FUNDAMENTALS OF CODES, GRAPHS, AND ITERATIVE DECODING

THE KLUWER INTERNATIONAL SERIES
IN ENGINEERING AND COMPUTER SCIENCE

FUNDAMENTALS OF CODES, GRAPHS, AND ITERATIVE DECODING

Stephen B. Wicker
Cornell University,
Ithaca, NY, U.S.A.

Saejoon Kim
Korea Institute for Advanced Study,
Seoul, Korea

KLUWER ACADEMIC PUBLISHERS
NEW YORK, BOSTON, DORDRECHT, LONDON, MOSCOW

eBook ISBN: 0-306-47794-7
Print ISBN: 1-4020-7264-3

Created in the United States of America

Visit Kluwer Online at: http://kluweronline.com
and Kluwer's eBookstore at: http://ebooks.kluweronline.com

Contents

List of Figures

List of Tables

We dedicate this book, with love and thanks, to our parents:

Louise Zeller Wicker,
Richard Fenton Wicker, Jr.,
Jung-ja Choi Kim,
and
Gu-ung Kim.

Preface

As with all scientific and engineering endeavors, the fifty-year history of error control coding can best be characterized as a mass of incremental research interrupted by occasional great leaps forward. The punctuated equilibrium model[1], developed by Niles Eldridge and the late Stephen Jay Gould to describe the process of natural evolution, is an equally apt model for the development of error control coding. Eldridge and Gould felt that the old models of speciation theory could not predict gradual transitions over millions of years, and that instead, the evolution of species was best characterized by the sudden appearance of new species, occasional eruptions in what would otherwise be an unbroken landscape of species stability. So we have found it in coding theory, but with one significant difference. Coding theorists have always had a well-defined goal – the performance bound set by Shannon's Noisy Channel Coding Theorem[2] – as well as useful metrics for assessing our progress toward that goal – signal to noise ratios, bit error rates, and computational complexity. Given the goal and metrics, we can safely state in the Summer of 2002 that error control coding has entered a fundamentally new and different stage in its development.

Looking back, several global tendencies can be seen to have been most helpful in punctuating the equilibrium and getting us where we are now. The most obvious lies in the exploitation of structure – the interpretation of error control codes in light of combinatorial, algebraic, and probabilistic models has allowed for the application of increasingly powerful design

---

[1]N. Elridge and S. J. Gould, "Punctuated Equilibria: An Alternative to Phyletic Gradualism," in *Models in Paleobiology,* T. J. M. Schopf (ed), San Francisco: Freeman Cooper, pp. 82 - 115, 1972.

[2]C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal,* Volume 27, pp. 379 - 423 and pp. 623 - 656, 1948.

tools. Slightly less obvious has been our dependence on and exploitation of other fields. The recognition of structure has allowed for the identification of connections to others fields of mathematics and engineering, and the subsequent looting of the other fields' respective toolboxes. Finally, coding theorists have shown great prescience over the years, a prescience so extreme that we often fail to appreciate our colleagues' results for several decades. Fortunately many of our colleagues have very good memories, and we can thus incorporate and build on results that were initially given short shrift.

To put this current volume in context, a quick review of that past fifty years will prove helpful.

The first significant error control codes – those due to Hamming[3] and Golay[4] – were based on linear algebra and some relatively simple combinatorial techniques. The early error control codes were developed as linear block codes – subspaces of vector spaces over finite fields. These subspaces have dual spaces, whose bases can be interpreted as explicit parity relations among the coordinates of the codewords that constitute the code. The creation and exploitation of parity relations is a major theme in this book, and the creative and intelligent inculcation of parity relations is clearly the key to the recent developments in error control. In the 1950's, however, the principle metric for the quality of an error control code was minimum distance, with the Hamming bound serving as the performance limit. This conflation of the sphere packing and error control problems was limiting, and the discovery of all of the perfect codes by 1950 (a result unknown at the time[5]) left little room for new results through the combinatorial approach.

Reed took the first step away from the combinatorial approach with his recognition that Muller's application of Boolean algebra to switching circuits could be re-interpreted as a construction technique for error control codes. Reed saw that by viewing codewords as truth tables of Boolean functions, various results in Euclidean geometry and Boolean algebra could be used as design tools for error control codes[6]. The resulting Reed-Muller codes were a significant step beyond the earlier work

[3]R. W. Hamming, "Error Detecting and Error Correcting Codes", *Bell System Technical Journal,* Volume 29, pp. 147 – 160, 1950.

[4]M. J. E. Golay, "Notes on Digital Coding," *Proceedings of the IRE,* Volume 37, pg. 657, June 1949.

[5]A. Tietäväinen, "On the Nonexistence of Perfect Codes over Finite Fields," *SIAM Journal of Applied Mathematics,* Volume 24, pp. 88 - 96, 1973.

[6]I. S. Reed, "A Class of Multiple-Error-Correcting Codes and a Decoding Scheme," *IEEE Transactions on Information Theory,* Volume 4, pp. 38 – 49, September 1954. See also D. E. Muller, "Application of Boolean Algebra to Switching Circuit Design," *IEEE Transactions on Computers,* Volume 3, pp. 6 - 12, September 1954.

of Hamming and Golay, but remained relatively weak in comparison to what was to come.

The next major step beyond the combinatorial approach was made by Reed and Solomon in 1960[7]. By interpreting the coordinates of codewords as the coefficients of polynomials, they opened up a world of structure that allowed for far more powerful and elegant codes. The concurrent development of the theory of cyclic codes by Prange, Bose, Ray-Chaudhuri, Hocquenghem and others led to an interpretation of Reed-Solomon, BCH and in general all cyclic codes as rings of polynomials over finite fields[8]. This led to several deep results in algebraic coding theory in the 1960's, culminating in Berlekamp's decoding algorithm for Reed-Solomon codes in 1967[9].

At virtually the same time that Reed was trying to move beyond Reed-Muller codes, Elias was focusing on the the use of shift registers for creating parity relations in an information stream[10]. The resulting convolutional encoders were a significant advance in that they constituted a means for recursively introducing parity constraints across an arbitrarily large information stream, and were thus the first significant step toward the codes that would provide the performance promised by Shannon's work. The subsequent development of sequential decoders by Fano[11] and others was even more promising, in hindsight, in that it constituted a suboptimal, yet efficient approach to decoding convolutional codes with extremely long constraint lengths.

The sequential decoding of convolutional codes gave way to Viterbi decoding[12] in the late 1960's. The "optimal," maximum-likelihood approach to decoding represented by the Viterbi algorithm works extremely

---

[7]I. S. Reed and G. Solomon, "Polynomial Codes over Certain Finite Fields," *SIAM Journal on Applied Mathematics,* Volume 8, pp.300 – 304, 1960. See also S. B. Wicker and V. K. Bhargava, (editors) , *Reed-Solomon Codes and Their Applications,* Piscataway: IEEE Press, 1994.

[8]See, for example, E. Prange, "Some Cyclic Error-Correcting Codes with Simple Decoding Algorithms," *Air Force Cambridge Research Center-TN-58-156,* Cambridge, Mass., April, 1958, R. C. Bose and D. K. Ray-Chaudhuri, "On a Class of Error Correcting Binary Group Codes," *Information and Control,* Volume 3, pp. 68 - 79, March 1960, A. Hocquenghem, "Codes Correcteurs d'Erreurs," *Chiffres,* Volume 2, pp. 147 - 156, 1959, and D. Gorenstein and N. Zierler, "A Class of Error Correcting Codes in p^m Symbols," *Journal of the Society of Industrial and Applied Mathematics,* Volume 9, pp. 207 - 214, June 1961.

[9]E. Berlekamp, "Nonbinary BCH Decoding," presented at the 1967 International Symposium on Information Theory, San Remo, Italy. See also E. R. Berlekamp, *Algebraic Coding Theory,* New York: McGraw-Hill, 1968. (Revised edition, Laguna Hills: Aegean Park Press, 1984.)

[10]P. Elias, "Coding for Noisy Channels," *IRE Conv. Record,* Part 4, pp. 37 – 47, 1955.

[11]R. M. Fano, "A Heuristic Discussion of Probabilistic Decoding," *IEEE Transactions on Information Theory,* IT-9, pp. 64 - 74, April 1963.

[12]A. J. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm," *IEEE Transactions on Information Theory,* IT-13, pp. 260 - 269, April 1967.

well for many applications, including deep space telecommunications[13], and has broader applications in operations research[14]. As the complexity of maximum-likelihood decoders increases exponentially with the constraint length of the codes in use (and thus with the extent of the parity relations across the information stream), the performance of convolutional encoding with Viterbi decoding is limited by the amount of computational power available to the decoder. This is not to say, however, that efforts were not made. Increasingly immense Viterbi decoders were developed for deep space telecommunications through the early 1990's[15], ending only when the advent of turbo error control offered a significantly better, less complex alternative.

As convolutional encoders impart parity relations through the use of a delay line, the encoding process can be described in terms of a sequence of transitions from one encoder state to another. It follows that the resulting code can be represented as a trellis[16]. In 1974 Bahl, Cocke, Jelinek, and Raviv showed that any linear block code can be represented as a trellis, making them amenable to "optimal" soft-decision decoding[17]. Bahl et al. also showed that a particularly powerful algorithm for solving the inference problem in hidden Markov models can be applied to the decoding of any code that can be represented as a trellis. This algorithm, known as the Baum-Welch (BW) algorithm in the statistics community, was developed in a classified research environment in the early 1960's. It was described in a series of articles[18] in the late 1960's, and was subsequently applied and duly referenced by Bahl et al. in 1974. The BW algorithm was a progenitor of the class of Expectation-Maximization

[13]S. B. Wicker, "Deep Space Applications," *Handbook of Coding Theory,* (Vera Pless and William Cary Huffman, ed.), Amsterdam: Elsevier, 1998.

[14]The structure of the Viterbi algorithm has its roots in earlier optimization algorithms. See, for example, G. J. Minty, "A Comment on the Shortest Route Problem," *Operations Research,* Volume 5, p. 724, October 1957.

[15]O. Collins, "The Subtleties and Intricacies of Building a Constraint Length 15 Convolutional Decoder," *IEEE Transactions on Communications,* Volume 40, Number 12, pp. 1810–1819, December 1992. See also S. B. Wicker, "Deep Space Applications," *Handbook of Coding Theory,* (Vera Pless and William Cary Huffman, ed.), Amsterdam: Elsevier, 1998.

[16]See, for example, G. D. Forney, Jr., "Convolutional Codes I: Algebraic Structure," *IEEE Transactions on Information Theory,* IT-16, pp. 720 – 738, November 1970 and G. D. Forney, Jr., "Convolutional Codes II: Maximum Likelihood Decoding," *Information and Control,* Volume 25, pp. 222 – 266, July 1974.

[17]L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory,* IT-20:284–287, 1974.

[18]L.E. Baum and T. Petrie, "Probabilistic functions of finite state Markov chains," *Ann. Math. Stat.* 37:1554–1563, 1966, L.E. Baum and G.R. Sell, *Growth transformations for functions on manifolds,"* Pac. J. Math. *27(2):211–227, 1968, and L. E. Baum, T. Petrie, G. Soules and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains"* Ann. Math. Stat. *41:164–171, 1970.*

(EM) algorithms, and remains a topic of research in statistics. When applied to convolutional codes, the BW algorithm[19] provides a means for iteratively generating maximum likelihood estimates of the information represented by the received codeword. This stands in contrast to the Viterbi algorithm, which provides a maximum likelihood estimate of the transmitted codeword, not the information used to generate that codeword. This distinction would prove crucial in the performance of turbo decoders.

The equilibrium that was coding theory was punctuated in the 1990's by two interconnected events. Over the past ten years these events have launched the field into a significantly different stage in its development. The first event was the invention of turbo error control by Berrou, Glavieux, and Thitimajshima[20]. What is now called "Turbo Coding" consists of two discrete elements: parallel concatenated encoding and iterative (turbo) decoding. Parallel concatenation is a clever means for generating very complicated codes using several "component" encoders. Turbo decoding exploits this component encoder substructure by using separate, relatively simple BW decoders to develop separate estimates of the transmitted information. A consensus is then obtained, when possible, by iterating between the estimates[21]. Turbo error control brought coding theory within a few tenths of a decibel of the Shannon limit. The only problem was that it was not at all clear how or why it worked.

The second critical event was the recognition by the teams of McEliece, MacKay, and Cheng and Kschischang and Frey that turbo decoding was actually an instance of belief propagation in a graphical model[22]. This was a critical discovery in that it freed research in coding theory from the specific, and in places ad hoc elements of turbo coding, and brought the focus to bear on the more general problems of algorithms on graphs. It is now clear that the best error control systems are to be developed through the systematic, recursive generation of parity connections and

[19]The portion of the BW algorithm relevant to the decoding of convolutional codes is often referred to in the coding community as the BCJR algorithm.

[20]C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turb Codes," *Proceedings of the 1993 International Conference on Communications,* 1064–1070, 1993.

[21] See, for example, C. Heegard, and S. B. Wicker, *Turbo Coding,* Boston: Kluwer Academic Press, 1999.

[22]R. J. McEliece, D. J. C. MacKay and J. -F. Cheng, "Turbo Decoding as an Instance of Pearl's 'Belief Propagation' Algorithm," *IEEE Journal on Selected Areas in Commun.,* vol. 16, pp. 140-152, Feb. 1998 and F.R. Kschischang and B.J. Frey, "Iterative Decoding of Compound Codes by Probability Propagation in Graphical Models," *IEEE Journal on Selected Areas in Commun.,* vol. 16, pp. 219-230, Feb. 1998.

iterative, suboptimal decoding that reduces complexity by exploiting repeated local structure across the code. The key tools are graph theory and probabilistic methods based on graphs.

With these ideas firmly in place, the systematic despoiling of graph theory, bayesian belief propagation, and coding theory's own deep archives began. The forty-year-old work of Gallager on low density parity check codes[23] was finally recognized for being a deeply prophetic work. In his 1961 Ph.D. thesis, Gallager had developed techniques for recursively generating very long codes, and then proposed several suboptimal techniques for decoding these codes. Given a graph-theoretic interpretation, this was exactly the direction that needed to be taken to realize the promise of Shannon. More recent work by Tanner[24] was also dusted off and recognized as critical to the construction of good long codes with local structure that lends itself to iterative decoding.

Sipser and Spielman showed a quick appreciation for the work of Gallager and Tanner, extending it to a class of low-density parity-check codes called *expander codes*[25] in 1996. MacKay then showed in 1999[26] that low-density parity-check codes can achieve the Shannon limit when decoded using a maximum-likelihood decoding algorithm. It was then only a matter of time, with Davey, MacKay, Luby, Mitzenmacher, Shokrollahi, Spielman, Richardson, Urbanke, and others trading results[27] in a last dash to the Shannon limit.

Fifty years of learning how to design good codes can now be reduced to a single sentence: good codes have high degrees of local connectivity, but must have simply structural descriptions to facilitate iterative decoding. This book is an explanation of how to introduce local connectivity, and how to exploit simple structural descriptions. Chapter 1 provides an overview of Shannon theory and the basic tools of complexity theory, communication theory, and bounds on code construction. Chapters 2 – 4 provide an overview of "classical" error control coding,

---

[23]R.G. Gallager, *Low-Density Parity-Check Codes.* The M.I.T. Press, Cambridge, MA, 1963.
[24]R. M. Tanner, "A recursive approach to low complexity codes," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 533-547, Sept. 1981.
[25]M. Sipser and D.A. Spielman, "Expander Codes," *IEEE Trans. Inform. Theory*, vol. IT-42, pp. 1710-1722, Nov. 1996.
[26]D.J.C. MacKay, "Good Error-Correcting Codes based on Very Sparse Matrices," *IEEE Trans. Inform. Theory*, vol. 45, pp. 399-431, Mar. 1999.
[27]See, for example, M.G. Luby, M. Mitzenmacher, M.A. Shokrollahi and D.A. Spielman, "Improved Low-Density Parity-Check Codes Using Irregular Graphs and Belief Propagation," *Proc. 1998 IEEE Int. Symp. on Inform. Theory,* Boston, USA, August 16-21, 1998, M.C. Davey and D.J.C. MacKay, "Low-Density Parity Check Codes over $GF(q)$," *IEEE Commun. Letters,* vol. 2., no. 6, June 1998, an T. Richardson, M.A. Shokrollahi and R. Urbanke, "Design of Provably Good Low-Density Parity-Check Codes," submitted to *IEEE Trans. Inform. Theory.*

with an introduction to abstract algebra, and block and convolutional codes. Chapter 5 – 9 then proceed to systematically develop the key research results of the 1990's and early 2000's with an introduction to graph theory, followed by chapters on algorithms on graphs, turbo error control, low density parity check codes, and low density generator codes.

This book is intended as a synthesis of recent research results with a recognition of where these results fit into the bigger picture of error control coding.

The authors have been very fortunate to have the active cooperation of several of those who have made key contributions in the last few years, including Alexander Barg, Sae-Young Chung, Venkat Guruswami, Amin Shokrollahi, and Yair Weiss. Special thanks go to Alexander Barg and Amin Shokrollahi for carefully reading early versions of the book and providing us with invaluable help and suggestions.

The authors would like to thank the National Science Foundation and the Defense Advanced Research Projects Agency of the United States, as well as Samsung Electronics Co. and the Korea Institute for Advanced Study for their long term support for our efforts.

The authors extend their thanks to their editor, Jennifer Evans, for her support, patience, and good humor.

The first author would also like to thank Toby Berger, Terrence Fine, Robert Thomas, and James Thorp for their able mentoring over the past few years. Their efforts have been greatly appreciated, though they should not be held responsible for the subsequent results.

The first author would like to extend warm thanks to the twenty-five doctoral students he has supervised over the past fifteen years. He is grateful for the intellectual stimulation that they have provided, and for their boundless energy, dedication, and patience. And of course, the first author is very grateful to have had the opportunity to "supervise" the research of the second author, Dr. Saejoon Kim. Graduate students are certainly one, if not the only unmixed blessing of an academic career.

Finally, the authors are forever indebted to their parents for gifts too numerous to mention. We wish that we could have dedicated a more readable work to them, but this is the best we could do. Thank you.