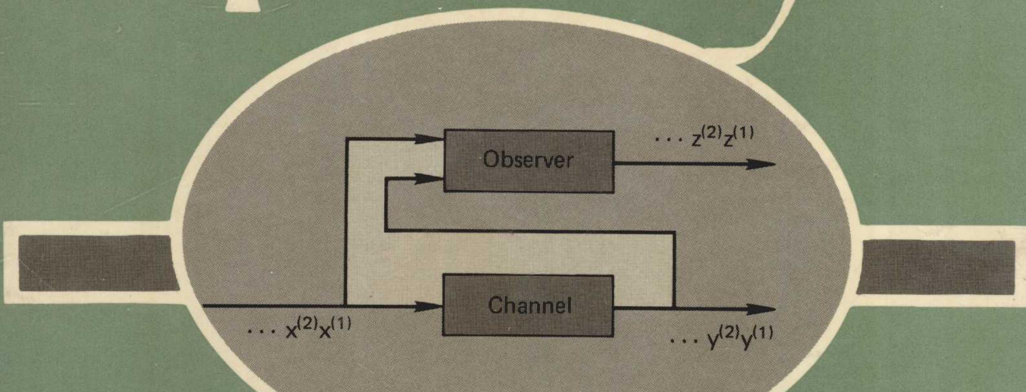
INTRODUCTION TO
Information Theory



MASUD MANSURIPUR

INTRODUCTION TO INFORMATION THEORY

Masud Mansuripur
College of Engineering
Boston University

PRENTICE-HALL, INC.
Englewood Cliffs, New Jersey 07632

Library of Congress Cataloging-in-Publication Data

Masud Mansuripur (date)
Introduction to information theory.

Bibliography: p.
Includes index.
1. Information theory. I. Title.
Q360.M3215 1987 001.53′9 86-17033
ISBN 0-13-484668-0

Cover Design: Diane Saxe
Manufacturing Buyer: Rhett Conklin

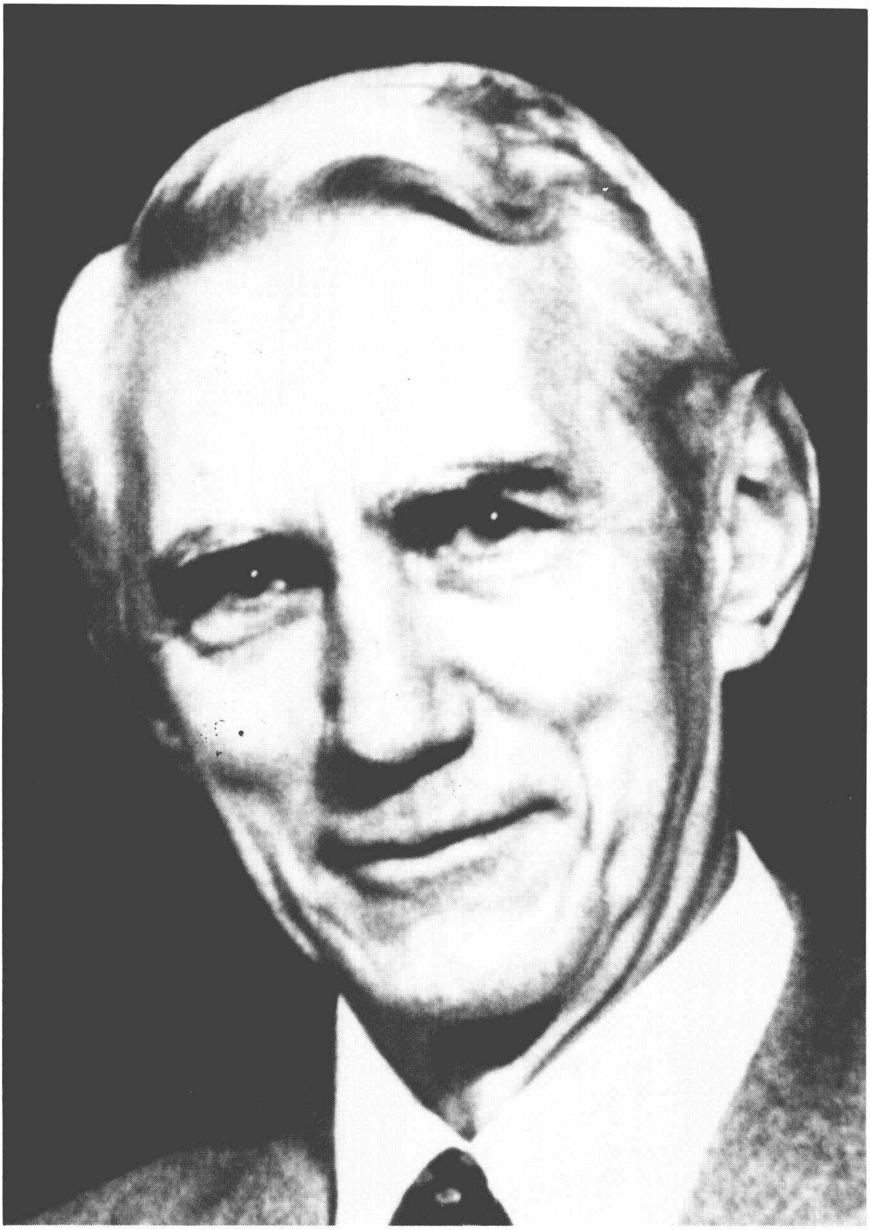
Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

ISBN 0-13-484668-0 025

Prentice-Hall International (UK) Limited, *London*
Prentice-Hall of Australia Pty. Limited, *Sydney*
Prentice-Hall Canada Inc., *Toronto*
Prentice-Hall Hispanoamericana, S.A., *Mexico*
Prentice-Hall of India Private Limited, *New Delhi*
Prentice-Hall of Japan, Inc., *Tokyo*
Prentice-Hall of Southeast Asia Pte. Ltd., *Singapore*
Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

INTRODUCTION TO INFORMATION THEORY



CLAUDE E. SHANNON
The Father of Information Theory
(Photo by Frank Ross, Cape May Courthouse, New Jersey)

PREFACE

This book has been developed out of a course in information theory that I have taught at Boston University during the past several years. The course is intended for beginning graduate students, as well as advanced undergraduates who desire a conceptual understanding of the foundations of information theory. It is thus designed to cover the major ideas involved in this important branch of modern science and to give the student an overview of the subject without getting nailed down in more advanced and sometimes less intuitive issues. The required background is probability theory at an undergraduate level. Although some basic concepts of probability are reviewed in the beginning, it is essential that the reader be familiar with the techniques of probability theory in general. A certain degree of mathematical maturity and capacity for abstract thinking on the part of the reader is also assumed.

Several excellent textbooks on information theory are available at this time, and it may seem inappropriate to write yet another one on the subject. I found it difficult, however, to find a text that addressed all the major concerns of the theory at the proper level for my intended audience. Some of the books were written for the Ph.D. level students and researchers and thus required a much higher knowledge base than introductory probability theory, while others, although appropriate for the less advanced student, were not comprehensive enough in their coverage. The present book is thus the outcome of my effort to bridge this gap; although it contains some of the more recent developments in information theory, which, to my knowledge, have not yet appeared in any textbook, it remains an introductory exposition of the basic concepts at the core. In organizing the book, I have been guided by the original paper of C. E. Shannon, "A Mathematical Theory of Communication" and by the monumental work of R. Gallager, *Information Theory and Reliable Communication*. I also benefited from the books *Information Theory and Coding*, by N. Abramson, *Information Theory*, by R. Ash, *The Theory of Information and Coding*, by R. McEliece, *Coding and Information Theory*, by R. Hamming, and *Rate Distortion Theory*, by T. Berger. The chapter on universal source coding is based on the original papers of B. Fitingof and the section on numerical computation of channel capacity is based on the papers of R. Blahut and S. Arimoto.

Although intuitive development of concepts has been my goal in this book, I have tried not to achieve this goal at the expense of mathematical rigor. All the results are derived from the basic principles, and every step in the derivation is carefully described. The problems are an important part of the book in that they either give a different perspective on a subject that has already been developed in the text or try to extend and generalize the concepts.

Chapter 1 is a review of the mathematical tools that are required for the understanding of the book. It is not necessary, however, to cover this material in the beginning; the reader can refer to it as the need arises. Chapter 2 concerns entropy and its properties; typical or likely sequences are introduced here, and a first hint of the usefulness of the concept in data-compression applications is given. Chapter 3 expands on the idea of source coding for data compression and introduces various properties of variable-length source codes. In Chapter 4 we look at source coding from a different point of view and introduce the idea of universal coding. This is a departure from the classical information theory in that a previous knowledge of the probabilistic structure of the source is no longer required for the construction of an optimum universal code. Chapters 5 and 6 are concerned with the discrete memoryless channel; after defining conditional entropy and mutual information and familiarizing the reader with some elementary discrete memoryless channels in Chapter 5, we prove the noisy channel theorem and its converse in Chapter 6. Chapter 7 is an elementary treatment of the rate-distortion theory. Here the concept of source coding with a fidelity criterion is introduced and the rate-distortion function is defined; the fundamental theorem of rate-distortion theory for the discrete memoryless source with single-letter fidelity criterion and its converse are then established.

I found it hard to completely ignore the problems associated with the more practical aspects of information theory, which, more often than not, are treated independently under the name of Coding Theory. I thus included some material on error-correcting codes, addressing the elementary aspects of linear codes for error correction in Chapter 8. Both block codes and convolutional codes are discussed here. It is hoped that through this chapter the reader will see the relation between theory and practice. I deliberately avoided any mention of cyclic codes in order to stay away from algebraic field theory; that would have been beyond the scope of this book. Finally, in Chapter 9, some advanced topics relating to stationary and ergodic sources and continuous channels are discussed. This can be viewed as an extension of some of the ideas developed in the previous chapters and is aimed at students who like to have a glimpse at what could be the subject of a second course in information theory.

I have used the book for a one-semester, four-credit course in information theory at Boston University. Chapters 1 through 6 are usually covered in the beginning and then, depending on the level of understanding and the interest of students, I have selected topics from the remaining three chapters. Chapter 8 on linear codes has always been a favorite of engineering students. If I were to teach the course in a school with the quarter system, I would probably teach the first six chapters in the first quarter and use some additional material with the last three chapters for the second quarter.

I would like to thank Professor Lev Levitin who inspired me and spent many of his precious hours discussing the more subtle points of information theory with me. He also honored me by agreeing to write an introductory chapter for this book. I would also like to thank Dean Louis Padulo of the College of Engineering who encouraged me in writing this book and provided me with the time needed for completing it. Thanks are due to Mr. Tim Bozik, the editor, for his support of the project. Finally, I would like to thank my wife, Annegret, without whose patience and encouragement this book would not have become a reality.

Acknowledgments

Problems 1.1, 2.1, 2.2, 2.6, 3.1, 3.4, 3.9, 3.10, 3.11, 3.12, 5.4, 5.5, 5.6, 6.4, 6.9, 7.4, 8.8, 8.9, 8.10, and 9.5 are from R. G. Gallagher, INFORMATION THEORY AND RELIABLE COMMUNICATION. New York: Wiley, 1968. Copyright © 1968 by John Wiley & Sons, Inc. Reprinted with permission.

Problems 2.4, 2.5, 2.9, 6.8, 8.2, 8.5, and 8.6 are from R. Ash, INFORMATION THEORY. New York: Interscience, 1965. Copyright © 1965 by John Wiley & Sons, Inc. Reprinted with permission.

Problems 2.11, 2.13, 3.15 (adapted), and 8.16 are from R. J. McEliece, THE THEORY OF INFORMATION AND CODING: A MATHEMATICAL FRAMEWORK FOR COMMUNICATION. New York: Cambridge, 1977. Copyright © 1977 by Cambridge University Press. Reprinted with permission.

INFORMATION THEORY AND THE MODERN SCIENTIFIC OUTLOOK

Lev B. Levitin

One of the greatest revolutions in the scientific world outlook in our century is the turn from Laplacian determinism to a probabilistic picture of nature. The development of statistical mechanics and (even in a more categorical way) of quantum theory has brought us to the appreciation of the fact that the world we live in is essentially probabilistic. A natural extension of this point of view is an understanding that our knowledge is of a probabilistic nature, too. Any information we obtain affects the probabilities of possible alternatives, rather than indicates uniquely one particular outcome (as "genuine" deterministic knowledge was supposed to do).

Therefore, it seems to be not just a sheer coincidence that information theory emerged after statistical and quantum mechanics had been developed, and that it shares with statistical physics the fundamental concept of entropy.

Mathematically, information theory is a branch of the theory of probabilities and stochastical processes. It has won its first victories by answering the most basic questions concerning information transmission over communication channels. In the past, communication engineers believed that the rate of information transmission over a noisy channel had to decline to zero, if we require the error probability to approach zero. Shannon was the first to show that the information-transmission rate can be kept constant for an arbitrarily small probability of error. Besides its technological importance, this result has a remarkable philosophical meaning. Information theory not only gives a quantitative measure of information common for both deterministic and probabilistic cases, but it also shows the qualitative equivalence of these two kinds of knowledge in the following sense: even if the input and the output of a communication channel are only statistically dependent, it is possible to transmit an amount of data that is arbitrarily close to the amount of information in the output about the input, with a vanishing error probability (i.e., in almost deterministic way).

The development of information theory is an excellent illustration of the statement that "nothing is more practical than a good theory." Indeed, at the time when the basic results of information theory related to communication channels had first been formulated, communication technology was not at all able to implement them in practice. It took a long time, about a quarter of a century, until the development of practical methods of data encoding and decoding together with new computer technology made it possible to process information in real time in accordance with the recommendations following from information theory, and thus to make the theoretical limits attainable. The results regarded once by some too "practical" people as "academic exercises" have become a matter of today's engineering practice. We are witnessing now the ongoing information revolution, the vigorous development of new means of information transmission, storage, and retrieval that may lead to profound changes in the nature of our society in ways that could hardly be envisioned by the most daring utopians. And information theory plays a crucial role in this development by providing not only a theoretical basis, but also a deep philosophical insight into new and challenging problems we have to encounter—today and tomorrow.

However, the importance and generality of information theory concepts and approaches go far beyond the area of communication engineering. The ideas of information theory were applied in a variety of diverse fields from physics to linguistics, from biology to computer science, and from psychology to chemistry and proved to be productive and innovative—of course, in the hands of those who knew how to handle them properly. Thus information theory has become not just another special branch, but an indispensable part of the modern scientific outlook. It should be borne in mind, however, that "seldom do more than a few of nature's secrets give way at one time," as the founder of information theory, Claude E. Shannon, observed, warning against immature attempts to use information theory just because it had become "something of a scientific bandwagon."

> A thorough understanding of the mathematical foundation and its communication application is surely a prerequisite to other applications. I personally believe that many of the concepts of information theory will prove useful in these other fields—and, indeed, some results are already quite promising—but the establishing of such applications is not a trivial matter of translating words to a new domain, but rather the slow tedious process of hypothesis and experimental verification [1].

Historically, the basic concepts of information theory, such as entropy, mutual information, equivocation, and redundancy were first introduced by Shannon in connection with cryptographic systems [2], rather than the usual communication channels. The modern development of cryptography added an important aspect of complexity to be taken into consideration. Information-theoretical analysis plays a significant role in the theory of computational and structural complexity and in the design of effective decision algorithms (e.g., [3]). (An elementary example of such a decision algorithm is that related to the famous counterfeit coin problem [4].)

classical [23, 24] and the quantum [25, 26] form. According to this principle, any information is represented by a certain ensemble of states of a physical system and associated with its deviation from the thermodynamic equilibrium. Thus the basic concepts of information theory can be defined on the basis of statistical physics, and a way is open to develop a consistent physical information theory. The subject of this theory is investigation of the physical nature of information transmission, storage, retrieval, and processing (so called physics of communication and physics of computation) (e.g., [27–29]). On the other hand, the information-theoretical approach has been applied to statistical physics [30–34]. Information theory shed a new light on the classical problems of Maxwell's demon, Gibbs' paradox [22, 34], and on the foundations of statistical physics in general. There is a hope that further development in this direction will eventually bridge the gap between physical and cybernetical descriptions of a complex system and will lead to formulation of a physical theory of high-organized systems, both artificial and natural (biological). The progress on this way is slow and difficult [35], but the goal is worth all the efforts.

Here we have to touch on another philosophical problem. Since the time of Newton (or even Democritus), scientists believed that the most fundamental laws, the most hidden secrets of nature, are those of elementary particles and elementary forces acting between them. Indeed, if everything that happens in the universe is no more than a combination of these elementary acts, then isn't the knowledge of laws that govern the interactions between elementary particles sufficient to describe any phenomenon in the world, to predict theoretically outcomes of any experiment? Today we have achieved incredible progress in discovering and describing the nature of elementary particles and interactions, and have learned that this knowledge is incredibly insufficient for the ambitious purpose of "explaining everything," of building an accomplished scientific picture of the world.

It seems that we know, indeed, how to derive the behavior of a system, even as large as stars and galaxies, from the "first principles," if we deal with a low-organized, chaotic system. But we find our knowledge almost irrelevant when we have to face a system at a higher level of organization (whatever it means, I should add, since we still lack even a good definition of "level of organization"). For instance, we have a wonderful theory of electromagnetic interactions that predicts the experimental results with an accuracy of 15 decimal digits, and we know that the processes in a human body are mostly electromagnetic, but this perfect theory tells us very little, if anything, about how our bodies function.

There has occurred another revolutionary shift in the minds of scientists: to become aware that the greatest secrets of nature, the hardest to discover — and the most important for us — are the laws of organization, the understanding of how a certain complex "combination" of particles and processes can emerge and persist as an organized system among the chaos of myriads of elementary

Another important area of the application of information theory to computer science is fault-tolerant computing. Although the first attempts in this direction were unsuccessful, more thorough investigations [5, 6, 7] have shown that it is possible to achieve arbitrarily small probability of error in data storage and computation at the expense of limited redundancy, exactly as in the case of communications.

Essential interconnections between information theory and statistics have been found [8] and new methods of statistical analysis based on information theory have been suggested [9].

A number of interesting attempts have been made to apply information theory in political economy and economics. For instance, the theory of optimal investment appears to be exactly parallel to the theory of optimal source coding [10].

Application of information theory to linguistics seem to be highly relevant (e.g., [11–15]). Indeed, a natural language gives us a remarkable example of a system used for generating long sequences of symbols (i.e., texts) that can be considered as realizations of a random process. But, in contrast with other random processes that exist in nature, this random process was developed, modified, and selected during a long period of evolution and "natural selection," being specially intended for meaningful communication between human beings. Information-theoretical studies of texts have revealed a number of significant linguistic features. For instance, they provide objective criteria for the characterization of different styles and different schools in poetry and prose, and even for identification of individual authors [16]. An illustration of the importance of the information-theoretical characteristics of a language is given by the fact (noted first by Shannon) that large crossword puzzles are only possible if the redundancy of a language does not exceed 50 percent (on the vocabulary level). If the entropy of a language were two times less than its actual value, poetry in its usual form (with rhymes and meters) would be impossible.

Application of information theory t experimental psychology made it possible to discover some remarkable facts related to sensory organs and neural systems [17]. It was found, for example, that the reaction time of a subject is a linear function of the amount of information contained in the stimulus [4, 18, 19]. Moreover, our sensory organs can be characterized by a certain information capacity, as engineering communication lines are.

Perhaps the most important and meaningful are interconnections between information theory and statistical physics. Long before information theory was founded, L. Boltzman and later L. Szilard [20] attributed an information meaning to the thermodynamical notion of entropy. On the other hand, D. Gabor [21] pointed out that "the communication theory should be considered as a branch of physics." In the classical work of L. Brillouin [22], a profound relationship between physical entropy and information was first formulated in a general form. Later the "entropy defect principle" was established in the quasiinteractions; how it can be formed and sustained by those interactions; how a "more organized" system can be built from "less organized" parts; why, when, and how such a system becomes able to display the properties of "high organization" such as self-regulation, self-organization, learning, adaptivity, expedient behavior, self-reproduction, and, eventually, intelligence. These are the crucial problems of life and death. Until we solve them, we remain, with all our knowledge and technology, helpless children in the cradle of nature.

Three geniuses of our time, J. von Neumann, N. Wiener, and C. E. Shannon, were the most influential in recognizing the problem with all its generality and consequence and bringing it to our attention [36–39]. And all of them stressed the importance of the concepts of entropy, information, and control for developing a general theory of high-organized systems, a new science for which Wiener coined the name cybernetics, but which is still waiting to be created. Today, information theory provides the only existing narrow bridge between the two different worlds of chaotic and organized systems. And I strongly believe that information theory will win its new triumphs by helping us on our way to the ultimate knowledge intellectual beings desire—the knowledge of ourselves.

References

1. C. E. Shannon, "The Bandwagon," *Trans. IRE,* IT-2, No. 1, 1956.
2. C. E. Shannon, "Communication Theory of Secrecy Systems," *BSTJ,* 28, No. 4, 1949.
3. C. R. P. Hartmann, and others, "Application of Information Theory to the Construction of Efficient Decision Trees," *IEEE Trans.,* IT-28, No. 4, 1982.
4. A. M. Yaglom and I. M. Yaglom, *Probability and Information.* Hingham, Mass.: D. Reidel Publishing Co., 1983.
5. S. Winograd and J. D. Cowan, *Reliable Computation in the Presence of Noise.* Cambridge, Mass.: MIT Press, 1963.
6. M. C. Taylor, "Reliable Information Storage in Memories Designed from Unreliable Components" and "Reliable Computation in Computing Systems Designed from Unreliable Components," *BSTJ,* 47, No. 10, 1968.
7. A. V. Kuznetsov, "Information Storage in a Memory Assembled from Unreliable Components," *Problems in Information Transmission,* 9, No. 3, 1973.
8. S. Kulback, *Information Theory and Statistics.* New York: Wiley, 1959.
9. S. Watanable, "Information-Theoretical Analysis of Multivariate Correlation," *IBM J. Res. Develop.,* 4, No. 1, 1960.
10. T. M. Cover, "Information Theory and Investment," *1985 IEEE Intern. Symp. Information Theory,* Brighton, England, June 1985. Also "An Algorithm for Maximizing Expected Log Investment," *IEEE Trans.,* IT-30, No. 2, 1984.
11. C. E. Shannon, "Prediction and Entropy of Printed English," *BSTJ,* 30, No. 1, 1951.

12. B. Mandelbrot, "An Informational Theory of the Statistical Structure of Language." In *Communication Theory,* W. Jackson, ed. New York: Academic Press, 1953.
13. I. M. Yaglom, R. L. Dobrushin, and A. M. Yaglom, "Information Theory and Linguistics," *Voprosy yazykoznaniya (Problems of Linguistics),* No. 1, 1960 (in Russian).
14. T. M. Cover and R. C. King, "A Convergent Gambling Estimate of the Entropy of English," *IEEE Trans.,* IT-24, No. 4, 1978.
15. L. Levitin and Z. Reingold, "Evaluation of the Entropy of a Language by an Improved Prediction Method," *Proc 11th IEEE Convention in Israel,* Tel-Aviv, 1979.
16. A. M. Kondratov, "Information Theory and Prosody (Entropy of the Rhythm of Russian Speech)," *Problemy Kibernetiki (Problems of Cybernetics),* 9, 1963 (in Russian).
17. H. Quastler ed., *Information Theory in Psychology.* New York: Free Press, 1955.
18. J. A. Leonard, "Choice Reaction Time Experiments and Information Theory." In *Information Theory,* C. Cherry ed., London, Eng.: Butterworth, 1961.
19. A. T. Welford, "The Measurement of Sensory-Motor Performance: Survey and Reappraisal of Twelve Years Progress," *Ergonomics,* 3, No. 3, 1960.
20. L. Szilard, "Uber die Entropieverminderung in einem Thermodynamischen System bei Eingriff Intelligenter Wesen," *Z. Physik,* 53, No. 5, 1929.
21. D. Gabor, "Communication Theory and Physics," *Phil. Mag.,* 41, No. 7, 1950.
22. L. Brillouin, *Science and Information Theory.* New York: Academic Press, 1956.
23. D. S. Lebedev and L. B. Levitin, "Information Transmission by Electromagnetic Field," *Information and Control,* 9, No. 1, 1966.
24. L. B. Levitin, "A Thermodynamic Characterization of Ideal Physical Information Channels," *Journal of Information and Optimization Sciences,* 2, No. 3, 1981.
25. L. B. Levitin, "On the Quantum Measure of the Amount of Information," *Proc. IV National Conf. Information Theory,* Tashkent, 1969 (in Russian).
26. L. B. Levitin, "The Amount of Information and the Quantum-Mechanical Irreversibility of Measurements," *Proc. II Intern. Symp. Information Theory,* Yerevan, 1971 (in Russian).
27. V. V. Mityugov, "Physical Grounds of Information Theory," *Sovietskoe Radio.,* Moscow, 1976 (in Russian).
28. R. Landauer, "Fundamental Physical Limitation of the Computational Process." In *Noise in Physical Systems,* P. H. E. Meijer, R. D. Mountain, R. J. Soulen, eds., NBS Spec. Pub. 614, Washington, D.C.: 1981.
29. L. B. Levitin, "Physical Limitations of Rate, Depth and Minimum Energy in Information Processing," *Intern. J. Theoretical Physics,* 2.1, No. 2/3, 1982.
30. E. T. Jaynes, "Information Theory and Statistical Mechanics," *Phys. Rev.,* Part I, 106, 620–630, 1957; Part II, 108, 171–190, 1959.
31. A. Katz, *Principles of Statistical Mechanics. The Information Theory Approach.* San Francisco: W. H. Freeman, 1967.
32. R. S. Ingarden, "Information Theory and Thermodynamics," Part I, Torun, Poland, 1974; Part II, Torun, Poland, 1975.
33. R. S. Ingarden, "Quantum Information Theory," Torun, Poland, 1975.

34. L. B. Levitin, "Quantum Amount of Information and Maximum Work," *Proc. 13th IUPAP Conf. Statistical Physics,* D. Cabib, D. G. Kuper, I. Riess, eds., Bristol, Eng.: A. Hilger, 1978.

35. H. P. Yockey, R. L. Platzman, and H. Quastler, eds., *Information Theory in Biology,* Elmsford, N.Y.: Pergamon Press, 1958.

36. N. Wiener, *Cybernetics, or Control and Communication in the Animal and Machine,* 2nd ed. Cambridge, Mass.: MIT Press, 1961.

37. J. von Neumann, "The General and Logical Theory of Automata," *The Hixon Symposium,* 1948, L. Jeffres, ed., Wiley, 1951.

38. C. E. Shannon, "Computers and Automata," *Proc. IRE,* 41, No. 10, 1953.

39. C. E. Shannon, "Von Neumann's Contributions to Automata Theory," *Bull. Amer. Math. Soc.,* 64, No. 2, 1958.

CONTENTS