

Book Reviews

Coding for Wireless Channel—Ezio Biglieri (Springer Science, 2006, 428 pp.). *Reviewed by Joachim Hagenauer*

Coding, especially forward error correcting coding (FEC), has become a key element in wireless and mobile system design. No wonder that numerous books exist on this still-developing subject. Some of the books are overloaded with system or channel details, while others are quickly outdated because they concentrate on too specific subjects. In the early days of coding theory the focus was mostly on algebraic structures of codes and algebraic decoders rather than on probabilistic decoding algorithms, which in a wireless system should use soft-decisions and channel state information. Professor Biglieri avoids all these pitfalls in his book and concentrates on the essentials. As he states in the introduction his leitmotiv is “soft-decodable codes described through graphical structures (trellises and factor graphs).” This leitmotiv is visible throughout the book. He avoids overloaded notations with several super- and subscripts and circumvents lengthy uninspiring proofs following Kac’s statement: “A proof is what convinces a reasonable reader, a rigorous proof is one which convinces an unreasonable reader.” I believe here he is in line with Shannon. And furthermore, he introduces the reader slowly into the deeper understanding of a subject by starting each chapter with a tour d’horizon, an introduction or some preliminaries. Prof. Biglieri’s book could be further characterized in most chapters by a statement attributed to Einstein (?) “Everything should be made as simple as possible but not simpler.” He also guides the reader throughout his book by marking with an asterisk those chapters about which the reader should gain a full understanding.

The book starts with a brief introduction on fading channels giving all the essential parameters of such channel which will be further detailed later in Chapter 4. Chapter 3 is the classical treatment of coding in signal space introducing all the facts about bandwidth, SNR, and constellations. The essentials of trellis and convolutional codes are explained in Chapters 5 and 6. Although he concentrates on the basics, he does not forget to include tailbiting and thus introduces the link to block codes. Trellis coded modulation in the sense of Ungerboeck is treated in Chapter 7. Biglieri almost downplays his own 1998 contribution of bit-interleaved coded modulation that performs very close to capacity and can be elegantly implemented by doing turbo iterations between the demapper and the decoder. The up-to-date coding theory starts in the following chapter “Codes on graphs” introducing factor graphs and Tanner graphs, the sum-product algorithm and its derivatives such as the BCJR and the min-sum algorithm. Everything in this chapter is explained beautifully with a minimal but sufficient notation and small but enlightening examples. Chapter 9 contains an up-to-date description of LDPC codes and parallel concatenated codes (so-called turbo codes) including the ten Brink’s EXIT charts, an important tool for analyzing the behavior in the waterfall region. In the last chapter where the important subject of multiple antennas is treated, the book comes closer to a research monograph rather than an introductory course for master students. This is especially true for Chapter 10.6 on nonergodic Rayleigh fading and Chapter 10.7 on imperfect channel state information. This might go beyond an introductory course, but an instructor might nevertheless benefit

from the treatment of practical codes such as the Alamouti scheme and some of the BLAST architectures.

What I found very useful are the compact appendices with an introduction to information theory, the very compact proofs of various MIMO capacities and a summary of facts about random matrix theory.

The whole book is basically aimed as a textbook for graduate studies. Consequently the chapters are interleaved with examples and they are concluded with not too sophisticated problems in order to encourage students to study the subject by themselves.

Probably some readers will miss certain subjects in this book which is solely oriented toward the physical layer. Therefore adaptive schemes such as hybrid ARQ, scheduling and access control rightly have no place in this book. On the other hand a subject such as equalization could be easily incorporated in future editions of this valuable book.

Ezio Biglieri is not only an impressive scientist in the information technology community, but one of the foremost university teachers in this field. His textbooks were some of the earliest in the field such as the book with Benedetto almost 18 years ago which is still popular in newer revisions. His considerable teaching experience, not only in Torino, but in Princeton, Los Angeles, Munich, and Barcelona, have made him aware of the needs of a variety of students. His lectures are delivered in a careful, mathematically deep and almost artistic way. All this is condensed and reflected in this book, which I believe will be a great success.

Codes for Mass Data Storage Systems (Second Edition)—Kees H. Schouhamer Immink (Eindhoven, The Netherlands: Shannon Foundation Publishers, 2004, 349 + xiv pp). *Reviewed by Paul H. Siegel*

One can trace the origins of constrained coding to Shannon’s classic 1948 work [1], specifically the chapter on discrete noiseless channels. During the past 50 years, constrained codes have arguably found their most important application, as well as the primary impetus for continued theoretical and practical progress, in mass data storage systems, in the context of which they are often called recording codes.

Despite their inherent scientific elegance and considerable technological impact, recording codes have largely been ignored in textbooks on coding and information theory. In 1991, Kees Immink published *Coding Techniques for Digital Recorders*, the first book devoted to the subject [2]. It became the standard reference in the field. A significantly updated version of the text, entitled *Codes for Mass Data Storage Systems*, appeared in 1999.

The volume under review, released in late 2004, is a second edition of that 1999 text. It provides a very accessible introduction to the theory and, especially, the practice of constrained coding for digital recording. Immink’s enthusiasm for the subject matter is evident, as is his desire to “spread the gospel” to a wide audience. (As he points out in the preface,

Manuscript received September 12, 2006.

The reviewer is with Lehrstuhl für Nachrichtentechnik, Technische Universität München, D-80290 München, Germany (e-mail: hagenauer@tum.de).

Communicated by S. Verdú, Associate Editor for Book Reviews.

Digital Object Identifier 10.1109/TIT.2006.885538

Manuscript received September 8, 2006; revised September 8, 2006.

The reviewer is with the Center for Magnetic Recording Research, University of California, San Diego, La Jolla, CA 92093-0401 (e-mail: psiegel@ucsd.edu).

Communicated by S. Verdú, Associate Editor for Book Reviews.

Digital Object Identifier 10.1109/TIT.2006.885529

the publication of a Chinese translation of the 1999 edition makes his message accessible to a considerably larger flock.) It is comprehensive in its survey of families of constraints, their properties, and techniques for designing efficient encoders and decoders. Boasting an extensive bibliography—with over 350 references to journal papers, conference proceedings articles, books, expository surveys, and roughly 80 patent-related sources—the text provides an exceptional entrée to the constrained coding literature.

The author is one of the most inventive, prolific, and highly recognized contributors to the art of constrained coding. He has received accolades from the scientific and engineering communities, the consumer electronics and entertainment industries, and the Queen of The Netherlands. The impact of his coding innovations upon audio and video recording, in particular the CD and DVD standards, has even earned him, remarkably, an Emmy Award! It is hard to imagine a more authoritative source for the material presented in this book.

Echoing the preface to his earlier book, Immink plainly spells out his pedagogical aim, namely, “to show how theoretical principles are applied to design encoders and decoders.” He also emphasizes the text’s pragmatic bent, declaring that “the practitioners view will predominate.” Immink achieves his stated goal and, true to his promised perspective, he always keeps within clear focus the tradeoffs involving code efficiency, performance, and implementation complexity.

The text is suitable for use as the primary reference in a “special topics” course in mathematics or engineering at an advanced undergraduate or beginning graduate level. The author’s expansive expository style makes for very pleasant reading, and the multitude of carefully worked examples illustrate the concepts and techniques very effectively. A shortcoming is the absence of exercises to reinforce understanding and guided problems to develop topics not covered in the body of the text.

This book can also serve as an excellent supplemental text for a course whose focus is on a rigorous mathematical development of the theory of constrained coding, such as that found in [3] and [4]. Of course, for professionals interested in the fields of coding theory and data storage, it represents an essential technical reference.

Shannon Foundation Publishers has produced a handsome volume, so it is somewhat surprising to discover that the text suffers from a profusion of errors in typography and, to a lesser degree, in grammar, sentence structure, and word usage. Generally, these defects represent only a mild distraction, but there are instances where editorial oversights thwart the intentions of the author and have a detrimental impact upon the effectiveness of the text. (For example, on pages 280 and 292, references to “adjacency matrix” and “parity” that were evidently intended to be reflected in the index were lost due to obvious proofreading lapses.)

The text begins with a high-level, somewhat rambling introduction to the role of recording codes in data storage and the many factors involved in choosing an appropriate code for a specific application. Chapters 2 and 3 then present a nice review of entropy of Markov sources, capacity of noiseless channels, and spectral analysis of signals. These concepts and tools are used frequently in the following chapters. As is the case throughout the book, numerous examples are presented in elaborate detail.

The next two chapters together provide a thorough review of codes satisfying runlength-limited (RLL) constraints, which have played a dominant role in recording technology. Chapter 4 looks first at combinatorial, information-theoretic, and spectral properties of the important family of (d, k) constrained sequences. Then a host of other interesting RLL-type constraints are considered. There is a brief discussion of maximum-transition-run (MTR) and $(0, G/I)$ constraints developed for channels using partial-response equalization and maximum-likelihood sequence detection, so-called PRML technology. RLL sequences

over nonbinary alphabets, as well as two-dimensional constraints, are briefly addressed. Chapter 5 covers a panoply of code construction methods, both algorithmic and ad hoc, that have been used to construct efficient, low-complexity block codes and block-decodable finite-state codes for RLL-type constraints.

Chapter 6 is a vastly expanded treatment of enumerative encoding and decoding techniques that offer a practical alternative as code rates and codeword lengths increase to the point where table look-up and boolean logic implementations are no longer feasible. Issues related to error propagation are addressed, followed by a description of alternative configurations of constrained codes and error-correcting codes that can combat the error propagation problem in systems using very long codewords.

In Chapter 7, the author examines several code design methods that produce encoders that are sliding-block decodable, thereby guaranteeing limited decoder error propagation. The treatments of variable-length synchronous RLL codes, look-ahead encoder constructions, and the celebrated Adler–Coppersmith–Hassner (ACH) sliding-block code construction algorithm are taken essentially verbatim from the 1991 edition. The presentation of the ACH algorithm, in particular, could be improved. The most important point, from the perspective of the practitioner, is that the state-splitting technique can be applied directly to a graph without passing to a higher-order edge graph. This elementary observation can vastly simplify the code design process, as illustrated, without explicit mention, by the two constructions in the text of a rate $2/3, (d, k) = (0, 1)$ RLL code. Another fact worth clarifying is that the ACH algorithm is guaranteed to produce a sliding-block decodable encoder for finite-memory constraints such as those in the RLL family, but not necessarily for others, such as the spectral null constraints addressed in subsequent chapters. The chapter concludes with two new constructions of practical, high-rate, sliding-block decodable codes satisfying $d = 1$ and $d = 2$ constraints.

The next three chapters focus largely on the important class of codes whose spectral content at dc (zero frequency) is constrained. Chapter 8 addresses properties of maxentropic dc-balanced sequences, and techniques for evaluating low-frequency behavior are presented. Properties and performance of several specific, simple coding schemes are analyzed in considerable detail. Chapter 9 studies codes that improve low-frequency rejection by introduction of higher order spectral zeros at dc. Chapter 10 describes a more recently proposed design technique based upon “guided scrambling.” As the name suggests, this method has a pseudorandom aspect that offers the possibility of trading off implementation complexity against the degree of control over spectral control. Application to the construction of “weakly constrained” (d, k) codes is also considered.

Chapter 11 looks at the popular class of codes that combine RLL constraints with spectral constraints at dc. Readers will appreciate the detailed discussion of several important codes, particularly the EFM and EFMPlus codes that are used in CD and DVD technologies, respectively.

The text closes with a brief but helpful guide to further reading, with pointers to relevant special issues, book chapters, and sources of patent literature. The full-blown bibliography follows.

No one can question the sincerity of the author’s efforts to provide extensive coverage of a vast subject area. Still, several interesting topics that appeared in [2] have been omitted from the present volume, and some important recent developments receive scant attention. For example, one notes with some dismay the elimination of a fascinating introductory chapter that, in an informal and conceptual manner, demonstrated by reference to the Compact Disc some of the system-level considerations that enter into the process of recording code design. Sections on constraints with spectral nulls at frequencies other than zero, constrained codes that generate pilot tones, RLL sequences with

error-detecting capabilities, and minimum distance properties of higher order dc-constrained codes were likewise discarded.

The text would surely benefit from an expanded treatment of recording codes designed for PRML, the channel technology that caused a sea change in magnetic recording during the late 1990s and is now in ascendance in optical recording. The advent of PRML spawned new directions in constrained coding research, including distance-enhancing codes, high-rate parity coding, and architectures for combined detection and decoding, all of which have found wide use in commercial storage devices.

Kees Immink deserves praise for this latest update of his classic 1991 book. Comprehensive in scope, thoughtfully organized, and written with both the motivated student and the practicing engineer in mind, the volume should find a place on the bookshelf of anyone with an interest in coding techniques for data storage systems. Visit the Shannon Foundation Publishers website for information on ordering a copy. And once you're there, follow the link to Kees Immink's personal website where you can enjoy the photo of him receiving a well-deserved Emmy for his pioneering contributions to recording code technology.

REFERENCES

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul.–Oct. 1948.
- [2] K. A. S. Immink, *Coding Techniques for Digital Recorders*. Englewood Cliffs, NJ: Prentice-Hall, 1991.
- [3] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding*. Cambridge, U.K.: Cambridge Univ. Press, 1995, reprinted in 1999.
- [4] B. H. Marcus, R. M. Roth, and P. H. Siegel, "Constrained systems and coding for recording channels," in *Handbook of Coding Theory*, V. S. Pless and W. C. Huffman, Eds. Amsterdam, The Netherlands: Elsevier, 1998, ch. 20, see also <http://www.cs.technion.ac.il/~ronny/constrained.html>.

Statistical Methods in Bioinformatics (Second Edition)—Warren J. Ewens and Gregory R. Grant (New York: Springer-Verlag, 2005). *Reviewed by M. Vidyasagar*

It is an oft-repeated truism that the first half of the 21st century belongs to biology, just as the last half of the twentieth century belonged to microelectronics in its various forms (transistors, VLSI, microprocessors, systems on a chip, the worldwide web, etc.). As with all truisms, there is an element of truth, but also a lot of hype. The popular image of biology as predominantly an *experiment-based* science has now given way to the realization that biology is also now as much *information-based* as it is experiment-based. Rapid advances in experimental methods, such as high throughput DNA sequencing, microarrays (gene chips and now protein chips) and other related techniques, now permit biologists to generate data at a truly unprecedented pace. Coping with this flood of data and extracting useful "information" out of raw data is one of the major challenges that confronts today's biologists. Bioinformatics is the name given to a loosely defined body of mathematical and computational methods that seek to address this challenge.

"Bioinformatics" is thus the intellectual successor to earlier branches of science that had names such as theoretical biology, mathematical biology, and so on. But there is an important difference. Bioinformatics

is perhaps the first theoretical subject that addresses not only the underlying mathematics of a biological problem, but also the *computational complexity* of any solution algorithms developed.

Scientists know that science proceeds mostly through evolution, and rarely if ever by revolution. However, the attention of the popular media and of society at large is captured by discrete events in time, which are then said to usher in a "revolution." In this particular instance, the mapping of the human genome in February 2001 was considered to be a seminal event that would magically unlock all of the mysteries of life, if only we had sufficiently powerful computers at our disposal. The mapping of the human genome was rapidly followed by the publication of the genomes of several other organisms; so much so that Genbank, the repository of all-known genome sequences, holds (as of September 5, 2006) 1,796,667 sequences, totalling 80,369,977,826 base pairs in all. Similarly, the Protein Database (PDB) holds 38,620 structures of proteins as of September 5, 2006. It is interesting to note that both the Genbank and PDB seem to be doubling in size every six to nine months. The well-known (and often misquoted) Moore's law of VLSI states that the computational power of computers will double, and their cost will halve, every eighteen months. Thus, the conclusion is unavoidable: The glut in "raw data" being generated by various biology laboratories around the world *cannot be addressed* merely by Moore's law alone. *New theoretical methods* continue to be needed.

The promise of a genomic revolution being ushered in the mapping of the human genome also turned out to be premature. The readers of the *Information Theory Transactions* cannot be expected to be familiar with this point, but practically all known drugs are aimed at just about 500 "drug targets," that is, proteins produced by the human body. New drugs are supposed to function by inhibiting or enhancing the action of the target protein (depending on whether the action of the protein is beneficial or not). During the initial days after the human genome was mapped, computational methods turned up as many as 5000 potential drug targets—roughly an increase by a factor of ten! However, *validating* these targets required a great deal of expensive and time-consuming laboratory work, thus belying to a large extent the supposed benefits of post-genomic biology.

It would however be wrong to swing to the opposite end of the spectrum and declare that computational biology is not important. Problems of determining which parts of a genome correspond to genes (and which parts are "junk DNA"), predicting the three-dimensional structure of a protein, determining the function of a gene or a protein by assessing its similarity to other known genes and proteins, are all computationally intensive problems whose solution continues to be relevant to biology, and its application to drug discovery which is the biggest commercial application of such work.

Thus, to repeat, bioinformatics is as relevant a topic today as it was during the late 1990s and early 2000s. Interestingly, much of the theoretical work and algorithmic development in this field took place as early as the 1980s, when computational power was strictly limited, and networking of computers practically unknown.

The book under review represents a serious attempt to gather, in a single book, many if not most of the popular *statistical* methods used in bioinformatics. The book is the second edition of an earlier book with the same title that was published in 2001. As there is no Preface to the second edition, it is difficult to determine the extent of the changes made from the first edition. At a superficial glance, it appears to me that about 90% of the material is the same. Thus the review can be applied practically in tact to the first edition as well.

In the preface (to the first edition) dated February 2001, the authors say "We take *bioinformatics* to mean the emerging field of science growing from the application of mathematics, statistics, and information technology, including computers and the theory surrounding them, to the study of very large biological, and particularly genetic, data sets."

Manuscript received September 12, 2006.

The reviewer is with Tata Consulting Services, Hyderabad 500081, India (e-mail: sagar@atc.tes.co.in).

Communicated by S. Verdú, Associate Editor for Book Reviews.

Digital Object Identifier 10.1109/TIT.2006.885536