Codes for Mass Data Storage Systems (Second Edition) – Kees A. 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, 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 non-binary 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 1990's 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. Schouhamer Immink, *Coding Techniques for Digital Recorders.* New York: Prentice Hall, 1991.

[3] D. Lind and B. Marcus, *An Introduction to Symbolic Dynamics and Coding.* New Jersey: Cambridge University Press, 1995, reprinted 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.), ch. 20, Elsevier Science, 1998. (See, also, http://www.cs.technion.ac.il/∼ronny/constrained.html)