 \left( [1, 1, 1], [2, 2, 2], -\frac{N_o}{\mathcal{E}_s} \right) \right] + \\ & 2 - 2 \left( \frac{N_o}{\mathcal{E}_s} \right) e^{N_o/\mathcal{E}_s} E_1 \left( \frac{N_o}{\mathcal{E}_s} \right) - e^{2N_o/\mathcal{E}_s} \left[ E_1 \left( \frac{N_o}{\mathcal{E}_s} \right) \right]^2 \end{aligned} \quad (3.56)$$

We do not know the pdf for the exact mutual-information rate with fading. However, we can view a graphical representation of the pdf by sampling the mutual-information rate random variable many times and plotting the resulting histogram of values. In Fig. 3.3 the resultant curves for  $10^7$  samples of the mutual-information rate are shown for  $n = \{1, 5, 10\}$  at  $\mathcal{E}_s/N_0 = 3/4$ , along with the corresponding pdfs of the Gaussian approximation. The shape of the  $n = 1$  case is interesting, as it appears to be a ‘bent’ version of the Laplacian pdf. Similar to the AWGN case, for  $n$  as small as 10, the exact and approximate densities are taking on the same appearance.

### 3.4 Chapter Summary

In this chapter, we reviewed the concept of mutual information and its related measures, the average mutual information and the mutual-information rate. First we applied a constraint on the input modulation, which forces a finite set of symbols on the input distribution. With a Gaussian distribution no longer applicable, we apply a simple discrete uniform distribution. With this distribution in mind, we derived expressions for the mutual information over a codeword for finite blocklength, as well as the average mutual information, as the blocklength goes to infinity. The average mutual information under modulation constraint is also known as the modulation-constrained capacity.

Next the modulation constraint was removed to allow for the use of the capacity-achieving Gaussian input distribution. With an AWGN channel, an exact distribution was found for the mutual-information rate. This distribution is given by a mean-shifted Bessel-K random variable, which is equivalent to the average of  $n$  i.i.d. Laplacian random variables. Because of the inconvenience of the exact distribution, we introduced a Gaussian approximation to the mutual-information rate. Due to the Central Limit Theorem, we know that the exact distribution will approach the Gaussian approximation as the blocklength approaches infinity. We saw graphically that the pdfs of the two distributions are similar for even very small values of blocklength for a specific value of SNR. We will see later how the distributions compare for different values of blocklength and SNR.

The same strategy was applied for the fading channel. Although an exact distribution

was not available for the mutual-information rate, we were able to graphically view the pdf by sampling the mutual information a large number of times. Similar to the AWGN channel case, a Gaussian approximation distribution was introduced for the mutual-information rate using the expressions for the mean and variance. Although the plots of the pdf of the exact distributions (found through sampling) were noticeably different from the AWGN case, as the blocklength increased, the exact distribution approached the Gaussian approximation.

## Chapter 4

# Information-Outage Probability

The capacity of a channel is equivalent to the average mutual information between the channel input and output when the input distribution is Gaussian and a power constraint is imposed on the codewords. This gives the rate at which communication can take place reliably, or with arbitrarily low probability of error. However it is assumed that the length of the input codeword goes to infinity so that the codeword can experience all channel realizations and the effects can be averaged. If the codeword blocklength is finite, then the channel capacity can no longer be used as a deterministic metric.

In this chapter we will apply information-outage concepts to the case of codewords with finite length. In section 4.1, we will take a brief look at how information-outage is used with block fading channels. In section 4.2, we will define the event of an information-outage in terms of the mutual-information rate. The information-outage probability can be found by using the CDF of the mutual-information rate. We will show that the Gaussian approximations introduced in chapter 3 are accurate representations of the mutual-information rates for sufficiently large blocklength. In section 4.3 we will compare the information-outage probability curves with frame-error rates of a capacity-approaching turbo code. Finally, a review of the chapter will be given in section 4.4.

## 4.1 Block Fading Channels

In recent literature, information-outage probability has been used primarily as a performance metric for *block fading channels*. The concept of block fading was introduced in [2] to model channels that encounter what is known as *slow fading*.

In a *fast fading* channel, the channel coherence time,  $T_c$ , is on the order of or less than the symbol period,  $T_s$ . This allows for each symbol to experience a unique multiplicative fading coefficient. However, in a slow fading channel,  $T_c \gg T_s$ , so blocks of several consecutive symbols in a codeword will undergo the same fade.

Let  $B$  be the number of blocks in a given codeword with rate  $R$ , with each block experiencing a constant fading coefficient. When  $B$  is finite, the channel is no longer ergodic, and therefore the capacity does not exist for the channel [27]. For this reason, the information-outage probability is used to describe block fading channels.

If we view the blocks independently, we can say that each block occupies  $1/B$  of the channel. We can then find the *instantaneous capacity* of each block,  $\mathcal{C}(\gamma_i)$ ,  $1 \leq i \leq B$ . By using *code-combining*, the capacity of each block adds, so the overall instantaneous capacity,  $\mathcal{C}(\gamma_1, \dots, \gamma_B)$ , is just the average of the instantaneous capacity of each block  $\mathcal{C}(\gamma_i)$  [27]. If the instantaneous capacity of the codeword is less than the per block rate  $R_B = R/B$ , then we say an outage occurs. The information-outage probability is then defined as the probability that the instantaneous capacity is less than  $R_B = R/B$

$$P_o = P[\mathcal{C}(\gamma_1, \dots, \gamma_B) < R_B] \quad (4.1)$$

From [27], we see that the information-outage probability is an information-theoretic bound on the frame-error rate (FER) in a block fading environment. In other words, no other system can perform better than the bounds determined by the information-outage probability.

## 4.2 Finite-Blocklength Codes

In Chapter 3, we were able to show expressions for the mutual-information rates of both AWGN and Rayleigh fading channels, when the input distribution is Gaussian. The mutual-

information rate was simply an average of  $n$  samples of the mutual information and was in itself a random variable. Let us think of a single realization of the mutual-information rate as the *instantaneous capacity* of the channel for that given codeword. Because the mutual-information rate is random, the instantaneous capacity will vary with each realization, and therefore reliable communication will not always be possible for a given code rate. We let an *information-outage*, be the event when the mutual-information rate (instantaneous capacity) is less than the code rate. The information-outage probability is then just the probability of an outage, or the probability that the mutual-information rate is less than the code rate.

In section 4.2.1, we will look at the information-outage probability for an AWGN channel. We have a pdf for the distribution of the mutual-information rate in Chapter 3, so we can look at the information-outage probability in terms of the cumulative distribution function (CDF). We will also compare the Gaussian approximation for various blocklengths to see when it becomes a good approximation. In section 4.2.2, we apply the same concepts to the fading channel. We did not derive an exact distribution for the mutual-information rate in fading, so we will have to rely on simulation to develop the outage probability curves. However we do have a Gaussian approximation for the fading channel, so we can use its CDF to plot the outage curves and compare them with the simulated curves for the exact distribution.

### 4.2.1 AWGN Channel

Let  $R_2 = k/n$  represent the code rate in bits per symbol, where  $k$  is the number of information bits per message and  $n$  is the number of modulated symbols per codeword. When natural logarithms are used, the rate should be converted to units of nats per symbol, i.e.  $R_e = \log(2)R_2$ .

The mutual-information rate,  $Z_n$ , for an AWGN channel is given in (3.25). The information-outage probability  $P_o$  is the probability that the mutual-information rate random variable is less than the code rate

$$\begin{aligned} P_o &= P[Z_n \leq R_e] \\ &= F_{Z_n}(R_e), \end{aligned} \tag{4.2}$$



which, as indicated by the second line, is the CDF of  $Z_n$  evaluated at the rate. This statistic is also known as the *distribution of mutual information* or the *mutual-information spectrum* [7].

From chapter 3, we have that  $Z_n$  is a mean-shifted version of the Bessel-K random variable,  $W_n$ . Using (3.25), the information-outage probability can be expressed in terms of the CDF of  $W_n$

$$\begin{aligned} P_o &= P \left[ \log \left( 1 + \frac{\mathcal{E}_s}{N_0} \right) + W_n \leq R_e \right] \\ &= F_{W_n} \left( R_e - \log \left( 1 + \frac{\mathcal{E}_s}{N_0} \right) \right). \end{aligned} \quad (4.3)$$

As derived in [9], the Bessel-K variable  $W_n$  has CDF

$$F_{W_n}(w) = 1 - \sum_{l=0}^{n-1} \frac{\Gamma(n+l)\Gamma(n-l, \sqrt{2}w/\sigma)}{\Gamma(n)\Gamma(n-l)\Gamma(l+1)} 2^{-n-l} \quad (4.4)$$

for  $w \geq 0$  and

$$F_{W_n}(w) = \sum_{l=0}^{n-1} \frac{\Gamma(n+l)\Gamma(n-l, -\sqrt{2}w/\sigma)}{\Gamma(n)\Gamma(n-l)\Gamma(l+1)} 2^{-n-l} \quad (4.5)$$

for  $w < 0$ , where

$$\Gamma(a, z) = \int_z^{\infty} t^{a-1} e^{-t} dt$$

is the upper incomplete Gamma function [28].

A Gaussian approximation to the mutual-information rate was introduced in (3.31). The CDF of the Gaussian approximation  $\tilde{Z}_n$  to  $Z_n$  can be found using the Q-function

$$F_{\tilde{Z}_n}(z) = Q \left( \frac{\log \left( 1 + \frac{\mathcal{E}_s}{N_0} \right) - z}{\sqrt{\frac{2\mathcal{E}_s}{n(\mathcal{E}_s + N_0)}}} \right). \quad (4.6)$$

A comparison of the CDF using the exact expression and the Gaussian approximation is shown in Fig. 4.1. In this curve,  $n = 20$  and  $\mathcal{E}_s/N_0 = 3/4$ . As can be seen, the two curves coincide quite closely.

Rather than fixing the SNR and varying the rate, it is often more insightful to fix the rate and plot information-outage probability as a function of SNR. In Fig. 4.2 the exact

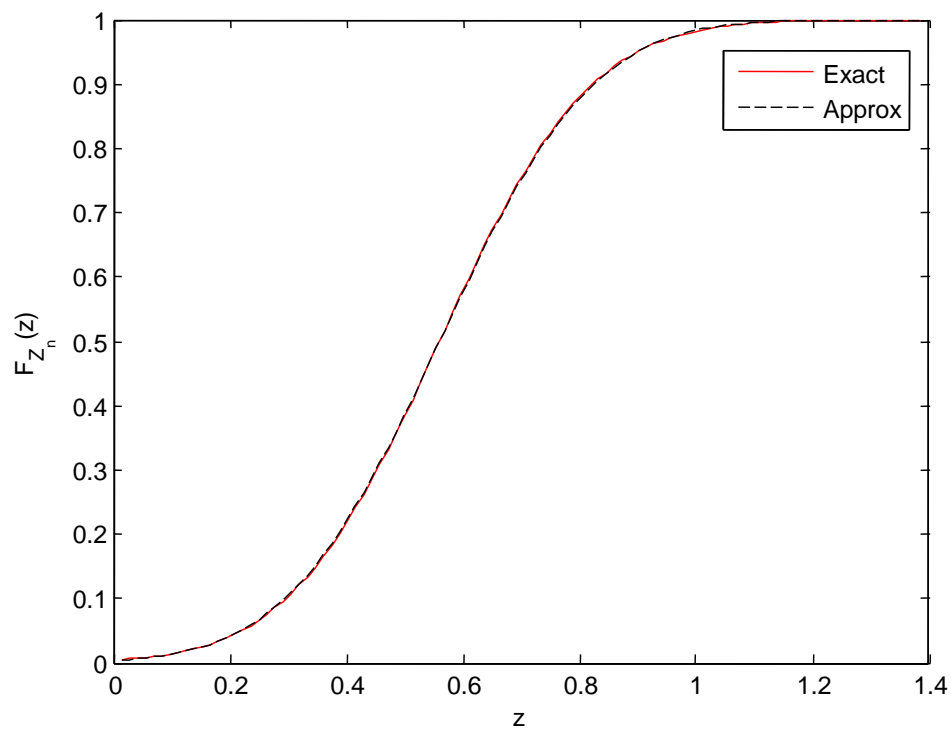


Figure 4.1: CDF of the exact and approximate mutual-information variables for  $n = 20$  and  $\mathcal{E}_s/N_0 = 3/4$ .

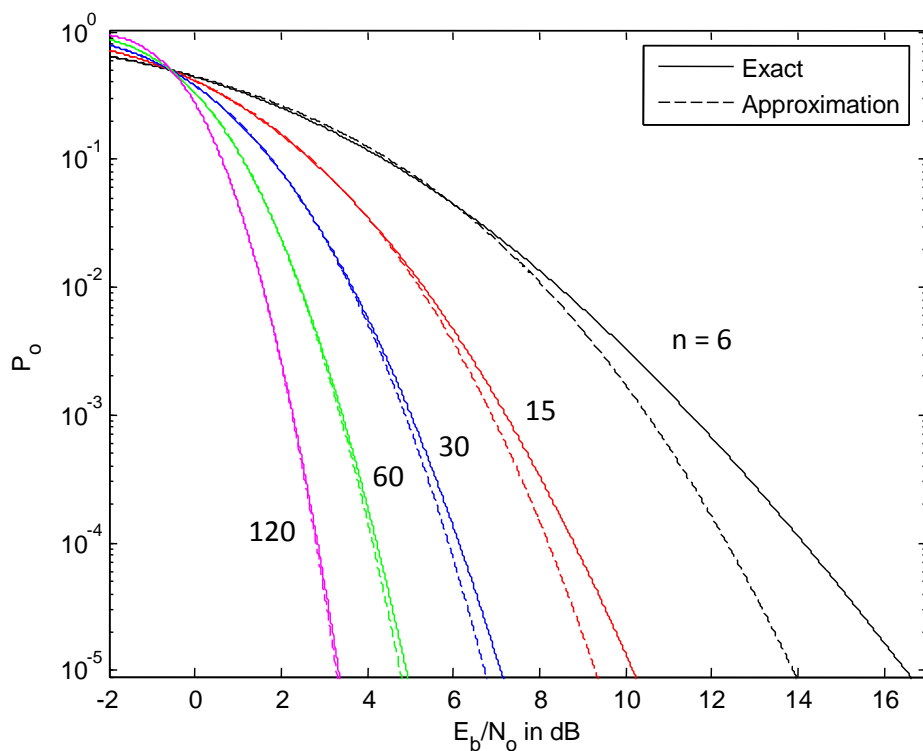


Figure 4.2: Information outage probability for AWGN vs. SNR for various blocklengths at binary rate  $R_2 = 2/3$ .

and approximate outage probabilities are shown as a function of  $\mathcal{E}_b/N_0 = (\mathcal{E}_s/N_0)/R_2$  for codewords of binary rate  $R_2 = 2/3$  over several different blocklengths. At low SNR, the Gaussian approximation seems to be a good fit, even for smaller blocklengths. However at higher SNR, the Gaussian approximation tends to diverge from the exact outage curve when  $n$  is small. As the blocklength increases, the approximation becomes a tighter fit. Based on this figure, it can be concluded that the Gaussian approximation is accurate for  $P_0 \geq 10^{-3}$  when  $n \geq 60$ .

Further insight into the tightness of the Gaussian approximation is provided by Table 4.1, which shows the minimum SNR required to achieve outage probabilities  $P_0 = \{10^{-4}, 10^{-5}\}$  and  $n = \{6, 15, 30, 60, 120\}$ . In agreement with Fig. 4.2, we see that the approximation diverges from the exact curve with decreasing  $P_0$  and  $n$ . However, even for blocklengths as small as  $n = 120$  and outage probabilities as small as  $P_0 = 10^{-5}$ , the difference in the two curves is less than 0.1 dB. Due to the Central Limit Theorem it is safe to assume that this gap

Table 4.1: Minimum SNR ( $E_b/N_o$ ) in dB Required for Various Outage Probabilities and Blocklengths - AWGN Channel (Binary Rate = 2/3)

Blocklength	N = 6		N = 15		N = 30		N = 60		N = 120	
$P_o$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$
Exact	14.14	16.51	8.8	10.18	6.14	7.11	4.23	4.92	2.85	3.35
Approximation	12.35	13.9	8.21	9.3	5.9	6.74	4.13	4.77	2.82	3.29
dB Difference	1.79	2.61	0.59	0.88	0.24	0.37	0.1	0.15	0.03	0.06

will decrease as the blocklength  $n$  increases until it becomes infinitesimally small as  $n \rightarrow \infty$ , when the AWGN channel capacity is reached. The SNR required to achieve  $P_0 = 10^{-4}$  is also shown in Fig. 4.3 for a range of blocklengths, again illustrating the tightness of the Gaussian approximation.

One complication which must be considered when using the exact mutual-information rate is the numerical instability created by the use of the Bessel-K CDF given in (4.4) and (4.5). This is due in part to the fact that for integer  $n$ ,  $\Gamma(n)$  can also be expressed as

$$\Gamma(n) = (n-1)! \quad (4.7)$$

For large  $n$  the Bessel-K CDF may be difficult to calculate with precision. In fact, using MATLAB, the Bessel-K CDF was not able to be calculated for  $n \geq 140$ . However, we have seen for sufficiently large  $n$ , the information-outage probability can be approximated by  $P_0 \approx F_{\tilde{Z}_n}(R_e)$ .

## 4.2.2 Rayleigh Fading Channel

We have an expression for the mutual-information rate of a channel with i.i.d. fading, given by (3.44). However we do not have a distribution associated with the expression that would give us access to a pdf or CDF. Because of this we will resort to simulation to determine the information-outage probability for the exact mutual-information rate.

A Gaussian approximation to the mutual-information rate in a fading channel was also introduced in (3.55). The information-outage probability using the Gaussian approximation,  $\tilde{Z}_n$  can be found by using the Q-function

$$F_{\tilde{Z}_n}(z) = Q\left(\frac{e^{N_o/\mathcal{E}_s} E_1\left(\frac{N_o}{\mathcal{E}_s}\right) - z}{\sqrt{\frac{\sigma_z^2}{n}}}\right), \quad (4.8)$$

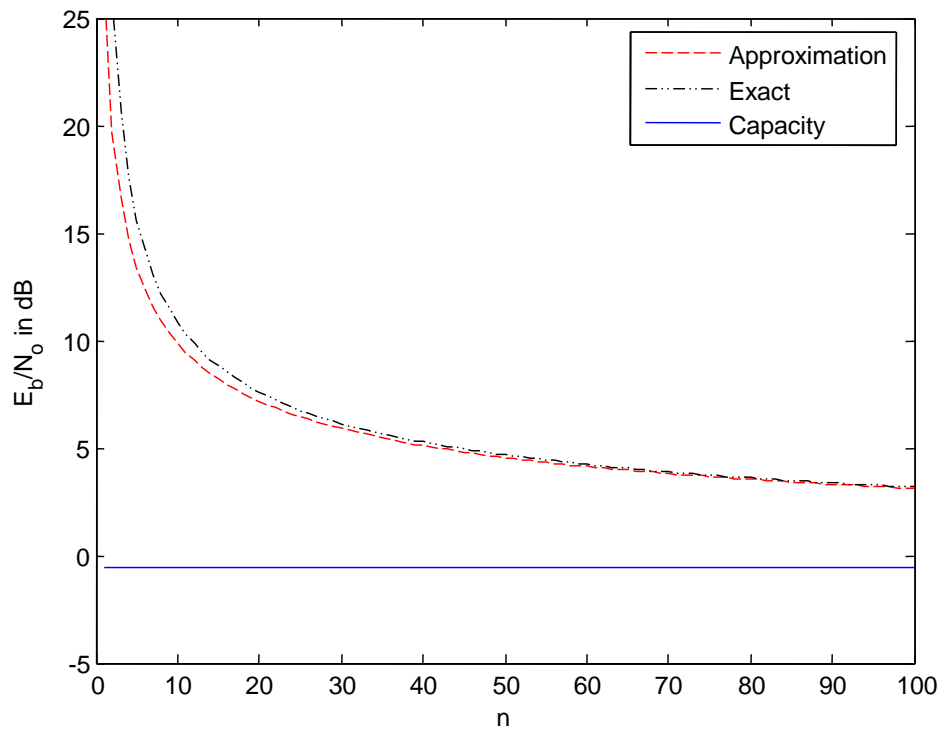


Figure 4.3: Minimum SNR ( $E_b/N_o$ ) in dB required to achieve  $P_o = 10^{-4}$  at binary rate  $R_2 = 2/3$  as a function of blocklength,  $n$ .

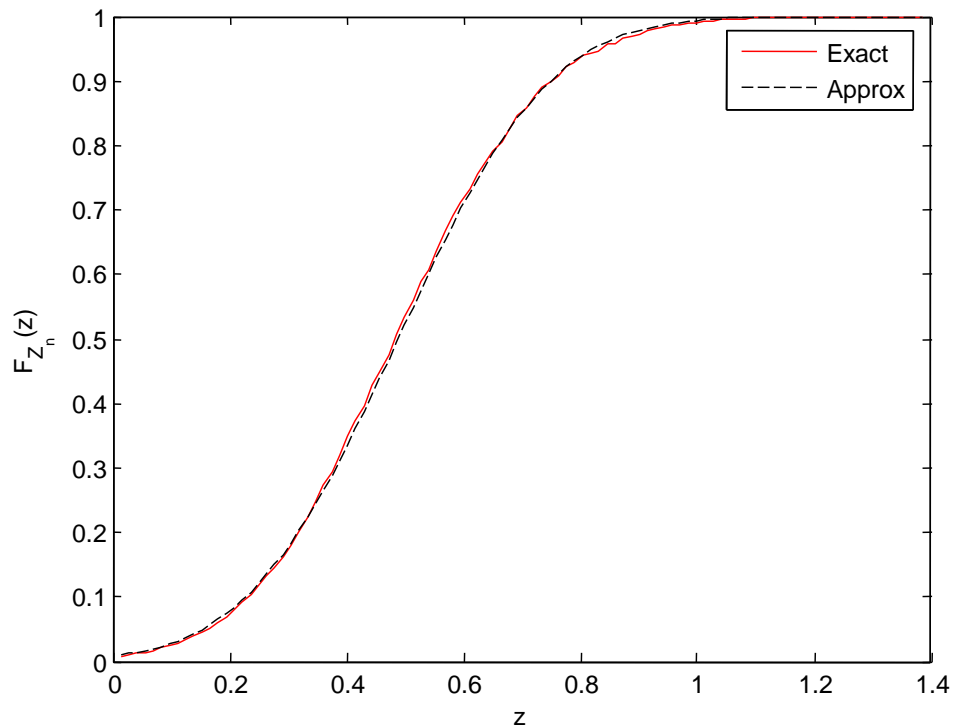


Figure 4.4: CDF of the exact and approximate mutual-information variables in Rayleigh fading for  $n = 20$  and  $\mathcal{E}_s/N_0 = 3/4$ .

where the variance,  $\sigma_i^2$ , is defined in (3.56).

The CDF of the exact mutual-information rate can be plotted by simulating the outage probability at a constant SNR and over a range of rates. A comparison of the CDF of the exact mutual-information rate (from simulation) and the Gaussian approximation is shown in Fig. 4.4. In this figure,  $n = 20$  and  $\mathcal{E}_s/N_0 = 3/4$ . Similar to the AWGN case, the curves match up very closely.

Similar to the AWGN case, we can see graphically how the approximate and exact outage probabilities compare. In Fig. 4.5 the exact and approximate outage probabilities are shown for codewords of binary rate  $R_2 = 2/3$  over the same blocklengths as were shown for the AWGN case. This figure is very interesting when compared with its counterpart for AWGN. For the AWGN plot, the approximation curves were consistently below the exact curves at higher outage probabilities, but were significantly distant at smaller  $n$ . However, in the fading plot, the approximation curves, in general, seem to hold closer to the exact curves, even for

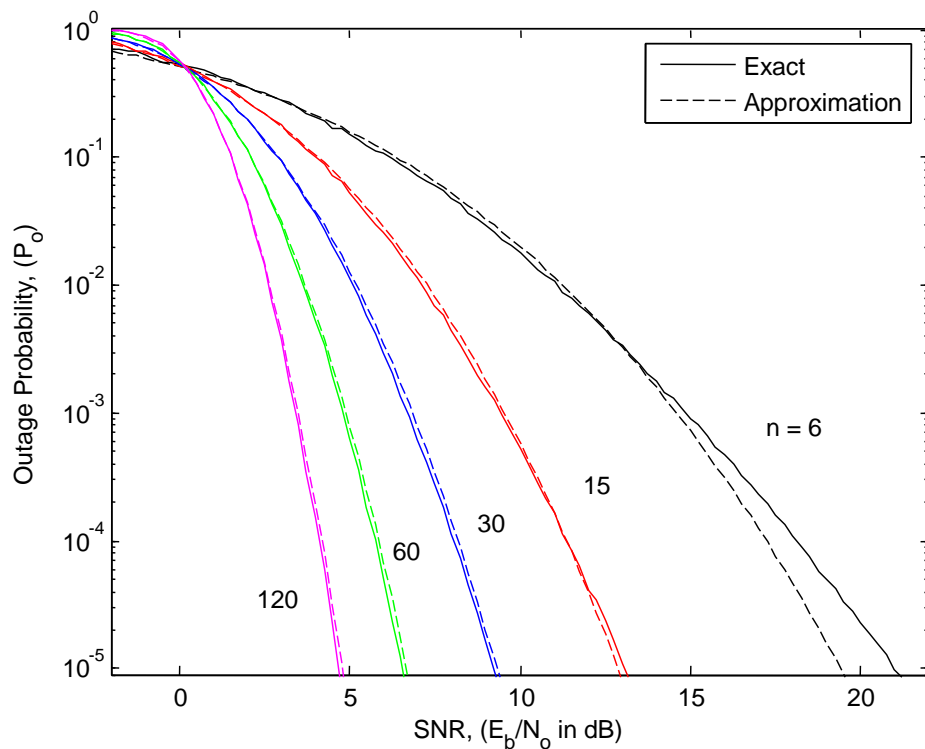


Figure 4.5: Information-outage probability for Rayleigh fading vs. SNR for various block-lengths at binary rate  $R_2 = 2/3$ .

smaller blocklength. While at smaller  $n$ , the approximation outperforms the exact curve, as the blocklength increases, the exact curve begins to perform better than the approximation. Keep in mind that the exact curves in the figure were created by simulation, and therefore will not be entirely smooth.

Table 4.2 shows the minimum SNR required to achieve outage probabilities  $P_o = \{10^{-4}, 10^{-5}\}$  and  $n = \{6, 15, 30, 60, 120\}$ . Again, we see the effects of the exact curves outperforming the approximation, as some dB differences are negative (shown in parentheses). The case of  $n = 15$  is particularly interesting. In the accompanying figure (4.5), the curves seem very close compared to AWGN. Even for blocklength this small, the exact and approximation curves are only 0.02 dB apart at  $P_o = 10^{-4}$ . It is close to this point when the approximation curves crosses the exact and begins to perform slightly better. Similar to the AWGN case, we again see that even for blocklengths as small as  $n = 120$  and outage probabilities as small as  $P_o = 10^{-5}$ , the difference in the two curves is less than 0.1 dB.

Table 4.2: Minimum SNR ( $E_b/N_o$ ) in dB Required for Various Outage Probabilities and Blocklengths - Rayleigh Fading Channel (Binary Rate = 2/3). Differences in parentheses indicate the approximation required higher SNR

Blocklength	N = 6		N = 15		N = 30		N = 60		N = 120	
$P_o$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$	$10^{-4}$	$10^{-5}$
Exact	18.13	21.02	11.35	13.04	8.06	9.23	5.76	6.58	4.12	4.72
Approximation	17.23	19.43	11.37	12.9	8.18	9.33	5.86	6.68	4.18	4.78
dB Difference	0.9	1.59	(0.02)	0.14	(0.12)	(0.1)	(0.1)	(0.1)	(0.06)	(0.06)

Therefore we can state that the Gaussian approximation can be used to accurately represent the mutual-information rate for sufficiently large  $n$  over both AWGN and Rayleigh fading channels.

### 4.3 Coded Performance

Having established the information-outage probability as a function of blocklength, it is worth discussing the meaning of the quantity. As seen in Figs. 4.2 and 4.5, the information-outage probability becomes steeper with increasing  $n$ . In the limit, as  $n \rightarrow \infty$ , the information-outage probability when plotted as a function of SNR for a fixed rate will converge to a vertical line at the location of the ergodic capacity limit for that rate. From the Shannon capacity theorem, it is well known that infinitely long codes exist that achieve arbitrarily low error rates for SNRs that are greater than the capacity limit. However, what if  $n$  is finite? Do codes exist that can achieve a frame-error rate that approaches the information-outage probability?

To provide some insight into these questions, it is worth comparing the block-error rate of some known good codes against the information-outage probability. To do this, we have simulated the performance of the turbo code from the UMTS-LTE (long term evolution) standard [29] with QPSK modulation. The binary rate of the code with QPSK modulation is  $R_2 = 2k/(3k + 12) \approx 2/3$  bits per symbol, where  $k$  is the number of data bits per frame. The LTE standard supports 188 distinct values of  $k$  ranging from 40 to 6144. In Fig. 4.6, frame-error rates of the LTE code with several values of  $k$  are compared against the corresponding information-outage probability (produced using the Gaussian approximation in (4.6)) for the AWGN channel. For the two longer blocks,  $P_0$  is approximately 0.7 dB



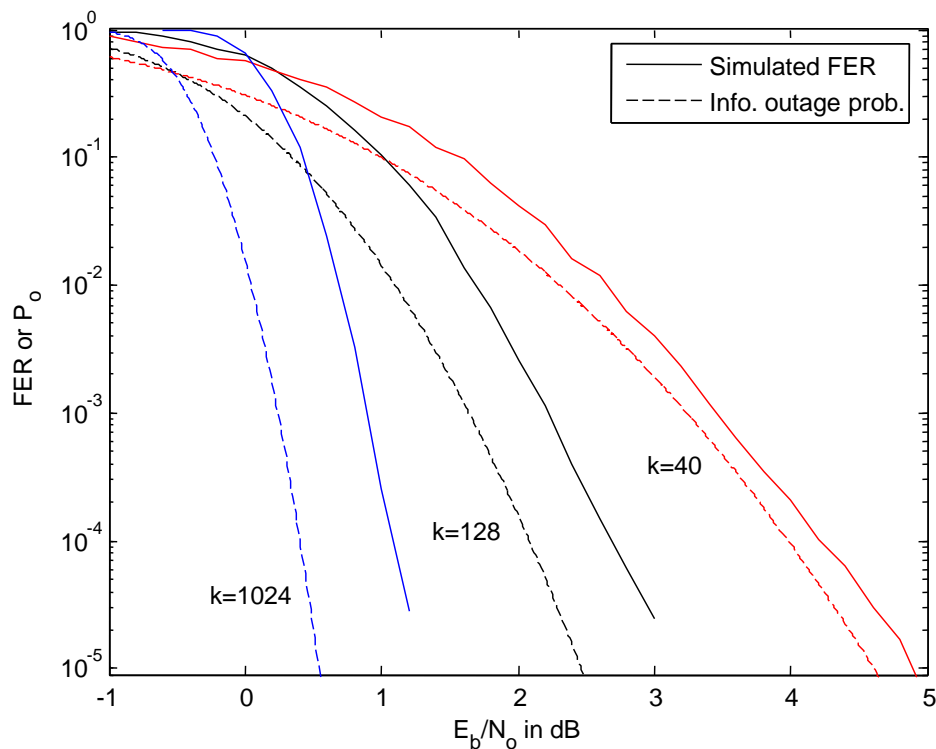


Figure 4.6: Frame-error rate (FER) of the turbo code from the LTE standard and the information-outage probability of the same blocklength over AWGN channel.  $k$  is the number of bits per block and the number of symbols per block is  $(3k + 12)/2$ .

better than the FER achieved with the actual turbo code, while for the shorter block,  $P_0$  is only 0.3 dB better. It is interesting to note that the slope of each outage probability curve is approximately the same as the corresponding turbo code FER.

Fig. 4.6 suggests that the information-outage probability is a useful quantity for predicting the performance of good finite-length codes over AWGN channels. Does the same hold true for the information-outage probability over Rayleigh fading channels? To answer this question we have also simulated the performance of the same UMTS-LTE turbo code over a fading channel.

In Fig. 4.7, frame-error rates of the LTE code over fading channels are compared against the corresponding information-outage probability (produced using the Gaussian approximation in (4.8)) for the Rayleigh fading channel. Similar to outage curves for the AWGN channel in Fig. 4.6, the information-outage curves in a fading environment share a common

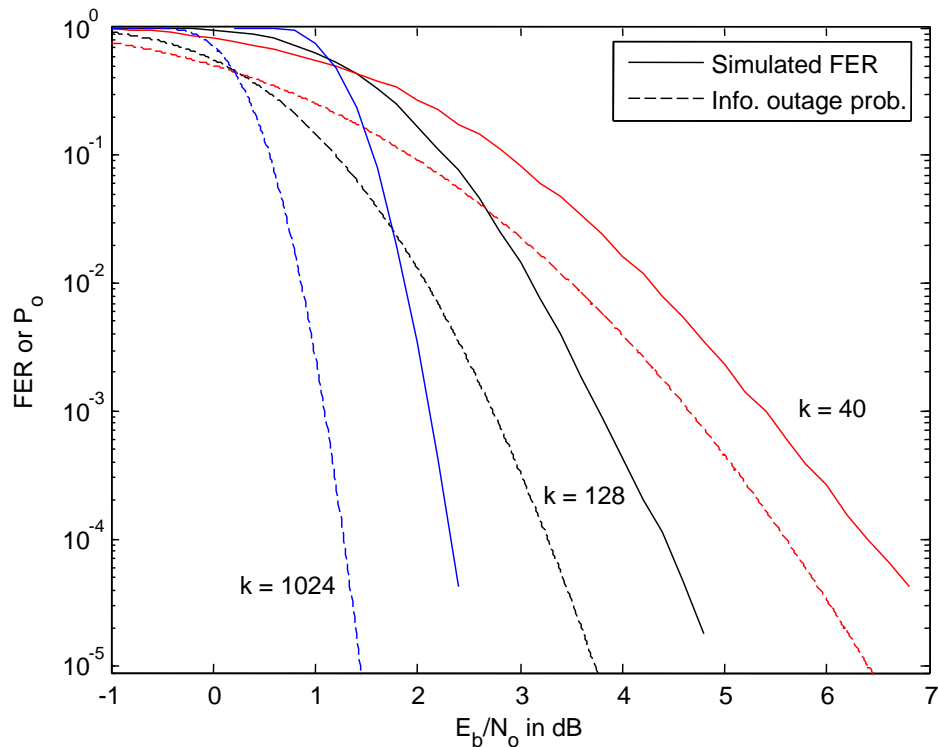


Figure 4.7: Frame-error rate (FER) of the turbo code from the LTE standard and the information-outage probability of the same blocklength over Rayleigh fading channel.  $k$  is the number of bits per block and the number of symbols per block is  $(3k + 12)/2$ .

slope with the turbo code FER curves. However the gap between the outage curves and the frame-error rate curves seems to be more constant. While the  $k = 40$  curves in fading are somewhat closer than the other two cases, they are much closer for AWGN.

## 4.4 Chapter Summary

In this chapter we discussed the quantity known as the information-outage probability. We first saw how the information-outage probability is used to analyze block fading channels. Although in the recent literature the outage probability has been primarily used to discuss block fading channels, our goal is to apply it to the case of finite blocklength over AWGN and fading channels.

In chapter 3, we provided expressions for the mutual-information rates over both AWGN and fading channels. Because the mutual-information rate is a random variable, we were able

to define the event of an information-outage based on the given code rate. The information-outage probability is the probability that the mutual-information rate is less than the code rate. For the AWGN channel, this was found using the CDF of a Bessel-K random variable. For the fading channel, no distribution was derived, so we had to rely on simulation to calculate the outage probabilities. However, for both the AWGN and fading channels, the corresponding Gaussian approximations required only the use of the Q-function to find the outage probabilities. It was then shown as the blocklength,  $n$ , increased, the Gaussian approximations became an accurate representation for the exact mutual-information rates.

The usefulness of the information-outage probability was tested by comparing the outage curves with the frame-error rate curves of the UMTS-LTE turbo code. It was shown that for both AWGN and fading channels, the curves for identical blocklengths shared a common slope. This suggests that the information-outage probability is a useful quantity for predicting the performance of capacity-approaching finite-length codes.

# Chapter 5

## Alternative Bounds

We have seen in Chapter 4, that the information-outage probability appears to be a useful quantity for predicting the performance of good finite-length codes. To further confirm its validity as a predictor of performance we will compare it with other previously derived bounds on codeword-error probability. Because the bounds were derived for the AWGN channel, we will only consider the information-outage probability for the AWGN channel.

In section 5.1, we will first compare the information-outage probability with a bound on maximal block error probability known as Feinstein's Lemma. Then we will compare the information-outage probability with two bounds derived by Shannon; namely a lower bound on codeword-error probability called the sphere-packing bound in section 5.2, and a bound on the average word-error probability named the random coding bound in section 5.3. Finally, in section 5.4, we will give a brief review of the chapter.

### 5.1 Feinstein's Lemma

Feinstein's Lemma is introduced in [30] as a bound on maximal block error probability. This bound is appropriate because of its direct relation to the mutual-information rate, which has been discussed in detail in Chapter 3. It was shown that the Gaussian approximation to the mutual-information rate was accurate for sufficiently large  $n$ . Therefore, we will use the Gaussian approximation to the mutual-information rate in the computation of Feinstein's Lemma.

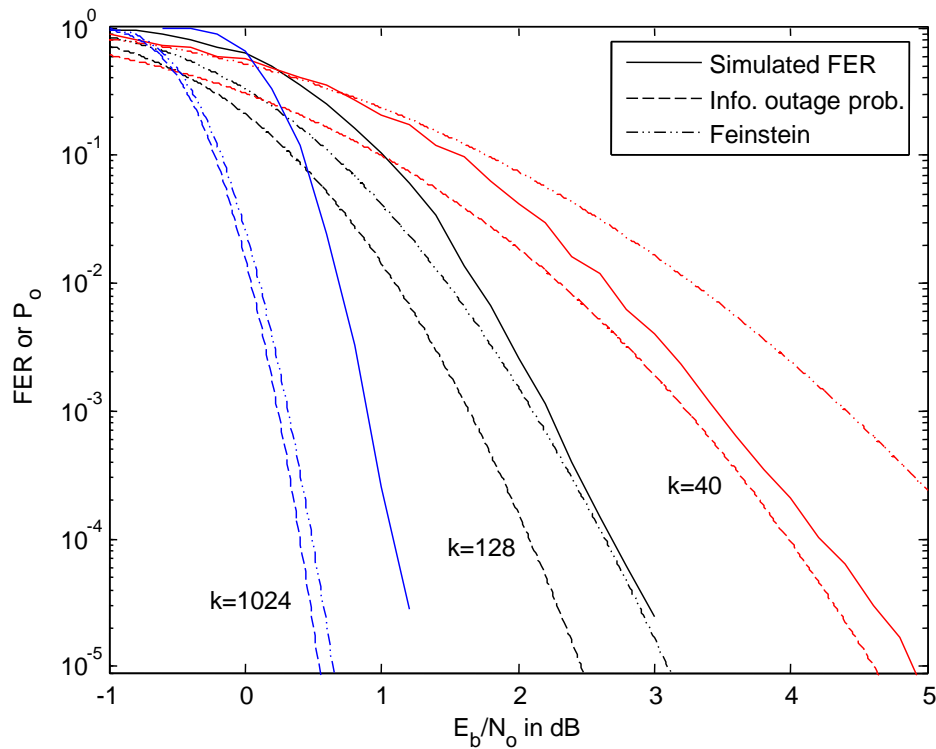


Figure 5.1: LTE Frame-error rate (FER) and IOP curves from Fig. 4.6 with the maximal error rate determined from Feinstein’s lemma superimposed.

Feinstein’s lemma [30] (see also [7, 9]) states that a length- $n$ , rate- $R_e$  code exists that can achieve a codeword error probability  $\epsilon$  that satisfies

$$\epsilon \leq P \left[ \frac{1}{n} \log \frac{p_{\mathbf{X},Y}(\mathbf{x}, \mathbf{y})}{p_{\mathbf{Y}}(\mathbf{y})p_{\mathbf{X}}(\mathbf{x})} \leq R + \gamma \right] + e^{-n\gamma}, \quad (5.1)$$

for any real valued  $\gamma : 0 < \gamma < 1$ . For the AWGN channel, this is related to the CDF of the mutual-information rate,  $Z_n$ , by

$$\epsilon \leq P_{Z_n} [R + \gamma] + e^{-n\gamma}. \quad (5.2)$$

Since  $P_{Z_n} [R + \gamma] \geq P_{Z_n} (R)$  for  $\gamma > 0$ , this bound is incrementally larger than the outage probability. Note that  $\epsilon$  is an upper bound on achievable codeword error rate performance, so codes may exist that can perform better than this bound. The bound is made sharpest by minimizing the function over  $\gamma$ .

The bound on maximal error rate computed using Feinstein’s lemma is shown in Fig. 5.1. As in Fig. 4.6, the LTE turbo code was used to produce the simulated FER curves

and the rate is  $R_2 = 2k/(3k + 12) \approx 2/3$ . As can be seen, the bound due to Feinstein's lemma is rather loose for the smallest blocksize. While the lemma predicts that there is a code with an error rate no worse than  $\epsilon$ , the LTE turbo code has actually beaten this bound at  $n = (3 * 40 + 12)/2 = 66$ . However, as  $n$  increases, the Feinstein bound approaches the information-outage probability.

## 5.2 Sphere-Packing Bound

Shannon derives a lower bound on codeword-error probability in [31], which is known as the *sphere-packing bound*. A similar bound was later developed for the discrete memoryless channel in [32], however we will concentrate on the bound in [31], which was presented for the Gaussian channel. As shown in [33], the bound takes the form

$$P_w \geq Q_n(\theta_s, A), \quad (5.3)$$

where  $P_w$  is the codeword error probability. The bound is based on the packing of a sphere in  $n$ -dimensional Euclidian space with  $N = 2^k$  circular cones, one for each possible codeword. The probability function,  $Q_n(\theta, A)$ , then represents the probability that an  $n$ -dimensional Gaussian random vector carries outside of a given cone of half-angle  $\theta$  around its mean.  $A = \sqrt{2\mathcal{E}_s/N_o}$ , is an SNR dependent amplitude. The half-angle,  $\theta_s$ , is selected so that it encloses  $1/N$  of the solid angle of the  $n$ -dimensional space.  $\theta_s$  is the solution to

$$\Omega_n(\theta_s) = \frac{1}{N}, \quad (5.4)$$

where

$$\Omega_n(\theta) = \int_0^\theta \frac{n-1}{n} \frac{\Gamma(\frac{n}{2} + 1)}{\Gamma(\frac{n+1}{2}) \sqrt{\pi}} (\sin \phi)^{n-2} d\phi, \quad (5.5)$$

is the solid-angle function. The probability function,  $Q_n(\theta, A)$ , is given in [33] as

$$Q_n(\theta, A) = \int_\theta^\pi \frac{(n-1)(\sin \phi)^{n-2}}{2^{n/2} \sqrt{\pi} \Gamma(\frac{n+1}{2})} \int_0^\infty s^{n-1} e^{-(s^2 + nA^2 - 2s\sqrt{n}A \cos \phi)/2} ds d\phi. \quad (5.6)$$

Even for small  $n$ , the expressions in (5.5) and (5.6) can lead to numerical accuracy problems. In [34], an alternate expression is given for the exact probability function in (5.6), as

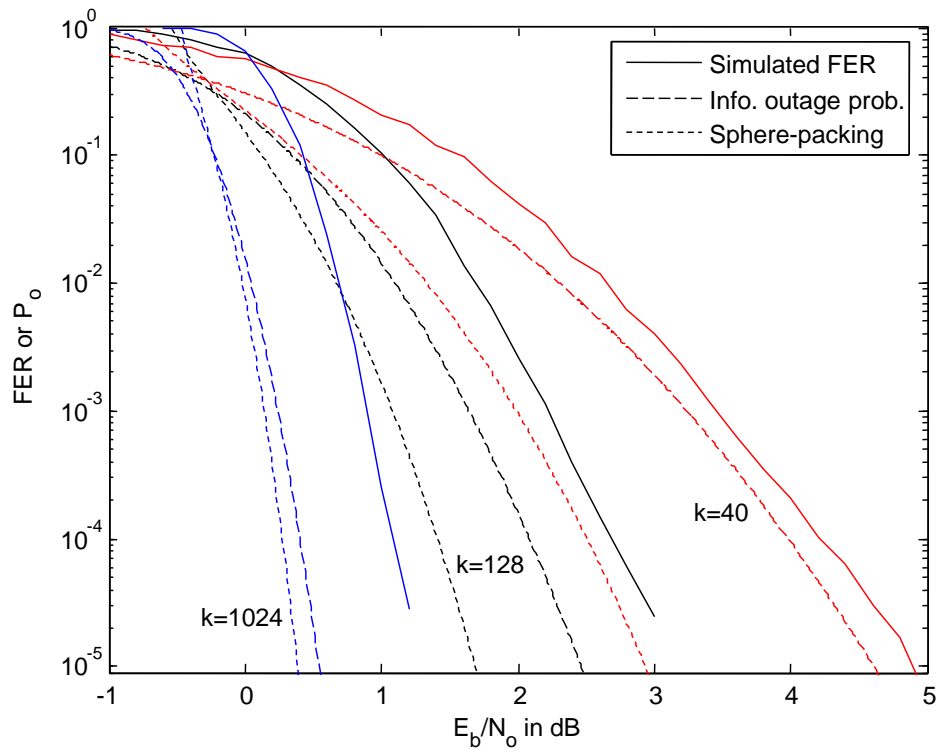


Figure 5.2: LTE Frame-error rate (FER) and IOP curves from Fig. 4.6 with sphere-packing bound superimposed.

well as a recursive algorithm that can be used to compute the probabilities while maintaining numerical accuracy. However, as  $n$  increases, the expressions become rather complicated and difficult to calculate with precision. In [35], a log-domain approach is proposed for calculating the sphere-packing bound, which alleviates the numerical precision issues that exist when the blocklength increases. For the purposes of this thesis, the blocklengths of interest are large enough that a good approximation to the sphere-packing bound would suffice.

Asymptotic approximations for (5.5) and (5.6) were also derived in [31], and are included in [33]

$$\Omega_n(\theta) \approx \frac{\Gamma\left(\frac{n}{2} + 1\right) (\sin \theta)^{n-1}}{n \Gamma\left(\frac{n+1}{2}\right) \sqrt{\pi} \cos \theta} \approx \frac{(\sin \theta)^{n-1}}{\sqrt{2\pi n} \cos \theta}, \quad (5.7)$$

and

$$Q_n(\theta, A) \approx \frac{\left[ G(\theta, A) \sin \theta e^{-(A^2 - AG(\theta, A) \cos \theta)/2} \right]^n}{\sqrt{n\pi} \sqrt{1 + G^2(\theta, A)} \sin \theta \left[ AG(\theta, A) \sin^2 \theta - \cos \theta \right]}, \quad (5.8)$$

where

$$G(\theta, A) = \frac{1}{2} \left[ A \cos \theta + \sqrt{A^2 \cos^2 \theta + 4} \right]. \quad (5.9)$$

In Chapter 4, we compared the information-outage probability using the Gaussian approximation to the mutual-information rate with frame error rates of the LTE turbo code. The plot in Fig. 4.6 shows curves for  $k = \{40, 128, 1024\}$ , or equivalently,  $n = \{66, 198, 1542\}$ . It is shown in [33], that for blocklengths as small as  $n = 60$ , the asymptotic approximation is very accurate. Therefore it is sufficient for us to exclusively use the asymptotic expression in (5.8).

In Fig. 5.2, the error rate and outage curves from Fig. 4.6 are shown with the sphere-packing bound now included. We recall that the sphere-packing bound is a lower bound on codeword error performance, as is shown in the figure. For the  $k = 40$  curves, the sphere-packing bound seems to predict that a code could perform much better than the turbo code. The outage probability curve predicts the performance of the code more closely, maintaining an almost constant gap and slope. As the blocklength increases, we can see the sphere-packing bound begin to converge with the outage curves. For the  $k = 1024$  curves, they are indeed much closer, and with similar slopes.

### 5.3 Random Coding Bound

In [31], Shannon also developed another bound on performance named the *random coding bound*. The bound is used to compute the *ensemble average* codeword-error probability,  $\bar{P}_w$ . The  $N = 2^k$  codewords of the codebook are selected at random. The ensemble average codeword-error probability is then averaged over the ensemble of all possible  $(n, k)$  spherical codes. The exact expression for the average codeword-error probability,  $\bar{P}_w$ , is given in [33] as

$$\begin{aligned} \bar{P}_w &= \int_0^\pi \left\{ 1 - [1 - \Omega_n(\theta)]^{N-1} \right\} \left[ \frac{\partial Q_n(\theta, A)}{\partial \theta} \right] d\theta \\ &= \int_0^\pi \left\{ 1 - [1 - \Omega_n(\theta)]^{N-1} \right\} \frac{(n-1)(\sin \theta)^{n-2}}{2^{n/2} \sqrt{\pi} \Gamma\left(\frac{n+1}{2}\right)} \int_0^\infty s^{n-1} e^{-(s^2 + nA^2 - 2s\sqrt{n}A \cos \theta)/2} ds d\theta \end{aligned} \quad (5.10)$$



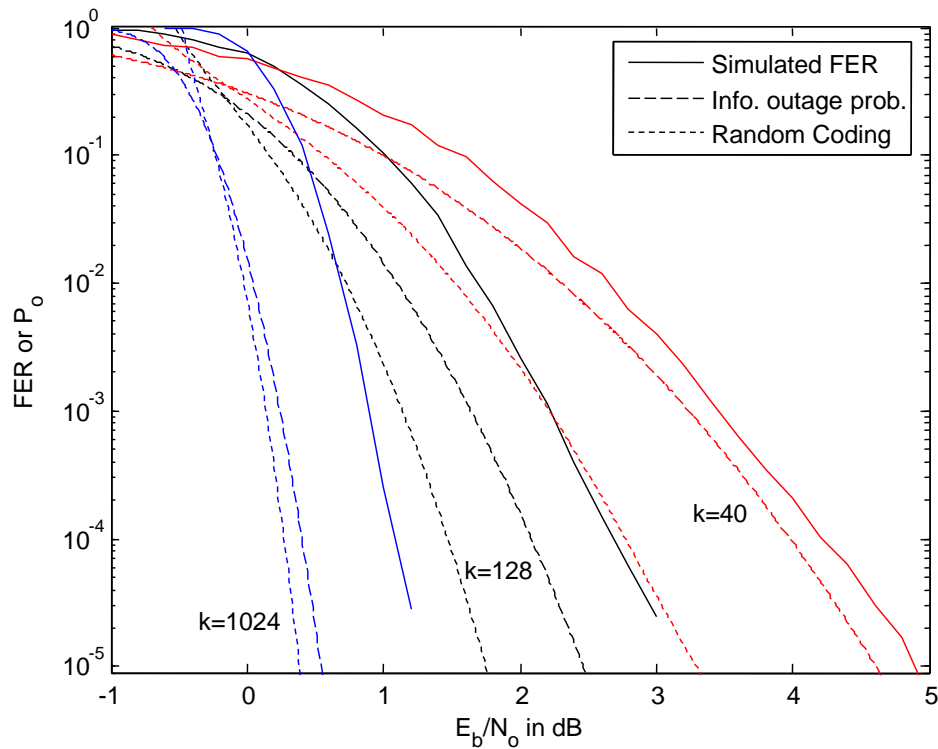


Figure 5.3: LTE Frame-error rate (FER) and IOP curves from Fig. 4.6 with random coding bound superimposed.

where  $\Omega_n(\theta)$  is the solid-angle function defined in (5.5), and  $Q_n(\theta, A)$  is the probability function defined in (5.6). Note that the limit of integration for the solid-angle function,  $\Omega_n(\theta)$ , is also the variable of integration in the expression for the average word-error probability in (5.10). Needless to say, this expression is also difficult to calculate while maintaining precision.

An asymptotic approximation for  $\bar{P}_w$  is also derived in [31], which is expressed as a function of the asymptotic expression for the probability function in (5.8)

$$\bar{P}_w \approx Q_n(\theta_s, A) \left( 1 + \frac{AG(\theta_s, A) \sin^2(\theta_s) - \cos(\theta_s)}{2 \cos(\theta_s) - AG(\theta_s, A) \sin^2(\theta_s)} \right) \quad (5.11)$$

In Fig. 5.3, we now compare the information-outage probability and LTE turbo code frame error rates with the random coding bound. As shown in the figure, the curves created by the random coding bound look similar to the curves generated by the sphere-packing bound shown in Fig. 5.2. The only difference being that the random coding curves are closer to the outage and frame error rate curves for lower  $n$ . Similar to Feinstein's Lemma

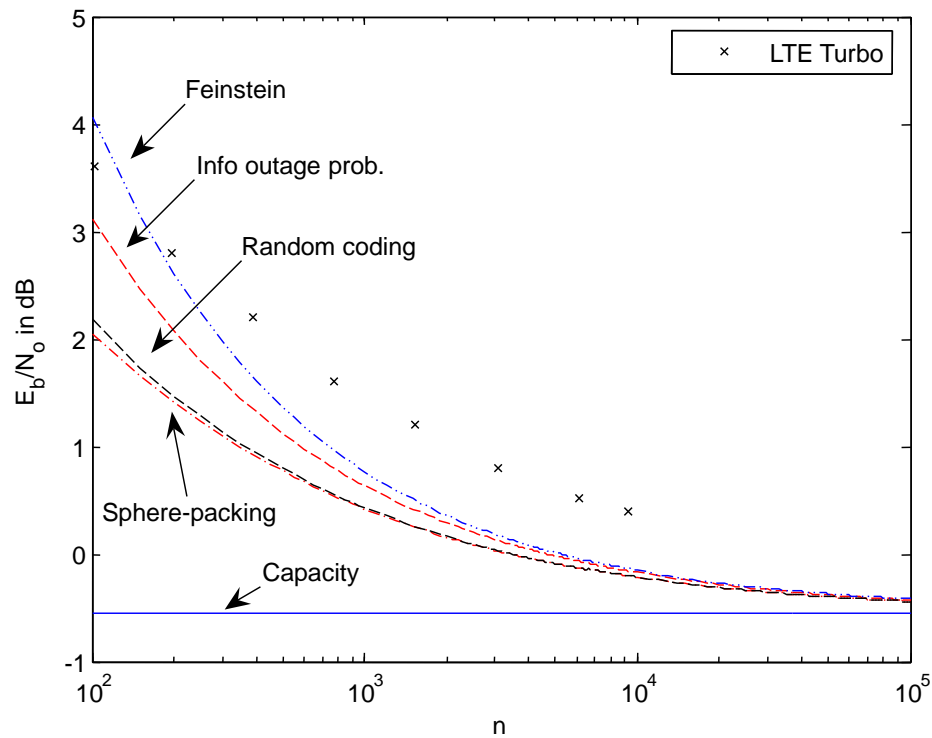


Figure 5.4: Minimum SNR ( $E_b/N_o$ ) in dB required to achieve  $P_o = 10^{-4}$  at binary rate  $R_2 = 2/3$  as a function of blocklength,  $n$ .

and the sphere-packing bound, the random coding bound appears to converge with the information-outage probability curve as  $n$  increases.

We have seen in Figures 5.1, 5.2 and 5.3 that each of the bounds appears to converge with the information-outage probability curves as the blocklength increases. As the blocklength approaches infinity, the information-outage probability when plotted as a function of SNR for a fixed rate will converge to a vertical line at the SNR required to achieve that rate. We can verify if the bounds do in fact converge with the information-outage probability by looking at the SNR required to achieve a given error rate as a function of blocklength.

The results are shown in Fig. 5.4. As shown in the figure, each bound converges to the same SNR required to achieve the binary rate of  $R_2 = 2/3$ , which is shown as the capacity curve. Also shown in the figure are the SNRs required for various lengths of the LTE turbo code. The sphere-packing and random coding bounds appear to share the same slope with the turbo code results for lower blocklengths, while the information-outage curve

seems to share the slope for larger blocklengths. As seen earlier, the turbo code outperforms Feinstein's Lemma for small  $n$ .

## 5.4 Chapter Summary

In this chapter we have compared the outage curves generated by the information-outage probability to previously derived bounds on codeword-error rate. In each case, the information-outage probability provided a more accurate prediction of the performance of the turbo code.

While the actual performance of the information-outage probability relative to the other bounds would seem to be enough to warrant its use in the stead of the bounds, we can also speak for its simplicity. The calculation of the information-outage probability requires only the use of the Q-function. Conversely, the sphere-packing and random coding bounds require integration of complicated expressions, often containing arguments which are raised to powers on the order of the blocklength,  $n$ . Feinstein's Lemma is directly based on the mutual-information rate. We saw how Feinstein's Lemma can be expressed in terms of the CDF of the mutual-information rate, which is used to calculate the information-outage probability. However, it also requires the minimization of the function over an arbitrary constant to make the bound tightest.

The information-outage probability offers the closest prediction (of the bounds presented in this chapter) of performance of good finite-length codes, while arguably being the easiest to calculate. Thus we have reaffirmed the information-outage probability's utility as a performance limit for moderate blocklengths (e.g.  $n > 1000$ , which is typical for practical systems).

# Chapter 6

## Conclusion

### 6.1 Summary

Information-outage probability has been shown in recent literature to be a useful metric when analyzing block-fading channels. It is necessary in this case because of the static nature of the fading coefficient over the duration of individual blocks. In information-outage analysis of a block fading channel, the mutual information of each block is set equal to the Shannon channel capacity ( $\log(1 + SNR)$ ), for the SNR experienced by each block. The instantaneous capacity is then found by averaging the mutual information of each block. However, there is arguably a contradiction in the assumptions behind the information-outage analysis of block-fading channels. Each block must be short enough to span only a single coherence interval, yet long enough to achieve Shannon capacity.

In this thesis, we applied information-outage probability to the case of codes with finite-length. The information-outage probability is simply the probability that the mutual-information rate is less than the code rate. For infinitely long codewords transmitted over an ergodic channel, the mutual-information rate is deterministic, and is equivalent to the capacity of the channel. However, when the length of the codewords is finite, the mutual-information rate is a random variable, and reliable communications is not guaranteed for a specific rate. We use the instantaneous capacity to represent a sample of the mutual-information rate, or the rate at which reliable communications can occur, for each codeword. If the instantaneous capacity is lower than the code rate, then reliable communication is

not possible, and we say an information outage occurs. To analyze the information-outage probability for a given channel, we must first determine the mutual-information rate random variable.

The mutual-information rate was found for both the AWGN and Rayleigh fading channels. With the AWGN channel, an exact distribution is known for the mutual information rate. Because the exact distribution is cumbersome, we introduced a Gaussian approximation to the mutual-information rate. Due to the Central Limit Theorem, we know that the exact distribution will approach the Gaussian approximation as the blocklength approaches infinity. We saw graphically that the pdfs of the two distributions are similar for even very small values of blocklength for a specific value of SNR.

For the fading channel, an exact distribution cannot be readily found for the mutual-information rate. However, we were able to estimate the pdf by sampling the mutual information a large number of times and compiling the results into a histogram. Similar to the AWGN channel case, a Gaussian approximation was introduced for the mutual-information rate using the expressions for the mean and variance. Although the pdf of the exact distributions were noticeably different from the AWGN case, as the blocklength increased, the exact distribution approached the Gaussian approximation.

At this point, we have an exact distribution for the mutual-information rate in AWGN, which is a mean-shifted Bessel-K random variable. The exact distribution in fading is not known. We also have an accurate Gaussian approximation for both distributions. When computing the information-outage probability, it is convenient to know the distribution of the mutual-information rate. This is due to the fact that the information-outage probability is equivalent to the CDF of the mutual-information rate evaluated at the code rate. Because there is no known distribution available for the exact mutual-information rate with fading, we had to use simulation to compute the outage probabilities. However, even with the Bessel-K CDF available for AWGN, we saw that it was inconvenient to manipulate, even for small  $n$ . Due to numerical stability issues associated with the complexity of the CDF, it was also unable to be computed for  $n \geq 140$ .

However, for both the AWGN and fading channels, the corresponding Gaussian approximations required only the use of the Q-function to find the outage probabilities. It was

then shown as the blocklength,  $n$ , increased, the Gaussian approximations became an accurate representation for the exact mutual-information rates. Because of the accuracy of the approximation for moderately small blocklength, we used it exclusively to calculate the information-outage probability in this thesis (except in cases where approximation was compared with exact).

The usefulness of the information-outage probability was tested by comparing the outage curves with the frame-error rate curves of the UMTS-LTE turbo code. It was shown that for both AWGN and fading channels, the curves for identical blocklengths shared a common slope. While this suggested that the information-outage probability is a useful quantity for predicting the performance of capacity-approaching finite-length codes, we further tested its utility by comparing it with some other previously derived bounds on codeword error probability.

The information-outage probability was compared with three other bounds: Feinstein's Lemma, the sphere-packing bound, and the random coding bound. In each case, the information-outage probability provided a more accurate prediction of the performance of the turbo code. This can be stated by considering that the outage curves consistently matched the slopes of the frame error rate curves of the turbo code, while also maintaining a more consistent horizontal offset.

The information-outage probability offers the closest prediction (compared with the other bounds presented in this thesis) of performance of good finite-length codes, while arguably being the easiest to calculate. Thus we conclude that the information-outage probability is a useful, yet simple, predictor of performance for coded systems of moderate blocklengths.

## 6.2 Future Research

We now suggest possible ideas for continuing or branching from the research from this thesis.

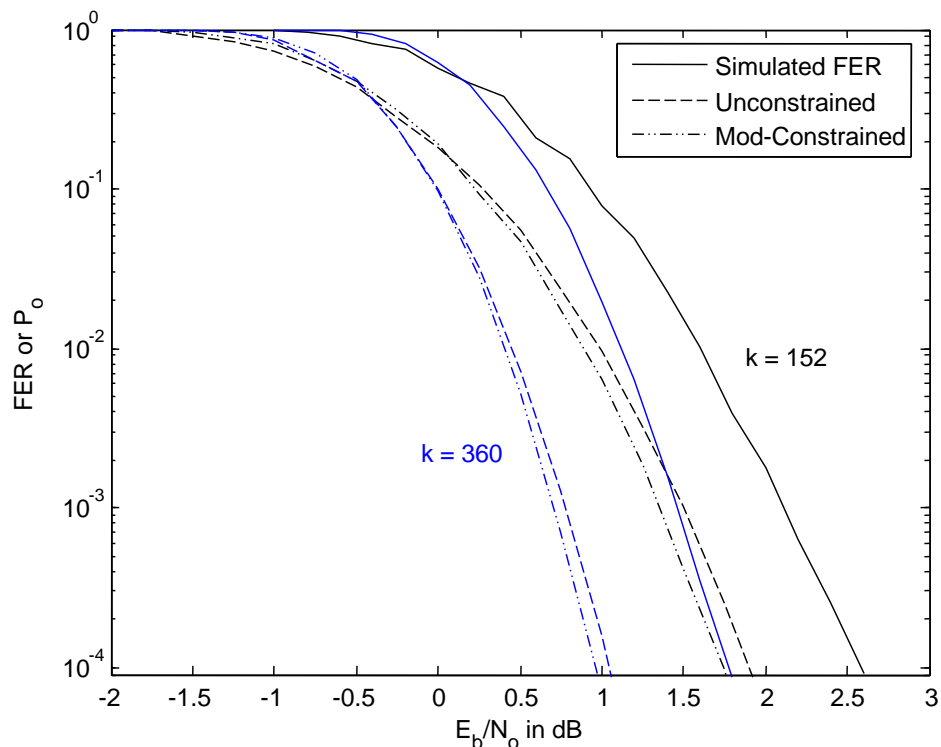


Figure 6.1: Information-outage probability of unconstrained Gaussian input and modulation-constrained input for AWGN channel. Frame error rates of turbo code included for reference with outage curves.

### Bridging the Gap of Coded Performance

We have seen that the information-outage curves share a common slope with the frame-error rate curves for the LTE turbo code for both AWGN and fading channels. While the slopes were similar, there existed a horizontal offset between the curves. If we make the assumption that the outage curves are optimal, then a combination of sub-optimal parameters must have been imposed on the coded system to create the horizontal gap in SNR. These would include a constraint on the modulation, which includes a constraint on the input distribution, and the actual channel coding technique which is used. So can we quantify the decrease in performance caused by imposing these constraints in terms of the additional SNR required to achieve a given outage probability or frame-error rate?

We saw earlier an expression for the modulation-constrained mutual-information rate in Chapter 3. This expression can be used to compute the information-outage probability in

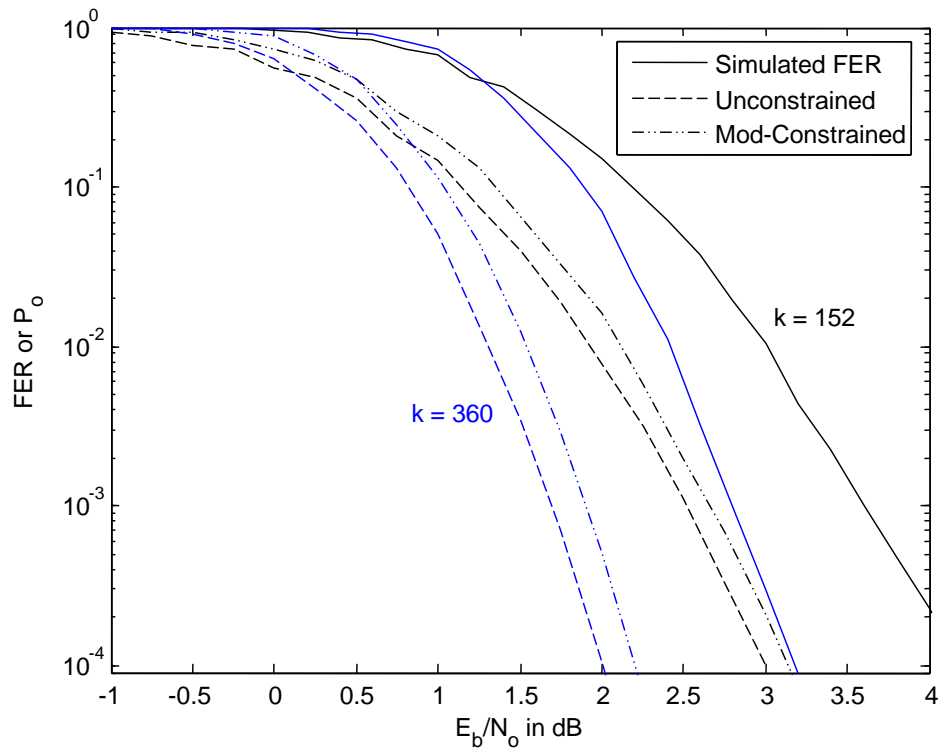


Figure 6.2: Information-outage probability of unconstrained Gaussian input and modulation-constrained input for Rayleigh fading channel. Frame error rates of turbo code included for reference with outage curves.

the same way we used the expressions for the mutual-information rate for unconstrained Gaussian input. Without a distribution, we rely on simulation to do so.

Figure 6.1 shows the outage curves for unconstrained Gaussian input and modulation-constrained input for two blocklengths over AWGN. The modulation-constrained outage curves were generated with QPSK modulation. Strangely, the modulation-constrained information-outage actually outperforms the unconstrained Gaussian input. It is unclear as to why this is the case. Perhaps this suggests that our earlier assumption that the information-outage with unconstrained Gaussian input is optimal is incorrect.

If we move to a fading environment, the results are different. Figure 6.2 shows the same outage curves from Fig. 6.1, only this time simulated over a fading channel. According to the figure, the outage curves with unconstrained input are now outperforming the modulation-constrained curves. This is what was expected with the AWGN curves. Perhaps the intro-



duction of a fading coefficient would allow for our earlier assumption to hold. Regardless, there is still a significant gap between the outage probability with modulation-constraint and the frame error rate curves of the LTE turbo code.

If the offset between the information-outage curves and the frame error rates of a given code can be better quantified or estimated, then the information-outage curves could be shifted to better predict the performance of an actual system. This remains a topic for future research.

# References

- [1] C. E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.
- [2] L. Ozarow, S. Shamai, and A. D. Wyner, “Information theoretic considerations for cellular mobile radio,” *IEEE Trans. Veh. Tech.*, vol. 43, pp. 359–378, May 1994.
- [3] R. Knopp and P. A. Humblet, “On coding for block fading channels,” *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 189–205, Jan. 2000.
- [4] A. Guillén i Fàbregas and G. Caire, “Coded modulation in the block-fading channel: Coding theorems and code construction,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 262–271, Jan. 2006.
- [5] G. Kaplan and S. Shamai, “Error exponents and outage probabilities for the block-fading gaussian channel,” *IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 329–334, 23-25 Sep 1991.
- [6] S. Verdù and T. S. Han, “A general formula for channel capacity,” *IEEE Trans. Inform. Theory*, vol. 40, pp. 1147–1157, July 1994.
- [7] T. S. Han, *Information Spectrum Methods in Information Theory*, Springer, 2003.
- [8] M. S. Pinsker, *Information and Information Stability of Random Variables and Processes*, Holiday-Day, 1964.
- [9] J. N. Laneman, “On the distribution of mutual information,” in *Proc. Workshop on Information Theory and its Applications*, San Diego, CA, Feb. 2006.
- [10] D. S. Buckingham and M. C. Valenti, “The information-outage probability of finite-length codes over AWGN channels,” in *Proc. Conf. on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2008.
- [11] J. Proakis, *Digital Communications*, McGraw-Hill, Inc., New York, NY, fourth edition, 2001.
- [12] B. P. Lathi, *Modern Digital and Analog Communication Systems*, Oxford University Press, third edition, 1998.

- [13] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall PTR, Upper Saddle River, NJ, second edition, 2002.
- [14] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, Addison-Wesley, second edition, 1994.
- [15] Rodger E. Ziermer and Roger L. Peterson, *Introduction to Digital Communications*, Prentice Hall, second edition, 2001.
- [16] David A. Huffman, "A method for the construction of minimum-redundancy codes," *Proceedings of the I.R.E.*, vol. 40, pp. 1098–1101, Sept. 1952.
- [17] Shu Lin and Jr. Daniel J. Costello, *Error Control Coding*, Prentice Hall, second edition, 2004.
- [18] I. S. Reed and G. Solomon, "Polynomial codes over certain fields," *J. Soc. Ind. Appl. Math.*, vol. 8, pp. 300–304, June 1960.
- [19] Y. Huang and J. A. Ritcey, "Optimal constellation labeling for iteratively decoded bit-interleaved space-time coded modulation," *IEEE Trans. Inform. Theory*, vol. 51, pp. 1865–1871, May 2005.
- [20] M. C. Valenti, R. Doppalapudi, and D. Torrieri, "A genetic algorithm for designing constellations with low error floors," in *Proc. Conf. on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2008.
- [21] G. Caire, G. Taricco, and E. Biglieri, "Bit-interleaved coded modulation," *IEEE Trans. Inform. Theory*, vol. 44, pp. 927–946, May 1998.
- [22] Andrea Goldsmith, *Wireless Communications*, Cambridge University Press, 2005.
- [23] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 260–264, Feb. 1998.
- [24] R. G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.
- [26] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, Academic Press, seventh edition, 2007.
- [27] Tarik Khaled Ghanim, "Analysis of hybrid-arq based relaying protocols under modulation constraints," M.S. thesis, West Virginia University, 2006.
- [28] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover, 1972.
- [29] Third Generation Partnership Project, "Evolved universal terrestrial radio access (E-UTRA): Multiplexing and channel coding," *3GPP TS 36.212 version 8.0.0*, Sept. 12 2007.

- [30] A. Feinstein, “A new basic theorem of information theory,” *IRE Trans. Info. Thy.*, vol. 4, pp. 2–22, Sept. 1954.
- [31] C. E. Shannon, “Probability of error for optimal codes in a gaussian channel,” *Bell Syst. Tech. J.*, vol. 38, pp. 611–656, 1959.
- [32] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels,” *Inform. Contr.*, vol. 10, no. 1, pp. 65–103, Feb. 1967.
- [33] S. Dolinar, D. Divsalar, and F. Pollara, “Code performance as a function of block size,” Tech. Rep., JPL TDA Progress Report, May 1998.
- [34] Antoine Valembois and Marc P.C. Fossorier, “Sphere-packing bounds revisited for moderate block lengths,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 2998–3014, Dec. 2004.
- [35] G. Wiechman and I. Sason, “An improved sphere-packing bound for finite-length codes on symmetric memoryless channels,” *to appear in IEEE Trans. Inform. Theory*, vol. 54, no. 5, May 2008.