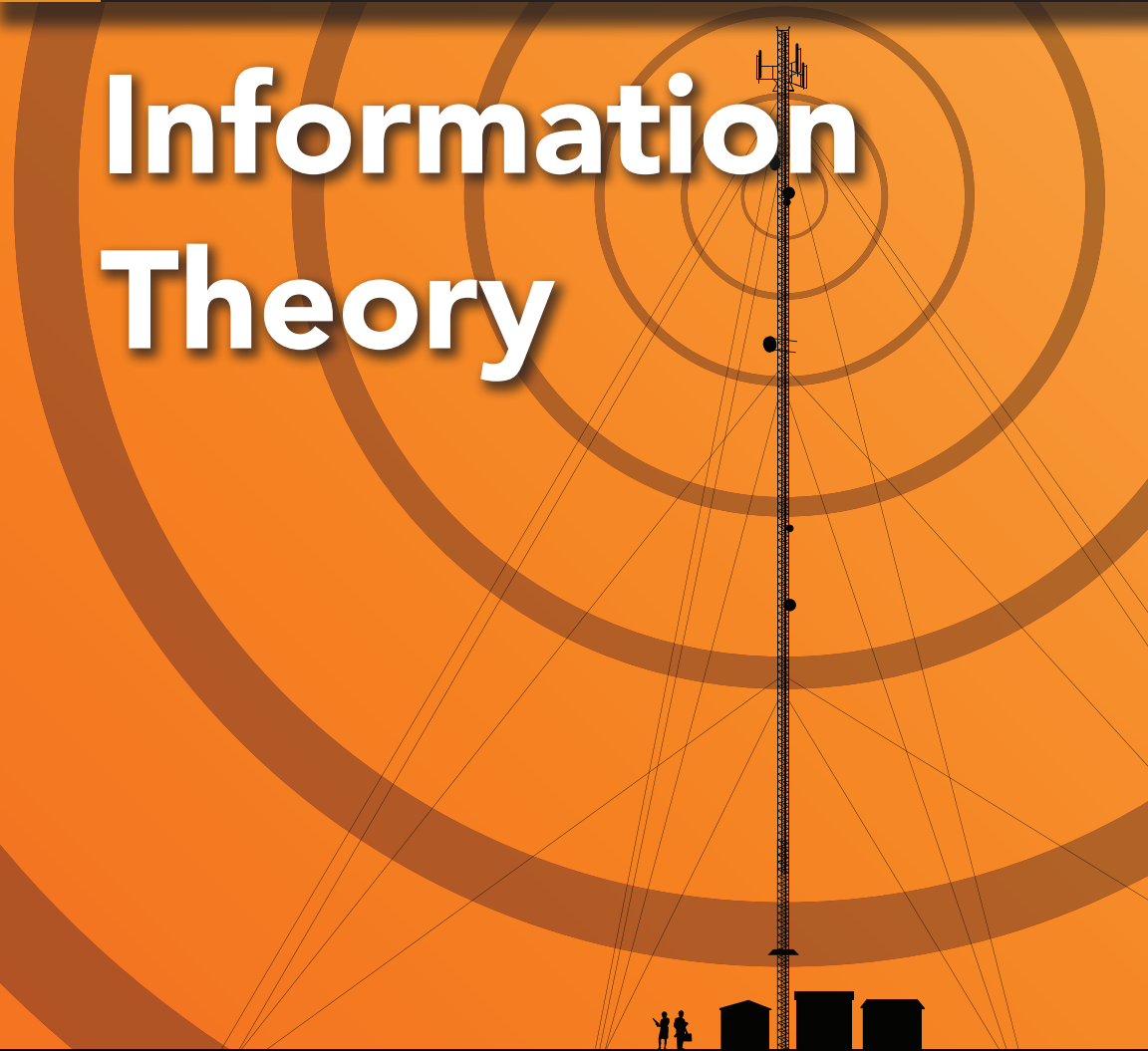


COMMUNICATIONS AND SIGNAL
PROCESSING COLLECTION

Orlando R. Baiocchi, *Editor*

Information Theory



Marcelo S. Alencar



MOMENTUM PRESS
ENGINEERING

INFORMATION THEORY

INFORMATION THEORY

MARCELO S. ALENCAR



MOMENTUM PRESS, LLC, NEW YORK

Information Theory

Copyright © Momentum Press®, LLC, 2015.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or any other—except for brief quotations, not to exceed 400 words, without the prior permission of the publisher.

First published by Momentum Press®, LLC

222 East 46th Street, New York, NY 10017

www.momentumpress.net

ISBN-13: 978-1-60650-528-1 (print)

ISBN-13: 978-1-60650-529-8 (e-book)

Momentum Press Communications and Signal Processing Collection

DOI: 10.5643/9781606505298

Cover and interior design by Exeter Premedia Services Private Ltd.,
Chennai, India

10 9 8 7 6 5 4 3 2 1

Printed in the United States of America

This book is dedicated to my family.

ABSTRACT

The book presents the historical evolution of Information Theory, along with the basic concepts linked to information. It discusses the information associated to a certain source and the usual types of source codes, the information transmission, joint information, conditional entropy, mutual information, and channel capacity. The hot topic of multiple access systems, for cooperative and noncooperative channels, is discussed, along with code division multiple access (CDMA), the basic block of most cellular and personal communication systems, and the capacity of a CDMA system. The information theoretical aspects of cryptography, which are important for network security, a topic intrinsically connected to computer networks and the Internet, are also presented. The book includes a review of probability theory, solved problems, illustrations, and graphics to help the reader understand the theory.

KEY WORDS

Code division multiple access, coding theory, cryptography, information theory, multiple access systems, network security

CONTENTS

LIST OF FIGURES	xi
LIST OF TABLES	xv
PREFACE	xvii
ACKNOWLEDGMENTS	xix
1. INFORMATION THEORY	1
1.1 Information Measurement	2
1.2 Requirements for an Information Metric	4
2. SOURCES OF INFORMATION	11
2.1 Source Coding	11
2.2 Extension of a Memoryless Discrete Source	12
2.3 Prefix Codes	14
2.4 The Information Unit	17
3. SOURCE CODING	19
3.1 Types of Source Codes	19
3.2 Construction of Instantaneous Codes	23
3.3 Kraft Inequality	24
3.4 Huffman Code	27
4. INFORMATION TRANSMISSION	31
4.1 The Concept of Information Theory	32
4.2 Joint Information Measurement	32
4.3 Conditional Entropy	34
4.4 Model for a Communication Channel	34
4.5 Noiseless Channel	35
4.6 Channel with Independent Output and Input	36
4.7 Relations Between the Entropies	37
4.8 Mutual Information	38
4.9 Channel Capacity	41

5. MULTIPLE ACCESS SYSTEMS	49
5.1 Introduction	49
5.2 The Gaussian Multiple Access Channel	51
5.3 The Gaussian Channel with Rayleigh Fading	54
5.4 The Noncooperative Multiple Access Channel	59
5.5 Multiple Access in a Dynamic Environment	62
5.6 Analysis of the capacity for a Markovian Multiple Access Channel	63
6. CODE DIVISION MULTIPLE ACCESS	71
6.1 Introduction	71
6.2 Fundamentals of Spread Spectrum Signals	74
6.3 Performance Analysis of CDMA Systems	76
6.4 Sequence Design	79
7. THE CAPACITY OF A CDMA SYSTEM	87
7.1 Introduction	87
7.2 Analysis of a CDMA System with a Fixed Number of Users and Small SNR	87
7.3 CDMA System with a Fixed Number of Users and High SNR	97
7.4 A Tight Bound on the Capacity of a CDMA System	103
8. THEORETICAL CRYPTOGRAPHY	117
8.1 Introduction	117
8.2 Cryptographic Aspects of Computer Networks	118
8.3 Principles of Cryptography	119
8.4 Information Theoretical Aspects of Cryptography	120
8.5 Mutual Information for Cryptosystems	123
APPENDIX A PROBABILITY THEORY	125
A.1 Set Theory and Measure	125
A.2 Basic Probability Theory	131
A.3 Random Variables	133
REFERENCES	139
ABOUT THE AUTHOR	147
INDEX	149

LIST OF FIGURES

Figure 1.1.	Graph of an information function	9
Figure 2.1.	Source encoder	12
Figure 2.2.	Decision tree for the code in Table 2.5	16
Figure 3.1.	Classes of source codes	23
Figure 3.2.	Probabilities in descending order for the Huffman code	28
Figure 3.3.	Huffman code. At each phase, the two least probable symbols are combined	28
Figure 3.4.	(a) Example of the Huffman coding algorithm to obtain the codewords. (b) Resulting code.	29
Figure 4.1.	Model for a communication channel	32
Figure 4.2.	A probabilistic communication channel	33
Figure 4.3.	Venn diagram corresponding to the relation between the entropies	40
Figure 4.4.	Memoryless binary symmetric channel	44
Figure 4.5.	Graph for the capacity of the memoryless binary symmetric channel	46
Figure 4.6.	Binary erasure channel	46
Figure 4.7.	Graph of the capacity for the binary erasure channel	47
Figure 5.1.	The multiple access channel	52
Figure 5.2.	Capacity region for the Gaussian multiple access channel, $M = 2$	54
Figure 5.3.	Average and actual capacity, for $\gamma = 0.5, 1.0,$ and 2.0	58
Figure 5.4.	Capacity region for the noncooperative channel, $M = 2$	61
Figure 5.5.	Markov model for the multiple access channel	64

Figure 5.6.	Capacity for the channel with Geometric accessibility, as a function of the signal-to-noise ratio, for different values of ρ	66
Figure 5.7.	Capacity for the channel with Geometric accessibility, as a function of the utilization factor, for different values of the signal-to-noise ratio	67
Figure 5.8.	Capacity for the channel with Poisson accessibility, as a function of the signal-to-noise ratio, for $\rho = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$	68
Figure 5.9.	Capacity for the channel with Poisson accessibility, as a function of the utilization factor, for some values of the signal-to-noise ratio	69
Figure 6.1.	Data signal and pseudo-noise sequence	72
Figure 6.2.	Direct sequence spread spectrum system	73
Figure 6.3.	Frequency hopped spread spectrum system	73
Figure 6.4.	Spread spectrum using random time windows	73
Figure 6.5.	Spectra of transmitted and received signals	75
Figure 6.6.	Pseudo-noise sequence generator	82
Figure 6.7.	Gold sequence generator	83
Figure 7.1.	Capacity approximations for the channel, as a function of the signal-to-noise ratio, for $M = 500$ and $N = 100$	96
Figure 7.2.	<i>Bound 1</i> on the capacity for the channel, as a function of the signal-to-noise ratio (E_b/N_0)	100
Figure 7.3.	<i>Bound 2</i> on the capacity for the channel, as a function of the signal-to-noise ratio (E_b/N_0)	101
Figure 7.4.	<i>Bound 3</i> on the capacity for the channel, as a function of the signal-to-noise ratio (E_b/N_0)	103
Figure 7.5.	Capacity for the channel, compared with the lower bound, as a function of the signal-to-noise ratio (E_b/N_0), for $M = 20$ and sequence length $N = 100$	106
Figure 7.6.	Approximate capacity for the channel, as a function of the signal-to-noise ratio (E_b/N_0), for $M = 20$, having N as a parameter	107
Figure 7.7.	Capacity for the channel, using the approximate formula, as a function of the sequence length (N), for different values of M	108

Figure 7.8.	Capacity for the channel, as a function of the number of users (M), using the approximation for the capacity	109
Figure 7.9.	Capacity for the channel, as a function of the number of users (M), including the case $M \gg N$	110
Figure 7.10.	Capacity for the channel, as a function of the probability of error (P_e)	111
Figure 7.11.	<i>Bound 4</i> on the capacity for the channel, as a function of the signal-to-noise ratio (E_b/N_0)	112
Figure 7.12.	Bounds on the capacity, as a function of the signal-to-noise ratio (E_b/N_0)	113
Figure 7.13.	Comparison between the new and existing bounds, as a function of the signal-to-noise ratio (E_b/N_0), for $M = 20$ and $N = 100$	114
Figure 8.1.	General model for a cryptosystem	119
Figure A.1.	A Venn diagram that represents two intersecting sets	127
Figure A.2.	A Venn diagram representing disjoint sets	127
Figure A.3.	Increasing sequence of sets	128
Figure A.4.	Decreasing sequence of sets	128
Figure A.5.	Partition of set B by a family of sets $\{A_i\}$	133
Figure A.6.	Joint probability density function	137

LIST OF TABLES

Table 1.1.	Symbol probabilities of a two-symbol source	4
Table 1.2.	Identically distributed symbol probabilities	5
Table 1.3.	Unequal symbol probabilities	5
Table 1.4.	Symbol probabilities of a certain source	9
Table 2.1.	A compact code	13
Table 2.2.	A compact code for an extension of a source	14
Table 2.3.	A prefix code for a given source	14
Table 2.4.	A source code that is not prefix	15
Table 2.5.	Example of a prefix code	15
Table 3.1.	A binary block code	20
Table 3.2.	A ternary block code	20
Table 3.3.	A nonsingular block code	20
Table 3.4.	A nonsingular block code	21
Table 3.5.	The second extension of a block code	21
Table 3.6.	Uniquely decodable codes.	22
Table 3.7.	Another uniquely decodable code	22
Table 3.8.	Selected binary codes	25
Table 3.9.	Discrete source with five symbols and their probabilities	28
Table 3.10.	Four distinct Huffman codes obtained for the source of Table 3.9	30
Table 6.1.	Number of maximal sequences	81
Table 6.2.	Relative peak cross-correlations of m -sequences, Gold sequences and Kasami sequences	83

PREFACE

Information Theory is a classic topic in the educational market that evolved from the amalgamation of different areas of Mathematics and Probability, which includes set theory, developed by Georg Cantor, and measure theory, fostered by Henri Lebesgue, as well as the axiomatic treatment of probability by Andrei Kolmogorov in 1931, and finally the beautiful development of Communication Theory by Claude Shannon in 1948.

Information Theory is fundamental to several areas of knowledge, including the Engineering, Computer Science, Mathematics, Physics, Sciences, Economics, Social Sciences, and Social Communication. It is part of the syllabus for most courses in Computer Science, Mathematics, and Engineering.

For Electrical Engineering courses it is a pre-requisite to some disciplines, including communication systems, transmission techniques, error control coding, estimation, and digital signal processing. This book is self-contained, it is a reference and an introduction for graduate students who did not have information theory before. It could also be used as an undergraduate textbook. It is addressed to a large audience in Electrical and Computer Engineering, and Mathematics and Applied Physics. The book's target audience is graduate students in these areas, who may not have taken basic courses in specific topics, who can find a quick and concise way to obtain the knowledge they need to succeed in advanced courses.

REASONS FOR WRITING THE BOOK

According to a study by the Institute of Electrical and Electronics Engineers (IEEE), the companies, enterprises, and industry are in need of professionals with a solid background on mathematics and sciences, instead of the specialized professional of the previous century. The employment market in this area is in demand of information technology professionals

and engineers who could afford to change and learn, as the market changes. The market needs professionals who can model and design.

Few books have been published covering the subjects needed to understand the very fundamental concepts of Information Theory. Most books, which deal with the subject, are destined to very specific audiences.

The more mathematically oriented books are seldom used by people with engineering, economics, or statistical background, because the authors are more interested in theorems and related conditions than in fundamental concepts and applications. The books written for engineers usually lack the required rigour, or skip some important points in favor of simplicity and conciseness.

The idea is to present a seamless connection between the more abstract advanced set theory, the fundamental concepts from measure theory and integration and probability, filling in the gaps from previous books and leading to an interesting, robust, and, hopefully, self-contained exposition of the Information Theory.

DESCRIPTION OF THE BOOK

The book begins with the historical evolution of Information Theory. Chapter 1 deals with the basic concepts of information theory, and how to measure information. The information associated to a certain source is discussed in Chapter 2. The usual types of source codes are presented in Chapter 3. Information transmission, joint information, conditional entropy, mutual information, and channel capacity are the subject of Chapter 4. The hot topic of multiple access systems, for cooperative and noncooperative channels, is discussed in Chapter 5.

Chapter 6 presents code division multiple access (CDMA), the basic block of most cellular and personal communication systems in operation. The capacity of a CDMA system is the subject of Chapter 7. The information theoretical aspects of cryptography are presented in Chapter 8, which are important for network security, a topic intrinsically connected to computer networks and the Internet. The appendix includes a review of probability theory. Solved problems, illustrations, and graphics help the reader understand the theory.

ACKNOWLEDGMENTS

The author is grateful to all the members of the Communications Research Group, certified by the National Council for Scientific and Technological Development (CNPq), at the Federal University of Campina Grande, for their collaboration in many ways, helpful discussions and friendship, as well as our colleagues at the Institute for Advanced Studies in Communications (Iecom).

The author also wishes to acknowledge the contribution of professors Francisco Madeiro, from the State University of Pernambuco, and Waslon T. A. Lopes, from the Federal University of Campina Grande, Brazil, to the chapter on source coding.

The author is also grateful to professor Valdemar Cardoso da Rocha Jr., from the Federal University of Pernambuco, Brazil, for technical communications, long-term cooperation, and useful discussions related to information theory.

The author is indebted to his wife Silvana, sons Thiago and Raphael, and daughter Marcella, for their patience and support during the course of the preparation of this book.

The author is thankful to professor Orlando Baiocchi, from the University of Washington, Tacoma, USA, who strongly supported this project from the beginning and helped with the reviewing process.

Finally, the author registers the support of Shoshanna Goldberg, Destiny Hadley, Charlene Kronstedt, Jyothi, and Millicent Treloar from Momentum Press, in the book preparation process.

CHAPTER 1

INFORMATION THEORY

Information Theory is a branch of Probability Theory, which has application and correlation with many areas, including communication systems, communication theory, Physics, language and meaning, cybernetics, psychology, art, and complexity theory (Pierce 1980). The basis for the theory was established by Harry Theodor Nyqvist (1889–1976) (Nyquist 1924), also known as Harry Nyquist, and Ralph Vinton Lyon Hartley (1888–1970), who invented the Hartley oscillator (1928). They published the first articles on the subject, in which the factors that influenced the transmission of information were discussed.

The seminal article by Claude E. Shannon (1916–2001) extended the theory to include new factors, such as the noise effect in the channel and the savings that could be obtained as a function of the statistical structure of the original message and the information receiver characteristics (Shannon 1948). Shannon defined the fundamental communication problem as the possibility of, exactly or approximately, reproducing, at a certain point, a message that has been chosen at another one.

The main semantic aspects of the communication, initially established by Charles Sanders Peirce (1839–1914), a philosopher and creator of Semiotic Theory, are not relevant for the development of the Shannon information theory. What is important is to consider that a particular message is selected from a set of possible messages.

Of course, as mentioned by John Robinson Pierce (1910–2002), quoting the philosopher Alfred Jules Ayer (1910–1989), it is possible to communicate not only information, but knowledge, errors, opinions, ideas, experiences, desires, commands, emotions, feelings. Heat and movement can be communicated, as well as, force, weakness, and disease (Pierce 1980).

Hartley has found several reasons why the natural logarithm should measure the information:

- It is a practical metric in Engineering, considering that various parameters, such as time and bandwidth, are proportional to the logarithm of the number of possibilities.
- From a mathematical point of view, it is an adequate measure, because several limit operations are simply stated in terms of logarithms.
- It has an intuitive appeal, as an adequate metric, because, for instance, two binary symbols have four possibilities of occurrence.

The choice of the logarithm base defines the information unit. If base 2 is used, the unit is the bit, an acronym suggested by John W. Tukey for binary digit, which is a play of words that can also mean a piece of information. The information transmission is informally given in bit(s), but a unit has been proposed to pay tribute to the scientist who developed the concept, it is called the shannon, or [Sh] for short. This has a direct correspondence with the unit for frequency, hertz or [Hz], for cycles per second, which was adopted by the International System of Units (SI).¹

Aleksandr Yakovlevich Khinchin (1894–1959) put the Information Theory in solid basis, with a more precise and unified mathematical discussion about the entropy concept, which supported Shannon's intuitive and practical view (Khinchin 1957).

The books by Robert B. Ash (1965) and Amiel Feinstein (1958) give the mathematical reasons for the choice of the logarithm to measure information, and the book by J. Aczél and Z. Daróczy (1975) presents several of Shannon's information measures and their characterization, as well as Alfréd Rényi's (1921–1970) entropy metric.

A discussion on generalized entropies can be found in the book edited by Luigi M. Ricciardi (1990). Lotfi Asker Zadeh introduced the concept of fuzzy set, an efficient tool to represent the behavior of systems that depend on the perception and judgement of human beings, and applied it to information measurement (Zadeh 1965).

1.1 INFORMATION MEASUREMENT

The objective of this section is to establish a measure for the information content of a discrete system, using Probability Theory. Consider a discrete random experiment, such as the occurrence of a symbol, and its associated sample space Ω , in which X is a real random variable (Reza 1961).

The random variable X can assume the following values

$$X = \{x_1, x_2, \dots, x_n\},$$

$$\text{in which } \bigcup_{k=1}^N x_k = \Omega, \quad (1.1)$$

with probabilities in the set P

$$P = \{p_1, p_2, \dots, p_n\},$$

$$\text{in which } \sum_{k=1}^N p_k = 1. \quad (1.2)$$

The information associated to a particular event is given by

$$I(x_i) = \log \left(\frac{1}{p_i} \right), \quad (1.3)$$

because the sure event has probability one and zero information, by a property of the logarithm, and the impossible event has zero probability and infinite information.

Example: suppose the sample space is partitioned into two equally probable spaces. Then

$$I(x_1) = I(x_2) = -\log \frac{1}{2} = 1 \text{ bit}, \quad (1.4)$$

that is, the choice between two equally probable events requires one unit of information, when a base 2 logarithm is used.

Considering the occurrence of 2^N equiprobable symbols, then the self-information of each event is given by

$$I(x_k) = -\log p_k = -\log 2^{-N} = N \text{ bits}. \quad (1.5)$$

It is possible to define the source entropy, $H(X)$, as the average information, obtained by weighing of all the occurrences

$$H(X) = E[I(x_i)] = -\sum_{i=1}^N p_i \log p_i. \quad (1.6)$$

Observe that Equation 1.6 is the weighing average of the logarithms of the probabilities, in which the weights are the real values of the probabilities of the random variable X , and this indicates that $H(X)$ can be interpreted as the expected value of the random variable that assumes the value $\log p_i$, with probability p_i (Ash 1965).

Table 1.1. Symbol probabilities of a two-symbol source

Symbol	Probability
x_1	$\frac{1}{4}$
x_2	$\frac{3}{4}$

Example: consider a source that emits two symbols, with unequal probabilities, given in Table 1.1.

The source entropy is calculated as

$$H(X) = -\frac{1}{4} \log \frac{1}{4} - \frac{3}{4} \log \frac{3}{4} = 0.81 \text{ bits per symbol.}$$

1.2 REQUIREMENTS FOR AN INFORMATION METRIC

A few fundamental properties are necessary for the entropy in order to obtain an axiomatic approach to base the information measurement (Reza 1961).

- If the event probabilities suffer a small change, the associated measure must change in accordance, in a continuous manner, which provides a physical meaning to the metric

$$H(p_1, p_2, \dots, p_N) \text{ is continuous in } p_k, k = 1, 2, \dots, N, \\ 0 \leq p_k \leq 1. \quad (1.7)$$

- The information measure must be symmetric in relation to the probability set P . That is, the entropy is invariant to the order of events.

$$H(p_1, p_2, p_3, \dots, p_N) = H(p_1, p_3, p_2, \dots, p_N). \quad (1.8)$$

- The maximum of the entropy is obtained when the events are equally probable. That is, when nothing is known about the set of events, or about what message has been produced, the assumption of a uniform distribution gives the highest information quantity, that corresponds to the highest level of uncertainty.

$$\text{Maximum of } H(p_1, p_2, \dots, p_N) = H\left(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right). \quad (1.9)$$

Table 1.2. Identically distributed symbol probabilities

Symbol	Probability
x_1	$\frac{1}{4}$
x_2	$\frac{1}{4}$
x_3	$\frac{1}{4}$
x_4	$\frac{1}{4}$

Table 1.3. Unequal symbol probabilities

Symbol	Probability
x_1	$\frac{1}{2}$
x_2	$\frac{1}{4}$
x_3	$\frac{1}{8}$
x_4	$\frac{1}{8}$

Example: consider two sources that emit four symbols. The first source symbols, shown in Table 1.2, have equal probabilities, and the second source symbols, shown in Table 1.3, are produced with unequal probabilities.

The mentioned property indicates that the first source attains the highest level of uncertainty, regardless of the probability values of the second source, as long as they are different.

- Consider that an adequate measure for average uncertainty was found $H(p_1, p_2, \dots, p_N)$ associated to a set of events. Assume that event $\{x_N\}$ is divided into M disjoint sets, with probabilities q_k , such that

$$p_N = \sum_{k=1}^M q_k, \quad (1.10)$$

and the probabilities associated to the new events can be normalized in such a way that

$$\frac{q_1}{p_n} + \frac{q_2}{p_n} + \dots + \frac{q_m}{p_n} = 1. \quad (1.11)$$

Then, the creation of new events from the original set modifies the entropy to

$$H(p_1, p_2, \dots, p_{N-1}, q_1, q_2, \dots, q_M) = H(p_1, \dots, p_{N-1}, p_N) + p_N H\left(\frac{q_1}{p_N}, \frac{q_2}{p_N}, \dots, \frac{q_M}{p_N}\right), \quad (1.12)$$

with

$$p_N = \sum_{k=1}^M q_k.$$

It is possible to show that the function defined by Equation 1.6 satisfies all requirements. To demonstrate the continuity, it suffices to do (Reza 1961)

$$\begin{aligned} H(p_1, p_2, \dots, p_N) &= -[p_1 \log p_1 + p_2 \log p_2 + \dots + p_N \log p_N] \\ &= -[p_1 \log p_1 + p_2 \log p_2 + \dots + p_{N-1} \log p_{N-1} \\ &\quad + (1 - p_1 - p_2 - \dots - p_{N-1}) \\ &\quad \cdot \log(1 - p_1 - p_2 - \dots - p_{N-1})]. \end{aligned} \quad (1.13)$$

Notice that, for all independent random variables, the set of probabilities p_1, p_2, \dots, p_{N-1} and also $(1 - p_1 - p_2 - \dots - p_{N-1})$ are contiguous in $[0, 1]$, and that the logarithm of a continuous function is also continuous. The entropy is clearly symmetric.

The maximum value property can be demonstrated, if one considers that all probabilities are equal and that the entropy is maximized by that condition

$$p_1 = p_2 = \dots = p_N. \quad (1.14)$$

Taking into account that according to intuition the uncertainty is maximum for a system of equiprobable states, it is possible to arbitrarily choose a random variable with probability p_N depending on p_k , and $k = 1, 2, \dots, N - 1$. Taking the derivative of the entropy in terms of each probability

$$\begin{aligned} \frac{dH}{dp_k} &= \sum_{i=1}^N \frac{\partial H}{\partial p_i} \frac{\partial p_i}{\partial p_k} \\ &= -\frac{d}{dp_k}(p_k \log p_k) - \frac{d}{dp_N}(p_N \log p_N) \frac{\partial p_N}{\partial p_k}. \end{aligned} \quad (1.15)$$

But, probability p_N can be written as

$$p_N = 1 - (p_1 + p_2 + \dots + p_k + \dots + p_{N-1}). \quad (1.16)$$

Therefore, the derivative of the entropy is

$$\frac{dH}{dp_k} = -(\log_2 e + \log p_k) + (\log_2 e + \log p_n), \quad (1.17)$$

that, using a property of logarithms, simplifies to,

$$\frac{dH}{dp_k} = -\log \frac{p_k}{p_n}. \quad (1.18)$$

But,

$$\frac{dH}{dp_k} = 0, \text{ which gives } p_k = p_n. \quad (1.19)$$

As p_k was chosen in an arbitrary manner, one concludes that to obtain a maximum for the entropy function, one must have

$$p_1 = p_2 = \dots = p_N = \frac{1}{N}. \quad (1.20)$$

The maximum is guaranteed because

$$H(1, 0, 0, \dots, 0) = 0. \quad (1.21)$$

On the other hand, for equiprobable events, it is possible to verify that the entropy is always positive, for it attains its maximum at (Csiszár and Kórner 1981)

$$H\left(\frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N}\right) = \log N > 0. \quad (1.22)$$

To prove additivity, it suffices to use the definition of entropy, computed for a two set partition, with probabilities $\{p_1, p_2, \dots, p_{N-1}\}$ and $\{q_1, q_2, \dots, q_M\}$,

$$\begin{aligned} & H(p_1, p_2, \dots, p_{N-1}, q_1, q_2, \dots, q_M) \\ &= -\sum_{k=1}^{N-1} p_k \log p_k - \sum_{k=1}^M q_k \log q_k \\ &= -\sum_{k=1}^N p_k \log p_k + p_N \log p_N - \sum_{k=1}^M q_k \log q_k \\ &= H(p_1, p_2, \dots, p_N) + p_N \log p_N \\ &\quad - \sum_{k=1}^M q_k \log q_k. \end{aligned} \quad (1.23)$$

But, the second part of the last term can be written in a way to display the importance of the entropy in the derivation

$$\begin{aligned}
 p_N \log p_N - \sum_{k=1}^M q_k \log q_k &= p_N \sum_{k=1}^M \frac{q_k}{p_N} \log p_N - \sum_{k=1}^M q_k \log q_k \\
 &= -p_N \sum_{k=1}^M \frac{q_k}{p_N} \log \frac{q_k}{p_N} \\
 &= p_N H\left(\frac{q_1}{p_N}, \frac{q_2}{p_N}, \dots, \frac{q_M}{p_N}\right), \quad (1.24)
 \end{aligned}$$

and this demonstrates the mentioned property.

The entropy is non-negative, which guarantees that the partitioning of one event into several other events does not reduce the system entropy, as shown in the following

$$H(p_1, \dots, p_{N-1}, q_1, q_2, \dots, q_M) \geq H(p_1, \dots, p_{N-1}, p_N), \quad (1.25)$$

that is, if one splits a symbol into two or more, the entropy always increases, and that is the physical origin of the word.

Example: consider a binary source, X , that emits symbols 0 and 1 with probabilities p and $q = 1 - p$. The average information per symbol is given by $H(X) = -p \log p - q \log q$, that is known as entropy function.

$$H(p) = -p \log p - (1 - p) \log (1 - p). \quad (1.26)$$

Example: for the binary source, consider that the symbol probabilities are $p = 1/8$ and $q = 7/8$, and compute the entropy of the source.

The average information per symbol is given by

$$H(X) = -1/8 \log 1/8 - 7/8 \log 7/8,$$

which gives $H(X) = 0.544$.

Note that even though 1 bit is produced for each symbol, the actual average information is 0.544 bits due to the unequal probabilities.

The entropy function has a maximum, when all symbols are equiprobable, of $p = q = 1/2$, for which the entropy is 1 bit/symbol. The function attains a minimum of $p = 0$ or $p = 1$.

This function plays an essential role in determining the capacity of a binary symmetric channel. Observe that the entropy function is concave, that is

$$H(p_1) + H(p_2) \leq 2H\left(\frac{p_1 + p_2}{2}\right). \quad (1.27)$$

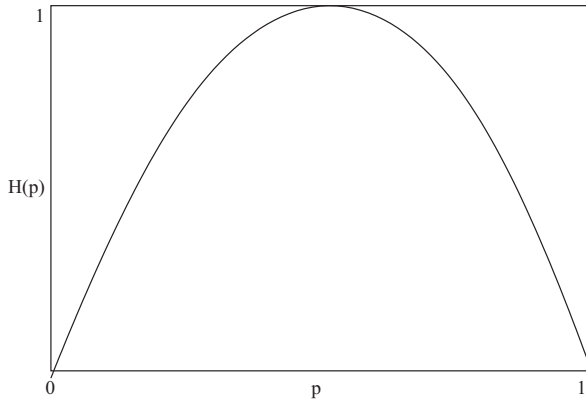


Figure 1.1. Graph of an information function.

The entropy function is illustrated in Figure 1.1, in which it is possible to notice the symmetry, concavity, and the maximum for equiprobable symbols. As consequence of the symmetry, the sample spaces with probability distributions obtained from permutations of a common distribution provide the same information quantity (van der Lubbe 1997).

Example: consider a source that emits symbols from an alphabet $X = \{x_1, x_2, x_3, x_4\}$ with probabilities given in Table 1.4. What is the entropy of this source?

The entropy is computed using Formula 1.6, for $N = 4$ symbols, as

$$H(X) = - \sum_{i=1}^4 p_i \log p_i,$$

or

$$H(X) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{4} - \frac{2}{8} \log \frac{1}{8} = 1.75 \text{ bits per symbol.}$$

Table 1.4. Symbol probabilities of a certain source

Symbol	Probability
x_1	$\frac{1}{2}$
x_2	$\frac{1}{4}$
x_3	$\frac{1}{8}$
x_4	$\frac{1}{8}$

NOTES

- 1 The author of this book proposed the adoption of the shannon [Sh] unit during the IEEE International Conference on Communications (ICC 2001), in Helsinki, Finland, shortly after Shannon's death.

CHAPTER 2

SOURCES OF INFORMATION

2.1 SOURCE CODING

The efficient representation of data produced by a discrete source is called source coding. For a source coder to obtain a good performance, it is necessary to take the symbol statistics into account. If the symbol probabilities are different, it is useful to assign short codewords to probable symbols and long ones to infrequent symbols. This produces a variable length code, such as the Morse code.

Two usual requirements to build an efficient code are:

1. The codewords generated by the coder are binary.
2. The codewords are unequivocally decodable, and the original message sequence can be reconstructed from the binary coded sequence.

Consider Figure 2.1, which shows a memoryless discrete source, whose output x_k is converted by the source coder into a sequence of 0s and 1s, denoted by b_k . Assume that the source alphabet has K different symbols, and that the k -ary symbol, x_k , occurs with the probability p_k , $k = 0, 1, \dots, K - 1$.

Let l_k be the average length, measured in bits, of the binary word assigned to symbol x_k . The average length of the words produced by the source coder is defined as (Haykin 1988)

$$\bar{L} = \sum_{k=1}^K p_k l_k. \quad (2.1)$$

The parameter \bar{L} represents the average number of bits per symbol from that is used in the source coding process. Let L_{\min} be the smallest possible

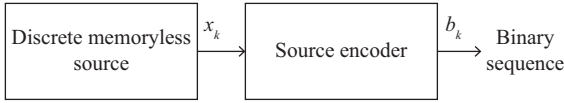


Figure 2.1. Source encoder.

value of \bar{L} . The source coding efficiency is defined as (Haykin 1988)

$$\eta = \frac{L_{\min}}{\bar{L}}. \quad (2.2)$$

Because $\bar{L} \geq L_{\min}$, $\eta \leq 1$. The source coding efficiency increases as η approaches 1.

Shannon's first theorem, or source coding theorem, provides a means to determine L_{\min} (Haykin 1988).

Given a memoryless discrete source with entropy $H(X)$, the average length of the codewords is limited by

$$\bar{L} \geq H(X).$$

Entropy $H(X)$, therefore, represents a fundamental limit for the average number of bits per source symbol \bar{L} , which is needed to represent a memoryless discrete source, and this number can be as small as, but never smaller than, the source entropy $H(X)$. Therefore, for $L_{\min} = H(X)$, the source coding efficiency can be written as (Haykin 1988)

$$\eta = \frac{H(X)}{\bar{L}}. \quad (2.3)$$

The code redundancy is given by (Abramson 1963).

$$1 - \eta = \frac{\bar{L} - H(X)}{\bar{L}}. \quad (2.4)$$

2.2 EXTENSION OF A MEMORYLESS DISCRETE SOURCE

It is useful to consider the encoding of blocks of N successive symbols from the source, instead of individual symbols. Each block can be seen as a product of an extended source with an alphabet X^N that has K^N distinct blocks. The symbols are statistically independent, therefore, the probability of an extended symbol is the product of the probabilities of the original symbols, and it can be shown that

$$H(X^N) = NH(X). \quad (2.5)$$

Example: consider the discrete memoryless source with alphabet

$$X = \{x_1, x_2, x_3\}.$$

The second order extended source has an alphabet

$$X^2 = \{x_1x_1, x_1x_2, x_1x_3, x_2x_1, x_2x_2, x_2x_3, x_3x_1, x_3x_2, x_3x_3\}.$$

For the second order extended source of the example, $p(x_i x_j) = p(x_i)p(x_j)$. In particular, if all original source symbols are equiprobable, then $H(X) = \log_2 3$ bits. The second order extended source has nine equiprobable symbols, therefore, $H(X^2) = \log_2 9$ bits. It can be noticed that $H(X^2) = 2H(X)$.

2.2.1 IMPROVING THE CODING EFFICIENCY

The following example illustrates how to improve the coding efficiency using extensions of a source (Abramson 1963).

Example: consider a memoryless source, $S = \{x_1, x_2\}$, with $p(x_1) = \frac{3}{4}$ and $p(x_2) = \frac{1}{4}$. The source entropy is given by $\frac{1}{4} \log_2 4 + \frac{3}{4} \log_2 \frac{4}{3} = 0.811$ bit.

A compact code for the source is presented in Table 2.1.

The average codeword length is one bit, and the efficiency is

$$\eta_1 = 0.811.$$

Example: to improve the efficiency, the second extension of the source is encoded, as shown in Table 2.2.

The average codeword length is $\frac{27}{16}$ bits. The extended source entropy is 2×0.811 bits, and the efficiency is

$$\eta_2 = \frac{2 \times 0.811 \times 16}{27} = 0.961.$$

The efficiency improves for each new extension of the original source but, of course, the codes get longer, which implies that they take more time to transmit or process.

Table 2.1. A compact code

x_i	$p(x_i)$	Compact code
x_1	$\frac{3}{4}$	0
x_2	$\frac{1}{4}$	1

Table 2.2. A compact code for an extension of a source

Symbol	Probability	Compact code
x_1x_1	9/16	0
x_1x_2	3/16	10
x_2x_1	3/16	110
x_2x_2	1/16	111

Example: the efficiencies associated to the third and fourth extensions of the source are

$$\eta_3 = 0.985$$

and

$$\eta_4 = 0.991.$$

As higher order extensions of the source are encoded, the efficiency approaches 1, a result that is proved in the next section.

2.3 PREFIX CODES

For a prefix code, no codeword is a prefix, of the first part, of another codeword. Therefore, the code shown in Table 2.3 is prefix. On the other hand, code shown in Table 2.4 is not prefix, because the binary word 10, for instance, is a prefix for the codeword 100.

To decode a sequence of binary words produced by a prefix encoder, the decoder begins at the first binary digit of the sequence, and decodes a codeword at a time. It is similar to a decision tree, which is a representation of the codewords of a given source code.

Table 2.3. A prefix code for a given source

Symbol	Code
x_1	1
x_2	01
x_3	001
x_4	000

Table 2.4. A source code that is not prefix

Symbol	Code
x_1	1
x_2	10
x_3	100
x_4	1000

Figure 2.2 illustrates the decision tree for the prefix code pointed in Table 2.5.

The tree has one initial state and four final states, which correspond to the symbols x_1 , x_2 , and x_3 . From the initial state, for each received bit, the decoder searches the tree until a final state is found.

The decoder, then, emits a corresponding decoded symbol and returns to the initial state. Therefore, from the initial state, after receiving a 1, the source decoder decodes symbol x_1 and returns to the initial state. If it receives a 0, the decoder moves to the lower part of the tree, in the following, after receiving another 0, the decoder moves further to the lower part of the tree and, after receiving a 1, the decoder retrieves x_2 and returns to the initial state.

Considering the code from Table 2.5, with the decoding tree from Figure 2.2, the binary sequence 011100010010100101 is decoded into the output sequence $x_1x_0x_0x_3x_0x_2x_1x_2x_1$.

By construction, a prefix code is always unequivocally decodable, which is important to avoid any confusion at the receiver.

Consider a code that has been constructed for a discrete source with alphabet $\{x_1, x_2, \dots, x_K\}$. Let $\{p_1, p_1, \dots, p_K\}$ be the source statistics, and l_k be the codeword length for symbol x_k , $k = 1, \dots, K$. If the binary code constructed for the source is a prefix one, then one can use the

Table 2.5. Example of a prefix code

Source symbol	Probability of occurrence	Code
x_0	0.5	1
x_1	0.25	01
x_2	0.125	001
x_3	0.125	000

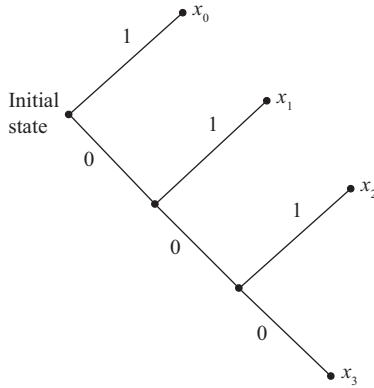


Figure 2.2. Decision tree for the code in Table 2.5.

Kraft-McMillan inequality

$$\sum_{k=1}^K 2^{-l_k} \leq 1, \quad (2.6)$$

in which factor 2 is the radix, or number of symbols, of the binary alphabet.

For a memoryless discrete source with entropy $H(X)$, the codeword average length of a prefix code is limited to

$$H(X) \leq \bar{L} < H(X) + 1. \quad (2.7)$$

The left hand side equality is obtained on the condition that symbol x_k be emitted from the source with probability $p_k = 2^{-l_k}$, in which l_k is the length of the codeword assigned to symbol x_k .

Consider the N th order extension of a memoryless discrete source. For the N th order extension of a code, the encoder operates on blocks of N symbols from the original source, instead of individual ones, and the source alphabet X^N has an entropy that is N times the entropy of the original source.

Let \bar{L}_N be the average length for the extended prefix code. For an unequivocally decodable code, \bar{L}_N is as small as possible, from Equation 2.7 it follows that

$$H(X^N) \leq \bar{L}_N < H(X^N) + 1, \quad (2.8)$$

therefore,

$$NH(X) \leq \bar{L}_N < NH(X) + 1, \quad (2.9)$$

or, in an equivalent way,

$$H(X) \leq \frac{\bar{L}_N}{N} < H(X) + \frac{1}{N}. \quad (2.10)$$

In the limit, as N goes to infinity, the inferior and superior limitants converge, and therefore,

$$\lim_{N \rightarrow \infty} \frac{1}{N} \bar{L}_N = H(X). \quad (2.11)$$

Therefore, for a prefix extended encoder, as the order N increases, the code represents the source as efficiently as desired, and the average code-word length can be as close to the source entropy as possible, according to Shannon's source coding theorem. On the other hand, there is a compromise between the reduction on the average codeword length and the increase in complexity of the decoding process (Haykin 1988).

2.4 THE INFORMATION UNIT

There is some confusion between the binary digit, abbreviated as bit, and the information particle, also baptized as bit by John Tukey and Claude Shannon.

In a meeting of the Institute of Electrical and Electronics Engineers (IEEE) the largest scientific institution in the World, the author of this book proposed the shannon [Sh] as a unit of information transmission, equivalent to bit per second. It is instructive to say that the bit, as used today, is not a unit of information, because it is not approved by the International System of Units (SI).

What is curious about that meeting was the misunderstanding that surrounded the units, in particular regarding the difference between the concepts of information unit and digital logic unit.

To make things clear, the binary digit is associated to a certain state of a digital system, and not to information. A binary digit 1 can refer to 5 volts, in TTL logic, or 12 volts, for CMOS logic.

The information bit exists independent of any association to a particular voltage level. It can be associated, for example, to a discrete information or to the quantization of an analog information.

For instance, the information bits recorded on the surface of a compact disk are stored as a series of depressions on the plastic material, which are read by an optical beam, generated by a semiconductor laser. But, obviously, the depressions are not the information. They represent a

means for the transmission of information, a material substrate that carries the data.

In the same way, the information can exist, even if it is not associated to light or other electromagnetic radiation. It can be transported by several means, including paper, and materializes itself when it is processed by a computer or by a human being.

CHAPTER 3

SOURCE CODING

3.1 TYPES OF SOURCE CODES

This chapter presents the classification of source codes, block codes, nonsingular codes, uniquely decodable codes, and instantaneous codes.

3.1.1 BLOCK CODES

Let $S = \{x_0, x_1, \dots, x_{K-1}\}$ be a set of symbols of a given source alphabet. A code is defined as a mapping of all possible symbol sequences from S into sequences of another set $X = \{x_0, x_1, \dots, x_{M-1}\}$, called the code alphabet.

A block code maps each of the symbols from the source set into a sequence of the code alphabet. The fixed sequences of symbols x_j are called codewords X_j . Note that X_j denotes a sequence of x_j 's (Abramson 1963).

Example: a binary block code is presented in Table 3.1, and a ternary block code is shown in Table 3.2.

3.1.2 NONSINGULAR CODES

A block code is said to be nonsingular if all codewords are distinct (Abramson 1963). Table 3.2 shows an example of a nonsingular code. The code shown in Table 3.3 is also nonsingular, but although the codewords are distinct there is a certain ambiguity between some symbol sequences of the code regarding the source symbol sequences.

Table 3.1. A binary block code

Source symbols	Code
x_0	0
x_1	11
x_2	00
x_3	1

Table 3.2. A ternary block code

Source symbols	Code
x_0	0
x_1	1
x_2	2
x_3	01

Table 3.3. A nonsingular block code

Source symbols	Code
x_0	0
x_1	01
x_2	1
x_3	11

Example: the sequence 1111 can correspond to $x_2x_2x_2x_2$, or $x_2x_3x_2$, or even x_3x_3 . Which indicates that it is necessary to define a more strict condition than nonsingularity for a code, to guarantee that it can be used in a practical situation.

3.1.3 UNIQUELY DECODABLE CODES

Let a block code map symbols from a source alphabet S into fixed symbol sequences of a code alphabet X . The source can be an extension of another source, which is composed of symbols from the original alphabet. The n -ary extension of a block code that maps symbols x_i into codewords X_i is the block code that maps symbol sequences from the source $(x_{i_1}x_{i_2} \dots x_{i_n})$ into the codeword sequences $(X_{i_1}X_{i_2} \dots X_{i_n})$ (Abramson 1963).

Table 3.4. A nonsingular block code

Source symbols	Code
x_0	1
x_1	00
x_2	11
x_3	10

Table 3.5. The second extension of a block code

Source symbols	Code	Source symbols	Code
x_0x_0	11	x_2x_0	111
x_0x_1	100	x_2x_1	1100
x_0x_2	111	x_2x_2	1111
x_0x_3	110	x_2x_3	1110
x_1x_0	001	x_3x_0	101
x_1x_1	0000	x_3x_1	1000
x_1x_2	0011	x_3x_2	1011
x_1x_3	0010	x_3x_3	1010

From the previous definition, the n -ary extension of a block code is also a block code. The second order extension of the block code presented in Table 3.4 is the block code of Table 3.5.

A block code is said to be uniquely decodable if and only if the n -ary extension of the code is nonsingular for all finite n .

3.1.4 INSTANTANEOUS CODES

Table 3.6 presents two examples of uniquely decodable codes. Code \mathcal{A} is a simpler method to construct a uniquely decodable set of sequences, because all codewords have the same length and it is a nonsingular code.

Code \mathcal{B} is also uniquely decodable. It is also called a comma code, because the digit zero is used to separate the codewords (Abramson 1963).

Consider the code shown in Table 3.7. Code \mathcal{C} differs from \mathcal{A} and \mathcal{B} from Table 3.6 in an important aspect. If a binary sequence composed of codewords from code \mathcal{C} occurs, it is not possible to decode the sequence.

Table 3.6. Uniquely decodable codes

Source symbols	Code \mathcal{A}	Code \mathcal{B}
x_0	000	0
x_1	001	10
x_2	010	110
x_3	011	1110
x_4	100	11110
x_5	101	111110
x_6	110	1111110
x_7	111	11111110

Table 3.7. Another uniquely decodable code

Source symbols	Code \mathcal{C}
x_0	1
x_1	10
x_2	100
x_3	1000
x_4	10000
x_5	100000
x_6	1000000
x_7	10000000

Example: if the bit stream 100000 is received, for example, it is not possible to decide if it corresponds to symbol x_5 , unless the next symbol is available. If the next symbol is 1, then the sequence is 100000, but if it is 0, then it is necessary to inspect one more symbol to know if the sequence corresponds to x_6 (1000000) or x_7 (10000000).

A uniquely decodable code is instantaneous if it is possible to decode each codeword in a sequence with no reference to subsequent symbols (Abramson 1963). Code \mathcal{A} and \mathcal{B} are instantaneous, and \mathcal{C} is not.

It is possible to devise a test to indicate when a code is instantaneous. Let $X_i = x_{i1}x_{i2} \dots x_{im}$ be a word from a certain code. The sequence of symbols $(x_{i1}x_{i2} \dots x_{ij})$, with $j \leq m$, is called the prefix of the codeword X_i .

Example: the codeword 10000 has five prefixes: 1, 10, 100, 1000, and 10000. A necessary condition for a code to be instantaneous is that no codeword is a prefix of another codeword.

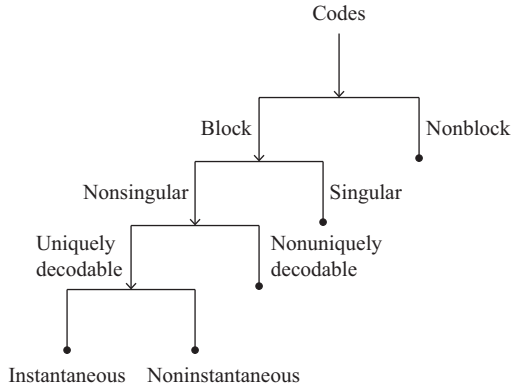


Figure 3.1. Classes of source codes.

The various classes of codes presented in this section are summarized in Figure 3.1.

3.2 CONSTRUCTION OF INSTANTANEOUS CODES

In order to construct a binary instantaneous code for a source with five symbols, one can begin by attributing the digit 0 to symbol s_0 (Abramson 1963)

$$x_0 \leftarrow 0.$$

In this case, the remaining source symbols should correspond to the codewords that begin with the digit 1. Otherwise, it is not a prefix code. It is not possible to associate x_1 to the codeword 1, because no other symbol would remain to begin the other codewords.

Therefore,

$$x_1 \leftarrow 10.$$

This codeword assignment requires that the remaining codewords begin with 11. If

$$x_2 \leftarrow 110$$

then, the only unused prefix with three bits is 111, which implies that

$$x_3 \leftarrow 1110$$

and

$$x_4 \leftarrow 1111.$$

In the previously constructed code, note that if one begins the code construction by making x_0 to correspond to 0, this restricts the available number of codewords, because the remaining codewords had to, necessarily, begin with 1.

On the other hand, if a two-digit word had been chosen to represent x_0 , there would be more freedom to choose the others, and there would be no need to assign very long codewords to the last ones.

A binary instantaneous code can be constructed to represent the five symbols (Abramson 1963). The first assignment is

$$x_0 \leftarrow 00.$$

Then, one can assign

$$x_1 \leftarrow 01$$

and two unused prefixes of length two are saved to the following codeword assignment:

$$x_2 \leftarrow 10$$

$$x_3 \leftarrow 110$$

$$x_4 \leftarrow 111.$$

The question of which code is the best is postponed for the next section, because it requires the notion of average length of a code, that depends on the symbol probability distribution.

3.3 KRAFT INEQUALITY

Consider an instantaneous code with source alphabet given by

$$S = \{x_0, x_1, \dots, x_{K-1}\}$$

and code alphabet

$$X = \{x_0, x_1, \dots, x_{M-1}\}.$$

Let X_0, X_1, \dots, X_{K-1} be the codewords, and let l_i be the length of the word X_i . The Kraft inequality establishes that a necessary and sufficient condition for the existence of an instantaneous code of length l_0, l_1, \dots, l_{K-1} is

$$\sum_{i=0}^{K-1} r^{-l_i} \leq 1, \tag{3.1}$$

in which r is the number of different symbols of the code.

For the binary case,

$$\sum_{i=0}^{K-1} 2^{-l_i} \leq 1. \quad (3.2)$$

The Kraft inequality can be used to determine if a given sequence of length l_i is acceptable for a codeword of an instantaneous code.

Consider an information source, with four possible symbols, x_0 , x_1 , x_2 , and x_3 . Table 3.8 presents five possible codes to represent the original symbols, using a binary alphabet.

Example: for code \mathcal{A} , one obtains

$$\sum_{i=0}^3 2^{-l_i} = 2^{-2} + 2^{-2} + 2^{-2} + 2^{-2} = 1.$$

Therefore, the codeword lengths of this code are acceptable for an instantaneous code. But, the Kraft inequality does not tell if \mathcal{A} is an instantaneous code. It is only a necessary condition that has to be fulfilled by the lengths.

For the example, the inequality states that there is an instantaneous code with four codewords of length 2. In this case, it is clear that the binary codewords of code \mathcal{A} satisfy the Kraft inequality and also form an instantaneous code.

For code \mathcal{B} ,

$$\sum_{i=0}^3 2^{-l_i} = 2^{-1} + 2^{-3} + 2^{-3} + 2^{-3} = 7/8 \leq 1.$$

In this case, the lengths of the codewords are suitable to compose an instantaneous code. Code \mathcal{B} is also a prefix code.

Table 3.8. Selected binary codes

Source symbols	Code \mathcal{A}	Code \mathcal{B}	Code \mathcal{C}	Code \mathcal{D}	Code \mathcal{E}
x_0	11	1	1	1	1
x_1	10	011	01	011	01
x_2	01	001	001	001	001
x_3	00	000	000	00	00

Code \mathcal{C} is similar to code \mathcal{B} , except for a discarded bit in the second codeword. For this code, one obtains

$$\sum_{i=0}^3 2^{-l_i} = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1.$$

The codeword lengths satisfy the Kraft inequality and, by inspection, one observes that this code is instantaneous.

Code \mathcal{D} is obtained from \mathcal{B} , discarding a bit in the fourth codeword. Although the lengths satisfy the Kraft inequality, code \mathcal{D} is not instantaneous, because it is not a prefix code. The fourth codeword is a prefix of the third one.

Finally, for code \mathcal{E} ,

$$\sum_{i=0}^3 2^{-l_i} = \frac{9}{8},$$

and the codeword lengths do not satisfy the Kraft inequality. Therefore, code \mathcal{E} is not instantaneous.

Consider a source with eight symbols to be encoded into an instantaneous ternary code, whose codeword lengths are 1, 2, 2, 2, 2, 3, 3, 3. Using the Kraft inequality,

$$\sum_{i=0}^9 3^{-l_i} = \frac{1}{3} + 4\frac{1}{9} + 3\frac{1}{27} = \frac{24}{27} < 1,$$

which indicates that this code is possible, as follows:

$$x_0 \leftarrow 0$$

$$x_1 \leftarrow 10$$

$$x_2 \leftarrow 11$$

$$x_3 \leftarrow 20$$

$$x_4 \leftarrow 21$$

$$x_5 \leftarrow 220$$

$$x_6 \leftarrow 221$$

$$x_7 \leftarrow 222$$

For a source with 11 symbols, if the codeword lengths are 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, it is not possible to obtain a ternary instantaneous code, because

$$\sum_{i=0}^{10} 3^{-l_i} = \frac{1}{3} + 6\frac{1}{9} + 4\frac{1}{27} = \frac{31}{27} > 1.$$

3.4 HUFFMAN CODE

This section describes the Huffman coding algorithm, and the procedure to construct the Huffman code when the source statistics are known.

The technique was developed by David Albert Huffman (1925–1999), in a paper for a course on Information Theory taught by Robert Mario Fano (1917–), at the Massachusetts Institute of Technology (MIT). The obtained sequences are called Huffman codes, and they are prefix codes.

Huffman procedure is based on two assumptions regarding the optimum prefix codes:

1. The most frequent symbols, those with higher probability, are represented by shorter codewords.
2. The least frequent symbols are assigned codewords of same length.

According to the first assumption, as the most probable symbols are also the most frequent, they must be as short as possible to decrease the average length of the code. The second assumption is also true, because for a prefix code a shorter codeword could not be a prefix of another one. The least probable symbols must be distinct and have same length (Sayood 2006).

Furthermore, the Huffman process is completed by the addition of a simple requisite. The longer codewords that correspond to the least frequent symbols differ only on the last digit.

3.4.1 CONSTRUCTING A BINARY HUFFMAN CODE

Given a discrete source, a Huffman code can be constructed along the following steps:

1. The source symbols are arranged in decreasing probability. The least probable symbols receive the assignments 0 and 1.
2. Both symbols are combined to create a new source symbol, whose probability is the sum of the original ones. The list is reduced by one symbol. The new symbol is positioned in the list according to its probability.
3. This procedure continues until the list has only two symbols, which receive the assignments 0 and 1.
4. Finally, the binary codeword for each symbol is obtained by a reverse process.

Table 3.9. Discrete source with five symbols and their probabilities

Symbols	Probabilities
x_0	0.4
x_1	0.2
x_2	0.2
x_3	0.1
x_4	0.1

In order to explain the algorithm, consider the source of Table 3.9.

The first phase is to arrange the symbols in a decreasing order of probability. Assign the values 0 and 1 to the symbols with the smallest probabilities. They are then combined to create a new symbol. The probability associated to the new symbol is the sum of the previous probabilities. The new symbol is repositioned in the list, to maintain the same decreasing order for the probabilities. The procedure is shown in Figure 3.2.

The procedure is repeated until only two symbols remain, which are assigned to 0 and 1, as shown in Figure 3.3.

The procedure is repeated to obtain all codewords, by reading the digits in inverse order, from Phase IV to Phase I, as illustrated in Figure 3.4. Following the arrows, for symbol x_4 one finds the codeword 011.

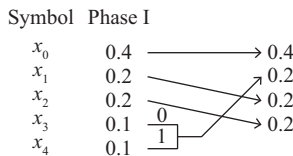


Figure 3.2. Probabilities in descending order for the Huffman code.

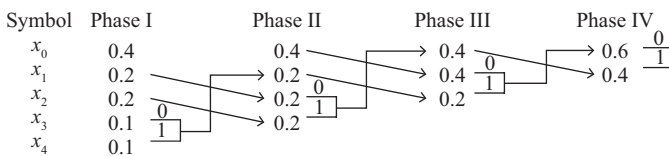


Figure 3.3. Huffman code. At each phase, the two least probable symbols are combined.

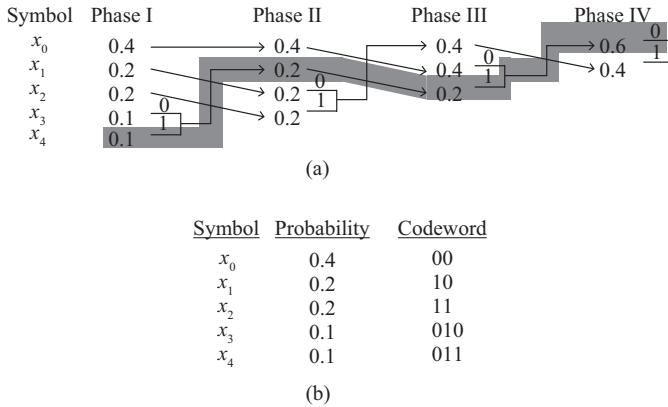


Figure 3.4. (a) Example of the Huffman coding algorithm to obtain the codewords. (b) Resulting code.

For the example, the average codeword length for the Huffman code is given by

$$\bar{L} = \sum_{i=0}^4 p_k l_k = 0,4(2) + 0,2(2) + 0,2(2) + 0,1(3) + 0,1(3) = 2.2 \text{ bits.}$$

The source entropy is calculated as

$$H(X) = \sum_{i=0}^4 p_k \log_2 \left(\frac{1}{p_k} \right)$$

$$H(X) = 0,4 \log_2 \left(\frac{1}{0,4} \right) + 0,2 \log_2 \left(\frac{1}{0,2} \right) + 0,2 \log_2 \left(\frac{1}{0,2} \right) + 0,1 \log_2 \left(\frac{1}{0,1} \right) + 0,1 \log_2 \left(\frac{1}{0,1} \right),$$

or,

$$H(X) = 2.12193 \text{ bits.}$$

The code efficiency is

$$\eta = \frac{H(X)}{\bar{L}} = \frac{2.12193}{2.2},$$

which is equal to 96.45 percent.

It is important to say that the Huffman procedure is not unique, and several variations can be obtained for the final set of codewords, depending

Table 3.10. Four distinct Huffman codes obtained for the source of Table 3.9

Symbols	Code I	Code II	Code III	Code IV
x_0	00	11	1	0
x_1	10	01	01	10
x_2	11	00	000	111
x_3	010	101	0010	1101
x_4	011	100	0011	1100

on the way the bits are assigned. But, in spite of how the probabilities are positioned, the average length is always the same, if the rules are followed.

The difference is the variance of the codeword lengths, defined as

$$V[L] = \sum_{k=0}^{K-1} p_k (l_k - \bar{L})^2, \quad (3.3)$$

in which p_k and l_k denote the probability of occurrence of the k -th source symbol, and length of the respective codeword.

Usually, the procedure of displacing the probability of the new symbol to the highest position in the list produces smaller values for $V[L]$, as compared to the displacement of the probability to the lowest position of the list.

Table 3.10 presents four Huffman codes obtained for the source of Table 3.9. Codes I and II were obtained shifting the new symbol to the highest position in the list of decreasing probabilities.

Codes III and IV were produced by shifting the new symbol to the lowest position in the list. Codes I and III used the systematic assignment of 0 followed by 1 to the least frequent symbols. Codes II and IV used the systematic assignment of 1 followed by 0 to the least frequent symbols. For all codes the average codeword length is 2.2 bits. For codes I and II, the variance of the codeword lengths is 0.16. For codes III and IV, the variance is 1.36.

CHAPTER 4

INFORMATION TRANSMISSION

Claude Elwood Shannon (1916–2001) is considered the father of Information Theory. In 1948, he published a seminal article on the mathematical concept of information, which is one of the most cited for decades. Information left the Journalism field to occupy a more formal area, as part of Probability Theory.

The entropy, in the context of information theory, was initially defined by Ralph Vinton Lyon Hartley (1888–1970), in the article “Transmission of Information”, published by *Bell System Technical Journal* in July 1928, 10 years before the formalization of the concept by Claude Shannon.

Shannon’s development was also based on Harry Nyquist’s work (Harry Theodor Nyqvist, 1889–1976), which determined the sampling rate, as a function of frequency, necessary to reconstruct an analog signal using a set of discrete samples.

In an independent way, Andrei N. Kolmogorov developed his Complexity Theory, during the 1960’s. It was a new information theory, based on the length of an algorithm developed to describe a certain data sequence. He used Alan Turing’s machine in this new definition. Under certain conditions, Kolmogorov’s and Shannon’s definitions are equivalent.

The idea of relating the number of states of a system with a physical measure, although, dates back to the 19th century. Rudolph Clausius proposed the term entropy for such a measure in 1895.

Entropy comes from the Greek word for transformation and in Physics, it is related to the logarithm of the ratio between the final and initial temperature of a system, or to the ratio of the heat variation and the temperature of the same system.

Shannon defined the entropy of an alphabet at the negative of the mean value of the logarithm of the symbols’ probability. This way, when the symbols are equiprobable, the definition is equivalent to that of Nyquist’s.

But, as a more generic definition, Shannon's entropy can be used to compute the capacity of communication channels. Most part of the researchers' work is devoted to either compute the capacity or to develop error-correcting codes to attain that capacity.

Shannon died on February 24, 2001, as a victim of a disease named after the physician Aloysius Alzheimer. According to his wife, he lived a quiet life, but had lost his capacity to retain information.

4.1 THE CONCEPT OF INFORMATION THEORY

The concept of information transmission is associated with the existence of a communication channel that links the source and destination of the message. This can imply the occurrence of transmission errors, caused by the probabilistic nature of the channel. Figure 4.1 illustrates the canonical model for a communication channel, proposed by Shannon in his seminal paper of 1948. This is a very simplified model of reality, but contains the basic blocks upon which the mathematical structure is built.

4.2 JOINT INFORMATION MEASUREMENT

Consider two discrete and finite sample spaces, Ω and Ψ , with the associated random variables X and Y ,

$$\begin{aligned} X &= x_1, x_2, \dots, x_N, \\ Y &= y_1, y_2, \dots, y_M. \end{aligned} \tag{4.1}$$

The events from Ω may jointly occur with events from Ψ . Therefore, the following matrix contains the whole set of events in the product

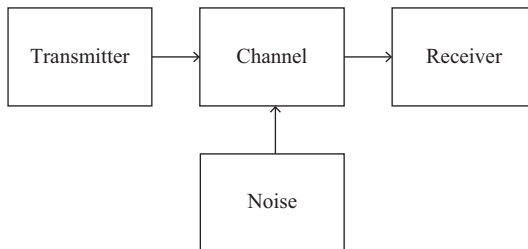


Figure 4.1. Model for a communication channel.

space $\Omega\Psi$,

$$[XY] = \begin{bmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_M \\ x_2y_1 & x_2y_2 & \cdots & x_2y_M \\ \cdots & \cdots & \cdots & \cdots \\ x_Ny_1 & x_Ny_2 & \cdots & x_Ny_M \end{bmatrix} \quad (4.2)$$

The joint probability matrix is given in the following in which no restriction is assumed regarding the dependence between the random variables.

$$[P(X, Y)] = \begin{bmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,M} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,M} \\ \cdots & \cdots & \cdots & \cdots \\ p_{N,1} & p_{N,2} & \cdots & p_{N,M} \end{bmatrix} \quad (4.3)$$

Figure 4.2 shows the relation between the input and output alphabets, which are connected by the joint probability distribution matrix $[P(X, Y)]$.

The joint entropy between the random variables from sources X and Y is given by

$$H(X, Y) = - \sum_{k=1}^N \sum_{j=1}^M p_{k,j} \log p_{k,j}, \quad (4.4)$$

which may be simplified to

$$H(X, Y) = - \sum_X \sum_Y p(x, y) \log p(x, y). \quad (4.5)$$

The marginal entropies may be written in terms of the marginal probabilities, $p(x)$ and $p(y)$

$$H(X) = - \sum_X p(x) \log p(x) \quad (4.6)$$

and

$$H(Y) = - \sum_Y p(y) \log p(y). \quad (4.7)$$

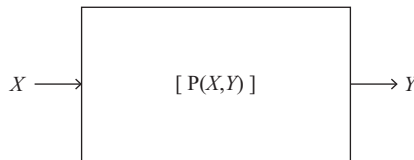


Figure 4.2. A probabilistic communication channel.

4.3 CONDITIONAL ENTROPY

The concept of conditional entropy is essential to model, and understand, the operation of the communication channel, because it provides information about a particular symbol, given that another symbol has occurred. The entropy of alphabet X conditioned to the occurrence of a particular symbol y is given by

$$\begin{aligned} H(X|y) &= - \sum_X \frac{p(x,y)}{p(y)} \log \frac{p(x,y)}{p(y)} \\ &= - \sum_X p(x|y) \log p(x|y). \end{aligned} \quad (4.8)$$

The expected value of the conditional entropy, for all possible values of y , provides the average conditional entropy of the system,

$$\begin{aligned} H(X|Y) &= E[H(X|y)] = \sum_Y p(y)[H(X|y)] \\ &= - \sum_Y p(y) \sum_X p(x|y) \log p(x|y), \end{aligned} \quad (4.9)$$

which can be written as

$$H(X|Y) = - \sum_Y \sum_X p(y)p(x|y) \log p(x|y), \quad (4.10)$$

or

$$H(X|Y) = - \sum_Y \sum_X p(x,y) \log p(x|y). \quad (4.11)$$

In the same way, the mean conditional entropy of source Y , given the information about source X , is

$$H(Y|X) = - \sum_X \sum_Y p(x)p(y|x) \log p(y|x), \quad (4.12)$$

or

$$H(Y|X) = - \sum_X \sum_Y p(x,y) \log p(y|x). \quad (4.13)$$

4.4 MODEL FOR A COMMUNICATION CHANNEL

A communication channel can be modeled based on the previous developments. Consider a source that has the given alphabet X . The source

transmits the information to the destiny using a certain channel. The system maybe described by a joint probability matrix, which gives the joint probability of occurrence of a transmitted symbol and a received one,

$$[P(X, Y)] = \begin{bmatrix} p(x_1, y_1) & p(x_1, y_2) & \cdots & p(x_1, y_N) \\ p(x_2, y_1) & p(x_2, y_2) & \cdots & p(x_2, y_N) \\ \cdots & \cdots & \cdots & \cdots \\ p(x_M, y_1) & p(x_M, y_2) & \cdots & p(x_M, y_N) \end{bmatrix} \quad (4.14)$$

There are five probability schemes to analyze:

1. $[P(X, Y)]$, joint probability matrix,
2. $[P(X)]$, marginal probability matrix of X ,
3. $[P(Y)]$, marginal probability matrix of Y ,
4. $[P(X|Y)]$, probability matrix conditioned on Y ,
5. $[P(Y|X)]$, probability matrix conditioned on X .

Those probability schemes produce five entropy functions, associated to the communication channel, whose interpretations are given as follows:

1. $H(X)$ —Average information per source symbol, or source entropy,
2. $H(Y)$ —Average information per received symbol, or receiver entropy,
3. $H(X, Y)$ —Average information associated to pairs of transmitted and received symbols, or average uncertainty of the communication system,
4. $H(X|Y)$ —Average information measurement of the received symbol, given that X was transmitted, or conditional entropy,
5. $H(Y|X)$ —Average information measurement of the source, given that Y was received, or equivocation.

4.5 NOISELESS CHANNEL

For the noiseless discrete channel, each symbol from the input alphabet has a one-to-one correspondence with the output. The joint probability matrix, as well as, the transition probability matrix, has the same diagonal format,

$$[P(X, Y)] = \begin{bmatrix} p(x_1, y_1) & 0 & \cdots & 0 \\ 0 & p(x_2, y_2) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & p(x_N, y_N) \end{bmatrix} \quad (4.15)$$

$$[P(X|Y)] = [P(Y|X)] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \quad (4.16)$$

The joint entropy equals the marginal entropies

$$H(X, Y) = H(X) = H(Y) = - \sum_{i=1}^N p(x_i, y_i) \log p(x_i, y_i), \quad (4.17)$$

and the conditional entropies are null

$$H(Y|X) = H(X|Y) = 0. \quad (4.18)$$

As a consequence, the receiver uncertainty is equal to the source entropy, and there is no ambiguity at the reception, which indicates that the conditional entropies are all zero.

4.6 CHANNEL WITH INDEPENDENT OUTPUT AND INPUT

For the channel with independent input and output, there is no relation between the transmitted and received symbols, that is, given that a given symbol has been transmitted, any symbol can be received, with no connection whatsoever with it. The joint probability matrix has N identical columns

$$[P(X, Y)] = \begin{bmatrix} p & p_1 & \cdots & p_1 \\ p_2 & p_2 & \cdots & p_2 \\ \cdots & \cdots & \cdots & \cdots \\ p_M & p_M & \cdots & p_M \end{bmatrix}, \quad \sum_i^M p_i = \frac{1}{N}. \quad (4.19)$$

The input and output symbol probabilities are statistically independent, that is,

$$p(x, y) = p(x)p(y). \quad (4.20)$$

Computing the entropy gives

$$H(X, Y) = -N \left(\sum_{i=1}^M p_i \log p_i \right), \quad (4.21)$$

$$H(X) = - \sum_{i=1}^M N p_i \log N p_i = -N \left(\sum_{i=1}^M p_i \log p_i \right) - \log N, \quad (4.22)$$

$$H(Y) = -N \left(\frac{1}{N} \log \frac{1}{N} \right) = \log N, \quad (4.23)$$

$$H(X|Y) = - \sum_{i=1}^M Np_i \log Np_i = H(X), \quad (4.24)$$

$$H(Y|X) = - \sum_{i=1}^M Np_i \log \frac{1}{N} = \log N = H(Y). \quad (4.25)$$

As a consequence, the channel with independent input and output does not provide information, that is, has the highest possible loss, contrasting with the noiseless channel.

4.7 RELATIONS BETWEEN THE ENTROPIES

It is possible to show, using Bayes rule for the conditional probability, that the joint entropy can be written in terms of the conditional entropy, in the following way

$$H(X, Y) = H(X|Y) + H(Y), \quad (4.26)$$

$$H(X, Y) = H(Y|X) + H(X). \quad (4.27)$$

Shannon has shown the fundamental inequality

$$H(X) \geq H(X|Y), \quad (4.28)$$

whose demonstration is given in the following.

The logarithm concavity property can be used to demonstrate the inequality, $\ln x \leq x - 1$, as follows,

$$\begin{aligned} H(X|Y) - H(X) &= \sum_Y \sum_X p(x, y) \log \frac{p(x)}{p(x|y)} \\ &\leq \sum_Y \sum_X p(x, y) \left(\frac{p(x)}{p(x|y)} - 1 \right) \log e. \end{aligned} \quad (4.29)$$

But, the right hand side of the inequality is zero, as shown in the following

$$\begin{aligned} \sum_Y \sum_X (p(x) \cdot p(y) - p(x, y)) \log e &= \sum_Y (p(y) - p(y)) \log e \\ &= 0. \end{aligned} \quad (4.30)$$

Therefore,

$$H(X) \geq H(X|Y). \quad (4.31)$$

In a similar manner, it can be shown that

$$H(Y) \geq H(Y|X). \quad (4.32)$$

The equality is attained if and only if X and Y are statistically independent.

4.8 MUTUAL INFORMATION

A measure of mutual information provided by two symbols (x_i, y_i) can be written as

$$\begin{aligned} I(x_i; y_j) &= \log_2 p(x_i|y_j) - \log_2 p(x_i) \\ &= \log_2 \frac{p(x_i|y_j)}{p(x_i)} = \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}. \end{aligned} \quad (4.33)$$

It can be noticed that the *a priori* information of symbol x_i is contained in the marginal probability $p(x_i)$. The *a posteriori* probability that symbol x_i has been transmitted, given that y_j was received is $p(x_i|y_j)$. Therefore, in an informal way, the information gain for the observed symbol equals the difference between the initial information, or uncertainty, and the final one.

The mutual information is continuous in $p(x_i|y_i)$, and also symmetric, or

$$I(x_i; y_j) = I(y_j; x_i), \quad (4.34)$$

which indicates that the information provided by x_i about y_j is the same provided by y_j about x_i .

The function $I(x_i; x_i)$ can be called the auto-information of x_i , or

$$I(x_i) = I(x_i; x_i) = \log \frac{1}{p(x_i)}, \quad (4.35)$$

because, for an observer of the source alphabet, the *a priori* knowledge of the situation is that x_i will be transmitted with probability $p(x_i)$, and the *a posteriori* knowledge is the certainty that x_i transmitted.

In conclusion,

$$I(x_i; y_j) \leq I(x_i; x_i) = I(x_i), \quad (4.36)$$

$$I(x_i; y_j) \leq I(y_j; y_j) = I(y_j). \quad (4.37)$$

The statistical mean of the mutual information per pairs of symbols provides an interesting interpretation of the mutual information concept,

$$I(X; Y) = E[I(x_i; y_j)] = \sum_i \sum_j p(x_i, y_j) I(x_i; y_j), \quad (4.38)$$

which can be written as

$$I(X; Y) = \sum_i \sum_j p(x_i, y_j) \log \frac{p(x_i|y_j)}{p(x_i)}. \quad (4.39)$$

The average mutual information can be interpreted as a reduction on the uncertainty about the input X , when the output Y is observed (MacKay 2003). This definition provides an adequate metric for the average mutual information of all pairs of symbols, and can be put in terms of the entropy, such as

$$I(X; Y) = H(X) + H(Y) - H(X, Y), \quad (4.40)$$

$$I(X; Y) = H(X) - H(X|Y), \quad (4.41)$$

$$I(X; Y) = H(Y) - H(Y|X). \quad (4.42)$$

Put that way, the average mutual information gives a measure of the information that is transmitted by the channel. Because of this, it is called transinformation, or information transferred by the channel. It is always non-negative, even if the individual information quantities are negative for certain pairs of symbols.

For a noiseless channel, the average mutual information equals the joint entropy.

$$I(X; Y) = H(X) = H(Y), \quad (4.43)$$

$$I(X; Y) = H(X, Y). \quad (4.44)$$

On the other hand, for a channel in which the output is independent of the input, the average mutual information is null, implying that no information is transmitted by the channel.

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y), \\ &= H(X) - H(X) = 0. \end{aligned} \quad (4.45)$$

It is possible to use the set theory, presented in the Appendix, to obtain a pictorial interpretation of the fundamental inequalities discovered by Shannon. Consider, the set of events A and B , associated to the set of symbols X and Y , and a Lebesgue measure m . It is possible to associate,

unambiguously, the measures of A and B with the entropies of X and Y (Reza 1961),

$$m(A) \longleftrightarrow H(X), \quad (4.46)$$

$$m(B) \longleftrightarrow H(Y). \quad (4.47)$$

In the same way, one can associate the joint and conditional entropies the union and intersection of sets, respectively,

$$m(A \cup B) \longleftrightarrow H(X, Y), \quad (4.48)$$

$$m(A\bar{B}) \longleftrightarrow H(X|Y), \quad (4.49)$$

$$m(\bar{B}A) \longleftrightarrow H(Y|X). \quad (4.50)$$

The average mutual information is, therefore, associated to the measure of the intersection of the sets,

$$m(A \cap B) \longleftrightarrow I(X; Y). \quad (4.51)$$

Those relations can be seen in Figure 4.3, which also serve as a means to memorize the entropy and mutual information properties.

Therefore, the fundamental inequalities can be written as a result of set operations (Reza 1961),

$$m(A \cup B) \leq m(A) + m(B) \longleftrightarrow H(X, Y) \leq H(X) + H(Y), \quad (4.52)$$

$$m(A\bar{B}) \leq m(A) \longleftrightarrow H(X|Y) \leq H(X), \quad (4.53)$$

$$m(\bar{B}A) \leq m(B) \longleftrightarrow H(Y|X) \leq H(Y), \quad (4.54)$$

and, finally,

$$m(A \cup B) = m(A\bar{B}) + m(\bar{B}A) + m(A \cap B), \quad (4.55)$$

which is equivalent to

$$H(X, Y) = H(X|Y) + H(Y|X) + I(X; Y). \quad (4.56)$$

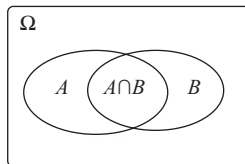


Figure 4.3. Venn diagram corresponding to the relation between the entropies.

For a noiseless channel, the two sets coincide, and the relations can be written as:

$$m(A) = m(B) \longleftrightarrow H(X) = H(Y), \quad (4.57)$$

$$m(A \cup B) = m(A) = m(B) \longleftrightarrow H(X, Y) = H(X) = H(Y), \quad (4.58)$$

$$m(A\bar{B}) = 0 \longleftrightarrow H(X|Y) = 0, \quad (4.59)$$

$$m(B\bar{A}) = 0 \longleftrightarrow H(Y|X) = 0, \quad (4.60)$$

and, finally,

$$\begin{aligned} m(A \cap B) &= m(A) = m(B) = m(A \cup B) \\ &\longleftrightarrow I(X; Y) = H(X) = H(Y) = H(X, Y). \end{aligned} \quad (4.61)$$

For a channel with output symbols independent from the input symbols, the sets A and B are considered mutually exclusive, therefore:

$$m(A \cup B) = m(A) + m(B) \longleftrightarrow H(X, Y) = H(X) + H(Y) \quad (4.62)$$

$$m(A\bar{B}) = m(A) \longleftrightarrow H(X|Y) = H(X) \quad (4.63)$$

$$m(A \cap B) = 0 \longleftrightarrow I(X; Y) = 0. \quad (4.64)$$

The same procedure can be applied to a multiple port channel. For a three port channel, one can obtain the following relations for the entropies:

$$H(X, Y, Z) \leq H(X) + H(Y) + H(Z), \quad (4.65)$$

$$H(Z|X, Y) \leq H(Z|Y). \quad (4.66)$$

In the same reasoning, it is possible to obtain the following relations for the average mutual information for a three-port channel:

$$I(X; Y, Z) = I(X; Y) + I(X; Z|Y), \quad (4.67)$$

$$I(Y, Z; X) = I(Y; X) + I(Z; X|Y). \quad (4.68)$$

4.9 CHANNEL CAPACITY

Shannon defined the discrete channel capacity as the maximum of the average mutual information, computed for all possible probability sets that

can be associated to the input symbol alphabet, that is, for all memoryless sources,

$$C = \max I(X; Y) = \max [H(X) - H(X|Y)]. \quad (4.69)$$

4.9.1 CAPACITY OF THE MEMORYLESS DISCRETE CHANNEL

Consider X as the alphabet of a source with N symbols. Because the transition probability matrix is diagonal, one obtains

$$C = \max I(X; Y) = \max [H(X)] = \max \left[- \sum_{i=1}^N p(x_i) \log p(x_i) \right]. \quad (4.70)$$

Example: the entropy attains a maximum when all symbols are equiprobable. Therefore, for the memoryless discrete channel, the capacity is

$$C = \max \left[- \sum_{i=1}^N \frac{1}{N} \log \frac{1}{N} \right],$$

which gives

$$C = \log N \quad \text{bits per symbol.} \quad (4.71)$$

The channel capacity can also be expressed in bits per second, or shannon (Sh), and corresponds to the information transmission rate of the channel, for symbols with duration T seconds,

$$C_T = \frac{C}{T} \quad \text{bits per second, or Sh.} \quad (4.72)$$

Therefore, for the noiseless channel,

$$C_T = \frac{C}{T} = \frac{1}{T} \log N \quad \text{bits per second, or Sh.} \quad (4.73)$$

4.9.2 RELATIVE REDUNDANCY AND EFFICIENCY

The absolute redundancy is the difference between the actual information transmission rate and $I(X; Y)$ and the maximum possible value,

$$\begin{aligned} \text{Absolute redundancy for a noisy channel} &= C - I(X; Y) \\ &= \log N - H(X). \end{aligned} \quad (4.74)$$

The ratio between the absolute redundancy and the channel capacity is defined as the system relative redundancy,

$$\begin{aligned} \text{Relative redundancy for a noiseless channel, } D &= \frac{\log N - H(X)}{\log N} \\ &= 1 - \frac{H(X)}{\log N}. \end{aligned} \quad (4.75)$$

The system efficiency is defined as the complement of the relative redundancy,

$$\begin{aligned} \text{Efficiency of the noiseless channel, } E &= \frac{I(X; Y)}{\log N} = \frac{H(X)}{\log N} \\ &= 1 - D. \end{aligned} \quad (4.76)$$

When the transmitted symbols do not occupy the same time interval, it is still possible to define the average information transmission rate for the noiseless channel, as

$$R_T = \frac{-\sum_{i=1}^N p(x_i) \log p(x_i)}{\sum_{i=1}^N p(x_i) T_i}, \quad (4.77)$$

in which T_i represent the symbol intervals.

For a discrete noisy channel the capacity is the maximum of the average mutual information, when the noise characteristic $p(y_i|x_i)$ is specified.

$$C = \max \left(\sum_{i=1}^N \sum_{j=1}^M p(x_i) p(y_j|x_i) \log \frac{p(y_j|x_i)}{p(y_j)} \right), \quad (4.78)$$

in which the maximum is over $p(x_i)$. It must be noticed that the maximization in relation to the input probabilities do not always leads to an admissible set of source probabilities.

Bayes' rule defines the relation between the marginal probabilities $p(y_j)$ and the *a priori* probabilities $p(x_i)$,

$$p(y_j) = \sum_{i=1}^N p_1(x_i) p(y_j|x_i), \quad (4.79)$$

in which the variable are restricted to the following conditions:

$$p(x_i) \geq 0 \quad i = 1, 2, \dots, N, \quad (4.80)$$

$$\sum_{i=1}^N p_1(x_i) = 1.$$

Example: determine the capacity of the memoryless Binary Symmetric Channel (BSC) (Blake 1987).

The channel is illustrated in Figure 4.4, in which the error probability, or the probability of transmitting a symbol and receiving another one, is indicated as p .

The channel capacity is given by Formula 4.69,

$$C = \max I(X; Y),$$

in which the average mutual information can be written as

$$I(X; Y) = \sum_{i=0}^1 \sum_{j=0}^1 p_{ij} \log \frac{p_{ij}}{p_i q_j}. \quad (4.81)$$

Assume that the symbol *a priori* probabilities are $p_0 = r$ and $p_1 = v$, with $r + v = 1$. Probabilities r and v are chosen in such a manner that, for each channel use, the maximum quantity of information is transferred.

The joint probabilities are

$$p_{00} = r(1 - p), \quad p_{01} = rp, \quad p_{10} = (1 - r)p, \quad p_{11} = (1 - r)(1 - p).$$

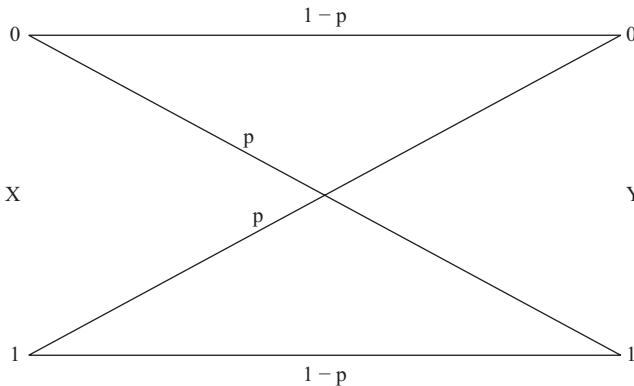


Figure 4.4. Memoryless binary symmetric channel.

and the average mutual information can be written as

$$\begin{aligned}
 I(X; Y) = & [(1-p)r] \log \left(\frac{(1-p)r}{r[(1-p)r + (1-r)p]} \right) \\
 & + [rp] \log \left(\frac{rp}{r[rp + (1-r)(1-p)]} \right) \\
 & + [(1-r)p] \log \left(\frac{(1-r)p}{(1-r)[r(1-p) + (1-r)p]} \right) \\
 & + [(1-p)(1-r)] \log \left(\frac{(1-p)(1-r)}{(1-r)[rp + (1-r)(1-p)]} \right),
 \end{aligned}$$

that can be put in the following way, after the simplification with logarithm properties,

$$\begin{aligned}
 I(X; Y) = & p \log p + (1-p) \log (1-p) \\
 & - [r(1-p) + (1-r)p] \log [r(1-p) + (1-r)p] \\
 & + [rp + (1-r)(1-p)] \log [rp + (1-r)(1-p)].
 \end{aligned}$$

The objective is to determine the value of r that maximizes the expression, taking into account that the logarithms are base two. The obtained expression for r is a complicated one, but the maximum that the average mutual information attains is given by

$$C = \max I(X; Y) = 1 - p \log p + (1-p) \log (1-p), \quad (4.82)$$

that represents the memoryless binary symmetric channel capacity. The graph for the capacity $C(p)$, as a function of the channel error probability, is shown in Figure 4.5.

Example: determine the capacity of the Binary Erasure Channel (BEC) shown in Figure 4.6, in which the parameter E represents the erasure and $1-p$ the occurrence probability (Blake 1987).

The average mutual information for this channel is given by

$$I(X; Y) = \sum_{i=0}^1 \sum_{j=0}^2 p_{ij} \log \frac{p_{ij}}{p_i q_j}, \quad (4.83)$$

and the probabilities p_{ij} are the following, for $p_0 = r$ and $p_1 = v$, with $r + v = 1$,

$$\begin{aligned}
 p_{00} = & rp, \quad p_{01} = r(1-p), \quad p_{10} = 0, \quad p_{11} = (1-r)(1-p), \quad p_{02} = 0, \\
 p_{12} = & (1-r)p.
 \end{aligned}$$

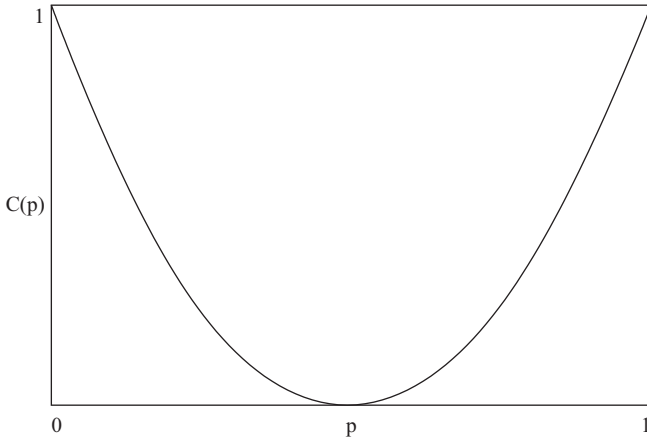


Figure 4.5. Graph for the capacity of the memoryless binary symmetric channel.

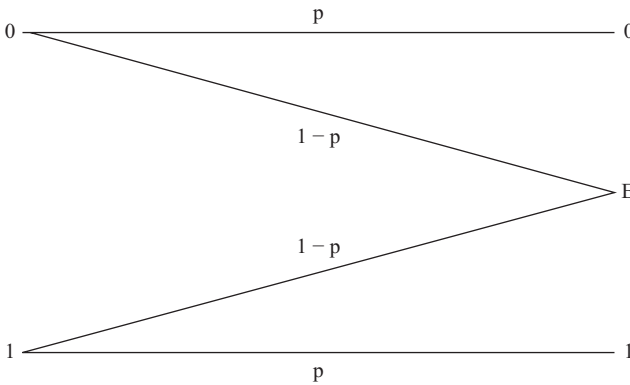


Figure 4.6. Binary erasure channel.

Substituting the probability values in Equation 4.83, one obtains

$$\begin{aligned}
 I(X; Y) &= [rp] \log \left(\frac{rp}{r^2p} \right) \\
 &\quad + [(1-p)r] \log \left(\frac{(1-p)r}{(1-p)r} \right), \\
 &\quad + [(1-r)(1-p)] \log \left(\frac{(1-r)(1-p)}{(1-p)^2} \right) \\
 &\quad + [(1-r)p] \log \left(\frac{(1-r)p}{(1-r)^2p} \right),
 \end{aligned}$$

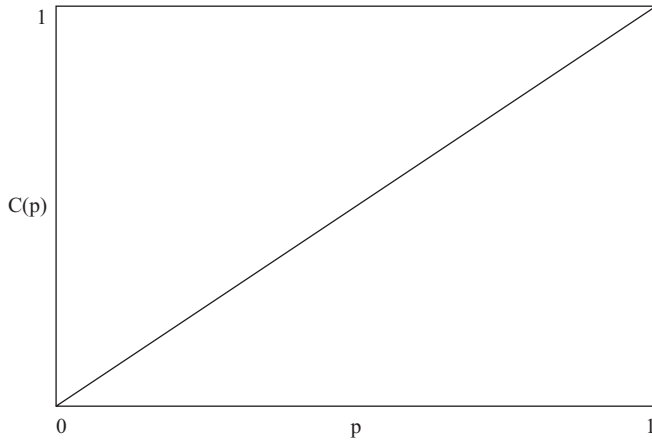


Figure 4.7. Graph of the capacity for the binary erasure channel.

Simplifying the terms in the expression, gives

$$I(X; Y) = p[r \log r - (1 - r) \log (1 - r)] = pH(r), \quad (4.84)$$

in which $H(r)$ is the entropy function.

Because p is determined, $I(X; Y)$ is maximized by the choice of a value for r that produces the highest $H(r)$, that is, $r = 1/2$, for which $H(r) = 1$. Therefore, the capacity of the binary erasure channel is simply p . Figure 4.7 shows the graph for the capacity as a function of the probability p .

MULTIPLE ACCESS SYSTEMS

5.1 INTRODUCTION

The multiple access channel can be analyzed through the use of an information theoretic approach. In this type of communications system, several users transmit simultaneously to a common receiver. The main multiplexing strategies that fall in that category include frequency division multiple access (FDMA), time division multiple access (TDMA), and code division multiple access (CDMA).

The main objective of the information theoretic approach of the multiple access channel is to find its capacity region, that is, the set of information rates at which the simultaneous reliable communication of the messages of each user is possible (Verdú 1989b).

The problem of finding the capacity region of a multiple access channel can be traced to a pioneering work on two-way communication channels by Shannon, in 1961 (Shannon 1961). A similar approach led to the multi-way solution for the discrete memoryless channel (Ahlsvede 1971). The related problem of simultaneous communication from one source to several receivers was proposed in 1971, and solutions were given to the cases of orthogonal, Gaussian, and compound channels (Cover 1972).

The usual assumption for the Shannon channel was a strategic cooperation between the senders and the receiver. This approach changed in 1981, when the lack of synchronization was proven not to reduce the capacity region for the multiple access channel (Cover, McEliece, and Posner 1981). Later it was shown, for the two-user case, that the only effect of frame asynchronism on the discrete memoryless channel is the removal of the convex hull operation from the expression of the capacity region.

The same result could be extended to the multiuser case and implies that the maximum achievable rate sum, or total capacity, of the memoryless multiple access channel is never decreased by the lack of frame synchronism. This is a direct implication of the convexity property, because the rate sum of any convex combination of rate pairs is equal to the convex combination of the respective rate sums (Hui and Humblet 1985).

Further study on the asynchronous Gaussian multiple access channel established the nonoptimality of the conventional receiver (independent detection of the signals), which is a result of the additive component of multiple access interference (MAI). Fortunately for CDMA systems, in which the designer is allowed to choose a signal constellation with large bandwidth, the cross-correlations between the signals can be kept low for all relative delays and an acceptable performance can be achieved (Verdú 1986).

A paper published in 1975 showed, by means of an example, that the use of feedback could increase the capacity region of the multiple access memoryless channel. But the problem of determining the actual capacity region, when feedback is available, remained unsolved (Gaarder and Wolf 1975). A point to be stressed here is that, despite the efforts of many researchers in the field, the capacity region of the multiterminal communication channel remains, in general, unknown. A general formula for the capacity of such channels is expressed by (Mandell and McEliece 1991).

The role of the memory in the multiple access environment was stressed in the article (Verdú 1989a). In fact, the study of asynchronous multiple access channels can be enriched with the theory of multiple access channels with memory, in which the effect on the current symbol depends on the previous state of the channel. The usual approach to the design of codes for channels with memory has been the assumption of interleaving, which provides an interesting trade-off between decoding complexity and storage capacity.

In one of the first works to suggest this technique, interleaving appeared as a type of time division multiplexing (Elliott 1965). The issue there was that interleaving could enhance the error-control effectiveness of error-correcting codes, when the separation between bits of a code word is on the order of several hundred bits. The purpose of interleaving is to provide randomization, in order to mitigate potential burst errors.

There are many different types of channels that can be characterized by the nature of errors that are generated between transmitter and receiver. On a bursty channel, given a burst of symbols, an interleaving scheme can be used to spread the errors to many codewords. A suitable error-correcting code could then be utilized to correct the remaining errors in

each codeword. An optimum interleaver is then one which utilizes the least amount of memory and has the smallest average delay between the input and output of a symbol from the interleaver, besides providing the necessary randomization of the data.

The symbol-asynchronous Gaussian multiple access channel encompasses many interesting models. The direct sequence spread spectrum multiple access channel represents a possible application for some of those models. The capacity region of the symbol-asynchronous Gaussian multiple access channel was obtained recently (Verdú 1989b). However, further study is necessary to obtain the capacity region for the asynchronous Gaussian multiple access channel with memory. This could be helpful in modeling the wireless digital communications channel as introduced in the last chapter.

Moreover, for channels with memory, frame asynchronism may drastically reduce the capacity region and, in particular, the maximum achievable rate sum. A useful model for a channel with memory will be introduced in the last section of this chapter and analyzed in some detail.

5.2 THE GAUSSIAN MULTIPLE ACCESS CHANNEL

The multiple access channel can be defined as a system in which M transmitters, $X_i \in \mathcal{X}_i$, simultaneously communicate to a common receiver, $Y \in \mathcal{Y}$. It is characterized by the transition probability matrix $\mathbf{P}(y|x_1, \dots, x_M)$, which represents the probabilities associated with output y when the M inputs are x_1, \dots, x_M . The channel is assumed memoryless, and the alphabets are considered continuous sets. When the noise affecting the transmitted signals is considered Gaussian, the system is called the Gaussian multiple access channel, which was studied in detail by (Wyner 1974).

The choice of Gaussian signals to convey information in the additive Gaussian channel is not a matter of chance. The use of a set of Gaussian waveforms maximizes the mutual information for this kind of channel (Sommer 1966). The multiple access channel is shown in Figure 5.1.

In the following, a general idea on the computation of the capacity region is presented. Consider that the output is $Y = \sum_i X_i + Z$, in which the X_i are inputs and Z is a Gaussian noise random variable independent of the inputs, with variance σ_Z^2 . The encoding constraint is that the messages, which are n -vectors, satisfy $E[(1/n)\|\mathbf{X}_i\|^2] \leq \sigma_i^2$. The Euclidean norm $\|\cdot\|$ is used for the vectors in the preceding formula. Besides, one imposes the

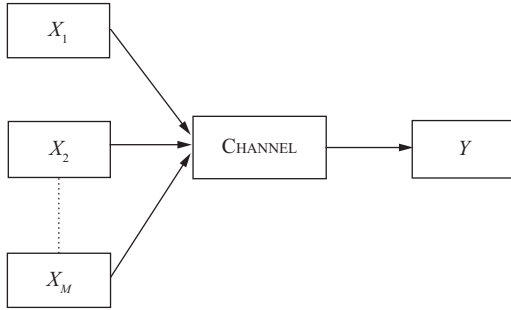


Figure 5.1. The multiple access channel.

additional condition that $E[X_i^2] \leq \sigma_i^2$, $i = 1, \dots, M$ (Wyner 1974). In the following, all logarithms are to the base two, which implies that information rates, and capacities, are measured in bits.

The mutual information can be defined for the set of inputs and output as the difference between the output entropy $H(Y)$ and the noise entropy $H(Z)$

$$I(X_1, \dots, X_M; Y) = H(Y) - H(Y|X_1, \dots, X_M) \quad (5.1)$$

$$= H(Y) - H(Z). \quad (5.2)$$

Considering the independence between the inputs and the Gaussian noise, the variance of the output can be written

$$V[Y] = \sum_{i=1}^M V[X_i] + V[Z] \quad (5.3)$$

$$\leq \sum_{i=1}^M \sigma_i^2 + \sigma_Z^2. \quad (5.4)$$

Therefore, the mutual information is bounded by

$$I(X_1, \dots, X_M; Y) \leq \frac{1}{2} \log 2\pi e \left(\sum_{i=1}^M \sigma_i^2 + \sigma_Z^2 \right) - \frac{1}{2} \log 2\pi e \sigma_Z^2 = \frac{1}{2} \log \left(\frac{\sum_{i=1}^M \sigma_i^2 + \sigma_Z^2}{\sigma_Z^2} \right) \quad (5.5)$$

$$= \frac{1}{2} \log(1 + S^2) \triangleq C. \quad (5.6)$$

in which the total signal-to-noise ratio (SNR) is defined as

$$S^2 = \frac{\sum_{i=1}^M \sigma_i^2}{\sigma_Z^2}. \quad (5.7)$$

Furthermore,

$$\begin{aligned} I(X_i; Y|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_M) \\ = H(Y|X_i) - H(Y|X_1, \dots, X_M) \end{aligned} \quad (5.8)$$

$$\text{for } i = 1, \dots, M \quad (5.9)$$

which is bounded as

$$\begin{aligned} I(X_i; Y|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_M) \\ \leq \frac{1}{2} \log \left(\frac{\sigma_i^2 + \sigma_Z^2}{\sigma_Z^2} \right) \\ = \frac{1}{2} \log(1 + s_i^2) \triangleq C_i \\ \text{for } i = 1, \dots, M \end{aligned} \quad (5.10)$$

in which s_i^2 is

$$s_i^2 = \frac{\sigma_i^2}{\sigma_Z^2}. \quad (5.11)$$

A multiuser discrete multiple access channel satisfies

$$\begin{aligned} R_i &\leq I(X_i; Y|X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_M), \\ &\text{for } i = 1, \dots, M \\ R_1 + \dots + R_M &\leq I(X_1, \dots, X_M; Y) \end{aligned} \quad (5.12)$$

in which the set of R_i represents the transmission rates for M independent input signals X_i . The output signal is Y . The boundary of this region is called the convex hull.

Example: the set of all achievable rates must lie inside the region defined by the inequalities $R_1 + \dots + R_M \leq C$ and $R_i \leq C_i$, as depicted in Figure 5.2, for the case of two users.

For the case of three users, the set of equations above define an upper bound on the capacity, because other constraints can be imposed on the partial sums of transmission rates.

The channel capacity, or the maximum possible transmission rate, can only be attained on the convex hull, or contour surface. Thus time-sharing could be necessary to achieve some of the points in the capacity region, for

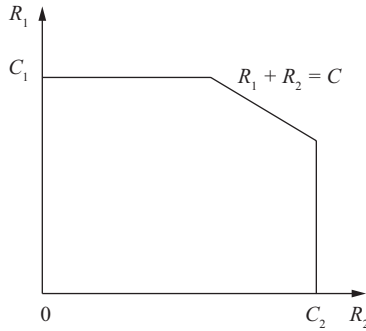


Figure 5.2. Capacity region for the Gaussian multiple access channel, $M = 2$.

instance if $C_i = C$ (Bierbaum and Wallmeier 1979). This model assumes cooperative users, which can be translated into a time-sharing environment. The transmission of information without any prearrangement between the users results in a degradation of the capacity region (Sommer 1966), which can be minimized with the use of multiplexing techniques such as CDMA.

5.3 THE GAUSSIAN CHANNEL WITH RAYLEIGH FADING

Consider that the output is $Y = \sum_i r_i X_i + Z$, in which the X_i are the transmitted signal, Z is a Gaussian noise random variable independent of the inputs, with variance σ_Z^2 , and r_i are the attenuation coefficients associated with the fading channel. The channel is assumed memoryless, and the alphabets are considered continuous sets.

The choice of Gaussian signals to convey information in the additive Gaussian channel does not optimize the transmission rate any more. The accuracy of a simple approximation to the capacity of a fading, additive Gaussian noise channel is discussed for the single user case. The special case of Rayleigh fading is considered in greater detail and the signal distribution that yields capacity for this case is found.

The random variables r_i are the attenuation parameters defined in Chapter 2, which are assumed to follow a Rayleigh distribution given by

$$p_R(r) = \frac{r}{\gamma^2} e^{-\frac{r^2}{2\gamma^2}} u(r). \quad (5.13)$$

The output is not guaranteed to be Gaussian, for a Gaussian input, in the fading channel. This is the reason for not assuming an input Gaussian distribution in the first hand. The product of the signal versions and the attenuation factors yield a distribution that is not Gaussian, in the general case, for the received signal.

This poses a problem for computing the capacity region: The capacity is obtained by maximizing the mutual information over all the input distributions and this is achieved, for the Gaussian channel, only by producing a Gaussian output (Shannon 1948). Thus, in order to attain the capacity one has to “Gaussianize” the output by choosing a suitable distribution for the set of input signals.

A fading additive Gaussian noise channel is assumed to have an output of the form $Y = RX + Z$ in which: N is a Gaussian random variable with mean zero and variance σ_Z^2 , that is, $N \sim N(0, \sigma_Z^2)$; X is a signal random variable with density function $p_X(x)$ and variance σ_X^2 ; R is a positive valued random variable with density function $p_R(r)$ and $E[R^2] = 2\gamma^2$; Y is the received random variable. It is assumed that X and therefore Y has mean zero. The capacity of this channel has been determined (Ericson 1970) as:

$$\begin{aligned} C(s^2) &= E_R \left[\frac{1}{2} \log(1 + R^2 s^2) \right], \quad s^2 = \sigma_X^2 / \sigma_Z^2 \\ &= \int_0^\infty p_R(r) \frac{1}{2} \log(1 + r^2 s^2) dr, \end{aligned} \quad (5.14)$$

in which the units will be bits per channel use and the parameter s^2 will be referred to as SNR. The equation results from the consideration of the parametric equations defining the reliability function for random coding for this channel and is valid for any positive valued random variable.

It is of interest here to know how this expression is related to the equation

$$C(s^2) = \frac{1}{2} \log(1 + \gamma^2 s^2), \quad (5.15)$$

which might be easier to use in many situations.

From the convexity of the logarithm function and Jensen’s inequality (Blahut 1987) it is clear that for all SNRs, s^2 , $C(s^2) \leq C(s^2)$. The magnitude of the difference $C(s^2) - C(s^2)$ will be shown, graphically, to be small for the case of a Rayleigh fading distribution. Implicit in the form of the capacity expression is the fact that Y should be a Gaussian random variable to achieve capacity. Consequently $U = RX$ should be a Gaussian random variable.

The question of determining the distribution of X , for a given distribution of R to yield U a Gaussian distribution is considered. It is noted that

since U and R are correlated with unknown joint distribution, this problem is quite different from the usual transformation problem for random variables. It is solved for the case of R Rayleigh distributed. The general problem seems difficult.

From the form of the equation for the capacity of the channel, it is argued that to achieve capacity, Y must be a zero mean Gaussian random variable. For a given fading random variable R , the signal random variable X should be chosen so that $U = RX$ is Gaussian. Since U and R are correlated it does not seem a simple matter to compute the distribution of $X = U/R$, using the standard techniques.

Likewise, it does not seem a simple matter to compute $I(X, Y)$ and the maximizing distribution to achieve capacity, making the technique of (Ericson 1970) the more interesting.

Example: for the case of R a Rayleigh distributed random variable with constant the distribution of X can be achieved in the following manner.

Since R and X are, by assumption, independent and $Y - N = U = RX \sim N(0, \sigma_U^2)$, as discussed, then X has a zero mean and a symmetric distribution. Odd moments of U and X are zero and $E[X^{2k}] = E[U^{2k}]/E[R^{2k}]$. The required moments are readily computed as

$$E[R^{2k}] = k! \gamma^{2k} \quad E[U^{2k}] = \frac{(2k)!}{2^k k!} \sigma_U^{2k}, \quad (5.16)$$

and so

$$E[X^{2k}] = \frac{E[U^{2k}]}{E[R^{2k}]} = \binom{2k}{k} \left(\frac{\sigma}{\sqrt{2}\gamma} \right)^{2k}, \quad k = 1, 2, 3, \dots \quad (5.17)$$

Let $a = \sigma/(\sqrt{2}\gamma)$. Let $p_X(x)$ be the density of X and $\phi_X(v)$ the characteristic function. The series expansion of $\phi_X(v)$ is

$$\begin{aligned} \phi_X(v) &= \sum_{k=0}^{\infty} (-1)^k \frac{E[X^{2k}]}{(2k)!} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{\binom{2k}{k}}{(2k)!} \\ &= \sum_{k=0}^{\infty} (-1)^k \frac{(av)^{2k}}{(k!)^2} \end{aligned} \quad (5.18)$$

This series is identified as (Abramowitz and Stegun 1965, p. 360) $J_0(av)$ in which $J_0(\cdot)$ is a Bessel function of the first kind and zeroth order. The

corresponding density function is then given as (Oberhettinger 1990)

$$\begin{aligned} p_X(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} J_0(av) e^{-jvx} dv \\ &= \frac{1}{\pi((2a)^2 - x^2)^{1/2}}, \quad |x| < 2a \end{aligned} \quad (5.19)$$

Thus for the random variable X with this density, R with the Rayleigh density, $U = RX$ has the required Gaussian density $N(0, \sigma_U^2)$. While the above technique is of interest, since success depends on the identification of an infinite series, it does not lend itself to generalization.

The effect of fading has been removed from the final expression for the capacity at a certain cost. In order to counterbalance the stochastic attenuation, it is necessary to design a circuit that provides the transformation of random variables necessary at the transmission end. It is not certain, at this point, that this can be achieved in general for any type of fading.

In physical terms, the introduction of fading in the channel decreases the noise margin of the system. Thus, the detection threshold remains constrained to a smaller region near the origin, for the case of binary communications. If the input distribution cannot be altered, it is possible to define another transmission rate, which will fall short of being the optimum in this case. The trade-off assumed for attaining the full capacity of the system resides in the design of a circuit to transform the input signal and shape its probability density function (pdf) accordingly.

The capacity evaluated by Ericson (1970) can be interpreted as an average channel capacity, as defined by Lee (1990), or as analyzed by Dobrushin (1961). The concept of average or mean capacity can be proven useful, anyway, and has been successfully applied to the case of dynamic allocation of users (Alencar and Blake 1993a).

In order to complete the investigation, in light of the results of Ericson, one presents the derivation of the formula for the average capacity for the single user case, beginning with the general expression below (Ericson 1970)

$$C = E_R \left[\frac{1}{2} \log(1 + R^2 s^2) \right] \quad (5.20)$$

in which $s^2 = \sigma_X^2 / \sigma_Z^2$ is the SNR and $E[\cdot]$ represents the expected value functional.

Example: introducing the Rayleigh pdf, from Equation 5.13, into the previous formula gives

$$C = \frac{1}{2} \int_{-\infty}^{\infty} \log(1 + r^2 s^2) \frac{r}{\gamma^2} e^{-\frac{r^2}{2\gamma^2}} u(r) dr. \quad (5.21)$$

The last equation can be simplified by the application of a known theorem of probability (Blake 1987), leading to

$$C = \frac{\log e}{4\gamma^2} \int_0^\infty \log(1 + \xi s^2) e^{-\frac{\xi}{2\gamma^2}} d\xi, \quad (5.22)$$

which can be solved by standard methods, yielding

$$C = -\frac{\log e}{2} \text{Ei} \left(-\frac{1}{2\gamma^2 s^2} \right) e^{\frac{1}{2\gamma^2 s^2}} \quad (5.23)$$

in which the Exponential Integral function is defined by

$$\text{Ei}(x) = \int_{-\infty}^x \frac{e^t}{t} dt, \quad x < 0. \quad (5.24)$$

Concluding this section, Figure 5.3 illustrates a comparison between Formulas 5.15 and 5.23, in the single user case, for three values of the fading parameter γ . The dotted curves represent the average capacity, for each value of the parameter.

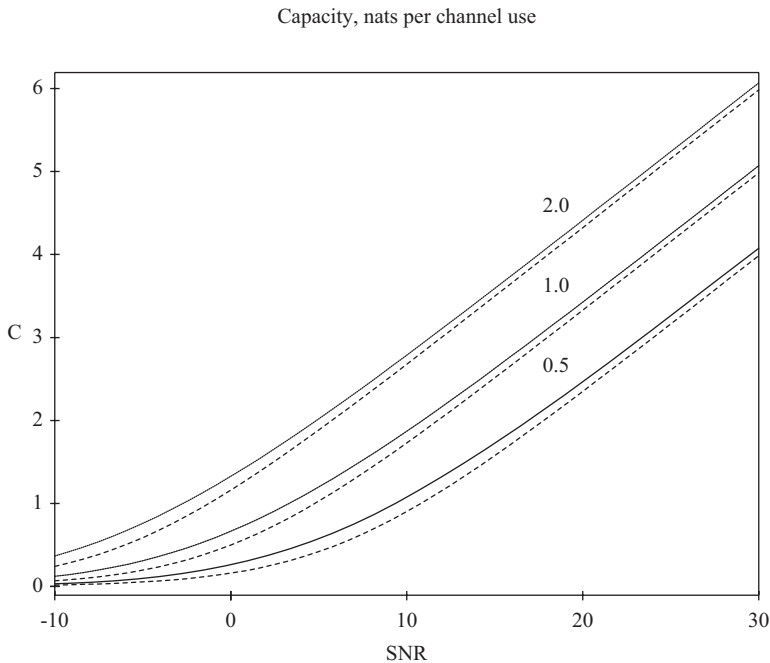


Figure 5.3. Average and actual capacity, for $\gamma = 0.5, 1.0,$ and $2.0.$

5.4 THE NONCOOPERATIVE MULTIPLE ACCESS CHANNEL

The noncooperative channel has asynchronous multiplexing, which implies that the M information sources transmit through a common medium without any prearrangement among themselves (Sommer 1966). Therefore, all components of the received signal are statistically independent, and so perform as interference with respect to any other given signal.

For the conditions mentioned, a slight generalization of a former work can be attempted (Sommer 1966). Based on the results stated in Shannon (1948), the mutual information between the set of inputs and the output, for the noncooperative additive Gaussian noise channel, can be written as

$$I(X_1, \dots, X_M; Y) = H(Y) - H(Y - G) \quad (5.25)$$

in which

$$G = \sum_{i=1}^M X_i. \quad (5.26)$$

This leads to the already evaluated expression

$$I(X_1, \dots, X_M; Y) = H(Y) - H(Z) \quad (5.27)$$

bounded as

$$\begin{aligned} I(X_1, \dots, X_M; Y) &\leq \frac{1}{2} \log 2\pi e \left(\sum_{i=1}^M \sigma_i^2 + \sigma_Z^2 \right) \\ &\quad - \frac{1}{2} \log 2\pi e \sigma_Z^2 \\ &= \frac{1}{2} \log \left(\frac{\sum_{i=1}^M \sigma_i^2 + \sigma_Z^2}{\sigma_Z^2} \right) \end{aligned} \quad (5.28)$$

$$= \frac{1}{2} \log (1 + S^2) \triangleq C. \quad (5.29)$$

By the same token, one can compute the individual mutual information

$$I(X_i; Y) = H(Y) - H(Y - X_i) \quad (5.30)$$

$$\begin{aligned} &\leq \frac{1}{2} \log 2\pi e (\sigma_Z^2 S^2 + \sigma_Z^2) \\ &\quad - \frac{1}{2} \log 2\pi e (\sigma_Z^2 S_i^2 + \sigma_Z^2) \end{aligned} \quad (5.31)$$

in which the interference to noise ratio (INR) is defined as

$$S_i^2 = \frac{\sum_{j=1, j \neq i}^M \sigma_j^2}{\sigma_Z^2}. \quad (5.32)$$

Rearranging the terms in 5.31 gives

$$I(X_i; Y) \leq \frac{1}{2} \log \left(\frac{S^2 + 1}{S_i^2 + 1} \right) \triangleq C_i. \quad (5.33)$$

The capacity region forms a square in two dimensions, a parallelepiped for three sources, and a hyper-parallelepiped for more than three sources. It can be seen that the hyperplane defined by expression 5.29 never touches the contour surface, which means that the capacity of the noncooperative channel is given by the sum of the individual rates,

$$\mathbf{C} = \sum_{i=1}^M C_i \quad (5.34)$$

or

$$\mathbf{C} = \frac{1}{2} \sum_{i=1}^M \log \left(\frac{S^2 + 1}{S_i^2 + 1} \right) \quad (5.35)$$

which can be put in the form

$$\mathbf{C} = \frac{1}{2} \log \left(\prod_{i=1}^M \frac{S^2 + 1}{S_i^2 + 1} \right). \quad (5.36)$$

Example: the last expression is maximized when all the transmitters carry the same power, that is, $\sigma_i^2 = \sigma^2$. This assumption simplifies the expression of the INR to

$$S_i^2 = \frac{M-1}{M} S^2 \quad (5.37)$$

and also simplifies the formula for the channel capacity

$$\mathbf{C} = \frac{1}{2} \log \left(\frac{S^2 + 1}{\frac{M-1}{M} S^2 + 1} \right)^M \quad (5.38)$$

or

$$\mathbf{C} = \frac{1}{2} \log \left(\frac{M(S^2 + 1)}{M(S^2 + 1) - S^2} \right)^M. \quad (5.39)$$

Example: an interesting result can be obtained by taking the limit of the capacity when the number of sources goes to infinity. First, the formula is rearranged to

$$\mathbf{C} = \frac{1}{2} \log \left(1 - \frac{S^2}{M(S^2 + 1)} \right)^{-M} \quad (5.40)$$

or, by noticing that $S^2 = Ms^2$

$$C = \frac{1}{2} \log \left(1 - \frac{s^2}{M(s^2 + 1/M)} \right)^{-M}. \quad (5.41)$$

Now, using the known property of limits

$$\lim_{M \rightarrow \infty} \left(1 + \frac{1}{M} \right)^M = e \quad (5.42)$$

one obtains finally

$$C^\infty = \frac{\log e}{2}. \quad (5.43)$$

Thus, for a large number of transmitters the capacity assumes the value $\log e/2$, which is bounded and low if compared to the capacity that can be achieved with the cooperative approach between the users. This must also be compared to the single user capacity, obtained by allowing $M = 1$ in expression 5.41

$$C = \frac{1}{2} \log (1 + s^2) \quad (5.44)$$

which is unbounded with respect to the SNR. The capacity region for the noncooperative channel can be seen in Figure 5.4, for two users, as a rectangle inside the diamond shaped capacity region for the cooperative channel, evaluated by Wyner (1974). The points C_1 , C'_1 , C_2 , and C'_2 are defined below

$$C_1 = \frac{1}{2} \log \left(1 + \frac{\sigma_1^2}{\sigma_2^2 + \sigma_Z^2} \right) \quad C'_1 = \frac{1}{2} \log \left(1 + \frac{\sigma_1^2}{\sigma_Z^2} \right) \quad (5.45)$$

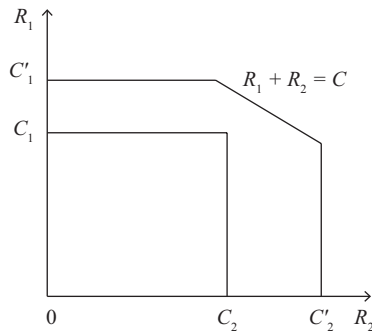


Figure 5.4. Capacity region for the noncooperative channel, $M = 2$.

and

$$C_2 = \frac{1}{2} \log \left(1 + \frac{\sigma_2^2}{\sigma_1^2 + \sigma_z^2} \right) \quad C_2' = \frac{1}{2} \log \left(1 + \frac{\sigma_2^2}{\sigma_z^2} \right). \quad (5.46)$$

As the number of users increases towards infinity the capacity region shrinks, and the channel capacity converges to a given value. There is no way to increase the capacity of the noncooperative channel by increasing the power of the transmitted signals. The limiting capacity C^∞ becomes independent of the SNR, as the number of senders goes to infinity. A possibility left to boost the capacity is by an increase in the available bandwidth. Spread spectrum techniques could prove useful in this case.

5.5 MULTIPLE ACCESS IN A DYNAMIC ENVIRONMENT

The selection of the more appropriate multiple access technique, for a given communications problem, depends on the channel characteristics. This section describes and analyzes a model for the Gaussian multiple access channel, in which the addition of new users is governed by a renewal process. An information theoretic approach is attempted, in order to find the set of transmission rates at which the reliable communication of the messages of each user is possible.

The problem of finding the capacity region of a multiple access channel can be linked to a seminal work by Shannon on two-way communication channels in 1961 (Shannon 1961). A usual assumption for the Shannon model is a strategic cooperation between the senders and the receiver.

The next section presents a different approach for the analysis of the multiple access channel. Instead of relying on static models for the channel, a dynamic structure is proposed and analyzed. The non-cooperative model in which the users compete for the channel resources is assumed. The access, or addition of new users to the channel, is governed by a Markovian process and the multiple access interference (MAI) is assumed proportional to the number of active users (Kleinrock 1975). This model assumes a variable assignment for the allocation of users in the channel, which is a reasonable assumption for CDMA systems.

5.6 ANALYSIS OF THE CAPACITY FOR A MARKOVIAN MULTIPLE ACCESS CHANNEL

Most models proposed for the multiple access channel assume that the number of users in the channel is fixed—but, in most cases, it is actually a random variable, governed by a distribution of probabilities. The channel capacity, for many channels, varies with the real time allocation of users. For instance, the capacity for the noncooperative model of accessibility is a function of the number of users which contribute to the multiple access interference (MAI). In the memoryless noncooperative channel the capacity decreases as more users are introduced, reaching a fixed value as this number increases to infinity (Alencar 1992b; Sousa 1989).

In order to account for the effect of a changing environment, with the users dropping in and out of the channel, a new model is proposed. Consider a Gaussian channel with k users. The variance of the MAI in the channel is assumed to be $k\sigma_X^2$ plus some background Gaussian noise of σ_Z^2 . Thus, the total noise variance depends on how many users are on the channel. It is meant to reflect a multiuser channel with other user interference modeled by Gaussian noise. Now suppose the channel is in state k if there are k users are on the channel, and a Markovian transition regime is imposed.

One model that might be interesting is a birth–death model, in which transitions occur only between adjacent states—so from state k you can go to either $k + 1$ or $k - 1$, with certain probabilities. Thus, for example, in a synchronous system this might reflect the model in which the probability of more than one user entering or leaving the system in a given data bit is negligible. The steady state probabilities are not difficult to compute and a formula for the average channel capacity for the model can be found.

The channel with extrinsic memory can be thought of as a dynamic system, governed by a birth–death process, with a variable number of active users and implying a dynamic capacity—contrary to the classical static capacity that assumes a fixed number of users. It can be seen as an infinite state channel (ISC), as opposed to the finite state models for channels found in the literature (Gilbert 1960; Fritchman 1967). In this new approach, the channel capacity changes as the number of users changes in a dynamic fashion. This last quantity is considered a stochastic process, and has an underlying Markov chain of states (Kleinrock 1975).

A transition probability matrix $\mathbf{P} = \{p_{ij}\}$ governs the behavior of the channel accessibility. The transition probabilities are obtained from the Markov model depicted in Figure 5.5, in which α_k and β_k are the birth and death parameters, respectively. The channel will be called the Alpha-Beta

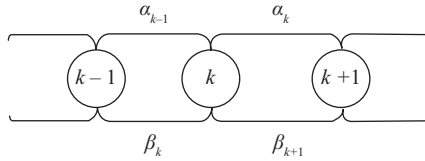


Figure 5.5. Markov model for the multiple access channel.

channel, or AB channel for short. The state of the channel k at a given time depends only on the previous state. In order to simplify the results, the input (X) and output (Y) alphabets are assumed to follow a Gaussian distribution. The model AB assumes all users transmitting with the same power σ_X^2 and the Gaussian plus MAI noise, at any state k , is obtained from the following equation

$$\sigma_k^2 = k\sigma_X^2 + \sigma_N^2. \quad (5.47)$$

After a certain number of interactions, the Markov chain reaches a steady state. The set of steady state probabilities $\Pi = \{\pi_k | k = 1, 2, 3 \dots\}$ can be computed by means of one of the well-known techniques (Kleinrock 1975; Adke and Manjunath 1984). Each state k defines the number of users, the multiple access interference and the conditional capacity given that the system is in that state, $C(k)$. This conditional capacity is given by the solution of the noncooperative model for the additive Gaussian channel (Alencar and Blake 1993b).

Therefore, the formula for the capacity of the channel with extrinsic memory is defined, in connection with the concept of mean capacity (Dobrushin 1961), as the supremum of the weighing of each state capacity, $C(k)$, by the associated steady state probability, π_k , given by Wang (1992)

$$\mathbf{C} = \sup_{\{\pi_k\}} \sum_{k=1}^M \pi_k C(k). \quad (5.48)$$

The conditional capacity of the noncooperative channel, given that the channel is in state k , is given by the sum of the individual capacities, C_{ik} , for each i -th user (Alencar and Blake 1993b),

$$C(k) = kC_{ik} \quad (5.49)$$

or

$$C_k = \frac{1}{2} \sum_{i=1}^k \log \left(\frac{S^2 + 1}{S_i^2 + 1} \right) \quad (5.50)$$

in which S^2 represents the total SNR. The expression can be put in the form

$$C_k = \frac{1}{2} \log \left(\prod_{i=1}^k \frac{S^2 + 1}{S_i^2 + 1} \right). \quad (5.51)$$

The last expression reaches a maximum when all the transmitters carry the same power, that is, $\sigma_i^2 = \sigma_X^2$. This assumption simplifies the formula of the interference to noise ratio (INR) to

$$S_i^2 = \frac{k-1}{k} S^2 \quad (5.52)$$

and also simplifies the formula for the channel conditional capacity, giving

$$C_k = \frac{1}{2} \log \left(\frac{S^2 + 1}{\frac{k-1}{k} S^2 + 1} \right)^k \quad (5.53)$$

or

$$C_k = \frac{1}{2} \log \left(\frac{k(S^2 + 1)}{k(S^2 + 1) - S^2} \right)^k, \quad (5.54)$$

which can be further simplified by noticing that the total SNR is $S^2 = ks^2$, in which $s^2 = \sigma_X^2 / \sigma_Z^2$.

Two special cases for the AB model are treated here. First, the case in which the birth and death parameters are considered constant for any state, and the access is unrestricted. Second, the case in which the users are discouraged by the amount of noise in the system. This occurs because every new user can be thought to degrade the system to a certain extent, by decreasing the SNR. For the first case, $\alpha_k = \lambda$ and $\beta_k = \mu$. For the second case, $\alpha_k = \lambda / (k + 1)$ and $\beta_k = \mu$, in which λ and β are fixed probabilities.

Example: the first case produces a Geometric distribution for the probabilities.

This distribution shows the likelihood of being in a certain state in the long run, after the transient dies out and is given by the equation

$$p_k = (1 - \rho)\rho^k, \quad k = 0, 1, 2, \dots, \quad \text{for } \rho < 1 \quad (5.55)$$

in which $\rho = \lambda / \mu$ is usually called the utilization factor of the system.

For the Geometric distribution considered, the statistical mean is given by $\rho / (1 - \rho)$ and the variance by $\rho / (1 - \rho)^2$. The probability of finding

Capacity, bits per channel use

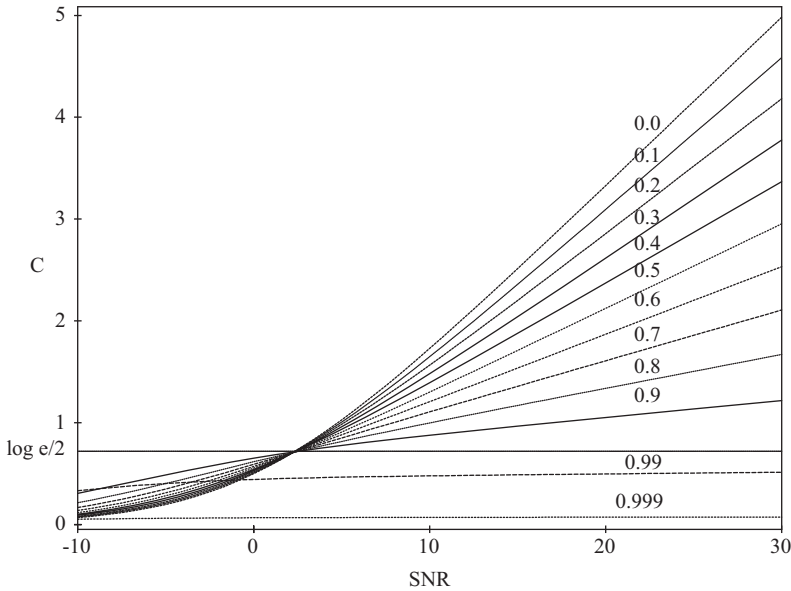


Figure 5.6. Capacity for the channel with Geometric accessibility, as a function of the signal-to-noise ratio, for different values of ρ .

more than L users at a time in the system is given by ρ^{L+1} . Using Equation 5.48, and substituting the appropriate terms, the capacity for this case is defined as the maximum, over the set of utilization factors

$$C = \max_{\rho} \sum_{k=1}^{\infty} (1 - \rho) \rho^k \log \left(\frac{(k+1)\sigma_X^2 + \sigma_N^2}{k\sigma_X^2 + \sigma_N^2} \right)^k, \quad (5.56)$$

whose plot is presented in Figures 5.6 and 5.7.

As can be seen, from the plots, the limiting value of the sum capacity, as the utilization factor increases, is the same obtained in the case of non-cooperative users of the previous section. The sum capacity decreases, for high SNR, to $\log e/2$, with an increase in the utilization factor. For low SNR, on the contrary, the curves show an increase to the limiting value, as the system tends to occupy states of higher order.

Example: the solution for the second case seems to be a better approach to reality. A Poisson distribution is obtained for the stationary probabilities of the ISC model, given by

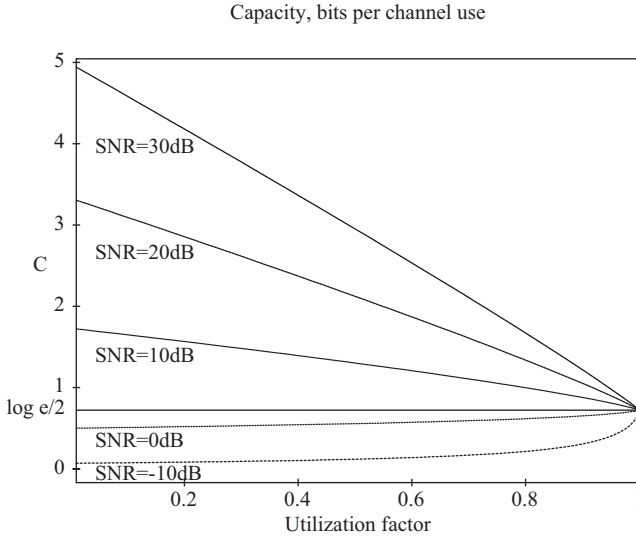


Figure 5.7. Capacity for the channel with Geometric accessibility, as a function of the utilization factor, for different values of the signal-to-noise ratio.

$$p_k = \frac{\rho^k}{k!} e^{-\rho}, \quad k = 0, 1, 2, \dots \tag{5.57}$$

with mean and variance equal to ρ .

The capacity, therefore, is given by the maximum, over the range of utilization factors, of the following formula

$$C = \max_{\rho} \sum_{k=1}^{\infty} \frac{\rho^k}{k!} e^{-\rho} \log \left(\frac{(k+1)\sigma_X^2 + \sigma_N^2}{k\sigma_X^2 + \sigma_N^2} \right)^k. \tag{5.58}$$

The results are shown in Figures 5.8 and 5.9 as a function of the SNR and utilization factor, respectively. Figure 5.9, in particular, gives a good example of the system behavior as the utilization factor increases. It can be seen that, as the distribution becomes more uniform, the sum capacity tends to the already computed limit $\log e/2$. As the factor increases there is a more pronounced tendency for the occupation of higher states in the Markov chain.

Several models were described in this chapter. The first modeled the case when the users acknowledge a certain protocol for transmission and cooperate to achieve the maximum transmission rates. In the second model, the users do not cooperate, in the sense that they transmit their signals

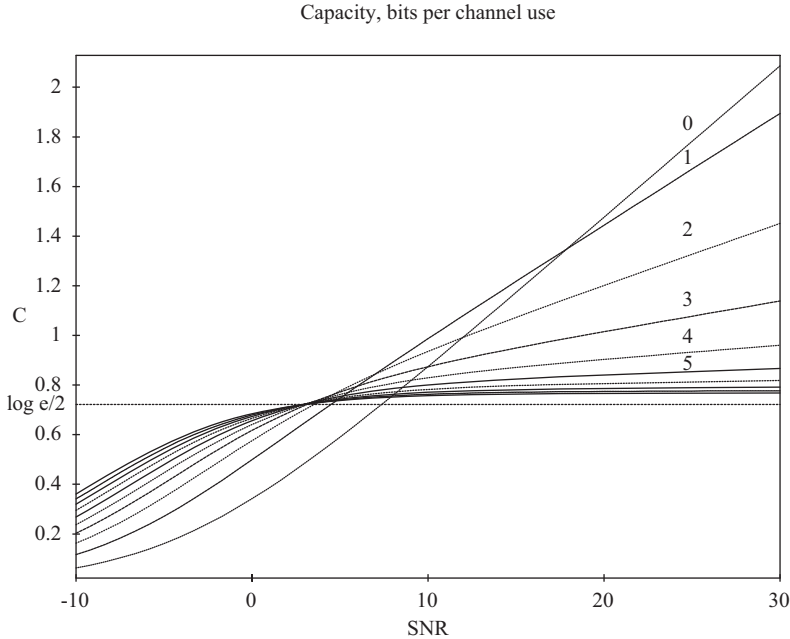


Figure 5.8. Capacity for the channel with Poisson accessibility, as a function of the signal-to-noise ratio, for $\rho = 0, 1, 2, 3, 4, 5, 6, 7, 8, 9$.

without concern to any approach to maximize the transmission of information. The third model included the role of accessibility on the evaluation of the capacity. The information theoretic approach is used in order to find the information rates at which the simultaneous and reliable communication of the messages for each user is possible.

For the cooperative model, the capacity region was obtained and the effect of fading was estimated for the case of Rayleigh attenuation factors. The capacity was shown to be attained by a process of “Gaussianization” of the output in terms of the input distribution. It is interesting to notice that the multipath effect, discussed in the last chapter, can be analyzed by the same techniques presented in this chapter and can provide a net gain for systems that use spread spectrum techniques (Alencar 1992a).

Spread spectrum thus provides the necessary diversity for the system, and the multipath phenomenon can be used to the advantage of the transmission. In a real channel, pseudo-noise spread spectrum signals are used to resolve the multipath introduced by the environment.

Two problems of interest associated with the closeness of an approximation to the capacity of a fading Gaussian noise channel have been

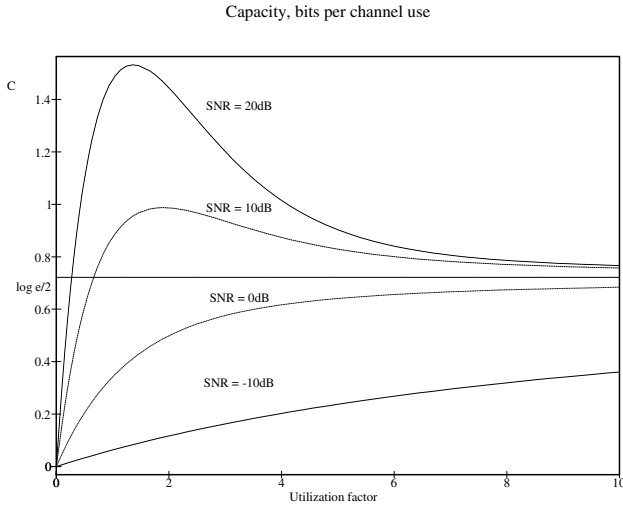


Figure 5.9. Capacity for the channel with Poisson accessibility, as a function of the utilization factor, for some values of the signal-to-noise ratio.

considered. The relative closeness of the capacity, as given by Ericson, to the expression proposed, independent of SNRs, is of interest, allowing the simpler expression to be used with some confidence in some analyses.

The more precise result available for the case of Rayleigh fading is also of interest. The consideration of capacity for this case led to the interesting distribution problem discussed in this chapter.

Some results were obtained from the analysis of the noncooperative model. First, as the number of users increases towards infinity, the capacity region shrinks and the channel capacity converges to a given value. Second, there is no way to increase the capacity of the noncooperative channel by increasing the power of the transmitted signals in this limiting case. The channel capacity C^∞ becomes independent of the SNR as the number of senders goes to infinity.

As discussed before, the classical Shannon capacity formula offers a tradeoff between signal power and bandwidth. In the noncooperative case, because of the inherent restriction on the transmitted power, the possibility left to enlarge the capacity is by an increase in the available bandwidth and spread spectrum techniques could prove useful in this case.

As discussed previously and inferred from the mathematical results and plots, the capacity for the Markovian multiple access channel is dependent on the utilization factor of the system. For both Geometric and Poisson distributions, when more users gain access to the *AB* channel the capacity

increases up to a point in which it reaches a maximum. From this point on, if the number of users entering the channel is proportionately larger than the number of drop outs, the capacity begins to decrease. The capacity will eventually reach zero as the utilization factor increases to the limit. In the case of a Geometrically distributed accessibility, the maximum capacity is achieved for a $\rho = 0.97$. The maximum value of ρ is one.

For the Poisson model the maximum is reached for $\rho = 1.3$, approximately. There is no limit for ρ if this distribution is used. The Poisson distribution models the case in which the incoming users are discouraged by the amount of noise in the system. As the SNR decreases, more users abandon the channel. Both the Geometric and Poisson models result in capacities that are upper bounded by the capacity of the noncooperative channel, with fixed assignment, as the number of users go to infinity (Alencar and Blake 1993b).

This model is suitable for describing the properties of systems that present a variable assignment for user allocation, as is the case with CDMA techniques.

CODE DIVISION MULTIPLE ACCESS

6.1 INTRODUCTION

This chapter presents some of the main characteristics of spread spectrum systems, including the indication of possible types of interference, anti-jamming advantages of the technique in terms of signal-to-noise (SNR) ratio and possible applications of the techniques. The usefulness of spread spectrum to overcome interference is stressed. Among others, the main sources of interference are multipath, multiple media, multiple user, and multiple cell (Viterbi 1991).

The first developments using concepts related with the current techniques of spread spectrum occurred in the mid-forties, and the applications involved military problems (Scholtz 1982). The techniques acquired great popularity in the communications community, for the last 20 years, since the papers on the subject began to appear after two decades of secrecy due to military interests.

The main feature of spread spectrum signals is related to their bandwidths, which are much larger than those assigned to baseband or modulated signals. The way in which the spectrum is spread is essential, and is often realized by a code that is independent of the transmitted data. The receiver needs a replica of the code, synchronized in time, in order to recover the data.

The spreading process introduces a great deal of redundancy in the transmitted signal. This allows the receiver to suppress part of the interference embedded in the signal, which is the chief advantage of the process.

It is possible to devise other advantages of the spread spectrum systems in relation to personal communications (Pickholtz, Schilling, and Milstein 1982; Scholtz 1982).

- Allows the addressing of a selective group of customers in the same channel, that is, code division multiple access (CDMA).
- Utilization of the signal to measure the propagation delay, which is important in personal communications systems to access the attenuation in the channel.
- It is possible to hide the signal in the background noise in order to make its reception difficult by those who do not know the spreading code (cryptographic capability).
- Useful as a means of diversity in multipath channels, which turns out to be the main problem of personal communications systems.
- Features a low probability of interception (jamming) of the signal, due to the randomness of the spreading code.
- Could improve the data rate, limited in wireless channels by multipath and fading problems.

One of the strategies in use to spread the spectrum is the Direct Sequence modulation (DS), in which a fast pseudo-random sequence causes amplitude switching in the data carrying signal, as suggested in Figure 6.1. In this Figure, the data signal is represented by $b(t)$, the pseudo-noise sequence is $c(t)$, T_b is the bit interval, and T_c is the chip interval.

Figure 6.2 shows a block diagram for the direct sequence spread spectrum system. The signal is randomized before reaching the modulator. At the receiver, a synchronization circuit retrieves the PN code to produce the original signal.

Another technique used is called Frequency Hopping (FH), in which the signal is subject to frequency switching, according to a predetermined pseudo-random sequence.

Figure 6.3 illustrates the FH technique, in which a pseudo-noise generator is used as input to a synthesizer to produce a pseudo-random sequence of frequencies that shifts the originally modulated signal frequency.

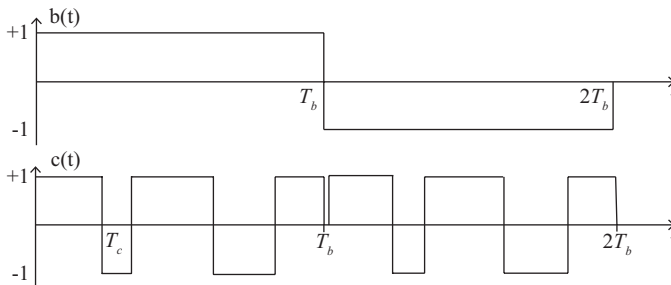


Figure 6.1. Data signal and pseudo-noise sequence.

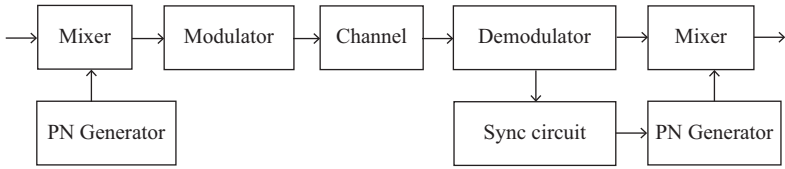


Figure 6.2. Direct sequence spread spectrum system.

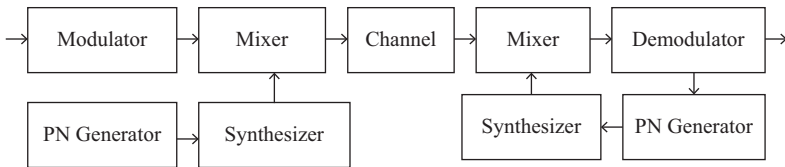


Figure 6.3. Frequency hopped spread spectrum system.

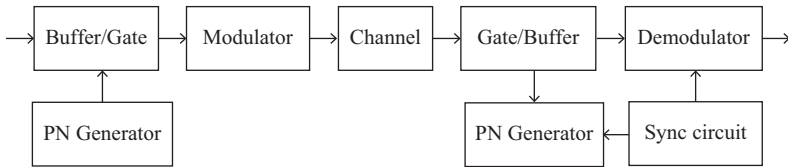


Figure 6.4. Spread spectrum using random time windows.

It is also possible to produce spread spectrum by transmitting bursts of the signal in random instants of time. This technique is known as Time Hopping (TH) and is illustrated in Figure 6.4.

Practical systems can combine the presented techniques, as is usually the case, to obtain certain properties. In a sense, the spread spectrum system is an antithesis of the usual communication system. Because, in contrast to the usual system premises, it requires a very large bandwidth and a very low power spectrum density for the transmitted signal.

Despite such advantages, the spread spectrum approach had some critiques as quoted below:

The mystique of spread spectrum communications is such that commercial enterprises, as well as academics, are often attracted by the novelty and cleverness of the technique. Also, in small artificially aided markets, there may be temporary economic advantages.

In the long run, though, it's the author's opinion that we must stand back and question the wisdom of squandering a precious resource such as bandwidth for reasons of expediency (Viterbi 1985).

6.2 FUNDAMENTALS OF SPREAD SPECTRUM SIGNALS

It was mentioned in the last section that the spread spectrum techniques can be used to protect the signal against jammers. The following discusses the most common types of jammers, and how it is possible to attain this objective.

An attempt to classify the jammers is sketched below (Pickholtz, Schilling, and Milstein 1982; Cook and Marsh 1983):

- Partial band jammer. The interference is concentrated in a portion of the signal bandwidth.
- Partial time jammer, in which the intentional interference occurs in only part of the transmission time. This can be accomplished by the transmission of narrow pulses or impulses.
- Tone jammer, which concentrates its power in a single frequency.
- Broad-band noise jamming. In this case the jammer tries to attack the whole signal bandwidth. One possible approach is the generation of a pseudo-noise sequence (PN) by the jammer.
- Repeater jamming. The jammer uses the very transmitted signal to cause the interference. This type of interference can produce a deleterious influence on the signal spectrum that mimics the multipath effect.

The vectorial, or signal space, approach is introduced here to explain how spread spectrum can be used in parallel with other existing systems and still protect the signals against interference to or from those systems (Wozencraft and Reiffen 1961). A data signal, with restricted dimensionality, is spatially expanded in such a way that prevents the degradation of the communication process by a limited power interfering signal.

The classical digital transmission problem is the one in which both transmitter and receiver know the complete set of M waveforms $x_i(t)$, each one with duration T_b seconds. The transmitter selects one of the M signals at each T_b seconds, which yields a data rate of $\log_2(M)/T_b$ bits/s. If, for example, signal $x_j(t)$ is transmitted, the received signal over the

interval $[0, T_b]$ will be $y(t) = x_j(t) + n_I(t) + n(t)$, in which $n(t)$ is the additive white Gaussian noise (AWGN) with power spectral density of $N_0/2$ W/Hz and $n_I(t)$ is the multiple access interference, also assumed to be Gaussian distributed.

It can be demonstrated that it is possible to represent the whole signal set by the use of no more than L ($L \leq M$) orthogonal basis functions. It must also be pointed out that only those noise waveforms inside the signal space are relevant. In theory, a noise signal would require an infinite number of components. It is said that the above defined set has dimension N , if this is the number of basis functions needed to characterize the set of signals (Wozencraft and Reiffen 1961). It can be shown that $L \approx 2WT_b$, in which W is the approximate bandwidth shared by the set of signals. The optimum receiver for an AWGN channel consists of a bank of correlators, or a bank of matched filters. The decision strategy is to choose the maximum correlator output. The performance of this system depends only on the ratio of the bit energy to the noise power spectrum density.

The key operation in the whole process is the spectral compression of the interfering signal accomplished by the receiver. This is done by correlating the received signal with a known spreading code. The information bearing signal is compacted and, simultaneously, the interfering signal power is spread over the frequency range. This implies an effective advantage for the data signal in relation to the interfering signal, but requires perfect synchronization of the pseudo-random code in the receiver. For a signal of duration T_b and bandwidth W , the approximate number of dimensions is $2WT_b$ and the processing gain, or sequence code length, can be written as

$$N = \frac{T_b}{T_c} \quad (6.1)$$

in which T_c is the chip interval.

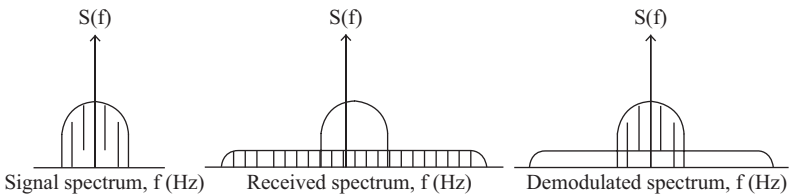


Figure 6.5. Spectra of transmitted and received signals.

6.3 PERFORMANCE ANALYSIS OF CDMA SYSTEMS

The main areas of concern on multiple access systems are multiaccess information theory, collision resolution, and spread spectrum (Gallager 1985). The approach followed by researchers on information theoretic multiaccess takes into consideration the effects of noise and interference, neglecting the random arrivals of messages. On the other hand, research on collision resolution focused primarily on the bursty arrivals of messages and the interference between transmitters ignoring the noise component (Hui 1984a; Hui 1984b).

For multiaccess communication using spread spectrum several sources can transmit at the same time, using different modulating sequences, and each will look like broadband noise to the others. Spread spectrum could also provide a solution to the problem of channel allocation.

One of the main problems in communications is the simultaneous allocation of multiple users to a channel. This must be accomplished without severe degradation in the performance of the system. Three techniques have been proposed to attack the problem, as mentioned in the introduction: FDMA, TDMA, and CDMA. The aim of the first two techniques is to separate the signals in frequency and time, respectively. The third one uses the properties of spread spectrum to separate the signals in code, that is, to modulate each message signal with a sequence selected from a set of orthogonal sequences. Each user has its own address embedded in one of the code sequences (Cook et al. 1983). CDMA relies on the ability of a spread spectrum system to reject independent interference by a factor approximately equal to its processing gain (Simon et al. 1985).

An ideal system should combine the features of all the mentioned multiple access techniques, for example, the orthogonality of the TDMA system leads to simultaneous achievability of single-user capacities by all users. In the case of FDMA or CDMA, this goal could be attained by a sufficient separation of the frequencies in the former, and by the use of very large spreading factors in the latter (Cheng and Verdú 1991).

Multiple access techniques such as FDMA and TDMA are susceptible to various types of interference, including jamming and multipath. FDMA also presents the intermodulation problem. On the other hand, TDMA must have all the users in perfect synchronism. Both FDMA and TDMA have their capacities limited by the available bandwidth. CDMA also has its drawbacks. Interference caused by transmitters that are close to the receiver cause the near-far problem. In a mobile environment, the near-far effect can be combated by the selection of sequences with very good

cross-correlation properties, as discussed in the next section, or by the use of adaptive power control (Schilling et al. 1991).

The second problem is the possibility of interference with existing systems (Kavehrad and Ramamurthi 1987). Unlike the former techniques, CDMA has its capacity solely limited by the amount of interference into the system, which allows any reduction in interference to be converted directly and linearly into and increase in capacity (Gilhousen et al. 1991).

As pointed out before, the use of spread spectrum allows new users to share the spectrum with existing ones. CDMA is also interference limited, that is, the number of users that share the same spectrum and still maintain an acceptable performance is determined by the interference generated by the set of remaining users. In CDMA systems, each user is given a distinct pseudo-random code. Hence, users in adjacent cells use the same bandwidth and therefore interfere, making a received signal appear noisier as the traffic increases in the system. The effect of the interfering energy depends on the processing gain of the spread spectrum technique.

Taking into account the adjacent cells interference, the total number of users M that can simultaneously access a given cell can be estimated from the following equation $M = 3G_P/8$, in which G_P is the processing gain. The computation of M above assumes all users simultaneously transmitting. For a speech activity factor p and an interference increase factor λ the heuristic formula above can be adjusted to give the net number of users U that can access a given cell in a given time period (Schilling et al. 1991)

$$U = \frac{M}{V(1 + \lambda)}. \quad (6.2)$$

A more formal approach can be followed with the use of algebraic coding theory and the definition of balanced codes. An (n, w, λ) binary balanced code is a set of binary n -tuples, each of weight w , such that any two words agree in at most λ positions. If the maximum number of words in a balanced code is U , it follows that (Blake and Mullin 1976; Chung and Kumar 1990)

$$U \leq \frac{\binom{n}{\lambda + 1}}{\binom{w}{\lambda + 1}}. \quad (6.3)$$

In the last expression it is evident that the maximum of the autocorrelation function is given by w , and a good set of codes requires that the minimum distance, $d = 2(w - \lambda)$, be kept at a maximum.

As implied above, spread spectrum systems are less subject to multipath signal variations than conventional systems. In a direct sequence receiver,

if the reflected signal is delayed, compared to the direct signal, by more than one code chip, the reflected signal is treated as any other uncorrelated input signal. The higher the code chip rate, for a particular system, the smaller its multipath problem. In the mobile environment, the result of multipath reception is that a moving receiver is subject to rapid fading and peaking of its input signal as the direct and reflected signals cancel and reinforce each other. Stationary receivers are also subject to multipath due to the reflecting and refracting surfaces (Dixon 1984).

A few topics have interested researchers, industries, and users of CDMA systems. The first one is the design of pseudo-noise sequences with good autocorrelation and cross-correlation properties. The second concerns the evaluation of the system capacity, or how many users a CDMA system can support under given circumstances. The number of allocated users dictates the type of distribution to characterize the Multiple Access Interference (MAI). A large number of subscribers favor the Gaussian approximation. For a small number of subscribers, considering that the MAI is additively composed of cross-correlations limited to a few chips of coincidence, the maximum interference is too small to apply the central limit theorem (Brady 1991).

The evaluation of system performance can be done in two ways: using the average probability of error or computing the outage of the system. The probability of error is averaged over the ensemble of channels obtained from the assumed model. The outage measures the probability of transmitting through a channel when the probability of error is greater than a predefined threshold. In the case of time-varying channel and stationary terminal, outage indicates the fraction of time the terminal operates with an error probability above the limit. For the time-invariant channel and mobile terminal, outage indicates the fraction of building area in which the terminal can operate (Kavehrad and McLane 1987).

One interesting peculiarity of Direct Sequence Spread Spectrum (DSSS) is that error detection and correction may not be advisable in some applications. This is due to the effect of the coding overhead, which increases the signal apparent data transmission rate and reduces the available processing gain (Dixon 1984). On the other hand, a substantial improvement in the capture performance of coded DSSS can be achieved, in comparison to an uncoded system. Capture phenomena characterize the ability of a receiver to successfully receive a packet, even though part or all of the packet arriving at the receiver overlaps with the other packets. The basic mechanism for capture is the ability of the receiver to synchronize with and lock on to one packet and subsequently reject other overlapping packets as noise (Soroushnejad and Geraniotis 1991).

In particular, time-frequency hopping modulation has a potential for application in those systems in which a large number of users, with widely variable distances or transmitted power, are to operate simultaneously in a single link. Such systems tend to employ a simple code sequence, primarily as an addressing medium, rather than to spread the spectrum specifically. This type of modulation is suitable for IWDC systems, in which random access and discrete addressing are the main concerns. It also offers one of the few viable solutions to the near-far problem. By the use of synchronization the users can be programmed to transmit on different frequencies as well as different times (Dixon 1984).

6.4 SEQUENCE DESIGN

Modeling a spread spectrum system implies the generation of random sequences. Those sequences must exhibit the following properties: *i*) each signal in the set is easy to distinguish from a time-shifted version of itself; *ii*) each signal in the set is easy to distinguish from a time-shifted version of every other signal in the set (Sarwate and Pursley 1980).

In real life, practical constraints dictate the use of pseudo-random, or pseudo-noise sequences (PN), whose properties are well known. The autocorrelation and the cross-correlation of the sequences play an important role in designing the system, and are responsible for the spectral characteristics of the spread spectrum signal (Sarwate and Pursley 1980). The definition of a pseudo-random sequence is done in terms of the periodic auto and cross-correlation functions. For a sequence x_k , $k = 0, \dots, N - 1$, the periodic autocorrelation is

$$r_X(i) = \sum_{k=0}^{N-1} x(k)x(k+i) \quad (6.4)$$

in which the addition is taken modulo N . Given two sequences \mathbf{x} and \mathbf{y} of length N , the periodic cross-correlation function is

$$r_{XY}(i) = \sum_{k=0}^{N-1} x(k)y(k+i). \quad (6.5)$$

A pseudo-random sequence is usually defined as a binary sequence such that for $i \neq 0$, $|r_X(i)|$ is small compared to $r_X(0)$. On the other hand, a pseudo-random ensemble of sequences of length N has the cross-correlation between any two distinct sequences small compared to N , or to the peak autocorrelation function.

The maximal length sequences (m -sequences) are individual codes with good autocorrelation properties, while the Gold and Kasami codes form ensembles that have good cross-correlation properties. The m -sequences are related to simplex signal sets, finite fields, and error-correcting codes. The Gold and Kasami sequences are constructed from certain subsets of the m -sequences (Blahut 1990).

When modulating a carrier with a code sequence one-zero balance can limit the degree of carrier suppression, which is dependent on the symmetry of the modulating sequence. The amount of offset produced by the sequence is inversely proportional to the sequence length. Therefore, a long code sequence can minimize the offset effect (Dixon 1984).

Pseudo-random sequences feature:

- Facility of circuit implementation.
- A controlled degree of randomness.
- Large periods for the chip sequence.

6.4.1 MAXIMAL-LENGTH LINEAR SEQUENCES

The binary pseudo-random sequences are produced by shift-registers and have been used for many years in different areas of electronics and communications. In much of the literature concerning periodic sequences, the terms pseudo-random sequence, pseudo-noise sequence, and maximal-length linear sequence are used synonymously. The sequence has a length of $N = 2^m - 1$ bits, and is produced by a shift-register with m stages. The periodic m -sequence is constructed using the Galois Field $GF(2^m)$ (Blahut 1990). Field $GF(2^m)$ is built using an irreducible polynomial $p(x)$ of degree m over $GF(2)$ (Blake and Mullin 1976).

In CDMA applications, a large number of code sequences is needed. Under these conditions, multiple feedback points are necessary since the maximum number of codes of any length available from a set of delay elements using single-tap feedback would be only $m - 1$. Unfortunately, many of the $m - 1$ codes may be very short cyclic sequences, or the generator may halt operation altogether by entering zeros in all elements in some configuration.

The maximum length sequences can be shown to possess a peak cross-correlation as high as $2^{(m+1)/2} - 1$, implying that their use in a CDMA application will lead to considerable multiaccess interference. The number of such sequences with the smallest possible value for cross-correlation is very limited and depends on the choice of N . The maximum length sequences can be regarded as codewords of minimum weight in the dual of

a Hamming code, generated by a primitive polynomial. The dual contains only one codeword and all its cyclic shifts.

Using all the possible linear combinations of feedback taps for an m -stage register, there are $[\phi(2^m - 1)]/m$ maximal sequences that can be generated. The last expression represents an Euler totient function, that is, the number of positive integers that are relatively prime to and less than $2^m - 1$.

Table 6.1 lists the number of maximal sequences available for register lengths 3 through 20. Among the listed sequences are also some Mersenne prime codes that are defined as those codes for which the code length $N = 2^m - 1$ is a prime number (Dixon 1984). What is interesting about the data presented in the table is the number of codes available, that is not a monotonic function of the register length. For instance, it is always preferable to use the least prime number for m as one progresses in the table because it gives the maximum number of sequences.

When several users share the same channel without phase coherence, as in the case of a totally asynchronous system, it is desirable that the cross-correlation functions be small in absolute value. In addition, the

Table 6.1. Number of maximal sequences

m	Number of codes	Prime factors of $2^m - 1$
3	2	7
4	4	3 and 5
5	6	31
6	4	3, 3, and 7
7	18	127
8	16	3, 5, and 17
9	48	7 and 73
10	60	3, 11, and 31
11	176	23 and 89
12	96	3, 3, 5, 7, and 13
13	630	8, 191
14	756	3, 43, and 127
15	1,800	7, 31, and 151
16	2,048	3, 5, 17, and 257
17	7,710	131, 071
18	1,728	3, 3, 3, 7, 19, and 73
19	27,594	524, 287
20	19,200	7, 7, 127, and 137

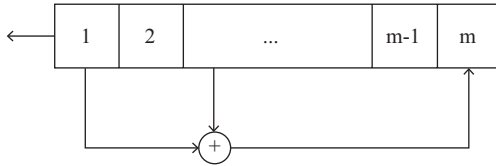


Figure 6.6. Pseudo-noise sequence generator

autocorrelation function must attain the maximum possible value. In matched filter detection, the output of the detector is the correlation of the input and a replica of the desired signal.

The investigation of lower bounds for the cross-correlation between pseudo-sequences is still a matter of research. One of the important results for the multichannel case is due to Welch, and states that the lower bound on the peak magnitude of the cross-correlation r_{MAX} is (Welch 1974)

$$r_{MAX} \geq N \sqrt{\frac{M-1}{MN-1}} \quad (6.6)$$

in which the set is composed of M sequences of length N .

Example: for large values of N and M the bound is well approximated as \sqrt{N} .

6.4.2 GOLD SEQUENCES

A method for choosing the linear maximal codes, to be used as components of Gold sequences, was introduced in 1967 (Gold 1967). The Gold sequences are generated by modulo-2 addition of a pair of maximal-length sequences, as shown in Figure 6.7. The code sequences are added on a chip basis, by synchronous clocking. The main feature of this type of sequence generator is the large number of codes it supplies, although it requires only one pair of feedback tap sets (Dixon 1984).

The Gold sequences are constructed from m -sequences. For a given m and a pair $\{\mathbf{a}, \mathbf{b}\}$ of distinct m -sequences, the Gold code is the set

$$C = \{\mathbf{a}, \mathbf{b}, \mathbf{a} + T^i \mathbf{b} : i = 0, \dots, 2^m - 2\}. \quad (6.7)$$

The Gold code contains $2^m + 1$ sequences of block-length $2^m - 1$. The elements of the code are called Gold sequences. Sequences \mathbf{a} and \mathbf{b} are selected so their cross-correlation function has a maximum value

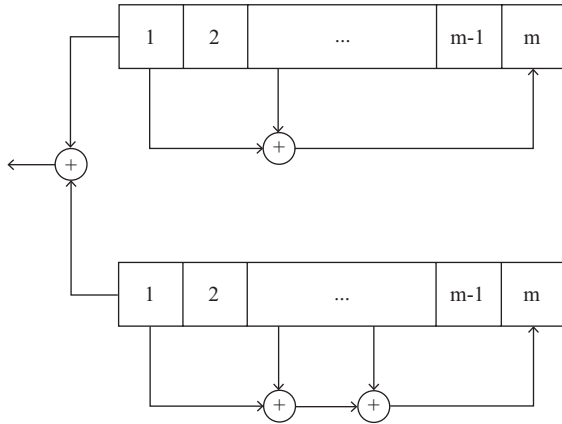


Figure 6.7. Gold sequence generator.

$2^{\lfloor(m+2)/2\rfloor} + 1$. Such a pair of sequences will always exist. The cross-correlation functions and, except of the main peak, the autocorrelation functions of a Gold code can only take values in the set $\{-1, -2^{\lfloor(m+2)/2\rfloor} - 1, 2^{\lfloor(m+2)/2\rfloor} - 1\}$.

Hence the largest magnitude of any cross-correlation of pairs of sequences from \mathcal{C} is $2^{\lfloor(m+2)/2\rfloor} + 1$ (Blahut 1990). Table 6.2 compares the relative peak cross-correlations for the m -sequences and Gold sequences (Proakis 1989). It can be seen that the periodic cross-correlation function between the pairs of m -sequences has large peaks when compared to the Gold sequences of the same period.

Table 6.2. Relative peak cross-correlations of m -sequences, Gold sequences and Kasami sequences

m	m -sequences	Gold sequences	Kasami sequences
3	0.71	0.71	–
4	0.60	0.60	0.33
5	0.35	0.29	–
6	0.36	0.27	0.14
7	0.32	0.13	–
8	0.37	0.13	0.07
9	0.22	0.06	–
10	0.37	0.06	0.03
11	0.14	0.03	–
12	0.34	0.03	0.03

Gold codes are useful for CDMA applications for they form a large set of codes. But, with the exception of a and b , the set of Gold sequences are not maximum-length shift-register sequences of length N . Hence their autocorrelation functions are not two-valued. Such codes can be viewed as the duals of the cyclic codes generated by the product of two primitive polynomials.

The advantage of Gold sequence generators is the possibility of producing a new sequence with every change in phase position between the two primary generators. In addition to this, the Gold codes may be chosen so that, over a set of codes available from a given generator, the cross-correlation between the codes is bounded. Thus, Gold codes are attractive for applications in CDMA.

Furthermore, the maximal sequences of the same length are not guaranteed to have bounded cross-correlation. But, for a given value of m , the Gold codes exhibit cross-correlation that is $\sqrt{2}$ greater than the maximal length sequences of the same length (Dixon 1984).

A method for choosing the linear maximal codes, to be used as components of Gold sequences, was introduced in 1967. The outputs form a set of sequences whose members display cross-correlation and autocorrelation side-lobes bounded by $2^{(m+1)/2} + 1$ for m odd, and by $2^{(m+1)/2} - 1$ for m even (Gold 1967).

6.4.3 KASAMI SEQUENCES

The Kasami ensemble of codes is also constructed from m -sequences. Each code is a collection of sequences that presents good cross-correlation properties. Consider a m -sequence \mathbf{a} of length $N = 2^m - 1$, for m even. This can be written as $N = (2^{m/2} - 1)(2^{m/2} + 1)$, which means that starting with any position of \mathbf{a} and taking every $(2^{m/2} + 1)$ th bit of sequence \mathbf{a} cyclically repeated gives a sequence \mathbf{b} of length N and period $(2^{m/2} - 1)$. There are $(2^{m/2} - 1)$ distinct shifts of \mathbf{b} .

The Kasami code is the set

$$\mathcal{C} = \{\mathbf{a}, \mathbf{a} + T^i \mathbf{b} : i = 0, \dots, 2^{m/2} - 2\} \quad (6.8)$$

in which the operator T denotes a cyclic shift by one bit position. The Kasami sequences are the elements of the Kasami code, each sequence having block-length N .

The cross-correlation functions and off-peak autocorrelation functions of elements of a Kasami code only take values in the set $\{-1, 2^{m/2} - 1, -2^{m/2} - 1\}$. Hence the largest magnitude of any cross-correlation function

pairs of sequences from the code is $2^{m/2} + 1$ (Blahut 1990). The Kasami sequences also satisfy the Welch bound, and are in this sense optimal (Proakis 1989).

It might be noted that if the sequences are allowed to be defined over the complex m th roots of unity, for some positive integer m , then the auto and cross-correlation properties of the sequences could be improved greatly, and still maintain the balanced and run properties of the m -sequences, that is, there is approximately an even distribution of each of the 2^m elements of the Galois Field, and short runs of identical symbols are more likely to appear than longer runs (Komo and Liu 1990). An extensive work on the subject of sequence design is found in (Sarwate and Pursley 1980).

Some attributes of spread spectrum, employed by CDMA systems, are important for personal communications systems, including the utilization of the voice activity factor, no need for equalization, use of one radio per site, soft handoff, no need for guard time, sectorization for capacity, less fading for the wide-band signal, easy transition from current systems, capacity advantage, no frequency management or assignment needed, soft capacity, coexistence with other systems and suitability for microcell and in-building systems (Lee 1991).

Modeling a spread spectrum system implies the generation of random sequences. Those sequences must exhibit the same properties, in order to be eligible for CDMA use. In real life, practical constraints dictate the use of pseudo-random, or pseudo-noise sequences (PN), whose properties are well-known.

CHAPTER 7

THE CAPACITY OF A CDMA SYSTEM

7.1 INTRODUCTION

The area of Personal Communications Network (PCN) is a huge success in the communications field. The main technology used in mobile communications is code division multiple access (CDMA) that has some advantages over the previous techniques. This chapter describes and analyzes a model for the discrete-input continuous-output, Direct-Sequence Spread Spectrum (DSSS), Gaussian multiple access channel.

In this chapter, the modulation scheme is specialized, with the choice of DPSK modulation. This modulation scheme deserves some comments. DPSK is regarded by many authors as a natural choice for CDMA, especially if one considers a fading multipath environment, in which synchronous carrier recovery is a difficult task (Misser, Wijffels, and Prasad 1991; Turin 1984b).

As a differentially coherent modulation technique, DPSK also compares well to similar systems using coherent PSK modulation and is less troublesome to demodulate in a complex environment such as an indoor wireless channel (Kavehrad and McLane 1985; Kavehrad and Ramamurthi 1987). The capacity of the PSK scheme, for the coding channel resulting from hard decision detection, was computed in Sousa (1992). In this chapter, upper and lower bounds on the sum capacity are obtained and compared.

7.2 ANALYSIS OF A CDMA SYSTEM WITH A FIXED NUMBER OF USERS AND SMALL SNR

One of the problems related to CDMA is the possibility of interference with existing systems (Kavehrad and McLane 1985). Unlike the former

techniques, CDMA has its capacity solely limited by the amount of interference into the system, which allows any reduction in interference to be converted directly and linearly into an increase in capacity (Gilhousen et al. 1991).

The use of spread spectrum allows new users to share the spectrum with existing ones. Because CDMA is interference limited, the number of users that share the same spectrum, and still maintain an acceptable performance, is determined by the interference generated by the set of users.

In CDMA systems, each user is given a distinct pseudo-random code, which either identifies the user to the central station in a star network, or addresses another user in a more fully connected network (Turin 1984b). Hence, the users utilize the same bandwidth and therefore interfere, making a received signal appear noisier as the traffic increases in the system. The effect of the interfering energy depends on the processing gain of the spread spectrum technique.

As implied, spread spectrum systems are less subject to multipath signal variations than conventional systems. In a direct sequence receiver, if the reflected signal is delayed compared to the direct signal by more than one code chip, the reflected signal is treated as uncorrelated to the direct signal.

The higher the code chip rate, for a particular system, the smaller its multipath problem. In the mobile environment, the result of multipath reception is that a moving receiver is subject to rapid fading and peaking of its input signal as the direct and reflected signals cancel and reinforce each other. Stationary receivers are also subject to multipath, due to the reflecting and refracting surfaces (Dixon 1984).

Researchers, industries, and prospective users of CDMA systems have been concerned with three major issues regarding the design of those systems. The first one is the design of pseudo-noise sequences with good autocorrelation and cross-correlation properties.

The second issue, is the evaluation of system performance, that can be done in two ways: using the average probability of error or computing the outage of the system. The probability of error is averaged over the ensemble of channels obtained from the assumed model.

In the case of time-varying channel and stationary terminal, outage indicates the fraction of time the terminal operates with an error probability above the limit. For the time-invariant channel and mobile terminal, outage indicates the fraction of building area where the terminal can operate (Kavehrad and Ramamurthi 1987).

The evaluation of the system capacity, or which transmission rate the system can support under given circumstances, is the third issue. The number

of allocated users dictates the type of distribution to characterize the multiple access interference.

A large number of subscribers favor the Gaussian approximation (Turin 1984b) and, for many applications, this approximation is satisfactory even for a processing gain as low as $N = 10$ and moderate values of the signal to interference ratio (Sousa 1990). On the other hand, the Gaussian approximation is typically more accurate for smaller values of the SNR, or large multiuser interference (Geraniotis and Ghaffari 1991).

For a small number of subscribers, considering that the MAI is additively composed of cross-correlations limited to a few chips of coincidence, the maximum interference is too small to apply the central limit theorem (Brady 1991), or the conditions for utilization of the theorem do not apply if certain parameters are held constant (Sadowsky and Bahr 1991).

But, even for a small number of users, accurate methods have been devised to calculate multiple access error probabilities using the Gaussian approximation (Holtzman 1992). This section is dedicated to the evaluation of the capacity of a CDMA system, when the number of users is held constant.

CDMA is a type of noncooperative channel, which has asynchronous multiplexing, implying that the M information sources transmit through a common medium without any prearrangement among themselves. This kind of channel has been investigated for some time, in a general sense, and some interesting results were found (Sommer 1966; Verdú 1989b).

In this type of communications system, several users transmit simultaneously and concurrently to a common receiver. Therefore, all components of the received signal are statistically independent and so, perform as interference in respect to any other given signal. The CDMA channel is analyzed through the use of an information theoretic approach in this section.

The concept of capacity is used regarding the limitation in the transmission rate due to bandwidth, or related aspects, while the idea of throughput is used when scheduling is the main concern in the performance evaluation of the system. In the case of CDMA, both factors are involved which gives rise to the concept of sum capacity, or limiting throughput per unit bandwidth (Sousa 1989).

Because of the reasoning and considerations developed in the following, the term capacity, used heretofore, is equivalent to the concept of sum capacity, derived in the cited article. The capacity is assumed for the case of noncooperation among the users and each user contributes a small fraction of the traffic to the channel (Sousa 1989).

The CDMA noncooperative multiple access channel discussed in this section has M binary inputs, $X_i \in \mathcal{X}_i$, and a single Gaussian output, $Y \in \mathcal{Y}$.

The output is the combination of the M inputs plus noise and can be assumed Gaussian, using the central limit theorem for a large number of users (Sousa 1989). The channel is assumed memoryless and, for the assumptions made, the optimum receiver for a particular signal will be the maximum likelihood receiver.

In the following, one considers the performance analysis of an asynchronous CDMA system using direct-sequence codes. Let $b_{ij} = \pm 1$, $i = 0, 1, \dots, M$, $j = -\infty, \dots, \infty$ be a binary bit string that modulates the i th transmitter and let the codes $c_i(t)$, $i = 1, \dots, M$, $0 \leq t \leq T_b$ form a set of quasi-orthogonal sequences, with cross-correlation given by

$$r_{ij}(\tau) = \int_0^{T_b} c_i(t)c_j(t - \tau) dt, \quad |\tau| \leq T_b, \quad i \neq j \quad (7.1)$$

which is uniformly small compared to their individual energy (Turin 1984b)

$$E_c = \int_0^{T_b} c_i^2(t) dt, \quad i = 1, \dots, M. \quad (7.2)$$

At the reception side, the desired signal is represented by

$$x_i(t) = \sum_{j=-\infty}^{\infty} b_{i\lfloor j/N \rfloor} c_i(t - jT_b) \phi_j(t) \quad (7.3)$$

in which $\{\phi_j(t)\}$ is a set of orthonormal functions and $\lfloor j/N \rfloor$ represents the greatest integer less than or equal to j/N . If modulation is used, along with rectangular chip pulses, the functions are given by

$$\phi_j(t) = \cos[\omega_c(t - jT_b)]. \quad (7.4)$$

Therefore, the signal whose reception is sought can be expressed as

$$x_i(t) = \sum_{j=-\infty}^{\infty} b_{i\lfloor j/N \rfloor} c_i(t - jT_b) \cos[\omega_c(t - jT_b)]. \quad (7.5)$$

However, the M transmitted signals combine additively, giving for any receiver a total signal $y(t)$, such that

$$y(t) = \sum_{i=1}^M x_i(t) + n(t) \quad (7.6)$$

in which $n(t)$ is a zero mean Gaussian noise of double-sided power density $N_0/2$. Consequently, part of the total signal is viewed as noise for each

user. The other $M - 1$ current users present a multiple access interference (MAI) at the i th receiver of the form

$$n_I(t) = \sum_{k=1, k \neq i}^M x_k(t) \quad (7.7)$$

or, upon substitution

$$n_I(t) = \sum_{k=1, k \neq i}^M \sum_{j=-\infty}^{\infty} b_{k\lfloor j/N \rfloor} c_k(t - jT_b - \tau_k) \cos[\omega_c(t - jT_b - \tau_k)], \quad (7.8)$$

in which $\{\tau_k\}$ is a set of independent random variables distributed over the interval $[0, T_b)$. The total received signal can be written as an explicit function of the MAI and Gaussian noise

$$y(t) = x_i(t) + n_I(t) + n(t). \quad (7.9)$$

Using the maximum likelihood receiver, or a matched filter followed by a detector, the filter output y_i at time $t = T_b$ is given by (Turin 1980; Kavehrad and McLane 1985)

$$y_i = \int_0^{T_b} y(t) c_i(t) \cos \omega_c t \, dt \quad (7.10)$$

which can be put in the form

$$\begin{aligned} y_i &= b_i E_b + \frac{1}{2} \sum_{k=1, k \neq i}^M \sum_{j=-\infty}^{\infty} b_{i\lfloor j/N \rfloor} \\ &\quad \times \left[\int_0^{T_b} c_i(t) c_k(t - jT_b - \tau_k) \, dt \right] \cos \phi_{kj} \\ &\quad + \int_0^{T_b} c_i(t) n(t) \cos \omega_c t \, dt \end{aligned} \quad (7.11)$$

in which

$$\phi_{kj} = \theta_k - \omega_c \tau_k - j\omega_c T_b \quad (7.12)$$

and $\{\theta_k\}$ form a set of independent random variables, uniformly distributed in the interval $[0, 2\pi)$. It is also considered that τ_k and θ_k are independent. The polarity of a given bit is determined by a comparison between the phases at the output of the filter at $t = (j + 1)T_b$ and jT_b .

It is assumed that $\omega_c T_b = 2\pi l$, in which l is an integer. This is equivalent to the narrow-band assumption, as discussed in (Turin 1984b). Also, the

mean and variance of the third term are 0 and $N_0E_b/2$, respectively, in which

$$E_b = \int_0^{T_b} c_i^2(t) \cos^2 \omega_c t \, dt, \quad i = 1, \dots, M. \quad (7.13)$$

The second term deserves some attention. It can be written as (Turin 1984a)

$$N_i = \frac{1}{2} \sum_{k=1, k \neq i}^M n_k \cos \phi_{kj} \quad (7.14)$$

in which

$$n_k = \sum_{j=-\infty}^{\infty} b_{i|j|/N} \int_0^{T_b} c_i(t) c_k(t - jT_b - \tau_k) \, dt. \quad (7.15)$$

But, the variance of the term n_k , defined as σ_k^2 , inside the summation can be expressed as

$$\sigma_k^2 = 2 \int_0^{T_b/N} \left(1 - \frac{N\tau}{T_b}\right)^2 (T_b - \tau) d\tau = 2 \left[\frac{T_b^2}{3N} - \frac{T_b^2}{12N^2} \right]. \quad (7.16)$$

Therefore, the variance of N_i , defined as σ_i^2 , is

$$\sigma_i^2 = \frac{M-1}{8} \sigma_k^2 = \frac{(M-1)T_b^2}{12N} - \frac{(M-1)T_b^2}{48N^2}. \quad (7.17)$$

Because the MAI term is assumed Gaussian, and $E_b = T_b/2$, the filter output is a Gaussian variable, with mean $\mu_I = b_i E_b$. Therefore, the multiple access variance σ_I^2 is given by (Borth and Pursley 1979; Turin 1984b; Pickholtz, Schilling, and Milstein 1982)

$$\sigma_I^2 = \frac{(M-1)E_b^2}{3N} - \frac{(M-1)E_b^2}{12N^2} + \frac{N_0E_b}{2}. \quad (7.18)$$

Based on the results stated by Shannon (1948), the mutual information between the i th individual input and the output, for the noncooperative additive channel, can be written as

$$\begin{aligned} I(X_i; Y) &= H(Y) - H(Y - X_i) \\ &= \sum_{j=0}^{q-1} \int_{-\infty}^{\infty} p(y|x_j) P(x_j) \log \frac{p(y|x_j)}{p(y)} dy \end{aligned} \quad (7.19)$$

in which

$$p(y) = \sum_{j=0}^{q-1} p(y|x_j) P(x_j), \quad (7.20)$$

the logarithms are to the base 2, unless otherwise stated, implying that information rates are given in bits and q represents the number of signal levels.

The maximization of Equation 7.19, over the set of input probabilities, gives the expressions for the individual capacities

$$C_i = \max_{P(x_j)} \sum_{j=0}^{q-1} \int_{-\infty}^{\infty} p(y|x_j)P(x_j) \log \frac{p(y|x_j)}{p(y)} dy. \quad (7.21)$$

The expression is set as the maximum of the average mutual information between the input $X = \{x_1, \dots, x_{q-1}\}$ and the output $Y = (-\infty, \infty)$.

For possible inputs $x_k = +1$ and $x_k = -1$, the average mutual information is maximized when the input probabilities are $P(x_k = +1) = P(x_k = -1) = 1/2$. Hence, the expression can be put in the form (Alencar and Blake 1993c)

$$C_i = \frac{1}{2} \int_{-\infty}^{\infty} p(y|+E_b) \log \frac{p(y|+E_b)}{p(y)} dy + \frac{1}{2} \int_{-\infty}^{\infty} p(y|-E_b) \log \frac{p(y|-E_b)}{p(y)} dy \quad (7.22)$$

in which, assuming that $n_I(t)$ is Gaussian,

$$p(y|\pm E_b) = \frac{1}{\sigma_I \sqrt{2\pi}} e^{-\frac{(y \mp E_b)^2}{2\sigma_I^2}}. \quad (7.23)$$

The noise variance σ_I^2 was obtained assuming that the interuser interference amounts to a Gaussian distribution, for a CDMA system using direct-sequence codes (Turin 1984b). It can be simplified by noting that for $c_{ik}(t) = \pm 1$, $k = 1, \dots, N$, the code energy is $E_c = T_b = 2E_b$, and for usual values of the processing gain (Pursley 1977; Geraniotis and Pursley 1985)

$$\sigma_I^2 = \frac{(M-1)E_b^2}{3N} + \frac{N_0 E_b}{2} \quad (7.24)$$

in which M is the number of users in the channel, N is the processing gain, $N_0/2$ is the Gaussian noise power spectrum density, and T_b is the bit time.

It can reasonably be argued, for the present model, that the sum capacity is a one-feature function containing most of the needed information about the behavior of the system. This is fair to say for spread spectrum techniques, which define a noncooperative type of channel whose capacity is given by the sum of the individual capacities,

$$C = \sum_{i=1}^M C_i. \quad (7.25)$$

The individual rates, upon substitution of the appropriate terms, are equal to

$$C_i = \frac{1}{2\sigma_I\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-E_b)^2}{2\sigma_I^2}} \log \left[\frac{2 \exp\left(-\frac{(y-E_b)^2}{2\sigma_I^2}\right)}{\exp\left(-\frac{(y-E_b)^2}{2\sigma_I^2}\right) + \exp\left(-\frac{(y+E_b)^2}{2\sigma_I^2}\right)} \right] dy \\ + \frac{1}{2\sigma_I\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y+E_b)^2}{2\sigma_I^2}} \log \left[\frac{2 \exp\left(-\frac{(y+E_b)^2}{2\sigma_I^2}\right)}{\exp\left(-\frac{(y-E_b)^2}{2\sigma_I^2}\right) + \exp\left(-\frac{(y+E_b)^2}{2\sigma_I^2}\right)} \right] dy$$

which can be simplified to

$$C_i = \frac{1}{\sigma_I\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-E_b)^2}{2\sigma_I^2}} \log \left[\frac{2}{1 + \exp\left(-\frac{2yE_b}{\sigma_I^2}\right)} \right] dy. \quad (7.26)$$

The remaining part of this section is concerned with establishing a suitable approximation for the sum capacity in Equation 7.26, in terms of upper and lower bounds, for small values of the signal to total noise ratio. Expanding the logarithm term in a power series in E_b , for low values of the signal-to-noise ratio (SNR), one obtains (Gallager 1968; Gradshteyn and Ryzhik 1990)

$$\log \left[\frac{2}{1 + \exp\left(-\frac{2yE_b}{\sigma_I^2}\right)} \right] = \log e \left[\frac{yE_b}{\sigma_I^2} - \frac{y^2E_b^2}{2\sigma_I^4} + \frac{y^4E_b^4}{12\sigma_I^8} + O(E_b^6) \right]. \quad (7.27)$$

The expansion is valid for the case in which the number of users in the system is much larger than the code length, or when the thermal noise is preponderant in the system, according to Equation 7.13.

Substituting into Equation 7.26 and integrating, gives

$$C_i = \log e \left[\frac{E_b^2}{2\sigma_I^2} - \frac{E_b^4}{4\sigma_I^4} + \frac{E_b^6}{2\sigma_I^6} + \frac{E_b^8}{12\sigma_I^8} + O\left(\frac{E_b^{10}}{\sigma_I^{10}}\right) \right] \quad (7.28)$$

in which the remaining terms become insignificant as the SNR goes to zero. It is appropriate to say that spread spectrum systems operate, in some cases, under conditions of very low SNR. Therefore, neglecting all but the first term, one obtains an upper bound on the individual capacity

$$C_i \leq \log e \left[\frac{E_b^2}{\sigma_I^2} \right] \quad (7.29)$$

Because the series in 7.27 is an alternating one, a lower bound can be found easily considering also the second term

$$C_i \geq \log e \left[\frac{E_b^2}{2\sigma_I^2} - \frac{E_b^4}{2\sigma_I^4} \right]. \quad (7.30)$$

Using the first term in Equation 7.27 as an approximation for the individual capacity and substituting the formula for the MAI variance one finds a formula for the capacity, under the assumption of very low SNR

$$C_i \approx \log e \left[\frac{E_b^2}{2 \left(\frac{(M-1)E_b^2}{3N} + \frac{N_0 E_b}{2} \right)} \right]. \quad (7.31)$$

Using now Equation 5.34, maximized for the case in which all the transmitters carry the same power, one obtains

$$\mathbf{C} = MC_i \approx \frac{3 \log e MN}{2(M-1) + 3N \frac{N_0}{E_b}}. \quad (7.32)$$

Example: interesting results are obtained when one varies the SNR E_b/N_0 or number of users in the system. One gets, as the SNR goes to zero

$$\lim_{\frac{E_b}{N_0} \rightarrow 0} \mathbf{C} = 0 \quad (7.33)$$

If the SNR is small, as compared to the ratio of processing gain to number of users, $E_b/N_0 \ll 3N/2(M-1)$, then the capacity can be approximated by

$$\mathbf{C} \approx \log e M \frac{E_b}{N_0} \quad (7.34)$$

and the capacity is proportional to the SNR, as well as to the number of channel users.

Example: another interesting result can be obtained by taking the limit of the capacity when the number of sources goes to infinity, for a rectangular chip pulse,

$$\lim_{M \rightarrow \infty} \mathbf{C} \approx \frac{3 \log e N}{2} \quad (7.35)$$

It is easy to conclude that, as the number of users increases towards infinity, the capacity region shrinks, and the channel capacity converges to a value that is proportional to the processing gain. In this case, by using a set of orthogonal spreading codes, the capacity becomes independent of the SNR and it is not possible to increase the capacity of the noncooperative channel by increasing the power of the transmitted signals.

Because the limiting capacity C becomes dependent on the processing gain, as the number of senders goes to infinity, the only possibility left to boost the capacity is by an increase in the available bandwidth. This is one of the reasons why spread spectrum techniques are useful for noncooperative type of systems. Figure 7.1 shows the relation between the exact (middle curve) and the approximate capacity (upper and lower bounds) in terms of the SNR, for a given code length ($N = 100$) and a number of users ($M = 500$).

The exact formula is obtained from 7.26, after substitution into 5.34. The approximate value 7.32 is used as the upper bound. It is possible to note that the upper and lower approximations are fairly close to the exact result, if the SNR is below -5 dB.

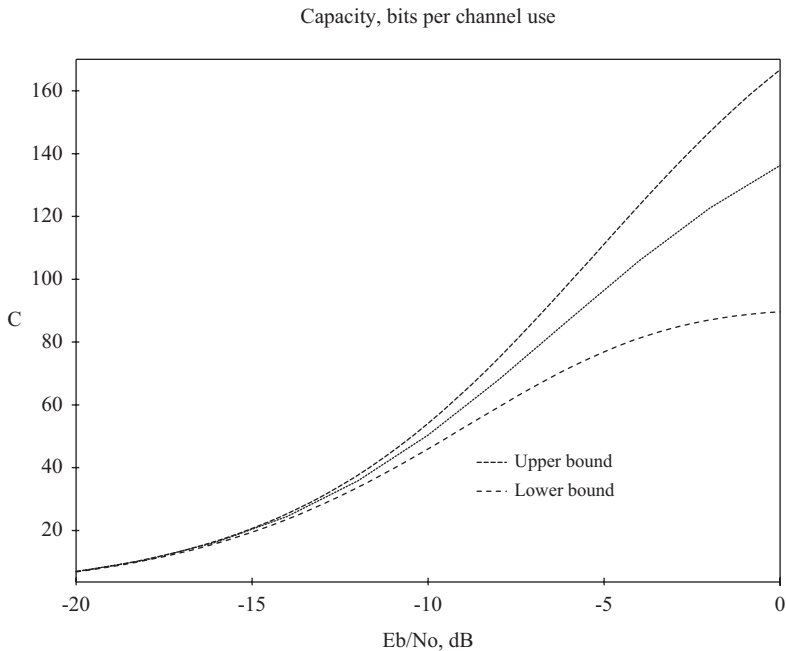


Figure 7.1. Capacity approximations for the channel, as a function of the SNR, for $M = 500$ and $N = 100$.

7.3 CDMA SYSTEM WITH A FIXED NUMBER OF USERS AND HIGH SNR

In the previous section, an approximation for the capacity of a CDMA system was obtained considering a small SNR. In that case, either the number of users was very large, as compared to the processing gain, or the signal power was completely embedded in the Gaussian noise. This section presents a set of approximations for the channel capacity, considering now a large SNR, or a large ratio of processing gain to number of users.

In order to analyze the sum capacity in detail, it is important to find both approximations and bounds for the function. The equation for the individual capacity 7.26 can be simplified by the application of logarithm properties to

$$C_i = \frac{1}{\sigma_I \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-E_b)^2}{2\sigma_I^2}} \left[1 - \log \left(1 + \exp \left(-\frac{2yE_b}{\sigma_I^2} \right) \right) \right] dy \quad (7.36)$$

which, upon integration of the first term, gives

$$C_i = 1 - \frac{1}{\sigma_I \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(y-E_b)^2}{2\sigma_I^2}} \log \left(1 + e^{-\frac{2yE_b}{\sigma_I^2}} \right) dy. \quad (7.37)$$

Expanding the exponent leads finally to

$$C_i = 1 - \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \frac{1}{\sigma_I \sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2\sigma_I^2}} e^{-\frac{yE_b}{\sigma_I^2}} \ln \left(1 + e^{-\frac{2yE_b}{\sigma_I^2}} \right) dy. \quad (7.38)$$

As seen from the analysis undergone in the last section, the integrand in the above equation is suitable for a series expansion, as long as $E_b/N_0 \ll 1$. In order to adjust the integrand, so that the expansion can hold for $E_b/N_0 \gg 1$ as well, one can resort to the Fourier transform properties. The functions inside the integral

$$g(y) = \frac{1}{\sigma_I \sqrt{2\pi}} e^{-\frac{y^2}{2\sigma_I^2}} \quad (7.39)$$

and

$$f(y) = e^{-\frac{yE_b}{\sigma_I^2}} \ln \left(1 + e^{-\frac{2yE_b}{\sigma_I^2}} \right) \quad (7.40)$$

can be Fourier transformed to (Gradshteyn and Ryzhik 1990),

$$G(\omega) = \mathcal{F}[g(y)] = e^{-\frac{\sigma_N^2 \omega^2}{2}} \quad (7.41)$$

and

$$F(\omega) = \mathcal{F}[f(y)] = \frac{\pi \csc\left(\frac{\pi}{2} + \frac{j\pi\sigma_I^2\omega}{2E_b}\right)}{\frac{E_b}{\sigma_I^2} + j\omega}. \quad (7.42)$$

Therefore, applying the Fourier transform property of the integral of the product of two functions,

$$\int_{-\infty}^{\infty} f(y)G(y)dy = \int_{-\infty}^{\infty} F(\omega)g(\omega) d\omega \quad (7.43)$$

one obtains

$$C_i = 1 - \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \int_{-\infty}^{\infty} e^{-\frac{\sigma_I^2\omega^2}{2}} \frac{\pi \csc\left(\frac{\pi}{2} + \frac{j\pi\sigma_I^2\omega}{2E_b}\right)}{\frac{E_b}{\sigma_I^2} + j\omega} d\omega. \quad (7.44)$$

This last integral can be simplified, using the following property of hyperbolic functions

$$\csc\left(\frac{\pi}{2} + \frac{j\pi\sigma_I^2\omega}{2E_b}\right) = \sec\left(\frac{j\pi\sigma_I^2\omega}{2E_b}\right) \quad (7.45)$$

then

$$C_i = 1 - \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \int_{-\infty}^{\infty} e^{-\frac{\sigma_I^2\omega^2}{2}} \frac{\pi \sec\left(\frac{j\pi\sigma_I^2\omega}{2E_b}\right)}{\frac{E_b}{\sigma_I^2} + j\omega} d\omega. \quad (7.46)$$

The application of yet another simple property gives

$$C_i = 1 - \pi \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \int_{-\infty}^{\infty} \frac{e^{-\frac{\sigma_I^2\omega^2}{2}}}{\cosh\left(\frac{\pi\sigma_I^2\omega}{2E_b}\right) \left(\frac{E_b}{\sigma_I^2} + j\omega\right)} d\omega. \quad (7.47)$$

Multiplying and dividing the integrand by the conjugate of the complex term in the denominator leads to

$$C_i = 1 - \pi \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \int_{-\infty}^{\infty} \frac{e^{-\frac{\sigma_I^2\omega^2}{2}} \left(\frac{E_b}{\sigma_I^2} - j\omega\right)}{\cosh\left(\frac{\pi\sigma_I^2\omega}{2E_b}\right) \left(\frac{E_b^2}{\sigma_I^4} + \omega^2\right)} d\omega. \quad (7.48)$$

The integral can now be split into two terms

$$C_i = 1 - \pi \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \left[\int_{-\infty}^{\infty} \frac{e^{-\frac{\sigma_I^2\omega^2}{2}} \left(\frac{E_b}{\sigma_I^2}\right)}{\cosh\left(\frac{\pi\sigma_I^2\omega}{2E_b}\right) \left(\frac{E_b^2}{\sigma_I^4} + \omega^2\right)} d\omega - \int_{-\infty}^{\infty} \frac{j\omega e^{-\frac{\sigma_I^2\omega^2}{2}}}{\cosh\left(\frac{\pi\sigma_I^2\omega}{2E_b}\right) \left(\frac{E_b^2}{\sigma_I^4} + \omega^2\right)} d\omega \right]. \quad (7.49)$$

But the second integrand is an odd function of ω and therefore integrates to zero. The first integrand is even, which yields finally

$$C_i = 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2} \frac{E_b}{\sigma_I^2}} \int_0^\infty \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\cosh\left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right) \left(\frac{E_b^2}{\sigma_I^4} + \omega^2\right)} d\omega. \quad (7.50)$$

The capacity can be lower and upper bounded by recognizing that the Cauchy function is upper bounded by 1 and lower bounded by $\exp(-\sigma_I^4 \omega^2 / E_b^2)$, respectively. The lower bound is

$$C_i \geq 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2} \frac{\sigma_I^2}{E_b}} \int_0^\infty \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\cosh\left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right)} d\omega \quad (7.51)$$

and the upper bound

$$C_i \leq 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2} \frac{\sigma_I^2}{E_b}} \int_0^\infty \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}} e^{-\frac{\sigma_I^4 \omega^2}{E_b^2}}}{\cosh\left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right)} d\omega. \quad (7.52)$$

Both expressions are good approximations for the actual value of the capacity if $E_b/\sigma_I^2 \gg 1$, which implies $N \gg M$. For very long code sequences, the hyperbolic cosine can also be simplified to

$$\cosh\left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right) = 1 + \frac{1}{2} \left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right)^2 + \dots \approx 1 \quad (7.53)$$

Using this simplification in Equation 7.51, one obtains the lower bound and approximation

$$C_i \geq 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2} \frac{\sigma_I^2}{E_b}} \int_0^\infty e^{-\frac{\sigma_I^2 \omega^2}{2}} d\omega. \quad (7.54)$$

and finally

$$C_i \geq 1 - \pi^{3/2} \log e \frac{\sigma_I}{E_b} e^{-\frac{E_b^2}{2\sigma_I^2}}. \quad (7.55)$$

This exponential approximation, referred to as *Bound 1*, is plotted along with the exact capacity in Figure 7.2, for $M = 20$ and $N = 100$, after the substitution into the formula for the sum capacity.

On the other hand, another approximation can be attempted using Equation 7.52, which can be put in the following form

$$C_i \leq 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2} \frac{\sigma_I^2}{E_b}} \int_0^\infty \frac{e^{-\left(\frac{\sigma_I^2}{2} + \frac{\sigma_I^4}{E_b^2}\right)\omega^2}}{\cosh\left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right)} d\omega. \quad (7.56)$$

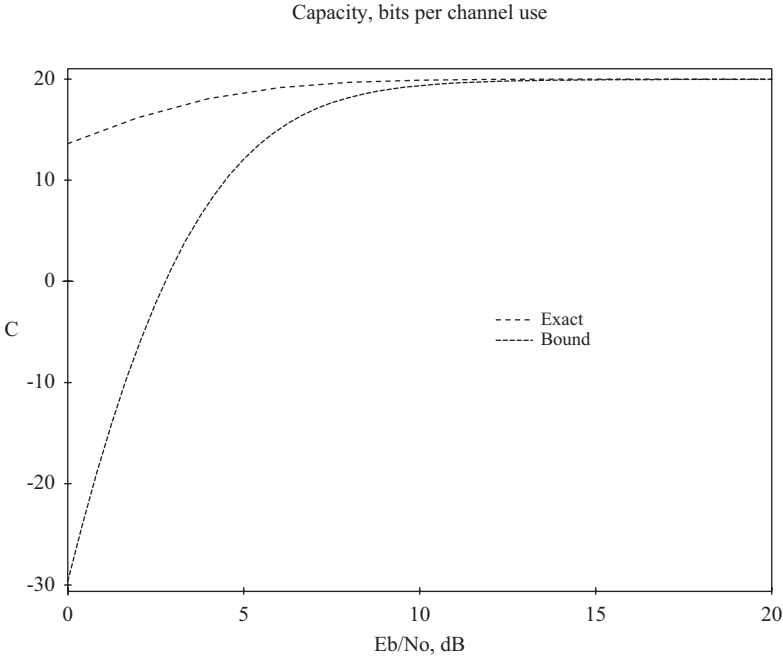


Figure 7.2. Bound 1 on the capacity for the channel, as a function of the SNR (E_b/N_0).

and integrated, after the simplification indicated by 7.53, to provide the approximation

$$C_i \approx 1 - \frac{\sqrt{2}\pi^{3/2}}{2} \log e \frac{\sigma_I}{E_b} \frac{1}{\sqrt{\frac{1}{2} + \frac{\sigma_I^2}{E_b^2}}} e^{-\frac{E_b^2}{2\sigma_I^2}} \quad (7.57)$$

It is interesting to observe that both 7.55 and 7.57 converge to the same limit as the SNR increases. This approximation, referred to as *Bound 2*, is plotted to compare with the exact capacity in Figure 7.3, for $M = 20$ and $N = 100$, after the substitution into the formula for the sum capacity.

Equation 7.51 can lead to another lower bound on the individual capacity. Consider for a while only the integral term of the mentioned equation

$$I = \int_0^\infty \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\cosh\left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right)} d\omega. \quad (7.58)$$

Capacity, bits per channel use

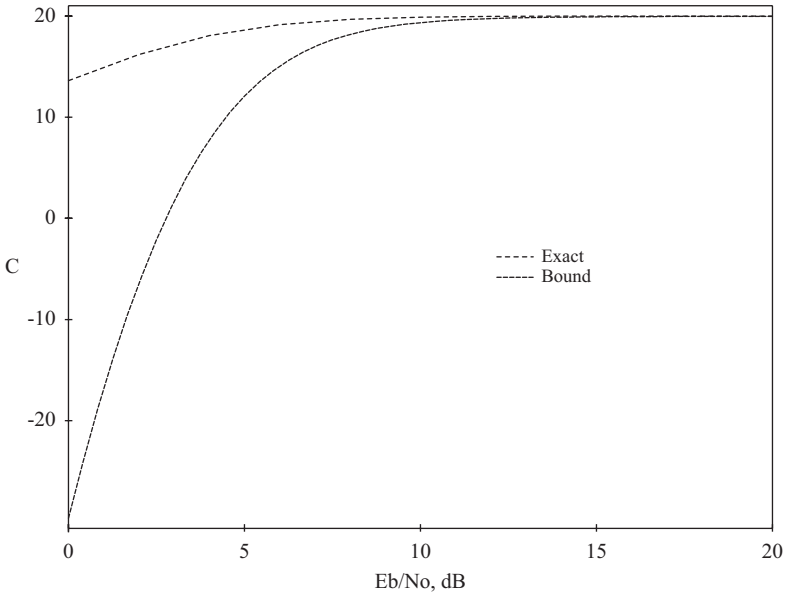


Figure 7.3. Bound 2 on the capacity for the channel, as a function of the SNR (E_b/N_0).

It is not difficult to obtain a simpler expression, making use of the property of integration by parts,

$$\begin{aligned}
 I = & \frac{2E_b e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\pi \sigma_I^2} \arctan \left[\sinh \left(\frac{\pi \sigma_I^2 \omega}{2E_b} \right) \right] \Bigg|_0^\infty \\
 & + \int_0^\infty \frac{2E_b}{\pi} \omega e^{-\frac{\sigma_I^2 \omega^2}{2}} \arctan \left[\sinh \left(\frac{\pi \sigma_I^2 \omega}{2E_b} \right) \right] d\omega. \quad (7.59)
 \end{aligned}$$

The first term on the right side of the expression vanishes, when the limits are applied, yielding

$$I = \int_0^\infty \frac{2E_b}{\pi} \omega e^{-\frac{\sigma_I^2 \omega^2}{2}} \arctan \left[\sinh \left(\frac{\pi \sigma_I^2 \omega}{2E_b} \right) \right] d\omega. \quad (7.60)$$

Using the usual assumption $E_b/\sigma_I^2 \gg 1$, and the inequality $\arctan(x) \leq x$ it is possible to upper bound the integrand, obtaining

$$I \leq \frac{2E_b}{\pi} \int_0^\infty \omega e^{-\frac{\sigma_I^2 \omega^2}{2}} \sinh \left(\frac{\pi \sigma_I^2 \omega}{2E_b} \right) d\omega. \quad (7.61)$$

which integrates to (Gradshteyn and Ryzhik 1990)

$$I \leq \frac{\sqrt{2\pi}}{2\sigma_I^2} e^{-\frac{\pi^2 \sigma_I^2}{8E_b^2}}. \quad (7.62)$$

Substituting the above equation into Equation 7.51 gives

$$C_i \geq 1 - \sqrt{2\pi}^{3/2} \log e \frac{\sigma_I}{E_b} e^{-\frac{E_b^2}{2\sigma_I^2} + \frac{\pi^2 \sigma_I^2}{8E_b^2}}. \quad (7.63)$$

The final version of the bound is found introducing the expression for the multiple access interference plus noise

$$\sigma_I^2 = \frac{(M-1)E_b^2}{3N} + \frac{N_0 E_b}{2} \quad (7.64)$$

into Equation 7.63. This yields

$$C_i \geq 1 - \sqrt{2\pi}^{3/2} \log e \sqrt{\frac{M-1}{3N} + \frac{N_0}{2E_b}} \times \exp \left[-\frac{3N}{2(M-1) + 3N \frac{N_0}{E_b}} + \frac{\pi^2}{8} \left(\frac{M-1}{3N} + \frac{N_0}{2E_b} \right) \right] \quad (7.65)$$

and the total capacity, or sum capacity, is lower bounded by

$$\mathbf{C} \geq M - \sqrt{2\pi}^{3/2} \log e M \sqrt{\frac{M-1}{3N} + \frac{N_0}{2E_b}} \times \exp \left[-\frac{3N}{2(M-1) + 3N \frac{N_0}{E_b}} + \frac{\pi^2}{8} \left(\frac{M-1}{3N} + \frac{N_0}{2E_b} \right) \right] \quad (7.66)$$

which is depicted, as a function of the SNR, in Figure 7.4 and referred to as *Bound 3*, for $M = 20$ and $N = 100$.

Example: two cases of interest are obtained by making $E_b/N_0 \rightarrow \infty$

$$\mathbf{C} \geq M - \sqrt{2\pi}^{3/2} \log e M \sqrt{\frac{M-1}{3N}} \exp \left[-\frac{3N}{2(M-1)} + \frac{\pi^2}{24} \frac{M-1}{N} \right] \quad (7.67)$$

and $N \rightarrow \infty$

$$\mathbf{C} \geq M - \sqrt{2\pi}^{3/2} \log e M \sqrt{\frac{N_0}{2E_b}} \exp \left[-\frac{E_b}{N_0} + \frac{\pi^2}{16} \frac{N_0}{E_b} \right] \quad (7.68)$$

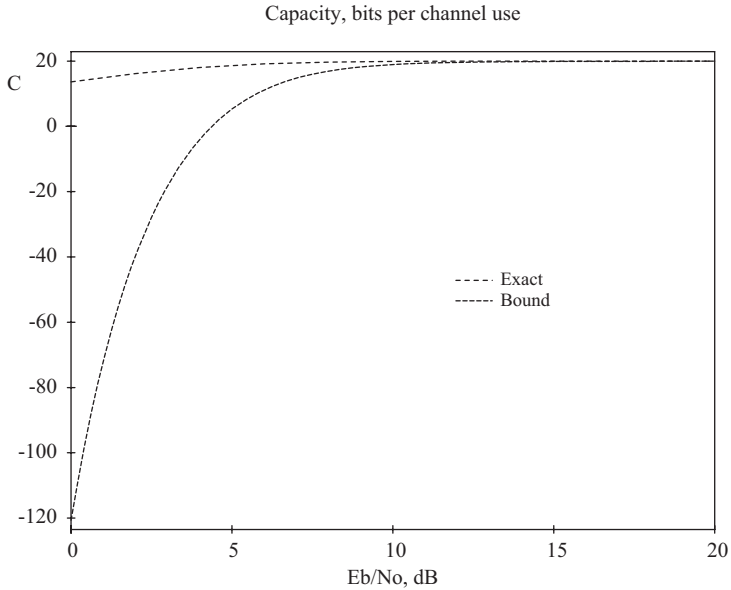


Figure 7.4. Bound 3 on the capacity for the channel, as a function of the SNR (E_b/N_0).

The bound for the sum capacity can be expressed in terms of the probability of error, using Equation 7.82

$$C \geq M - 2\pi^{3/2} \log e P_e M \sqrt{\frac{1}{\ln(1/2P_e)}} \exp\left(\frac{\pi^2}{8} \frac{1}{2 \ln(1/P_e)}\right) \quad (7.69)$$

7.4 A TIGHT BOUND ON THE CAPACITY OF A CDMA SYSTEM

In this section a lower bound on the sum capacity, for a CDMA system, is obtained. The bound is shown to be very tight, although not an exponential one, and therefore amenable for computation in place of the actual formula for the system capacity, which involves a time consuming numerical integration. Some results are reported in this section, in terms of the processing gain, number of users and SNR. An interesting threshold is discovered, relating the sequence length, or processing gain, and the number of users in the channel. Conditions for attaining the capacity are established and the capacity is plotted as a function of the probability of error.

The following derivation presents a lower bound for the channel capacity, considering a large SNR, or a large ratio of processing gain to number of users. This is likely to be the case in real systems, because both assumptions decrease the probability of error. In order to set the stage for the evaluation of the lower bound, it is interesting to present a slightly different formulation from that one presented in the past section. One begins with Equation 7.47, repeated here for reference

$$C_i = 1 - \pi \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \int_{-\infty}^{\infty} \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\cosh\left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right) \left(\frac{E_b}{\sigma_I^2} + j\omega\right)} d\omega. \quad (7.70)$$

The integral inside Equation 7.50, in last section, is not amenable for an analytic solution. One form of attacking the problem is to find a bound for the capacity, by a convenient substitution implied by the inequality $\cosh(x) \geq 1 + \frac{x^2}{2}$. Then,

$$C_i \geq 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \frac{E_b}{\sigma_I^2} \int_0^{\infty} \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\left[1 + \frac{1}{2} \left(\frac{\pi \sigma_I^2 \omega}{2E_b}\right)^2\right] \left(\frac{E_b}{\sigma_I^4} + \omega^2\right)} d\omega. \quad (7.71)$$

Rearranging the equation, in order to come out with a standard definite integral, one obtains

$$C_i \geq 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \frac{8E_b^3}{\pi^2 \sigma_I^6} \int_0^{\infty} \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\left(\frac{8E_b^2}{\pi^2 \sigma_I^4 \omega} + \omega^2\right) \left(\frac{E_b}{\sigma_I^4} + \omega^2\right)} d\omega \quad (7.72)$$

which can be separated by a partial fraction expansion to

$$C_i \geq 1 - 2\pi \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \frac{8E_b}{(\pi^2 - 8)\sigma_I^2} \times \left[\int_0^{\infty} \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\left(\frac{8E_b^2}{\pi^2 \sigma_I^4 \omega} + \omega^2\right)} d\omega - \int_0^{\infty} \frac{e^{-\frac{\sigma_I^2 \omega^2}{2}}}{\left(\frac{E_b}{\sigma_I^4} + \omega^2\right)} d\omega \right].$$

The above integrals can then be evaluated by noting that (Gradshteyn and Ryzhik 1990)

$$\int_0^{\infty} \frac{e^{-\mu^2 x^2}}{x^2 + \beta^2} dx = [1 - \operatorname{erfc}(\beta\mu)] \frac{\pi}{2\beta} e^{\beta^2 \mu^2} \quad (7.73)$$

for $\operatorname{Re}\beta > 0$ and $|\arg\mu| < \pi/4$.

Therefore,

$$\begin{aligned}
 C_i \geq & 1 - \log e e^{-\frac{E_b^2}{2\sigma_I^2}} \frac{2\pi}{\pi^2 - 8} \left[\left[1 - \operatorname{erf} \left(\frac{2E_b}{\pi\sigma_I} \right) \right] \right. \\
 & \left. \times \frac{\pi}{\sqrt{2}} e^{\frac{4E_b^2}{\pi^2\sigma_I^2}} - \left[1 - \operatorname{erf} \left(\frac{E_b}{\sqrt{2}\sigma_I} \right) \right] 2e^{\frac{E_b^2}{2\sigma_I^2}} \right] \quad (7.74)
 \end{aligned}$$

in which $\operatorname{erf}(x)$ is defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \quad (7.75)$$

This equation can be further simplified to

$$\begin{aligned}
 C_i \geq & 1 - \log e \frac{2\pi}{\pi^2 - 8} \left[\frac{\pi}{\sqrt{2}} \left[1 - \operatorname{erf} \left(\frac{2E_b}{\pi\sigma_I} \right) \right] e^{\frac{8-\pi^2}{2\pi^2} \frac{E_b^2}{\sigma_I^2}} \right. \\
 & \left. - 2 \left[1 - \operatorname{erf} \left(\frac{E_b}{\sqrt{2}\sigma_I} \right) \right] \right]. \quad (7.76)
 \end{aligned}$$

Substituting now the expression for the total noise variance into Equation 7.76 leads to the final expression for the lower bound on the sum capacity

$$\begin{aligned}
 C \geq & M - \log e \frac{2\pi M}{\pi^2 - 8} \quad (7.77) \\
 & \times \left[\frac{\pi}{\sqrt{2}} \left[1 - \operatorname{erf} \left(\frac{2}{\pi} \sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0 + 3N}} \right) \right] \right. \\
 & \times e^{\frac{8-\pi^2}{2\pi^2} \frac{6NE_b/N_0}{2(M-1)E_b/N_0 + 3N}} \\
 & \left. - 2 \left[1 - \operatorname{erf} \left(\frac{1}{\sqrt{2}} \sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0 + 3N}} \right) \right] \right].
 \end{aligned}$$

Equation 7.77 is plotted in Figure 7.5, along with the exact solution (upper curve), as a function of the SNR ($SNR = E_b/N_0$, dB) for $M = 20$ and sequence length $N = 100$. It is remarkable to notice that the difference does not exceed 5 percent from -10 dB to 20 dB and actually converges to zero as the SNR increases. The lower bound will thus be used as an approximation for the actual capacity in the following, for values of SNR of -10 dB or more.

One of the reasons for having a bound on the capacity, of the form expressed by Equation 7.77, is the possibility of interpreting correctly the role of the many parameters on the performance of the system, because

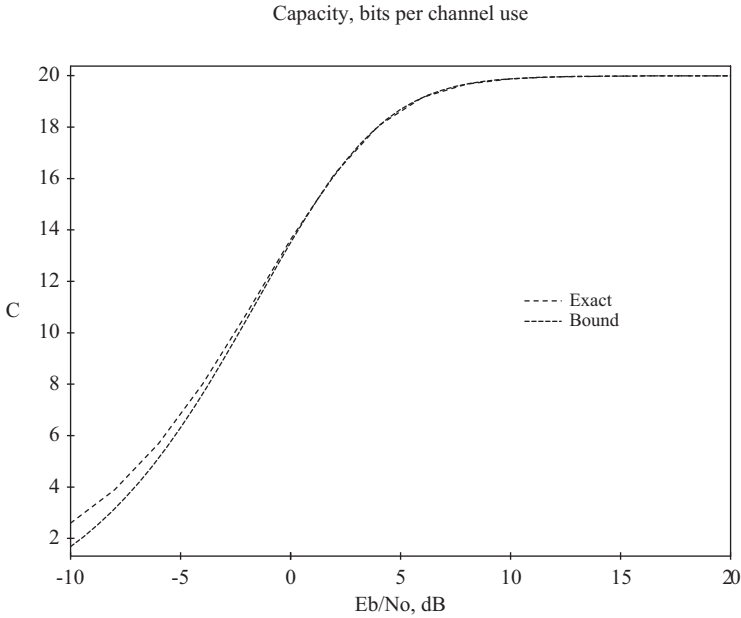


Figure 7.5. Capacity for the channel, compared with the lower bound, as a function of the SNR (E_b/N_0), for $M=20$ and sequence length $N=100$.

those parameters, and their relationship, appear explicitly in the formula for the bound. As will be seen in what follows, some parameters have the property of influencing and even disrupting the system.

The second reason involves the possibility of combining different formulas, to attain a certain objective, as in the case of expressing the capacity as a direct function of the probability of error and vice-versa.

The third reason is related to computational resources. Evaluation of the exact solution is time and memory consuming and is justified only when no feasible approximation is available. It is instructive to recall that even in the case when the exact solution is being computed, the computer system will always resort to some form of approximation or truncation of the formulas.

Equation 7.77 is also plotted in Figure 7.6, as a function of the SNR, for $M=20$, having the sequence length as a parameter. It is interesting to notice that for a high processing gain there is no further significant gain in capacity if $N \geq 10M$.

In Figure 7.7, the capacity is displayed as a function of the sequence length, for a SNR of 10 dB, parametrized by the number of users. It must

Capacity, bits per channel use

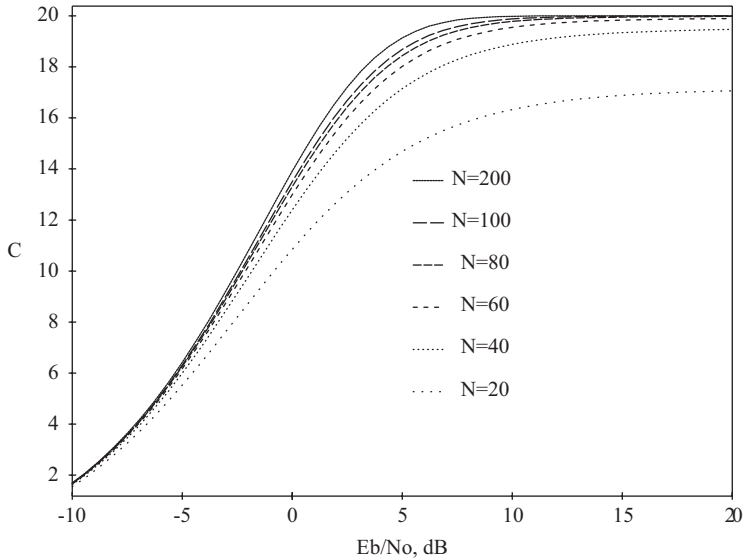


Figure 7.6. Approximate capacity for the channel, as a function of the signal-to-noise ratio (E_b/N_0), for $M = 20$, having N as a parameter.

be noticed that the capacity of the system is achieved for a sequence length of $N = 100$ and $N = 200$, for $M = 10$ and $M = 20$, respectively.

In Figure 7.8, the capacity is shown in terms of the number of users, for a given SNR (10 dB), and two values of the processing gain $N = 100$ and $N = 200$ (upper curve). This result shows that the capacity for the binary case, in bits per channel use, is numerically equal to the number of users, as long as this number is small compared to the processing gain. Any increase in the number of users above $M = 0.2N$ produces a transmission rate that falls short of the system capacity. This is in agreement with previous results (Turin 1984b; Pickholtz, Schilling, and Milstein 1982), contradicting (Hui 1984b), which suggested that using long PN sequences would decrease the sum capacity. Figure 7.9 illustrates the previous point, by showing a decrease in the capacity with an increase in the number of users, for a fixed value of N . It is important to mention that if both M and N increase to infinity, the limiting capacity will be given by $\log e/2$, as predicted in Chapter 2 (Sousa 1989).

The importance of finding a lower bound on the capacity is due to the nature of the noncooperativeness of CDMA. In this case, the lower bound

Capacity, bits per channel use

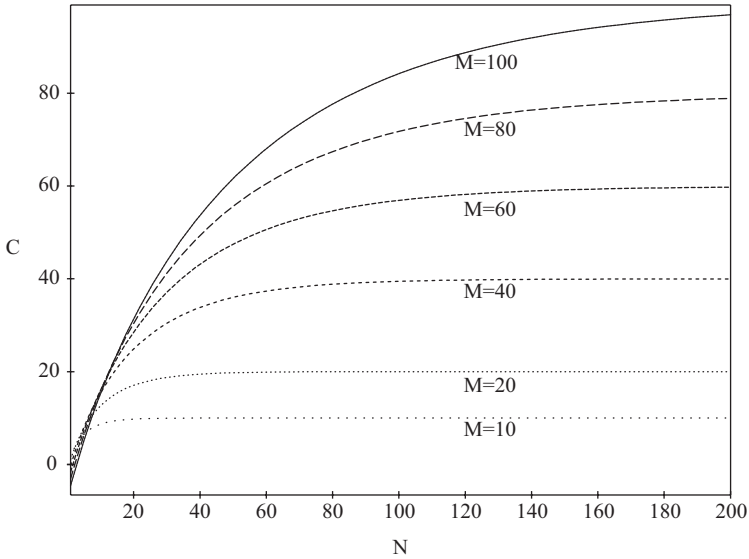


Figure 7.7. Capacity for the channel, using the approximate formula, as a function of the sequence length (N), for different values of M .

represents the worst possible scenario for the system, which is incidentally close to reality.

Example: from Equation 7.77 it is also possible to draw some limiting results, which are supposed to hold for the exact formula for the capacity. If the processing gain increases to infinity

$$\lim_{N \rightarrow \infty} \mathbf{C} = M. \quad (7.78)$$

In case the SNR goes to infinity

$$\lim_{\frac{E_b}{N_0} \rightarrow \infty} \mathbf{C} = M. \quad (7.79)$$

Finally, if the number of users increases without bound the sum capacity displays a peculiar behavior. If $M \geq 36N$

$$\lim_{M \rightarrow \infty} \mathbf{C} = 0 \quad (7.80)$$

otherwise

$$\lim_{M \rightarrow \infty} \mathbf{C} = \infty. \quad (7.81)$$

Capacity, bits per channel use

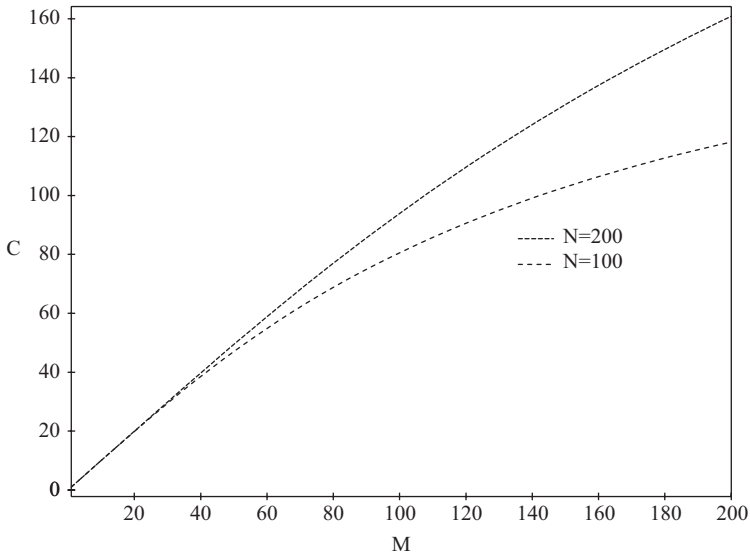


Figure 7.8. Capacity for the channel, as a function of the number of users (M), using the approximation for the capacity.

The relation between the number of users and the sequence length, $\chi = M/N$, has been explored in a recent paper (Sousa 1989). It is remarkable that $\chi = 36$ sets a threshold for the capacity, as the number of users increases to infinity. This channel behavior demands further study. It is rather interesting to observe that the capacity can be achieved, for $M \rightarrow \infty$, as long as the sequence length increases accordingly.

It is useful to express the sum capacity in terms of the error probability. A complete analysis of a DSSS with DPSK modulation, in terms of probability of error, can be found in Kavehrad and Ramamurthi (1987) or in Geraniotis and Pursley (1986). In the following, a simplified analysis is pursued. This can be done, by rewriting the formula for the probability of error for the DPSK system, considering now the effect of multiple access interference (Turin 1984b).

$$P_e = \frac{1}{2} \exp \left[-\frac{1}{2} \left(\frac{M-1}{3N} + \frac{N_0}{2E_b} \right)^{-1} \right] \quad (7.82)$$

solving the expression for E_b/N_0

$$\ln 2P_e = -\frac{1}{2} \left(\frac{M-1}{3N} + \frac{N_0}{2E_b} \right)^{-1} \quad (7.83)$$

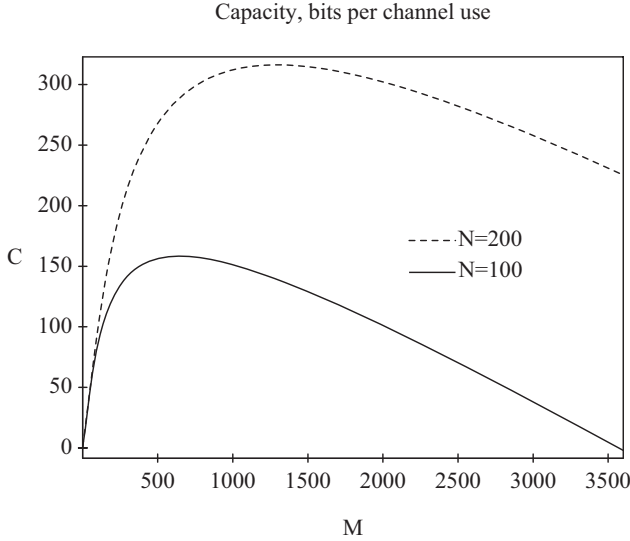


Figure 7.9. Capacity for the channel, as a function of the number of users (M), including the case $M \gg N$.

and finally

$$\frac{E_b}{N_0} = \left[\left(\ln \frac{1}{2P_e} \right)^{-1} - \frac{2(M-1)}{3N} \right]^{-1}. \tag{7.84}$$

Substituting this last result into the formula for the sum capacity one finds the relation between the capacity, in bits per channel use, and the probability of error for high values of the signal to noise error. This relation is expressed in the following formula

$$\begin{aligned} C \geq M - \log e \frac{2\pi M}{\pi^2 - 8} \left[\frac{\pi}{\sqrt{2}} \left[1 - \operatorname{erf} \left(\frac{2}{\pi} \sqrt{2 \ln \frac{1}{2P_e}} \right) \right] \right. \\ \left. \times e^{\frac{8-\pi^2}{\pi^2} \ln \frac{1}{2P_e}} - 2 \left[1 - \operatorname{erf} \left(\sqrt{\ln \frac{1}{2P_e}} \right) \right] \right], \end{aligned} \tag{7.85}$$

which is depicted in Figure 7.10, for $M = 100$ users.

From this graph, it is clear that an error probability $P_e \geq 10^{-2}$ will produce a decrease in available capacity of 3 percent or more. On the other hand, a probability of error $P_e \leq 10^{-3}$ will leave the capacity almost intact. Of course, the probability of error can be decreased only at the expense of an increase in the SNR of the channel, or through the use of long and better code sequences. This represents a lower bound on the sum capacity,

Capacity, bits per channel use

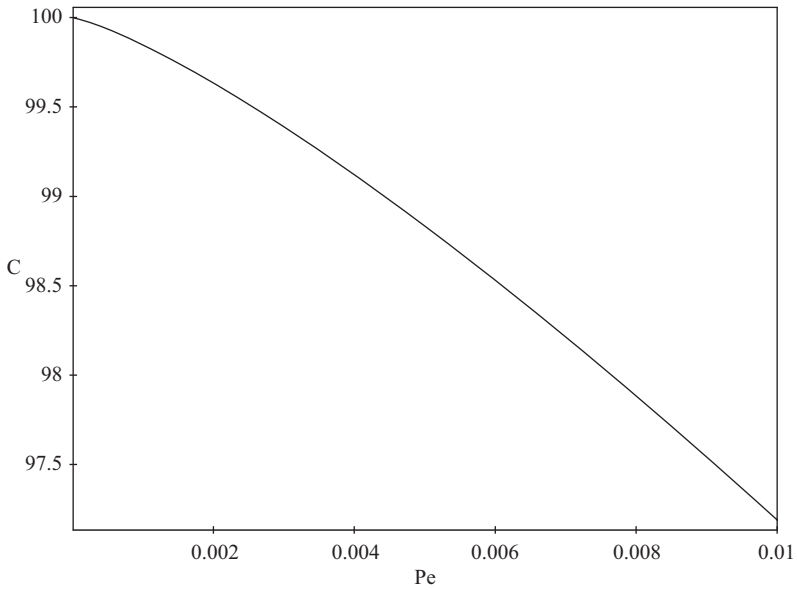


Figure 7.10. Capacity for the channel, as a function of the probability of error (P_e).

for a given probability of error, which also implies the worst case for the system operation.

In order to conclude this section, it is useful to express Equation 7.77 in a more simplified form, as an exponential bound, to sum up the results from the previous section. It is not difficult to recognize in Equation 7.77 the complementary error function, $\text{erfc}(x) = 1 - \text{erf}(x)$, which can be approximated by (Schwartz 1970)

$$\text{erfc}(x) \approx \frac{e^{-x^2}}{x\sqrt{\pi}}, \quad x \gg 1. \quad (7.86)$$

Therefore, after the appropriate substitution, one finds the exponential approximation below

$$C_i \approx 1 - \frac{\sqrt{2\pi}}{2} \log e \frac{\sigma_I}{E_b} e^{-\frac{E_b^2}{2\sigma_I^2}}. \quad (7.87)$$

This formula will be referred to as *Bound 4* in the following. It is plotted in Figure 7.11 for $M = 20$ users and gain $N = 100$.

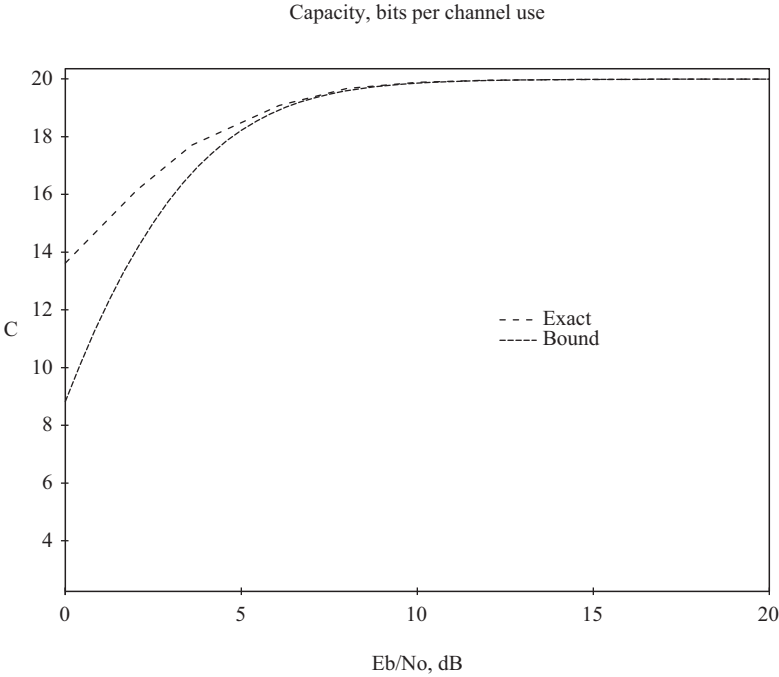


Figure 7.11. *Bound 4* on the capacity for the channel, as a function of the SNR (E_b/N_0).

A summary of the exponential bounds found in this chapter is shown in Figure 7.12. It is clear that *Bound 4*, although a simpler form of Equation 7.77, is vastly superior to all the remaining bounds.

The chapter discussed the performance analysis of a direct-sequence CDMA system, in terms of the sum capacity. Upper and lower bounds on the sum capacity, for a CDMA system, were obtained in this chapter. Some of the bounds were shown to be very tight and therefore amenable for computation in place of the actual formula for the system capacity, which involves a time-consuming numerical integration.

Some results were reported in terms of the processing gain, number of users, and SNR. An interesting threshold was discovered, relating the sequence length and the number of users in the channel. Conditions for attaining the capacity were established and the capacity was plotted as a function of the probability of error, for a DPSK modulation scheme. One of the reasons for the preference for DPSK instead of a coherent system is that the wireless mobile channel introduces random phase shift (Yue 1983).

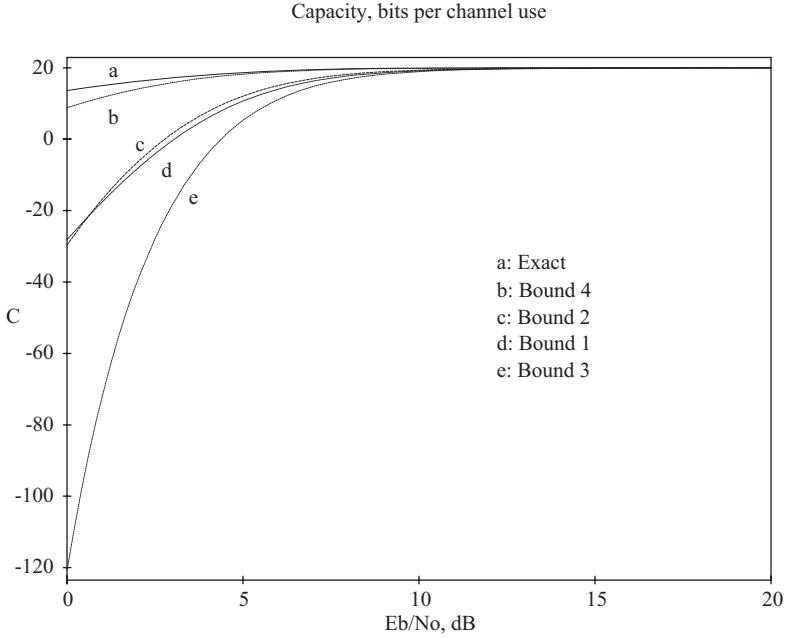


Figure 7.12. Bounds on the capacity, as a function of the SNR (E_b/N_0).

It is instructive to compare the new bound in Equation 7.77 with existing bounds and approximations on the sum capacity and with the exact solution. This is done in Figure 7.13, in which the upper curve (solid line) is the exact solution and the previous bounds are given below, for $M = 20$ users and $N = 100$

$$C \approx -M \log \left(\frac{1}{2} + \frac{1}{2} e^{-E_b/\sigma_I} \right), \quad (7.88)$$

$$C \geq \frac{M}{2} \log \left(1 + \frac{E_b}{\sigma_I} \right) - \frac{M}{2} \log \left(1 + \frac{E_b}{5\sigma_I} \right) - \frac{M}{2} \log \left(\frac{5\pi e}{24} \right) \quad (7.89)$$

and

$$C \geq M - M e^{-3E_b/4\sigma_I} \left(\frac{3\pi E_b}{4\sigma_I} \right)^{-1/2} - M h \left(e^{-3E_b/4\sigma_I} \left(\frac{3\pi E_b}{4\sigma_I} \right)^{-1/2} \right), \quad (7.90)$$

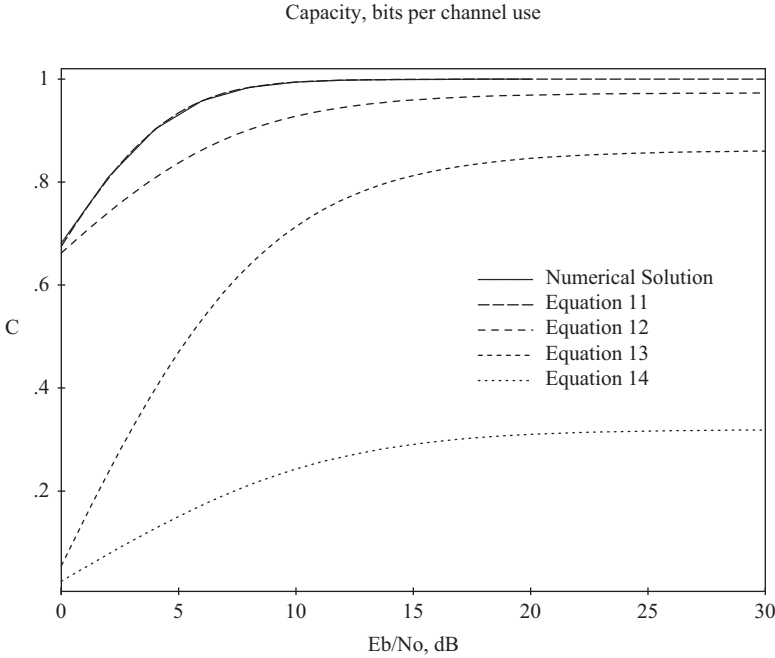


Figure 7.13. Comparison between the new and existing bounds, as a function of the SNR (E_b/N_0), for $M = 20$ and $N = 100$.

in which σ_I is given by Equation 7.24. Equation 7.13 is the corrected version of the bound given by Shamai, Ozarow, and Wyner (1991) and $h(x) = -x \log x - (1 - x) \log (1 - x)$ is the binary entropy function (Ozarow and Wyner 1990).

Substitution of the appropriate parameters and simplification of the resulting expressions yields

$$C \approx -M \log \left(\frac{1}{2} + \frac{1}{2} e^{-\sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0 + 3N}}} \right), \quad (7.91)$$

$$\begin{aligned}
 C \geq & \frac{M}{2} \log \left[1 + \sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0 + 3N}} \right] \\
 & - \frac{M}{2} \log \left[1 + \frac{1}{5} \sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0 + 3N}} \right] \\
 & - \frac{M}{2} \log \left(\frac{5\pi e}{24} \right)
 \end{aligned} \quad (7.92)$$

and

$$\begin{aligned}
 C \geq & M - Me^{-\frac{3}{4}\sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0+3N}}} \left[\frac{3\pi}{4} \sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0+3N}} \right]^{-1/2} \\
 & - Mh \left[e^{-\frac{3}{4}\sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0+3N}}} \right. \\
 & \left. \times \left(\frac{3\pi}{4} \sqrt{\frac{6NE_b/N_0}{2(M-1)E_b/N_0+3N}} \right)^{-1/2} \right] \quad (7.93)
 \end{aligned}$$

Equation 7.77 compares favorably with respect to the former bounds and approximations represented by Equations 7.91, 7.92, and 7.93. In fact, Equation 7.77 fits the exact solution very well, even for SNRs as low as 0 dB.

The objective of this chapter was the evaluation of the sum capacity of a CDMA system with a fixed number of users. A more realistic analysis of a CDMA system should include a variable allocation of users (Alencar and Blake 1993a).

New approaches, for the analysis of CDMA systems, were introduced in this chapter. One of the important new results is related to the evaluation of the sum capacity for the CDMA channel. Also, new limits were found with explicit formulas for the upper and lower bounds on the capacity of the channel in terms of the probability of error. A tight bound was found, that allows a computer efficient evaluation of the capacity.

CHAPTER 8

THEORETICAL CRYPTOGRAPHY

8.1 INTRODUCTION

Cryptography is the basis for network security. It comes from the Greek word *kryptos*, for tomb, hidden, or secret, combined with *graphein*, or writing. A free translation is hidden writing, which defines the proposal of cyphering a message. Cryptology, combines the same previous prefix with *logia*, meaning study, to give the translation of study of secret coding.

Cryptography is the practice and study of techniques for secure communication in the presence of third parties. It is also a collection of techniques for construction and analysis of protocols to overcome the influence of jammers and which are related to various aspects of information security, such as data confidentiality, data integrity, authentication, and nonrepudiation.

Along the centuries, governments used cryptography to cypher messages, supposed to remain in secrecy, while the spies tried to decipher them. During the wars, cryptography becomes more important, considering the nature of the operations.

Bletchley Park was the secret information and counter-information center in Great Britain during the Second World War. It was so secret that it remained as a legend after the end of the war, while England continued to break the codes of other countries, either enemies or friends (Haykin 1999).

The first computer, the Colossus, was built in Bletchley Park, developed by Alan Mathison Turing (1912–1954), one of the greatest computer geniuses of the World. Destroyed after the end of the war, the Colossus was rebuilt from the original drawings.

The German Enigma code was broken there with the help of Turing, who benefited from the information passed by a Hungarian mathematician

and a German officer, who provided information to French spies in exchange for money. The French did not know how to decipher the code, and passed on the information to the British.

8.2 CRYPTOGRAPHIC ASPECTS OF COMPUTER NETWORKS

The Internet has revolutionized the ways in which companies do business, since the Internet Protocol (IP) is undeniably efficient, inexpensive, and flexible. This section presents the main cryptographic aspects of modern TCP and IP computer networks, which include digital signature technology based on asymmetrical cryptographic algorithms, data confidentiality by applying symmetrical cryptographic systems, and system Public Key Infrastructure (PKI) (Tanenbaum 2003).

The possible vulnerabilities of TCP and IP computer networks, and possible techniques to eliminate them, are considered. It seems that only a general and multilayered security infrastructure could cope with possible attacks to the computer network systems.

8.2.1 *POTENTIAL VULNERABILITIES OF COMPUTER NETWORKS*

The IP is efficient, inexpensive, and flexible. However, the existing methods used to route IP packets leave them vulnerable to security risks, such as spoofing, sniffing, and session hijacking, and provide no form of nonrepudiation for contractual or monetary transactions.

Organizations need to secure communications between remote offices, business partners, customers, and travelling and telecommuting employees, besides securing the internal environment. The transmission of messages over the Internet or intranet poses a risk, given the lack of protection at the existing Internet backbone.

Control and management of security and access between the entities in a company's business environment is important. Without security, both public and private networks are susceptible to unauthorized monitoring and access. Internal attacks might be a result of minimal or nonexistent intranet security.

Risks from outside the private network originate from connections to the Internet and extranets. Password-based user access controls alone do not protect data transmitted across a network. Without security measures and controls in place, the data might be subjected to attack. Some attacks

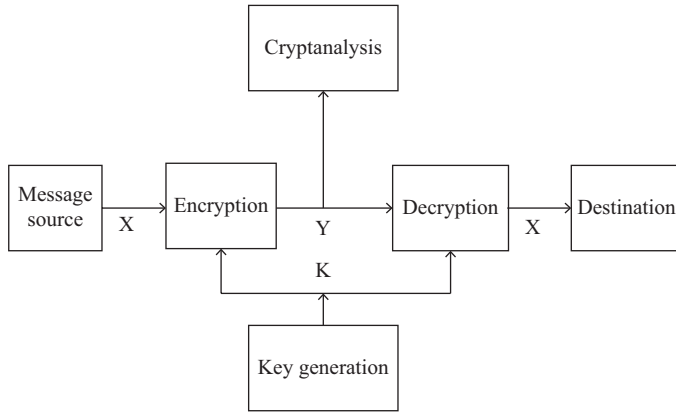


Figure 8.1. General model for a cryptosystem.

are passive, and the information is only monitored. Other attacks are active, and the information is altered to corrupt or destroy the data or the network itself.

8.3 PRINCIPLES OF CRYPTOGRAPHY

The general model for an encryption system is shown in Figure 8.1. The original message, also called plaintext and denoted by the letter X , is converted to an apparently random sequence called ciphertext and denoted by the letter Y . The encryption process consists of an encipherment algorithm that uses a key K , which is independent of the plaintext. The algorithm produces a different output depending on the key used (Stallings 1999).

It is important to consider that a cryptanalyst, or hacker, could observe the encrypted message and try to decode it. The algorithm must, therefore, be powerful enough to resist the attack.

Encryption can be seen as a transformation, parametrized by K ,

$$Y = E(X, K), \quad (8.1)$$

which converts the original text X to the ciphertext Y , to be transmitted.

At the reception side, the inverse transformation is performed on the ciphertext to produce the plaintext

$$X = D(Y, K). \quad (8.2)$$

Of course, the destination must possess the key to be able to invert the transformation. Therefore, the transmission of the key is an important part of the cryptobusiness.

The security of a cryptosystem depends on certain assumptions:

- The encryption algorithm must be robust enough that is difficult to decipher a message based only on the ciphertext.
- The key must be kept secret, and a secure channel should be provided for transmission.
- There is no need to keep the algorithm secret.

The application of cryptographic techniques depends on the knowledge of the security of the employed system. A cryptographic system can be classified as:

1. How to transform the plaintext to the ciphertext, as the encryption algorithms are based on two principles (Stallings 1999):
 - a. Substitution, in which the individual symbols of the original text (bit, byte, letter, words) are mapped into other elements.
 - b. Transposition, in which the elements of the original text are rearranged.
2. How the original text is processed, which involves a block cipher, that processes a block of plaintext at a time, or a convolutional cipher, also called stream cipher that processes the input symbols as they arrive at the encryption subsystem.
3. How the key is generated, which can lead to the production of a single, or symmetric, key to be used by both the transmitter and receiver, or use different keys for the sender and receiver, also called asymmetric or public key encryption.

8.4 INFORMATION THEORETICAL ASPECTS OF CRYPTOGRAPHY

It is possible to use the terminology of information theory to describe a secure encryption system. The measure of information, or entropy, of a plaintext, considered as a subset of symbols selected from the set of all possible messages, is given by (van der Lubbe 1997)

$$H(X) = - \sum_{l=1}^L p(x_l) \log p(x_l), \quad (8.3)$$

in which $p(x_l)$, $l = 1, 2, \dots, L$ represent the uncertainty of occurrence of the possible plaintexts, considering that the texts are mutually independent.

The entropy of the set of keys is given by

$$H(K) = - \sum_{m=1}^M p(k_m) \log p(k_m), \quad (8.4)$$

in which $k_m, m = 1, 2, \dots, M$ are the possible keys.

By the same token, it is possible to introduce a measure of information for the ciphertext

$$H(Y) = - \sum_{n=1}^N p(y_n) \log p(y_n), \quad (8.5)$$

in which $p(y_n), n = 1, 2, \dots, N$ represent the uncertainty of occurrence of the possible ciphertexts, considering that the texts are mutually independent. Usually, regarding the bijection from the plaintext into the ciphertext set, one considers $N = L$.

The conditional entropy, or key equivocation, is a measure of the uncertainty or information with respect to the key, when the ciphertext is available is defined as

$$H(K|Y) = \sum_{m=1}^M \sum_{n=1}^N p(k_m, y_n) \log p(k_m|y_n). \quad (8.6)$$

The uncertainty with respect to the plaintext, or conditional entropy when the ciphertext is available, also known as message equivocation, is defined as

$$H(X|Y) = \sum_{l=1}^L \sum_{n=1}^N p(x_l, y_n) \log p(x_l|y_n). \quad (8.7)$$

The conditional entropy, or uncertainty with respect to the key, for a given plaintext and corresponding ciphertext, is defined as

$$H(K|X, Y) = \sum_{m=1}^M \sum_{l=1}^L \sum_{n=1}^N p(k_m, x_l, y_n) \log p(k_m|x_l, y_n), \quad (8.8)$$

which is also known as key appearance equivocation.

Finally, the conditional entropy, or uncertainty with respect to the plaintext, when the key and the ciphertext are known, is defined as

$$H(X|Y, K) = \sum_{l=1}^L \sum_{n=1}^N \sum_{m=1}^M p(x_m, y_l, k_n) \log p(x_m|y_l, k_n). \quad (8.9)$$

Since there is a bijection between the plaintext and ciphertext sets, it is always true that

$$H(X|Y, K) = 0. \quad (8.10)$$

As expected, when the ciphertext and the key are available it is possible to retrieve the plaintext correctly, because the uncertainty with respect to X is null. There is no loss of information whatsoever at the receiver, and the full original message is recovered.

8.4.1 RELATIONS BETWEEN THE ENTROPIES

A cipher system user requires that the key appearance equivocation, or $H(K|X, Y)$, be as high as possible, because if a cryptanalyst manages to obtain both the plaintext and the ciphertext, then the uncertainty with respect to the key must be as large as possible.

This can be seen by the following argument. From the previous results in information theory, the joint measure of information in the plaintext, ciphertext, and key is

$$H(X, Y, K) = H(X|Y, K) + H(Y, K) = H(K|X, Y) + H(X, Y). \quad (8.11)$$

Also, consider the following relations

$$H(Y, K) = H(K|Y) + H(Y), \quad (8.12)$$

and

$$H(X, Y) = H(X|Y) + H(Y). \quad (8.13)$$

Combining the equations, one obtains

$$H(X|Y, K) + H(K|Y) = H(K|X, Y) + H(X|Y). \quad (8.14)$$

But, as discussed, $H(X|Y, K) = 0$, which results in

$$H(K|Y) = H(K|X, Y) + H(X|Y), \quad (8.15)$$

which leads to

$$H(K|X, Y) = H(K|Y) - H(X|Y). \quad (8.16)$$

Therefore, to obtain a large key appearance equivocation the message equivocation $H(X|Y)$ must be small. However, a small message equivocation implies that the uncertainty with respect to the plaintext, when the ciphertext is known, is small. But this is what must be avoided by the cryptosystem.

The uncertainty with respect to the key must be large to decrease the uncertainty with respect to the plaintext, and an increase in the uncertainty with respect to the plaintext decreases the uncertainty with respect to the key (van der Lubbe 1997).

8.5 MUTUAL INFORMATION FOR CRYPTOSYSTEMS

The information theoretical approach provides additional results. The mutual information of the plaintext and the ciphertext is defined as

$$I(X; Y) = H(X) - H(X|Y) = H(Y) - H(Y|X). \quad (8.17)$$

As expected, the objective of the cryptodesigner is to minimize the mutual information $I(X; Y)$. If the ciphertext provides no information about the original message, then

$$H(X|Y) = H(X),$$

and the mutual information between the plaintext and the encoded message is zero, or $I(X; Y) = 0$. This is referred to as the absolutely secure cryptosystem.

It is possible to obtain a lower limit for the mutual information between the plaintext and the ciphertext. First, consider Equation 8.16, and since $H(K|X, Y) \geq 0$, it follows that

$$H(K|Y) \geq H(X|Y). \quad (8.18)$$

The use of one of the properties of entropy gives

$$H(K) \geq H(K|Y), \quad (8.19)$$

and therefore,

$$H(K) \geq H(X|Y). \quad (8.20)$$

Substituting this last inequality into Equation 8.17 gives

$$I(X; Y) \geq H(X) - H(K). \quad (8.21)$$

Inequality 8.21 implies that a decrease in the uncertainty of a set of keys improves, on average, the independence between the plaintext and the ciphertext.

It is implicit in the derivation that absolute security of a cryptosystem can only be achieved for

$$H(K) \geq H(X). \quad (8.22)$$

For uniformly distributed keys

$$H(K) = - \sum_{m=1}^M p(k_m) \log p(k_m) = \log(M), \quad (8.23)$$

and for uniformly distributed plaintexts

$$H(X) = - \sum_{l=1}^L p(v_l) \log p(v_l) = \log(L). \quad (8.24)$$

Therefore,

$$H(K) = \log(M) \geq H(X) = \log(L), \text{ or } M \geq L, \quad (8.25)$$

because of the monotonicity of the logarithm function.

The last condition implies that the key must be at least the same length as the message to be transmitted. However, the inequality can be avoided with the introduction of security events. When a security event happens the system can be considered absolutely secure. If the event does not occur the cryptosystem may not be fully secure (Maurer 1989; Massey 1990).

Even when there is a small chance of a security breach, an absolutely secure cipher is still feasible, for which $H(K) < H(X)$. A security event could be the transmission of a certain key K , followed by the plaintext, provided that the hacker does not type the secret key at the beginning. Without typing K , the sequence is obscure to the intruder, and the system is secure. If the hacker happens to type K , the ciphertext can be deciphered, and the system has become unreliable.

Therefore, the introduction of the concept of security event gives a new way of regarding the system security based on information theorems. Information security can be assured, even if the length of the key is smaller than the original sequence, given that a security event is defined for the system, and provided that the event actually occurs (van der Lubbe 1997).

APPENDIX A

PROBABILITY THEORY

A.1 SET THEORY AND MEASURE

Georg Cantor (1845–1918) developed the modern theory of sets at the end of the 19th century, and established the mathematical and logical basis for the theory and demonstrated several important results. The concept of set cardinality, defined by Cantor, was fundamental to the theory. Cantor was born in Saint Petersburg, but lived most of his academic life in the city of Halle, Germany (Boyer 1974). The basic ideas of universal set, empty set, set partition, discrete systems, continuous systems, and infinity are as old as humanity itself.

Over the years, philosophers and mathematicians had tried to characterize the infinite, with no success. In 1872, J. W. R. Dedekind (1831–1916) indicated the universal property of infinite sets. He stated that a set is called infinite when it is similar to a part of itself, on the contrary the set is finite (Boyer 1974).

Cantor also investigated the properties of infinite sets but, different from Dedekind, he noticed that the infinite sets are not the same. This led to the concept of cardinal numbers, to establish a hierarchy of infinite sets in accordance with their respective powers. Cantor ideas established the set theory as a complete subject. As a consequence of his published results on transfinite arithmetic, considered advanced for his time, Cantor suffered attacks from mathematicians like Leopold Kronecker (1823–1891), who barred him for a position at the University of Berlin.

Cantor found a position at the ancient and small University of Halle, in the medieval city of Halle in Germany famous for its mines of rock salt, and died there in an institution for mental health treatment, following his attempts to apply his theory to justify religious paradigms in scientific events.

A.1.1 BASIC SET THEORY

The notion of a set, as simple as it can be, is axiomatic, in a sense that as it does not admit a definition that does not resort to the original notion of a set. The mathematical concept of a set is fundamental for all known mathematics, and is used to build important concepts, such as relation, cartesian product and function. It is also the basis for the modern measure theory, developed by Henry Lebesgue (1875–1941).

The set theory is based on a set of fundamental axioms: Axiom of Extension, Axiom of Specification, Peano's Axioms, Axiom of Choice, besides Zorn's Lemma, and Schröder-Bernstein's Theorem (Halmos, 1960).

The objective of this section is to present the theory of sets in an informal manner, just quoting those fundamental axioms, since this theory is used as a basis to establish a probability measure. Some examples of common sets are given in the following.

- The binary set: $\mathbb{B} = \{0, 1\}$;
- The set of natural numbers, including zero: $\mathbb{N} = \{0, 1, 2, 3, \dots\}$;
- The set of odd numbers: $\mathbb{O} = \{1, 3, 5, 7, 9, \dots\}$;
- The set of integer numbers: $\mathbb{Z} = \{\dots, -3, -2, -1, -2, 0, 1, 2, 3, \dots\}$;
- The set of real numbers: $\mathbb{R} = (-\infty, \infty)$.

There are two really important relations in set theory: the belonging relation, denoted as $a \in A$, in which a is an element of the set A , and the inclusion relation, $A \subset B$, which is read " A is a subset of the set B ", or B is a super set of the set A . Sets may also be specified by propositions. For example, the empty set can be written as $\emptyset = \{a \mid a \neq a\}$, that is, the set in which the elements are different from themselves.

A universal set contains all other sets of interest. An example of a universal set is provided by the sample space in probability theory, usually denoted as S or Ω . The empty set is that set which contains no element, usually denoted as \emptyset or $\{\}$. It is implicit that the empty set is contained in any set, that is, $\emptyset \subset A$, for any given set A . However, the empty set is not in general an element of any other set.

A usual way to represent sets is by means of the Venn diagram, as illustrated in Figure A.1.

Sets are said to be *disjoint* if they have no element in common, as illustrated in Figure A.2. Thus, for example, the set of even natural numbers and the set of odd natural numbers are disjoint.

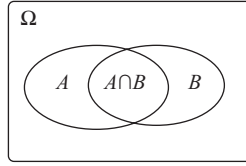


Figure A.1. A Venn diagram that represents two intersecting sets.

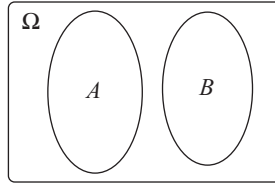


Figure A.2. A Venn diagram representing disjoint sets.

A.1.2 SOME OPERATIONS ON SETS

It is possible to operate on sets to produce new sets, or families of sets. The basic set operations are the complement, the union, the intersection, the subtraction, and the symmetric difference.

- The operation \bar{A} represents the complement of A with respect to the sample space Ω ;
- The union of two sets is composed of elements that belong to A or to B , and is written as $A \cup B$;
- The intersection of two sets is composed of elements that belong to A and to B , and is written as $A \cap B$;
- The subtraction of sets, denoted by $C = A - B$, gives as a result the set in which the elements belong to A and do not belong to B .
Note: If B is completely contained in A then $A - B = A \cap \bar{B}$;
- The symmetric difference is defined as the set of elements that belong to A and to B , but do not belong to $(A \cap B)$. It is written commonly as $A \Delta B = A \cup B - A \cap B$.

The generalization of these concepts to families of sets, as for example $\bigcup_{i=1}^N A_i$ and $\bigcap_{i=1}^N A_i$, is immediate. The following properties are usually employed as axioms in developing the theory of sets (Lipschutz 1968).

- Idempotent
 $A \cup A = A, \quad A \cap A = A$

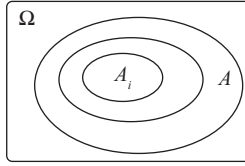


Figure A.3. Increasing sequence of sets.

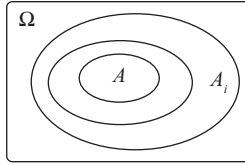


Figure A.4. Decreasing sequence of sets.

- Associative
 $(A \cup B) \cup C = A \cup (B \cup C), \quad (A \cap B) \cap C = A \cap (B \cap C)$
- Commutative
 $A \cup B = B \cup A, \quad A \cap B = B \cap A$
- Distributive
 $A \cup (B \cap C) = (A \cup B) \cap (A \cup C),$
 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- Identity
 $A \cup \emptyset = A, \quad A \cap U = A$
 $A \cup U = U, \quad A \cap \emptyset = \emptyset$
- Complementary
 $A \cup \bar{A} = U, \quad A \cap \bar{A} = \emptyset \quad \overline{\bar{A}} = A$
 $U = \emptyset, \quad \bar{\emptyset} = U$
- de Morgan laws
 $\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad \overline{A \cap B} = \bar{A} \cup \bar{B}$

A.1.3 FAMILIES OF SETS

The concept of family is important to characterize finite or infinite combinations of sets. The increasing sequence of sets, such that $\lim_{i \rightarrow \infty} \cup A_i = A$, is one of the most useful families of sets. This sequence is used in proofs of limits over sets.

The decreasing sequence of sets is defined in a similar manner, as $\lim_{i \rightarrow \infty} \cap A_i = A$, and is also used in proofs of limits over sets.

A.1.4 INDEXING OF SETS

The Cartesian product is useful to express the idea of indexing of sets. Indexing expands the possibilities for the use of sets, and permits to generate new entities, such as vectors and signals.

Example: consider $A_i = \{0, 1\}$. Starting from this set it is possible to construct an indexed sequence of sets by defining its indexing: $\{A_{i \in I}\}$, $I = \{0, \dots, 7\}$. This family of indexed sets A_i constitutes a finite discrete sequence, that is, a vector.

Example: again, let $A_i = \{0, 1\}$, but now use $I = \mathbb{Z}$, the set of positive and negative integers plus zero. It follows that $\{A_{i \in \mathbb{Z}}\}$, which represents an infinite series of 0s and 1s, that is, it represents a binary digital signal. For example, $\dots 0011111000 \dots$.

Example: using the same set $A_i = \{0, 1\}$, but now indexing over the set of real numbers, $\{A_{i \in I}\}$, in which $I = \mathbb{R}$, it is possible to form a signal which is discrete in amplitude but continuous in time.

Example: considering $A = \mathbb{R}$ and $I = \mathbb{R}$, the resulting set represents an analog signal—a signal that is continuous in time and in amplitude.

A.1.5 AN ALGEBRA OF SETS

For the construction of an algebra of sets or, equivalently, for the construction of a field over which operations involving sets make sense, a few properties have to be obeyed.

1. If $A \in \mathcal{F}$ then $\bar{A} \in \mathcal{F}$. \bar{A} is the set containing desired results, or over which one wants to operate;
2. If $A \in \mathcal{F}$ and $B \in \mathcal{F}$ then $A \cup B \in \mathcal{F}$.

The properties guarantee the closure of the algebra with respect to finite operations over sets. It is noticed that the universal set Ω always belongs to the algebra, that is, $\Omega \in \mathcal{F}$, because $\Omega = A \cup \bar{A}$. The empty set also belongs to the algebra, that is, $\emptyset \in \mathcal{F}$, since $\emptyset = \bar{\Omega}$, follows by property 1.

Example: the family $\{\emptyset, \Omega\}$ complies with the above properties and therefore represents an algebra. In this case, $\emptyset = \{\}$ and $\bar{\emptyset} = \Omega$. The union is also represented, as can be easily checked.

Example: given the sets $\{C_H\}$ and $\{C_T\}$, representing the faces of a coin, respectively, if $\{C_H\} \in \mathcal{F}$ then $\{\bar{C}_H\} = \{C_T\} \in \mathcal{F}$. It follows that $\{C_H, C_T\} \in \mathcal{F} \Rightarrow \Omega \in \mathcal{F} \Rightarrow \emptyset \in \mathcal{F}$.

The previous example can be explained by the following argument. If there a measure for heads then there must be also a measure for tails, if the algebra is to be properly defined. Whenever a probability is assigned to an event then a probability must also be assigned to the complementary event.

The cardinality of a finite set is defined as the number of elements belonging to this set. Sets with an infinite number of elements are said to have the same cardinality if they are equivalent, that is, $A \sim B$ if $\#A = \#B$. Some examples of sets and their respective cardinals are presented next.

- $I = \{1, \dots, k\} \Rightarrow C_I = k$
- $\mathbb{N} = \{0, 1, \dots\} \Rightarrow C_{\mathbb{N}}$ or \aleph_0
- $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\} \Rightarrow C_{\mathbb{Z}}$
- $\mathbb{Q} = \{\dots, -1/3, 0, 1/3, 1/2, \dots\} \Rightarrow C_{\mathbb{Q}}$
- $\mathbb{R} = (-\infty, \infty) \Rightarrow C_{\mathbb{R}}$ or \aleph

For the above examples the following relations are verified: $C_{\mathbb{R}} > C_{\mathbb{Q}} = C_{\mathbb{Z}} = C_{\mathbb{N}} > C_I$. The notation \aleph_0 , for the cardinality of the set of natural numbers was first employed by Cantor.

The cardinality of the power set, that is, of the family of sets consisting of all subsets of a given set I , $\mathcal{F} = 2^I$, is 2^{C_I} .

A.1.6 THE BOREL ALGEBRA

The Borel algebra, established by Félix Edouard Juston Émile Borel (1871–1956), and written as \mathcal{B} , or σ -algebra, is an extension of the algebra so far discussed to operate with limits at infinity. The following properties are required from a σ -algebra.

1. $A \in \mathcal{B} \Rightarrow \bar{A} \in \mathcal{B}$
2. $A_i \in \mathcal{B} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$

The properties guarantee the closure of the σ -algebra with respect to enumerable operations over sets. They allow the definition of limits in the Borel field.

Example: considering the above properties it can be verified that $A_1 \cap A_2 \cap A_3 \cdots \in \mathcal{B}$. In effect, it is sufficient to notice that

$$A \in \mathcal{B} \quad \text{and} \quad B \in \mathcal{B} \Rightarrow A \cup B \in \mathcal{B},$$

and

$$\bar{A} \in \mathcal{B} \quad \text{and} \quad \bar{B} \in \mathcal{B} \Rightarrow \overline{A \cap B} \in \mathcal{B},$$

and finally

$$\overline{\overline{A \cup B}} \in \mathcal{B} \Rightarrow A \cap B \in \mathcal{B}.$$

In summary, any combination of unions and intersections of sets belongs to the Borel algebra. In other words, operations of union or intersection of sets, or a combination of these operations, produce a set that belongs to the σ -algebra.

A.2 BASIC PROBABILITY THEORY

The first known published book on probability is *De Ludo Aleae* (About Games of Chance) by the Italian medical doctor and mathematician Girolamo Cardano (1501–1576), which came out in 1663, almost 90 years after his death. This book was a handbook for players, containing some discussion on probability.

The first mathematical treatise about the Theory of Probability, published in 1657, was written by the Dutch scientist Christian Huygens (1629–1695), a folder titled *De Ratiociniis in Ludo Aleae* (About Reasoning in Games of Chance).

Abraham de Moivre (1667–1754) was an important mathematician who worked on the development of Probability Theory. He wrote a book of great influence in his time called *Doctrine of Chances*. The law of large numbers was discussed by Jacques Bernoulli (1654–1705), Swiss mathematician, in his work *Ars Conjectandi* (The Art of Conjecturing).

The study of probability was improved in the 18th and 19th centuries, being worth of mentioning the works of French mathematicians Pierre-Simon de Laplace (1749–1827) and Siméon Poisson (1781–1840), as well as the German mathematician Karl Friedrich Gauss (1777–1855).

A.2.1 THE AXIOMS OF PROBABILITY

The basic axioms of probability were established by Andrei Nikolaevich Kolmogorov (1903–1987), and allowed the development of the complete theory. The three statements are as follows (Papoulis 1983):

1. Axiom 1 – $P(\Omega) = 1$, in which Ω denotes the sample space or universal set and $P(\cdot)$ denotes the associated probability;
2. Axiom 2 – $P(A) \geq 0$, in which A denotes an event belonging to the sample space;
3. Axiom 3 – $P(A \cup B) = P(A) + P(B)$, in which A and B are mutually exclusive events and $A \cup B$ denotes the union of events A and B .

Kolmogorov established a firm mathematical basis on which other theories rely, including the Theory of Stochastic Processes, the Communications Theory, and the Information Theory, that use his axiomatic approach to Probability Theory.

Kolmogorov's fundamental work was published in 1933 in Russian, and soon afterwards was translated to German with the title *Grundbegriffe der Wahrscheinlichkeits Rechnung* (Fundamentals of Probability Theory) (James 1981). In this work, Kolmogorov managed to combine Advanced Set Theory, developed by Cantor, with Measure Theory, established by Lebesgue, to produce what to this date is the modern approach to Probability Theory.

The application of the axioms makes it possible to deduce all results relative to Probability Theory. For example, the probability of the empty set, $\emptyset = \{\}$, is easily calculated as follows. First it is noticed that

$$\emptyset \cup \Omega = \Omega,$$

since the sets \emptyset and Ω are disjoint. Thus, it follows that

$$P(\emptyset \cup \Omega) = P(\Omega) = P(\emptyset) + P(\Omega) = 1 \Rightarrow P(\emptyset) = 0.$$

In the case of sets A and B which are not disjoint, it follows that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

A.2.2 BAYES' RULE

Bayes' rule, which is essential for the development of Information Theory, concerns the computation of conditional probabilities and can be expressed by the following definition,

$$P(A|B) = \frac{P(A \cap B)}{P(B)},$$

assuming $P(B) \neq 0$.

An equivalent manner of expressing the same result is the following,

$$P(A \cap B) = P(A|B) \cdot P(B), \quad P(B) \neq 0.$$

Some important properties of sets are presented next, in which A and B denote events from a given sample space.

- If A is independent of B , then $P(A|B) = P(A)$. It then follows that $P(B|A) = P(B)$ and that B is independent of A .
- If $B \subset A$, then $P(A|B) = 1$.

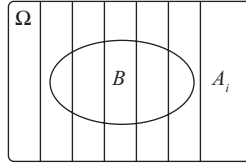


Figure A.5. Partition of set B by a family of sets $\{A_i\}$.

- If $A \subset B$, then $P(A|B) = \frac{P(A)}{P(B)} \geq P(A)$.
- If A and B are independent events then $P(A \cap B) = P(A) \cdot P(B)$.
- If $P(A) = 0$ or $P(A) = 1$, then event A is independent of itself.
- If $P(B) = 0$, then $P(A|B)$ can assume any arbitrary value. Usually in this case one assumes $P(A|B) = P(A)$.
- If events A and B are disjoint, and nonempty, then they are dependent.

A partition is a possible splitting of the sample space into a family of subsets, in a manner that the subsets in this family are disjoint and their union coincides with the sample space. It follows that any set in the sample space can be expressed by using a partition of that sample space, and thus be written as a union of disjoint events.

The following property can be illustrated by means of the Venn diagram, as illustrated in Figure A.5.

$$B = B \cap \Omega = B \cap \bigcup_{i=1}^M A_i = \bigcup_{i=1}^N B \cap A_i.$$

It now follows that

$$P(B) = P\left(\bigcup_{i=1}^N B \cap A_i\right) = \sum_{i=1}^N P(B \cap A_i),$$

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^N P(B \cap A_i)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_{i=1}^N P(B|A_i) \cdot P(A_i)}.$$

A.3 RANDOM VARIABLES

A random variable (r.v.) X represents a mapping of the sample space on the line, that is, the set of real numbers. A random variable is usually characterized by a cumulative probability function (CPF) $P_X(x)$, or by a probability density function (pdf) $p_X(x)$.

Example: a random variable with a uniform pdf, in the interval $[0, 1]$, is described by the formula $p_X(x) = u(x) - u(x - 1)$. It follows, by Axiom 1, that

$$\int_{-\infty}^{+\infty} p_X(x) dx = 1. \quad (\text{A.1})$$

In general, for a given probability distribution, the probability that X belongs to the interval $(a, b]$ is given by

$$P(a < x \leq b) = \int_a^b p_X(x) dx. \quad (\text{A.2})$$

The cumulative probability function $P_X(x)$, of a random variable X , is defined as the integral of $p_X(x)$,

$$P_X(x) = \int_{-\infty}^x p_X(t) dt. \quad (\text{A.3})$$

A.3.1 EXPECTED VALUE OF A RANDOM VARIABLE

Let $f(X)$ denote a function of a random variable X . The average value, or expected value, of the function $f(X)$ with respect to X is defined as

$$E[f(X)] = \int_{-\infty}^{+\infty} f(x) p_X(x) dx. \quad (\text{A.4})$$

The following properties of the expected value follow from (A.4).

$$E[\alpha X] = \alpha E[X], \quad (\text{A.5})$$

$$E[X + Y] = E[X] + E[Y] \quad (\text{A.6})$$

and if X and Y are independent random variables then

$$E[XY] = E[X] \cdot E[Y]. \quad (\text{A.7})$$

A.3.2 MOMENTS OF A RANDOM VARIABLE

The k th moment of a random variable X is defined as

$$m_k = E[X^k] = \int_{-\infty}^{+\infty} x^k p_X(x) dx. \quad (\text{A.8})$$

Various moments of X have special importance and physical interpretation, as defined in the following:

- $E[X]$, arithmetic mean, average value, average voltage, statistical mean;
- $E[X^2]$, quadratic mean, total power;
- $E[X^3]$, measure of asymmetry of the pdf;
- $E[X^4]$, measure of flatness of the pdf.

A.3.3 VARIANCE OF A RANDOM VARIABLE

The variance of a random variable X is an important quantity in communication theory, usually meaning AC power, and defined as follows,

$$V[X] = \sigma_X^2 = E[(X - E[X])^2]. \quad (\text{A.9})$$

The standard deviation σ_X is defined as the square root of the variance of X .

A.3.4 CHARACTERISTIC FUNCTION

The characteristic function $P_X(w)$, also called moment generating function, of a random variable X is usually defined based on the Fourier transform of the pdf of X , which is equivalent to substitute $f(x) = e^{-j\omega x}$ in (A.4), that is,

$$P_X(w) = E[e^{-j\omega x}] = \int_{-\infty}^{+\infty} e^{-j\omega x} p_X(x) dx, \text{ in which } j = \sqrt{-1}. \quad (\text{A.10})$$

The statistical moments of a random variable X can also be obtained directly from then characteristic function, as follows,

$$m_i = \frac{1}{(-j)^i} \left. \frac{\partial^i P_X(w)}{\partial w^i} \right|_{w=0}. \quad (\text{A.11})$$

Given that X is a random variable, it follows that $Y = f(X)$ is also a random variable, obtained by the application of the transformation $f(\cdot)$. The pdf of Y is related to that of X by the formula (Blake 1987)

$$p_Y(y) = \left. \frac{p_X(x)}{|dy/dx|} \right|_{x=f^{-1}(y)}, \quad (\text{A.12})$$

in which $f^{-1}(\cdot)$ denotes the inverse function of $f(\cdot)$. This formula assumes the existence of the inverse function of $f(\cdot)$ as well as its derivative in all points.

A.3.4.1 Two Important Distributions

1. Gaussian random variable

The random variable X with pdf

$$p_X(x) = \frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{(x-m_X)^2}{2\sigma_X^2}} \quad (\text{A.13})$$

is called a Gaussian (or Normal) random variable. The Gaussian random variable plays an extremely important role in engineering, considering that many well-known processes can be described or approximated by this pdf. The noise present in either analog or digital communications systems usually can be considered Gaussian as a consequence of the influence of many factors (Leon-Garcia 1989). In Formula (A.13), m_X represents the average value and σ_X^2 represents the variance of X .

2. Sinusoidal random variable

A sinusoidal tone $X(t) = V \cos(\omega_0 t + \phi)$, in which V represents a constant amplitude, ω_0 is a constant frequency, and ϕ is a uniformly distributed random variable, has the following pdf

$$p_X(x) = \frac{1}{\pi \sqrt{V^2 - x^2}}, \quad |x| < V. \quad (\text{A.14})$$

A.3.5 JOINT RANDOM VARIABLES

Consider that X and Y represent a pair of real random variables, with joint pdf $p_{XY}(x, y)$, as illustrated in Figure A.6. The probability of x and y being simultaneously in the region defined by the polygon [abcd] is given by the expression (Alencar 2014).

$$\text{Prob}(a < x < b, c < y < d) = \int_a^b \int_c^d p_{XY}(x, y) dx dy. \quad (\text{A.15})$$

The individual pdf's of X and Y , also called marginal distributions, result from the integration of the joint pdf as follows,

$$p_X(x) = \int_{-\infty}^{+\infty} p_{XY}(x, y) dy, \quad (\text{A.16})$$

and

$$p_Y(y) = \int_{-\infty}^{+\infty} p_{XY}(x, y) dx. \quad (\text{A.17})$$

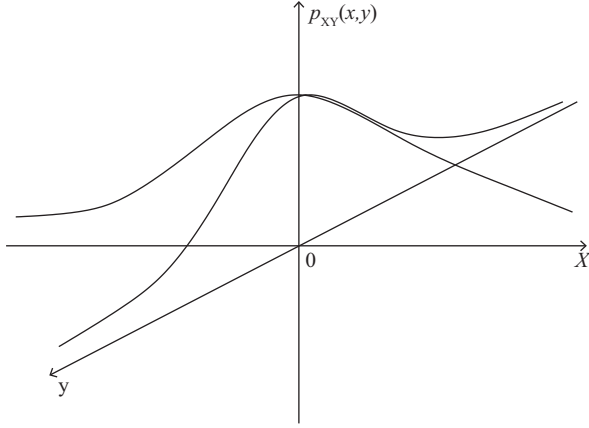


Figure A.6. Joint probability density function.

The joint average $E[f(X, Y)]$ is calculated as

$$E[f(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) p_{XY}(x, y) dx dy, \quad (\text{A.18})$$

for an arbitrary function $f(X, Y)$ of X and Y .

The joint moments m_{ik} , of order ik , are calculated as

$$m_{ik} = E[X^i, Y^k] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^i y^k p_{XY}(xy) dx dy. \quad (\text{A.19})$$

The two-dimensional characteristic function is defined as the two-dimensional Fourier transform of the joint probability density $p_{XY}(x, y)$

$$P_{XY}(\omega, \nu) = E[e^{-j\omega X - j\nu Y}]. \quad (\text{A.20})$$

When the sum $Z = X + Y$ of two statistically independent r.v.'s is considered, it is noticed that the characteristic function of Z turns out to be

$$P_Z(\omega) = E[e^{-j\omega Z}] = E[e^{-j\omega(X+Y)}] = P_X(\omega) \cdot P_Y(\omega). \quad (\text{A.21})$$

As far as the pdf of Z is concerned, it can be said that

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(\rho) p_Y(z - \rho) d\rho, \quad (\text{A.22})$$

or

$$p_Z(z) = \int_{-\infty}^{\infty} p_X(z - \rho) p_Y(\rho) d\rho. \quad (\text{A.23})$$

Equivalently, the sum of two statistically independent r.v.'s has a pdf given by the convolution of the respective pdf's of the r.v.'s involved in the sum.

The random variables X and Y are called uncorrelated if $E[XY] = E[X] \cdot E[Y]$. The criterion of statistical independence of random variables, which is stronger than correlation, is satisfied if $p_{XY}(x, y) = p_X(x) \cdot p_Y(y)$.

REFERENCES

- Abramowitz, M., and I.A. Stegun, eds. 1965. *Handbook of Mathematical Functions*. New York: Dover Publications Inc.
- Abramson, N. 1963. *Information Theory and Coding*. New York: McGraw-Hill.
- Aczél, J., and Z. Daróczy. 1975. *On Measures of Information and Their Characterizations*. New York: Academic Press.
- Adke, S.R., and S.M. Manjunath. 1984. *An Introduction to Finite Markov Processes*. New Delhi: John Wiley and Sons.
- Ahlsvede, R. 1971. "Multi-way Communication Channels." In *Proceedings of the 2nd International Symposium on Information Theory*, pp. 103–35. Tsahkad-sor, USSR: Akadémiai Kiadó..
- Alencar, M.S. 1992a. "The Capacity Region for the Gaussian Channel with Random Multipath." In *Proceedings of the Third IEEE International Symposium on Personal, Indoor and Mobile Radio Communications – PIMRC'92*, pp. 483–87. Boston, MA.
- Alencar, M.S. 1992b. "The Capacity Region of the Non-Cooperative Multiple Access Channel with Rayleigh Fading." In *Proceedings of the Annual Conference of the Canadian Institute for Telecommunications Research*, pp. 45–46. Sainte Adèle, Canada.
- Alencar, M.S. 2014. *Teoria de Conjuntos, Medida e Probabilidade*. São Paulo, Brasil: Edi-toraÉrica Ltda. ISBN 978-85-365-0715-6.
- Alencar, M.S., and I.F. Blake. 1993a. "Analysis of the Capacity for a Markovian Multiple Access Channel." In *IEEE Pacific Rim Conference on Communications, Computers and Signal Processing*, pp. 81–84. Victoria, Canada.
- Alencar, M.S., and I.F. Blake. 1993b. "Analysis of the Capacity Region for the Non-Cooperative Multiaccess Channel with Rician Fading." In *IEEE International Conference on Communications – ICC'93*, pp. 282–86. Geneva, Switzerland.

- Alencar, M.S., and I.F. Blake. 1993c. "The Capacity of a Code Division Multiple Access Channel." In *1993 International Symposium on 139 Communications – ISCOM'93*, Vol.1, pp. 1.1–1.9. Hsinchu, Taiwan.
- Ash, R.B. 1965. *Information Theory*. New York: Dover Publications Inc.
- Bierbaum, M., and H.-M. Wallmeier. 1979. "A Note on the Capacity Region of the Multiple Access Channel." *IEEE Transactions on Information Theory* 25, no. 4, p. 484. doi: <http://dx.doi.org/10.1109/tit.1979.1056064>
- Blahut, R.E. 1987. *Principles and Practice of Information Theory*. Reading, MA: Addison-Wesley Publishing Co.
- Blahut, R.E. 1990. *Digital Transmission of Information*. Reading, MA: Addison-Wesley Publishing Co.
- Blake, I.F. 1987. *An Introduction to Applied Probability*. Malabar, FL: Robert E. Krieger Publishing Co.
- Blake, I.F., and R.C. Mullin. 1976. *An Introduction to Algebraic and Combinatorial Coding Theory*. New York: Academic Press Inc.
- Borth, D.E., and M.B. Pursley. 1979. "Analysis of Direct-Sequence Spread-Spectrum Multiple-Access Communications Over Rician Fading Channels." *IEEE Transactions on Communications* 27, no. 10, pp. 1566–77. doi: <http://dx.doi.org/10.1109/tcom.1979.1094291>
- Boyer, C. 1974. *História da Matemática*. São Paulo, Brasil: Editora Edgard Blucher Ltda.
- Brady, D.P. 1991. "A Semiclassical Analysis of Optical Code Division Multiple Access." *IEEE Transactions on Communications* 39, no. 1, pp. 85–93. doi: <http://dx.doi.org/10.1109/26.68279>
- Cheng, R.S., and S. Verdú. 1991. "The Effect of Asynchronism on the Total Capacity of Gaussian Multiple-access Channels." In *Proceedings of the IEEE International Symposium on Information Theory*, p. 211. Budapest, Hungary.
- Chung, H., and P.V. Kumar. 1990. "Optical Orthogonal Codes – New Bounds and an Optimal Construction." *IEEE Transactions on Information Theory* 36, no. 4, pp. 866–73. doi: <http://dx.doi.org/10.1109/18.53748>
- Cook, C.E., F.W. Ellersick, L.B. Milstein, and D.L. Schilling. 1983. *Spread-Spectrum Communications*. New York: IEEE Press.
- Cook, C.E., and H.S. Marsh. 1983. "An Introduction to Spread Spectrum." *IEEE Communications Magazine* 21, no. 2, pp. 8–16. doi: <http://dx.doi.org/10.1109/mcom.1983.1091346>
- Cover, T.M. 1972. "Broadcast Channels." *IEEE Transactions on Information Theory* 18, no. 1, pp. 2–14. doi: <http://dx.doi.org/10.1109/tit.1972.1054727>

- Cover, T.M., R.J. McEliece, and E.C. Posner. 1981. "Asynchronous Multiple-Access Channel Capacity." *IEEE Transactions on Information Theory* 27, no. 4, pp. 409–13. doi: <http://dx.doi.org/10.1109/tit.1981.1056382>
- Csiszár, I., and J. Körner. 1981. *Information Theory: Coding Theorems for Discrete Memoryless Systems*. New York: Academic Press.
- Dixon, R.C. 1984. *Spread Spectrum Systems*. New York: John Wiley & Sons.
- Dobrushin, R.L. 1961. "Mathematical Problems in the Shannon Theory of Optimal Coding of Information." In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, Vol. 1, pp. 211–52. Berkeley, CA.
- Elliott, E.O. 1965. "A Model of the Switched Telephone Network for Data Communications." *The Bell System Technical Journal* 44, no. 1, pp. 88–109. doi: <http://dx.doi.org/10.1002/j.1538-7305.1965.tb04139.x>
- Ericson, T. 1970. "A Gaussian Channel with Slow Fading." *IEEE Transactions on Information Theory* 16, no. 3, pp. 353–55. doi: <http://dx.doi.org/10.1109/tit.1970.1054448>
- Feinstein, A. 1958. *Foundations of Information Theory*. New York: McGraw-Hill Book Company Inc.
- Fritchman, B.D. 1967. "A Binary Channel Characterization Using Partitioned Markov Chains." *IEEE Transactions on Information Theory* 13, no. 2, pp. 221–27. doi: <http://dx.doi.org/10.1109/tit.1967.1053975>
- Gaarder, N.T., and J.K. Wolf. 1975. "The Capacity of a Multiple-Access Discrete Memoryless Channel Can Increase with Feedback." *IEEE Transactions on Information Theory* 21, no. 1, pp. 100–02. doi: <http://dx.doi.org/10.1109/tit.1975.1055312>
- Gallager, R.G. 1968. *Information Theory and Reliable Communication*. New York: John Wiley and Sons Inc.
- Gallager, R.G. 1985. "A Perspective on Multiaccess Channels." *IEEE Transactions on Information Theory* 31, no. 2, pp. 124–42. doi: <http://dx.doi.org/10.1109/tit.1985.1057022>
- Geraniotis, E.A., and B. Ghaffari. 1991. "Performance of Binary and Quaternary Direct-Sequence Spread Spectrum Multiple-Access Systems with Random Signature Sequences." *IEEE Transactions on Communications* 39, no. 5, pp. 713–24. doi: <http://dx.doi.org/10.1109/26.87162>
- Geraniotis, E.A., and M.B. Pursley. 1985. "Performance of Coherent Direct-Sequence Spread Spectrum Communications Over Specular Multipath Fading Channels." *IEEE Transactions on Communications* 33, no. 6, pp. 502–08. doi: <http://dx.doi.org/10.1109/tcom.1985.1096335>

- Geraniotis, E.A., and M.B. Pursley. 1986. "Performance of Noncoherent Direct-Sequence Spread Spectrum Communications over Specular Multipath Fading Channels." *IEEE Transactions on Communications* 34, no. 6, pp. 219–26. doi: <http://dx.doi.org/10.1109/tcom.1986.1096528>
- Gilbert, E.N. 1960. "Capacity of a Burst-Noise Channel." *The Bell System Technical Journal* 39, no. 5, pp. 1253–65. doi: <http://dx.doi.org/10.1002/j.1538-7305.1960.tb03959.x>
- Gilhousen, K.S., I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, and C.E. Wheatley. 1991. "On the Capacity of a Cellular CDMA System." *IEEE Transactions on Vehicular Technology* 40, no. 2, pp. 303–12. doi: <http://dx.doi.org/10.1109/25.289411>
- Gold, R. 1967. "Optimal Binary Sequences for Spread Spectrum Multiplexing." *IEEE Transactions on Information Theory* 13, no. 4, pp. 619–21. doi: <http://dx.doi.org/10.1109/tit.1967.1054048>
- Gradshteyn, I.S., and I.M. Ryzhik. 1990. *Table of Integrals, Series, and Products*. San Diego, CA: Academic Press Inc.
- Halmos, P.R. 1960. *Naive Set Theory*. Princeton, NJ: D. Van Nostrand Company Inc.
- Hartley, R.V.L. 1928. "Transmission of Information." *Bell Systems Technical Journal* p. 535. doi: <http://dx.doi.org/10.1002/j.1538-7305.1928.tb01236.x>
- Haykin, S. 1988. *Digital Communications*. New York: John Wiley and Sons.
- Haykin, S. 1999. *The Code Book – The Science of Secrecy from Ancient Egypt to Quantum Cryptography*. New York: Doubleday.
- Holtzman, J.M. 1992. "A Simple, Accurate Method to Calculate Spread-Spectrum Multiple-Access Error Probabilities." *IEEE Transactions on Communications* 40, no. 3, pp. 461–64. doi: <http://dx.doi.org/10.1109/26.135712>
- Hui, J.Y.N. 1984a. "Multiple Accessing for the Collision Channel Without Feedback." *IEEE Journal on Selected Areas in Communications* 2, no. 4, pp. 575–82. doi: <http://dx.doi.org/10.1109/jsac.1984.1146089>
- Hui, J.Y.N. 1984b. "Throughput Analysis for Code Division Multiple Accessing of the Spread Spectrum Channel." *IEEE Journal on Selected Areas in Communications* 2, no. 4, pp. 482–86. doi: <http://dx.doi.org/10.1109/jsac.1984.1146083>
- Hui, J.Y.N., and P.A. Humblet. 1985. "The Capacity Region of the Totally Asynchronous Multiple-Access Channel." *IEEE Transactions on Information Theory* 31, no. 2, pp. 207–16.
- James, B.R. 1981. *Probabilidade: Um Curso Em Nível Intermediário*. CNPq, Rio de Janeiro, Brasil: Instituto de Matemática Pura e Aplicada.

- Kavehrad, M., and P.J. McLane. 1985. "Performance of Low-Complexity Channel Coding and Diversity for Spread Spectrum in Indoor, Wireless Communication." *AT&T Technical Journal* 64, no. 8, pp. 1927–64.
- Kavehrad, M., and P.J. McLane. 1987. "Spread Spectrum for Indoor Digital Radio." *IEEE Communications Magazine* 25, no. 6, pp. 32–40.
- Kavehrad, M., and B. Ramamurthi. 1987. "Direct Sequence Spread Spectrum with DPSK Modulation and Diversity for Indoor Wireless Communications." *IEEE Transactions on Communications* 35, no. 2, pp. 224–36.
- Khinchin, A.I. 1957. *Mathematical Foundations of Information Theory*. New York: Dover Publications, Inc.
- Kleinrock, L. 1975. *Queueing Systems*. New York: John Wiley & Sons.
- Komo, J.J., and S.C. Liu. 1990. "Maximal Length Sequences for Frequency Hopping." *IEEE Journal on Selected Areas in Communications* 8, no. 5, pp. 819–22.
- Lee, W.C.Y. 1990. "Estimate of Channel Capacity in Rayleigh Fading Environment." *IEEE Transactions on Vehicular Technology* 39, no. 3, pp. 187–89.
- Lee, W.C.Y. 1991. "Overview of Cellular CDMA." *IEEE Transactions on Vehicular Technology* 40, no. 2, pp. 291–302.
- Leon-Garcia, A. 1989. *Probability and Random Processes for Electrical Engineering*. Reading, Massachusetts: Addison-Wesley Publishing Co.
- Lipschutz, S. 1968. *Teoria de Conjuntos*. Rio de Janeiro, Brasil: Ao Livro Técnico S.A.
- MacKay, D.J.C. 2003. *Information Theory, Inference, and Learning Algorithms*. Cambridge, UK: Cambridge University Press.
- Mandell, M., and R. McEliece. 1991. Some Properties of Memory-less Multiterminal Interference Channels. *Proceedings of the IEEE International Symposium Information Theory*, p. 212. Budapest, Hungary: IEEE.
- Massey, J.L. 1990. The Relevance of Information Theory to Modern Cryptography. *Proceedings of the Bilkent International Conference on New Trends in Communications, Control and Signal Processing (NILCON'90)*, pp. 176–82. Ankara, Turkey: Elsevier Science Publisher.
- Maurer, U.M. 1989. A Provably-Secure Strongly-Randomized Cipher. *Proceedings of the Monte Verita Seminar on Future Directions in Cryptography*. Ascona, Switzerland.
- Misser, H.S., C.A.F.J. Wijffels, and R. Prasad. 1991. "Throughput Analysis of CDMA with DPSK Modulation and Diversity in Indoor Rician Fading Radio Channels." *Electronics Letters* 27, no. 7, pp. 601–603.
- Nyquist, H. 1924. "Certain Factors Affecting Telegraph Speed." *Bell Systems Technical Journal*, 3, p. 324.

- Oberhettinger, F. 1990. *Tables of Fourier Transforms and Fourier Transforms of Distributions*. Berlin, Germany: Springer-Verlag.
- Ozarow, L.H., and A.D. Wyner. 1990. "On the Capacity of the Gaussian Channel with a Finite Number of Input Levels." *IEEE Transactions on Information Theory* 36, no. 6, pp. 1426–28.
- Papoulis, A. 1983. "Random Modulation: A Review." *IEEE Transactions on Acoustics, Speech and Signal Processing* 31, no. 1, pp. 96–105.
- Pickholtz, R.L., D.L. Schilling, and L.B. Milstein. 1982. "Theory of Spread-Spectrum Communications—A Tutorial." *IEEE Transactions on Communications, COM* 30, no. 5, pp. 855–84.
- Pierce, J.R. 1980. *An Introduction to Information Theory—Symbols, Signals & Noise*. 2nd ed. New York: Dover Publications Inc.
- Proakis, J.G. 1989. *Digital Communications*. McGraw-Hill.
- Pursley, M.B. 1977. "Performance Evaluation for Phase-Coded Spread Spectrum Multiple-Access Communications—Part I: System Analysis." *IEEE Transactions on Communications* 25, no. 8, pp. 795–99.
- Reza, F.M. 1961. *An Introduction to Information Theory*. New York: McGraw-Hill Book Co.
- Ricciardi, L.M. 1990. *Lectures in Applied Mathematics and Informatics*. Manchester, UK: Manchester University Press.
- Sadowsky, J.S., and R.K. Bahr. 1991. "Direct-Sequence Spread-Spectrum Multiple-Access Communications with Random Signature Sequences: A Large Deviations Analysis." *IEEE Transactions on Information Theory* 37, no. 3, pp. 514–27.
- Sarwate, D.V. and M.B. Pursley. 1980. "Crosscorrelation Properties of Pseudorandom and Related Sequences." *Proceedings of the IEEE* 68, no. 5, pp. 593–619.
- Sayood, K. 2006. *Introduction to Data Compression*. San Francisco, CA: Morgan Kaufmann.
- Schilling, D.L., L.B. Milstein, R.L. Pickholtz, M. Kullback, and F. Miller. 1991. "Spread Spectrum for Commercial Communications." *IEEE Communications Magazine* 29, no. 4, pp. 66–79.
- Scholtz, R.A. 1982. "The Origins of Spread-Spectrum Communications." *IEEE Transactions on Communications, COM* 30, no. 5, pp. 822–54.
- Schwartz, M. 1970. *Information Transmission, Modulation, and Noise*. New York: McGraw-Hill.
- Shannon, C.E. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27, pp. 379–423.
- Shannon, C.E. 1961. "Two-way Communications Channels." *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, Vol. 1, pp. 611–44. Berkeley, CA: University of California Press.

- Shamai, S., L.H. Ozarow, and A.D. Wyner. 1991. "Information Rates for a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs." *IEEE Transactions on Information Theory* 37, no. 6, pp. 1527–39.
- Simon, M.K., J.K. Omura, R.A. Scholtz, and B.K. Levitt. 1985. Vol. 1. *Spread Spectrum Communications*. Rockville, MD: Computer Science Press.
- Sommer, R.C. 1966. "Asynchronously Multiplexed Channel Capacity." *Proceeding of the IEEE* 54, no. 1, pp. 79–80.
- Soroushnejad, M., and E. Geraniotis. 1991. "Probability of Capture and Rejection of Primary Multiple-Access Interference in Spread-Spectrum Networks." *IEEE Transactions on Communications* 39, no. 6, pp. 986–94.
- Sousa, E.S. 1989. "Throughput of Spread-Spectrum Systems with a Large Number of Users." *IEE Proceedings* 136, no. 3, pp. 220–26.
- Sousa, E.S. 1990. "Interference Modeling in a Direct-Sequence Spread-Spectrum Packet Radio Network." *IEEE Transactions on Communications* 38, no. 9, pp. 1475–82.
- Sousa, E.S. 1992. "Performance of a Spread Spectrum Packet Radio Network Link in a Poisson Field of Interferers." *IEEE Transactions on Information Theory* 38, no. 6, pp. 1743–54.
- Stallings, W. 1999. *Cryptography and Network Security—Principles and Practice*. Upper Saddle River, NJ: Prentice Hall.
- Tanenbaum, A.S. 2003. *Computer Networks*. Englewood Cliffs, NJ: Prentice-Hall, PTR.
- Turin, G.L. 1980. "Introduction to Spread-Spectrum Antimultipath Techniques and Their Application to Urban Digital Radio." *Proceedings of the IEEE* 68, no. 3, pp. 328–53.
- Turin, G.L. 1984a. "Commutation Signaling—An Antimultipath Technique." *IEEE Journal on Selected Areas in Communications* 2, no. 4, pp. 548–62.
- Turin, G.L. 1984b. "The Effects of Multipath and Fading on the Performance of Direct-Sequence CDMA Systems." *IEEE Journal on Selected Areas in Communications* 2, no. 4, pp. 597–603.
- Van der Lubbe, J.C.A. 1997. *Information Theory*. Cambridge, UK: Cambridge University Press.
- Verdú, S. 1986. "Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels." *IEEE Transactions on Information Theory* 32, no. 1, pp. 85–96.
- Verdú, S. 1989a. "Multiple-Access Channels with Memory with and without Frame Synchronization." *IEEE Transactions on Information Theory* 35, no. 3, pp. 605–19.

- Verdú, S. 1989b. "The Capacity Region of the Symbol Asynchronous Gaussian Multiple-Access Channel." *IEEE Transactions on Information Theory* 35, no. 4, pp. 733–51.
- Viterbi, A.J. 1985. "When Not to Spread Spectrum—A Sequel." *IEEE Communications Magazine* 23, no. 4, pp. 12–17.
- Viterbi, A.J. 1991. "Wireless Digital Communication: A View Based on Three Lessons Learned." *IEEE Communications Magazine* 29, no. 9, pp. 33–36.
- Wang, H.S. *Finite-State Modeling, Capacity, and Joint Source/Channel Coding for Time-Varying Channels*. [PhD Thesis]. New Brunswick, NJ: Graduate School – New Brunswick Rutgers, The State University of New Jersey; 1992
- Welch, L.R. 1974. "Lower Bounds on the Maximum Cross Correlation of Signals". *IEEE Transactions on Information Theory*, 20, no. 3, pp. 397–99.
- Wozencraft, J.M., and B. Reiffen. 1961. *Sequential Decoding*. Cambridge, US: MIT Press.
- Wyner, A.D. 1974. "Recent Results in the Shannon Theory". *IEEE Transactions on Information Theory*, 20, no. 1, pp. 2–9.
- Yue, O.-C. 1983. "Spread Spectrum Mobile Radio, 1977–1982". *IEEE Transactions on Vehicular Technology*, 32, no. 1, pp. 98–105.
- Zadeh, L.A. 1965. "Fuzzy Sets". *Information and Control*, 8, no. 3, pp. 338–53.

ABOUT THE AUTHOR

Marcelo Sampaio de Alencar was born in Serrita, Brazil in 1957. He received his bachelor's degree in Electrical Engineering from Federal University of Pernambuco (UFPE), Brazil in 1980; his master's degree from Federal University of Paraiba (UFPB), Brazil 1988; and his PhD from University of Waterloo, Canada in 1993. He is currently an emeritus member of the Brazilian Telecommunications Society (SBrT), IEEE senior member, chair professor at the Department of Electrical Engineering, Federal University of Campina Grande, Brazil. He also worked for the State University of Santa Catarina (UDESC), Embratel, and University of Toronto, as visiting professor.

He is founder and president of the Institute for Advanced Studies in Communications (Iecom). He has been awarded several scholarships and grants, including three scholarships and several research grants from the Brazilian National Council for Scientific and Technological Research (CNPq), two grants from the IEEE Foundation, a scholarship from the University of Waterloo, a scholarship from the Federal University of Paraiba, an achievement award for contributions to the Brazilian Telecommunications Society (SBrT), an award from the Medicine College of the Federal University of Campina Grande (UFCG), an achievement award from the College of Engineering of the Federal University of Pernambuco, and the Medal Professor Attilio Jose Giarola from the Brazilian Microwave and Optoelectronics Society (SBMO). He has published over 350 engineering and scientific papers and 15 books. He is a columnist of the traditional Brazilian newspaper *Jornal do Comercio*, and vice president for external relations, SBrT.

INDEX

A

Absolutely secure event, 124
Algebra, 127, 129
 Borel, 130
 closure, 129
Algebraic coding theory, 77
Alzheimer, Aloysius, 32
Appearance
 equivocation, 121
Asymmetric key, 120
Axiom
 choice, 126
 Peano, 126
 specification, 126
Axioms, 126
Ayer, Alfred Jules, 1

B

Belonging, 126
Binary
 Huffman, 27
Bit, 17
Bletchley Park, 117
Borel
 algebra, 130
Borel, Félix Edouard Juston
 Émile, 130
BSC, 44

C

Cantor, Georg, 125
Capacity

bound, 107
channel, 41
sum, 94

Cardinal

number, 130
numbers, 125

Cardinality, 130

Carrier suppression, 80

CDMA, 72, 76, 87

non-cooperative, 89
performance analysis, 76,
90

Channel

binary symmetric, 44
communication, 34
discrete, 43
independent output, 36
noiseless, 35
noisy, 43
non-cooperative, 89

Ciphertext

entropy, 121

Clausius, Rudolph, 31

Closure, 129

CMOS, 17

Code

chip rate, 88
classification, 19
comma, 21
decision tree, 14

- decoding tree, 15
- Huffman, 27
- instantaneous, 21, 23
- prefix, 14, 16
 - extended, 17
- uniquely decodable, 20
- Coder
 - codewords, 11
 - source, 13
- Codes
 - block, 19
 - non-singular, 19
- Codewords, 11
- Coding
 - efficiency, 12
 - source, 11
- Communications
 - personal, 71
- Computer network, 118
- Cryptography, 72, 118
 - history, 117
 - information theory, 121
 - model, 119
 - principles, 120
 - uncertainty, 121
- Cryptosystem
 - absolute security, 124
 - information theory, 123
- D**
- Dedekind, J. W. R., 125
- Direct sequence, 72
- Disjoint, 126
- DPSK, 87
- DS, 72
- DSSS, 78, 87
- E**
- Efficiency, 43
- Empty, 126
- Enigma, 117
- Entropy, 3
- conditional, 34
- cryptography, 120
- cryptosystem, 122
- extended source, 12
- inequality, 37
- joint, 33, 39
- properties, 6
- relations, 37
- Error probability, 109
- Extension
 - source, 12
- F**
- Fading, 72
- Families, 128
- Fano, Robert, 27
- FDMA, 76
- FH, 72
- Fractals, 125
- Frequency hopping, 72
- Fuzzy set, 2
- G**
- Gold sequence, 82
- H**
- Hartley, Ralph Vinton Lyon, 1, 31
- Hilbert, David, 125
- Huffman
 - non-uniqueness, 29
 - procedure, 28
- Huffman, David, 27
- I**
- Inclusion, 126
- Indexing, 129
- Inequality
 - Kraft, 24
 - Kraft-McMillan, 16
- Information
 - average mutual, 39

- channel, 34
- entropy, 2, 3
- joint, 32
- measure of, 2
- mutual, 38
- semantics, 1
- theory, 1
- Internet
 - backbone, 118
 - protocol, 118
- J**
- Jammer, 74
 - broad band noise, 74
 - partial band, 74
 - partial time, 74
 - repeater, 74
 - tone, 74
- Jamming, 72
- Joint entropy
 - cryptosystem, 122
- Joint random Variables, 136
- K**
- Kasami
 - sequence, 84
- Key
 - entropy, 120
 - equivocation, 121
- Khinchin, Aleksandr
 - Yakovlevich, 2
- Kolmogorov, Andrei
 - Nikolaevich, 31, 131
- Kraft
 - inequality, 24
- Kraft-McMillan
 - inequality, 16
- Kronecker, Leopold, 125
- L**
- Laser, 18
- Lebesgue, Henri Léon, 40
- M**
- MAI, 78
- Matched filter, 91
- Maximal length
 - linear sequence, 80
 - sequence, 79
- Maximum likelihood receiver, 91
- Measure, 126
 - Lebesgue, 40
 - probability, 126
- Message
 - equivocation, 121
 - uncertainty, 121
- Moments, 134
- Multipath, 72, 88
- Multiple access, 71
 - DSSS, 78
 - interference, 78
- N**
- Non-cooperative channel, 89
- Non-repudiation, 118
- Nyquist, Harry, 1, 31
- P**
- Password, 118
- PCN, 87
- Peano
 - axiom, 126
- Peirce, Charles Sanders, 1
- Personal communications, 71
- Photon, 17
- Pierce, John Robinson, 1
- Plaintext
 - entropy, 120
- Prefix
 - code, 14, 16
- Probability, 126
 - joint random variables, 136
 - moments, 134
 - random variables, 133

Probability theory, 133
Pseudo-random, 80
Public key infrastructure,
118

R

Rényi, Alfréd, 2
Radiation
electromagnetic, 18
Random variables, 133
Rate
transmission, 43
Redundancy, 12
absolute, 42
relative, 43

S

Schröder-Bernstein
theorem, 126
Sequence
Gold, 82
Kasami, 84
maximal length, 81
maximal-length, 79, 80
Mersenne, 81
Welsh bound, 82
Sequence design, 79
Session
hijacking, 118
Set, 125
algebra, 127
disjoint, 126
families, 128
fuzzy, 2
infinite, 125
operations, 127
universal, 126
universal set, 125
Set theory, 125, 126
Sets
algebra, 129
indexing, 129

Shannon
Claude Elwood, 39, 41
first theorem, 12
Shannon, Claude Elwood,
1, 31
Signal
space, 74
Sniffing, 118
Source
efficiency, 12
extended, 12
Source coding
theorem, 12
Spectral
compression, 75
Spoofing, 118
Spread spectrum, 71
m-sequence, 79
carrier suppression, 80
direct sequence, 72
DSSS, 87
frequency hopping, 72
interference, 77
performance Analysis, 76
pseudo-random, 80
sequence design, 79
time hopping, 73
time-frequency hopping,
78
Substitution, 120
Sum capacity, 94
Symmetric key, 120
T
TCP/IP, 118
TDMA, 76
TH, 73
Theorem
Schröder-Bernstein, 126
Time hopping, 73
Transfinite arithmetic, 125

Transformation
 cryptography, 119
Transposition, 120
TTL, 17
Turing, Alan Mathison, 117

U

Universal, 126
University of Halle, 125

V

Vectorial
 space, 74

Venn
 diagram, 126
Venn diagram, 126

W

Welsh bound, 82

Z

Zadeh, Lotfi Asker, 2
Zenon, 125
Zorn
 lemma, 126
Zorn's lemma, 126

FORTHCOMING TITLES IN OUR COMMUNICATIONS AND SIGNAL PROCESSING COLLECTION

Orlando Baiocchi, University of Washington Tacoma, Editor

Signal Integrity: The Art of Interconnect Design For High Speed and
High Reliability Circuits by Joel Jorgenson

Cryptography Explained by Raj Katti

Momentum Press publishes several other collections, including: Industrial, Systems, and Innovation Engineering; Manufacturing and Processes; Engineering Management; Electrical Power; Fluid Mechanics; Acoustical Engineering; Aerospace Engineering; Biomedical Engineering; and Healthcare Administration

Momentum Press is actively seeking collection editors as well as authors. For more information about becoming an MP author or collection editor, please visit <http://www.momentumpress.net/contact>

Announcing Digital Content Crafted by Librarians

Momentum Press offers digital content as authoritative treatments of advanced engineering topics by leaders in their field. Hosted on ebrary, MP provides practitioners, researchers, faculty, and students in engineering, science, and industry with innovative electronic content in sensors and controls engineering, advanced energy engineering, manufacturing, and materials science.

Momentum Press offers library-friendly terms:

- perpetual access for a one-time fee
- no subscriptions or access fees required
- unlimited concurrent usage permitted
- downloadable PDFs provided
- free MARC records included
- free trials

The **Momentum Press** digital library is very affordable, with no obligation to buy in future years.

For more information, please visit www.momentumpress.net/library or to set up a trial in the US, please contact mpsales@globalepress.com.

EBOOKS FOR THE ENGINEERING LIBRARY

*Create your own
Customized Content
Bundle—the more
books you buy,
the greater your
discount!*

THE CONTENT

- *Manufacturing Engineering*
- *Mechanical & Chemical Engineering*
- *Materials Science & Engineering*
- *Civil & Environmental Engineering*
- *Advanced Energy Technologies*

THE TERMS

- *Perpetual access for a one time fee*
- *No subscriptions or access fees*
- *Unlimited concurrent usage*
- *Downloadable PDFs*
- *Free MARC records*

For further information, a free trial, or to order, contact:
sales@momentumpress.net

Information Theory

Marcelo S. Alencar

Information Theory covers the historical evolution of information theory, along with the basic concepts linked to information. It discusses the information associated to a certain source and the usual types of source codes, information transmission, joint information, conditional entropy, mutual information, and channel capacity.

The hot topic of multiple access systems for cooperative and noncooperative channels is also discussed, along with code division multiple access (CDMA), the basic block of most cellular and personal communication systems, and the capacity of a CDMA system. Also presented is the information theoretical aspects of cryptography, which are important for network security, a topic intrinsically connected to computer networks and the Internet.

To help the reader understand the theory, the text also includes a review of probability theory, solved problems, illustrations, and graphics.

Marcelo Sampaio de Alencar received his bachelor's degree in Electrical Engineering from Federal University of Pernambuco (UFPE), Brazil in 1980; his master's degree from Federal University of Paraiba (UFPB), Brazil in 1988; and his PhD from the University of Waterloo, Canada in 1993. He is currently emeritus member of the Brazilian Telecommunications Society (SBTr), IEEE senior member, chair professor at the Department of Electrical Engineering, Federal University of Campina Grande, Brazil. He is founder and president of the Institute for Advanced Studies in Communications (Iecom), and has been awarded several scholarships and grants including three scholarships and several research grants from the Brazilian National Council for Scientific and Technological Research (CNPq), two grants from the IEEE Foundation, a scholarship from the University of Waterloo, a scholarship from the Federal University of Paraiba, and many others. He has published over 350 engineering and scientific papers and 15 books.



MOMENTUM PRESS
ENGINEERING

