Book Reviews

Elements of Information Theory, Thomas M. Cover and Joy A. Thomas (New York: John Wiley, 1991, 542 + xxii pages, U.S. $64.95)

Reviewer: Frans M. J. Willems, *Member, IEEE*

I. General Comments

Elements of Information Theory is probably the first book that covers the subject of Information Theory (in Shannon's sense) in full generality and nothing more. It does not focus on coding theorems only, nor is it a combination of subjects from both information theory and coding theory. Instead, it contains besides the classical results in information theory, i.e., the fundamental limits in communication, also the contributions of information theory made to intersecting research areas such as statistics, probability theory, economics, physics, computer science and mathematics. The discussions on this broad class of topics are perfectly balanced. Each subject gets the attention it deserves in the perspective of information theory.

Based on lecture notes, this book serves as an introductory text for senior and graduate level students of communication theory, statistics and computer science. It can also be very useful for professionals in these fields who feel the desire to broaden their view. It will certainly give all of its readers a firm foundation in the field of information theory.

The fact that the book is self-contained and rigorous does not make it boring. In their proofs the authors have consistently aimed at simplicity and elegance. It must have been a pleasant experience for them that nearly always this could be accomplished using information-theoretical arguments. This resulted in a coherent text which is exciting not only for readers not very familiar with information theory but also for specialists.

The book contains sixteen chapters including the introduction. All chapters are followed by a brief summary, a collection of problems and a set of historical notes. At the end of the book we find an overall summary, a biography and an index.

II. Specific Comments

The book starts out with the "Introduction and Preview." In their introduction, unlike many predecessors, the authors do not picture the classical communication situation in which a source generates data that is to be transmitted over a channel to the destination via an encoder and a decoder. Instead, they describe the relationship between information theory and the research areas it intersects with. This indicates that the authors consider finding fundamental performance limits in communication systems as not the only motivation for doing information theory. In the preview, the authors introduce quantities like entropy, mutual information, channel capacity, relative entropy and doubling rate in a brief and informal manner.

In the second chapter on "Entropy, Relative Entropy, and Mutual Information," these quantities are defined, and their properties and relations to another are investigated. The reader is also informed about useful inequalities like Jensen's inequality and the log-sum inequality. It is interesting to see that now already the authors can discuss the second law of thermodynamics in an information-theoretical perspective. This interesting digression is not contained in existing textbooks on information theory. The end of this chapter is devoted to the data processing theorem, but surprisingly enough also to more specific concepts such as sufficient statistics and Fano's inequality.

"The Asymptotic Equipartition Property (AEP)" is the title of the third chapter. By relating the cardinality of the ϵ-typical set (but also that of any other set with high enough probability) to the entropy of an independent and identically distributed generating random variable, this chapter provides a first justification for introducing entropy. The source coding theorem is a direct consequence of the AEP. A proof of this theorem in terms of the AEP including a converse result completes Chapter 3.

Chapter 4 ("Entropy Rates of a Stochastic Process") links the block definition to the conditional definition of entropy rate in a nonstandard but very natural way, using a property of the Césaro mean. Entropy rates of Markov chains and hidden Markov chains are considered, however, the authors do not mention the special class of "hidden" Markov chains for which the state (X_i) together with the output (Y_i) determine the new state (X_{i+1}). For such models the entropy rate can easily be found. The significance of entropy rate is demonstrated by a reference to the AEP for stationary ergodic processes (the Shannon–McMillan–Breiman theorem), which is proved in Chapter 15.

The next chapter ("Data Compression") contains the classical ideas on entropy coding for independent and identically distributed random variables. This chapter should be regarded as the constructive justification for the definition of entropy (rate). Apart from the well-known Huffman coding algorithm and the by now standard Shannon–Fano–Elias coding techniques the authors discuss the competitive optimality of the Shannon code and the generation of random variables from fair coins. Bounds on the expected code length are given for the case where the code is designed according to a distribution different from the source distribution. From here, it would only be a small step to universal source coding, yet the authors have decided to treat this topic in Chapter 12.

Chapter 6 on "Gambling and Data Compression" is an information-theoretical investigation of gambling techniques. The relation between optimal doubling rate and entropy is discussed and the increase in doubling rate due to side information is shown to be the mutual information between the side information and the outcome of the gamble. At the end of this chapter, upper bounds for the entropy of English text are determined from guessing and gambling estimates.

"Kolmogorov Complexity" is the topic of Chapter 7. It starts out with the definition of the Kolmogorov complexity of a string. The difference from information theory is that there are no probabilities involved now. Instead of minimizing the average codeword length over all strings as in information theory, we are interested here in the shortest-length program that makes a universal computer produce a given string. A first and crucial result is that Kolmogorov complexity is universal: The complexity cannot increase a lot if instead of a universal computer we use an alternative machine. Then, relations are derived between average Kolmogorov complexity and entropy of strings of independent and identically distributed random variables. An interesting section in this chapter is the one that is devoted to universal probability. This section and the next one on Kolmogorov

0018–9448/93$03.00

sufficient statistics will surely appeal to researchers in universal source coding theory. However, the goal of this chapter is much broader and the authors have made a good choice when they decided to include this topic in their book.

Chapter 8 is a very thorough discussion on "Channel Capacity." The authors give two definitions. The first one, the information capacity, is the maximum mutual information over all input distributions. The second one, the operational capacity, is the supremum over all achievable code rates. Both definitions are linked together in the noisy channel coding theorem whose achievability part is proved in terms of ϵ-typical sequences. The authors' approach to this central result in information theory, which was already outlined by Shannon, can be considered as the most natural one. Fano's inequality is used to obtain a weak converse result with message-error as the criterion. The chapter concludes with the converse for channels with feedback and the joint source-channel coding theorem. More involved topics such as error exponents, channels with memory, the strong converse, the symbol-error converse, the maximal code lemma, etc., are not covered.

The ninth chapter is entirely devoted to "Differential Entropy." The AEP for a continuous (-amplitude) random variable is stated. It is interesting to find out that instead of cardinality, volume is now the crucial parameter. Subsequently, the properties of entropy, relative entropy and mutual information of continuous ensembles are discussed. The chapter concludes with Massey's differential entropy bound on discrete-entropy.

Chapter 10 deals with "The Gaussian Channel." The first two sections treat the discrete-time case. Apart from a geometrical plausibility argument these sections contain a natural generalization of the channel coding theorem for discrete channels to the Gaussian case with an average power constraint on the input. Results obtained in Chapter 9 are frequently used. In the third section, band-limited channels are treated. This is the only section in which the channel is time-continuous. The sampling theorem and a suggestive argument lead to Shannon's formula. It is always a little bit disappointing that precisely this famous result cannot be proved rigorously, without making a reference to the papers of Slepian, Landau, and Pollak. In the last sections of this chapter, parallel Gaussian channels (water-filling argument), nonwhite Gaussian channels, and Gaussian channels with feedback are investigated.

The eleventh chapter on "Maximum Entropy and Spectral Estimation" contains the information-theoretical approach to the maximum entropy principle, culminating in Burg's maximum entropy theorem. The relevance of the maximum entropy principle is motivated only by pointing at the (Maxwell–Boltzmann) distribution of the velocities of molecules in a gas. Although they refer to the conditional limit theorem in Chapter 12, the authors fail to explain how this theorem justifies the minimum discrimination principle and consequently the maximum entropy principle.

Chapter 12 is a comprehensive discussion on "Information Theory and Statistics." Applying Csiszár and Körner's method of types, the law of large numbers is strengthened, a universal source coding method (allowing errors) is discussed, and Sanov's theorem and the conditional limit theorem is proved. After that the authors turn to hypothesis testing and treat the Neyman–Pearson lemma and Stein's lemma. They finally discuss Chernoff's Bayesian approach to hypothesis testing.

It comes more or less as a surprise to find a section on Lempel–Ziv coding here. Although Lempel–Ziv coding is today's most popular method for file compression, the (brilliant but lengthy) proof that it approaches entropy, does not give the reader a deeper level of understanding, unlike the other proofs in this book. Therefore, it is a bit strange that the authors have chosen the Lempel–Ziv algorithm to introduce the reader to the field of (noiseless) universal source coding instead of the classical, more fundamental, results. It is even stranger if one realizes that these fundamental results on (asymptotic) redundancy would fit perfectly in the purely information-theoretical framework of the book.

The remaining sections of the chapter on statistics deal with Fisher information and the Cràmer–Rao inequality. The reader is referred to the exercises to find out how Fisher-information relates to relative entropy.

Chapter 13 on "Rate Distortion Theory" begins with a section on quantization. Then, after the necessary definitions and some examples are given, the weak converse to the rate-distortion theorem is proved. Subsequently, a natural achievability proof in terms of typical sequences is given. The Gaussian source with squared-error distortion is considered briefly. The last sections of the chapter are devoted to the computation of channel capacity and rate distortion functions (Arimoto–Blahut algorithms).

The next chapter is an 86-page treatment of "Network Information Theory." First the authors introduce the Gaussian multiple access channel, broadcast channel, relay channel, interference channel and two-way channel. The discrete multiple access channel is investigated in more detail, just like the degraded broadcast channel and the physically degraded relay channel. In addition to that, the authors thoroughly discuss Slepian–Wolf encoding of correlated sources, source coding with side information and rate distortion with side information. The chapter ends with a section on general multiterminal networks for which it is demonstrated that the source-channel separation principle does not work and that feedback can increase the capacity.

Cover's idea of superposition coding has had a strong impact on multiuser information theory and is therefore discussed in detail in this chapter. Its counterpart, Gallager's converse for the degraded broadcast channel, is equally important. Therefore, it would be more appropriate to treat this striking result in the main text and not in an exercise. A final remark concerns the proof of Theorem 14.3.3. The authors seem to have overlooked the difficulty which is treated adequately in Appendix I of Slepian and Wolf's paper on the multiple access channel.

"Information Theory and the Stock Market" is the topic that is treated in Chapter 15. Log-optimal portfolio's are extensively studied. It is nice to find here a section on Kuhn–Tucker conditions and one with Shannon–McMillan–Breiman theorem before the book comes to an end.

Chapter 16 contains a collection of important "Inequalities in Information Theory" with their proofs. We mention only the entropy-power inequality (and the related Brunn–Minkowski inequality) and a series of inequalities for determinants.

The exercises at the end of each chapter are sufficiently instructive. Sometimes they lead to extensions of the theory which is developed in the main text. The level is adequate. A detailed solutions manual is available.

III. Conclusion

Although originally I was not happy with all the choices made by the authors, my final conclusion is that they have written an excellent book. It is extremely homogeneous both in choice of subjects and in their treatment. Its quality is very high, strong results are rigorously proved in a concise way. It gives the reader a deep understanding of the subject and, last but not least, it is easy to read.

Without any reservation, this book can be regarded as today's standard work on information theory. I am quite sure that we have to wait until the next millennium for a book that could possibly replace *Elements of Information Theory.*

F. M. J. Willems was a Research Assistant at the Catholic University of Louvain from 1979 to 1982. Since 1982, he has been a staff member at the Electrical Engineering Department of Eindhoven University of Technology. His research contributions are in the areas of multiuser information theory and noiseless source coding.

Dr. Willems received the Marconi Young Scientist Award in 1982. From 1988 to 1990, he served as Associate Editor for Shannon Theory for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Finite Fields for Computer Scientists and Engineers, Robert J. McEliece (Boston: Kluwer Academic Publishers, 1987, 207+x pages, U.S. $50.50)

Reviewer: Paul H. Siegel, *Senior Member, IEEE*

Finite fields play a fundamental role in the theory and application of error-control coding, multiuser communication, cryptography, and digital signal processing. Consequently, technical professionals in computer-related science and engineering often need to develop some facility in applying them. Most textbooks devoted to one or more of the aforementioned engineering disciplines include at most an appendix or possibly a chapter that summarizes the basic definitions and results pertaining to the arithmetic of finite fields as required for later discussion and applications. Proofs are sometimes given for major theorems, but typically the objective is to pave a direct route to the applications that form the heart of the book.

For the student who wishes to obtain a deeper understanding of the underlying mathematics that relate to the broader range of finite field applications, such a streamlined review—referred to in the Preface as a "Volkswagen" treatment—is unsatisfying.

The author's objective, then, is to offer to the student a "Cadillac" treatment of the subject, expanding upon the streamlined versions, while adding detailed discussion of several more advanced topics in finite fields that have close connections to applications, but are not found together, if at all, in other texts.

In my opinion, the text admirably achieves its goal, providing a self-contained, thorough, and mathematically rigorous presentation of the basics of finite field theory in a clear and interesting manner that should appeal to students and professionals in computer science, electrical engineering, and applied mathematics. Moreover, the book contains material that should be of interest even to those already familiar with the elements of finite field theory but with a taste for applications.

The book grew out of notes developed for a graduate-level class. The exposition retains the personal feel and the clarity of explanation one might find in a fine lecture series, qualities that characterize the author's other books [1], [2]. The author's enthusiastic approach to the subject, as well as his appealing sense of humor, are also in evidence, beginning with the book's dedication to Gus Solomon.

The book is replete with "worked" and, often, extended examples. These are very effectively used both to motivate new concepts and tools and to illustrate general results. They nicely complement the "mathematical" side of the book—the definitions, lemmas, theorems, and proofs—and emphasize the value (and satisfaction) of doing computations in finite fields.

The problems at the end of each section serve to reinforce the material, providing to the student opportunities to try his or her hand at calculations, as well as to extend some of the results in the chapter. There are also several (but not an excessive number of) "details left to the reader" from proofs or examples in the chapters.

Whether used as the basis for a complete course, as a source of material to supplement a class on coding theory, cryptography, or digital communications, or as a text for self-study, this volume is most worthwhile. The computer scientist will learn to enjoy the arithmetic of Galois fields as much as that of the integers; the engineer should end up feeling as comfortable with fields of nonzero characteristic as with the real and complex numbers; and for the mathematician who likes an interesting application, his or her appreciation for the richness of these algebraic objects will likely be enhanced.

Here is a synopsis of the book's contents.

Chapters 1–6 cover the fundamentals that would be relevant to a course on algebraic coding theory.

The first chapter is a very brief Prologue that gives the definition and a few elementary examples of fields, infinite and finite. Chapter 2 addresses Euclidean Domains and Euclid's algorithm (in its "extended" form) for finding greatest common divisors. Chapter 3 builds up the elements necessary to prove the basic theorem about unique factorization in Euclidean Domains. Finally, Chapter 4 demonstrates how to build fields from Euclidean domains. The material in these chapters is standard, but the presentation, including the discussion of numerous examples, makes palatable what might otherwise seem somewhat dry and abstract.

Chapter 5 provides a very lucid development of the abstract properties of finite fields, covering the concepts of order, primitive roots, minimal polynomials, conjugate elements, degree, and primitive polynomials. There is a nice discussion of Euler's "ϕ-function", which is used to prove the existence of primitive roots, as well as a treatment of Gauss's algorithm for finding primitive roots, a useful tool rarely found (explicitly) in books on coding theory. There is also a nice problem that relates the period of decimal expansions of rational numbers of the form k/p, for odd prime p not equal to 5, to the order of 10 in the finite field of size p.

Chapter 6 proves the existence and uniqueness of finite fields, making elegant use of some tools that may be familiar to computer scientists from other contexts: The Euler product technique and the Möbius inversion formula. Euler's method is used in a generating function approach to deriving an important identity involving I_d, the number of monic, irreducible polynomials of degree d over a finite field. Following an excellent treatment of the Möbius function $\mu(d)$ and the related inversion formula, the author applies the inversion method to derive useful expressions for $\phi(n)$ and I_d. Later, in Chapter 7, the inversion method also is used to investigate $V_d(x)$—the product of the monic, irreducible, degree d polynomials—and $\Phi_n(x)$—the nth (complex) cyclotomic polynomial. It therefore provides a satisfying link between the material in Chapters 5, 6, and 7. I might add that for myself, who as a graduate student dealt more with the Möbius strip than with the like-named function, this discussion, expanding upon that found in Berlekamp's book on algebraic coding theory, was most appealing.

Chapter 7 gives a thorough treatment of practical methods for factoring polynomials over finite fields. The factorization of $V_d(x)$ is reduced to the problem of factoring cyclotomic polynomials. Follow-

0018–9448/93$03.00

ing a discussion of the values of n for which $\Phi_n(x)$ is irreducible, the author addresses the factorization of the reducible cyclotomic polynomials, providing a very lucid discussion of Berlekamp's general polynomial factorization algorithm, complete with several detailed examples of its application.

Chapter 8 introduces the important trace function (and its multiplicative analogue, the norm function). The theoretical properties are presented at length, including Hilbert's constructive characterization of field elements with trace 0. These ideas are then used to solve the general quadratic equation over any finite field. The chapter concludes with the detailed development of a wonderful application of the trace function: Berlekamp's bit-serial Galois field multiplication circuit based upon dual bases.

Chapters 9, 10, and 11, building upon the previous chapters, give what must be called a "Rolls Royce" treatment of the theory of linear recurrences over finite fields, the theory of m-sequences, and cross-correlation properties of m-sequences.

Chapter 9 exploits the trace function to study the general solution to linear recurrences over finite fields. Results about periodicity, (projective) cyclic equivalence, and symbol distributions in sequences satisfying the recursion are developed for the case of irreducible characteristic polynomial. The situation for general characteristic polynomial is examined in a progressive manner, beginning with the case when the polynomial factors into distinct linear factors, then extending to the case when the distinct factors need not be linear, and finishing with a look at the problem of repeated roots when the characteristic polynomial is a power of a monomial.

Chapter 10 provides a very clear discussion of m-sequences—the binary sequences generated by linear recurrences over GF(2) with primitive characteristic polynomial. After a brief mention of the practical relevance of these sequences, generated by linear feedback shift registers and endowed with pseudo randomness properties, the author derives all of their standard and important features: the unique appearance of each nonzero m-gram; their pseudorandom run distributions; basic correlation and autocorrelation characteristics; the cycle-and-add property; the existence of canonical cyclic shifts; and the enumeration of all distinct m-sequences by means of decimation of a single one. For the student of algebraic coding, it would have been nice to see, perhaps in an example or exercise, a reference to the dual relationship between Hamming codes and m-sequences.

The final chapter, Chapter 11, treats in considerable depth the cross-correlation properties of m-sequences. It begins with a discussion of an idealized multiuser communications system and the practical significance of correlation functions. The exponential sum for the cross-correlation function of two m-sequences is derived, and the remainder of the chapter is devoted to developing the tools necessary to evaluate it. Several results related to quadratic forms over finite fields are then addressed, along the lines of the treatment by Dickson, including transformations to canonical forms and the Chevalley–Warning theorem on zero representations. The chapter concludes by proving the key theorem that gives the histogram of cross-correlation values for an m-sequence and its 2^e+1 decimations. Although the material in this chapter is mathematically challenging, it is very well presented, and the efforts of the practical-minded reader are rewarded in the end with an attractive "payoff": the application of the results to the analysis of the famous "Gold sequences" that were used in one of the early military spread spectrum radio systems and, more recently, in NASA's code-division multiple-access (CDMA) Tracking and Data Relay Satellite System (TDRSS).

The book ends with a very brief, annotated bibliography that lists five books dealing with the theory of finite fields. Included are three texts on algebra (many others could be added, of course), among which the most recent, *Finite Fields* by Lidl and Niederreiter (1983, Addison Wesley, publisher) is identified as an encyclopedic source providing a deeper and more comprehensive analysis of finite fields. Unfortunately, this volume is no longer in print, and is not readily available. A more recent publication by the same authors, entitled *An Introduction to Finite Fields and Their Applications* (1986, Cambridge University Press), might be a partial substitute. The remaining two books are apt choices: the classic texts on coding theory by Berlekamp and by Peterson and Weldon.

In future editions of this book, an expanded bibliography would certainly be an asset, providing to the student who has mastered the basic material some directions for further exploration. One can imagine a still modest but nevertheless useful guide to related texts and research literature on coding theory, shift-register sequences, multiuser and spread spectrum communications, cryptography, digital signal processing, and applied number theory.

The book was designed by the author and prepared in camera-ready form using a version of TEX. Physically, it is attractive and comfortable to read. However, there are some minor typographical errors; see, e.g., pp. 7, 47 (two), 55, 69, 81, 102, 110, 137, 144, 202 and Figs. 8.6 and 8.7. (Details are left to the reader.)

In summary, this book provides a pedagogically superb and mathematically substantial introduction to finite fields that will appeal to students, teachers, and professionals in computer science, engineering, and applied mathematics.

REFERENCES

[1] R. J. McEliece, *The Theory of Information and Coding, Encyclopedia of Mathematics and its Applications, vol. 3.* Reading, MA: Addison-Wesley, 1977.

[2] R. J. McEliece, R. Ash, and C. Ash, *Introduction to Discrete Mathematics.* New York: Random House, 1989.

P. H. Siegel is Manager of the Signal Processing and Coding project in the Storage Systems and Technology area at the Almaden Research Center, IBM Research Division, San Jose, CA. His main research interests are in coding, modulation, and detection, especially as applicable to digital data storage channels.