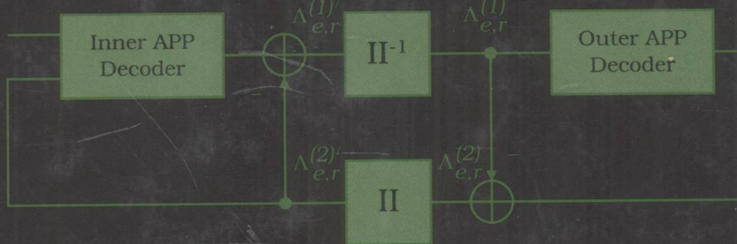
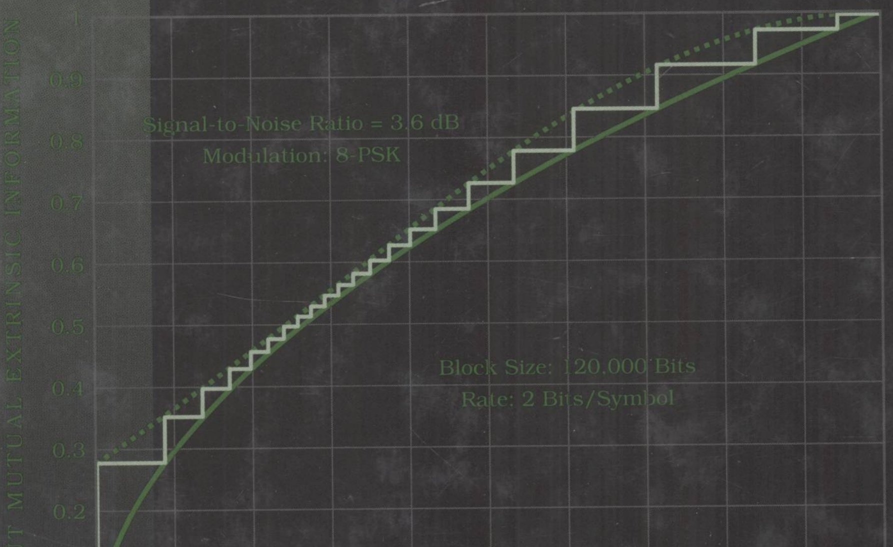
Trellis and Turbo Coding

CHRISTIAN B. SCHLEGEL
LANCE C. PÉREZ





IEEE SERIES ON DIGITAL & MOBILE COMMUNICATION
John B. Anderson, *Series Editor*

TRELLIS AND TURBO CODING

Christian B. Schlegel

Lance C. Pérez

IEEE SERIES ON DIGITAL & MOBILE COMMUNICATION

John B. Anderson, *Series Editor*

IEEE

IEEE PRESS

WILEY-INTERSCIENCE

A JOHN WILEY & SONS, INC., PUBLICATION

IEEE Press
445 Hoes Lane
Piscataway, NJ 08855

Library of Congress Cataloging-in-Publication Data

Schlegel, Christian.
Trellis and Turbo coding / Christian B. Schlegel, Lance C. Pérez.
p. cm.
Includes bibliographical references and index.
ISBN 0-471-22755-2 (cloth)
1. Error-correcting codes (Information theory) 2. Trellis-coded modulation. 3. Digital communications. 4. Coding theory. I. Perez, Lance. II. Title.

TK5102.96.S33 2004
621.382–dc22

2003063355

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

TRELLIS AND TURBO CODING

IEEE Press Series on Digital & Mobile Communication

The IEEE Press Digital & Mobile Communication Series is written for research and development engineers and graduate students in communication engineering. The burgeoning wireless and personal communication fields receive special emphasis. Books are of two types—graduate texts and the latest monographs about theory and practice.

Books in the IEEE Press Series on Digital & Mobile Communication

John B. Anderson, *Digital Transmission Engineering*
Rolf Johannesson and Kamil Sh. Zigangirov, *Fundamentals of Convolutional Coding*
Raj Pandya, *Mobile and Personal Communication Systems and Services*
Lajos Hanzo, Peter J. Cherriman, and Jürgen Streit, *Wireless Video Communications: Second to Third Generation Systems and Beyond*
Lajos Hanzo, F. Clare A. Somerville and Jason P. Woodard, *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*
Mansoor Shafi, Shigeaki Ogose and Takeshi Hattori (Editors), *Wireless Communications in the 21st Century*
Raj Pandya, *Introduction to WLLs: Application and Deployment for Fixed and Broadband Services*

To our children to whom belongs the new world:

Christina, Vincent, Isabelle,
Patrick, and Lindsey

PREFACE

In 1997 I completed my book *Trellis Coding*, which contained a single chapter on turbo coding. Since then, turbo coding and its associated iterative decoding method, the turbo principle, has swept through the communications society like a wildfire and established itself as the error coding method of choice. Its ease of implementation and its phenomenal performance, which in many cases pushes right up to Shannon's theoretical limits, lets one speculate that turbo coding is *the* right way to encode and decode digital data. Any potentially new method would have to measure itself against turbo coding and thus have virtually no chance of providing further gains. In recent years the implementation of turbo codes has also made great strides, and practically any decoding speeds can now be achieved by suitable parallelization of the decoding algorithms, leaving data storage and memory access as the limiting functions. Storage, however, is always necessary according to the requirements of Shannon's theory to process large blocks of data in order to approach a channel's capacity limit.

The turbo principle has found application even outside the narrow field of error control coding, in multiple access channel communications, in signalling over channels with intersymbol or interchannel interference, and, more recently, in source coding. In all these applications the principle has established itself quickly as a very powerful processing method, leading to the design of receivers which are far superior to conventional methods.

Since the publication of *Trellis Coding*, it has become evident that turbo coding had to be given central treatment, and Lance Pérez and I have embarked on writing up the material on turbo coding to fit together with the already existing material on trellis coding. Some of the less central parts of the old book have been discarded, the remainder has been updated and prepared to match the new chapters on turbo coding. *Trellis and Turbo Coding* is the result and reflects the fact that these two methods closely operate together. Most turbo codes use small trellis codes as their constituent building blocks. What we have learned from the last decade of intensive research is that we only want to use trellis codes with small state-space complexities, and if stronger codes are needed, to wire up trellis codes into concatenated turbo systems.

The first part of the book follows largely the original material in *Trellis Coding*, enhanced with many updates. It not only covers independent trellis coding methods, such as trellis coded modulation and its use in voiceband modems, but also lays the foundation for the component decoding methodology for turbo coded systems. The second part is completely new and deals with all major forms of turbo coding, viz. parallel and serial concatenated systems, low-density parity check coding, and iterative decoding of product codes. These methods and the material on trellis coding comprise the majority of channel coding strategies in use today, and likely for a long time to come.

Since turbo coding has achieved theoretical limits with manageable effort, any possible future advances are necessarily very limited, and we will undoubtedly see a migration of research and development from designing codes to building codecs. In that sense we believe that the topic of error control coding has reached a level of maturity and completion which will elevate it from a research domain to a classic theory. We hope that this book covers this theory in a comprehensive and accessible manner, and that the reader will experience the same awe and fascination that we felt researching the material for it.

ACKNOWLEDGMENTS

I am grateful to my co-author Lance Pérez for helping me conquer the vast material on turbo codes which has emerged during the last decade, to Sheryl Howard for co-authoring the chapter on factor graphs, and to Christopher Winstead for co-authoring the chapter on low-density parity check codes. I also wish to acknowledge my colleagues Robert Hang, Paul Goud, Sheryl Howard, and Christopher Winstead at the High-Capacity Digital Communications Laboratory at the University of Alberta for their careful reviewing of the chapters of this book, as well as Barbara Krzymien for her cover design drafts.

CHRISTIAN B. SCHLEGEL

Edmonton, Alberta, Canada
September, 2003

CONTENTS

CHAPTER 1

Introduction

1.1 MODERN DIGITAL COMMUNICATIONS

With the advent of high-speed logic circuits and very large-scale integration (VLSI), data processing and storage equipment has inexorably moved towards employing digital techniques. In digital systems, data is encoded into strings of zeros and ones, corresponding to the on and off states of semiconductor switches. This has brought about fundamental changes in how information is processed. Real-world data is typically in analog form; this is the only way we can perceive it with our senses. This analog information needs to be encoded into a digital representation—for example, into a string of ones and zeros. The conversion from analog to digital and back are processes that have become ubiquitous, as, for example, in the digital encoding of speech.

Digital information is treated differently in communications than analog information. Signal estimation becomes signal detection; that is, a communications receiver need not look for an analog signal and make a "best" estimate, it only needs to make a decision between a finite number of discrete signals, say a one or a zero in the most basic case. Digital signals are more reliable in a noisy communications environment. They can usually be detected perfectly, as long as the noise levels are below a certain threshold. This allows us to restore digital data, and, through error-correcting techniques, even correct errors made during transmission. Digital data can easily be encoded in such a way as to introduce dependency among a large number of symbols, thus enabling a receiver to make a more accurate detection of the symbols. This is called *error control coding*.

The digitization of data is convenient for a number of other reasons too. The design of signal processing algorithms for digital data seems much easier than designing analog signal processing algorithms. The abundance of such digital algorithms, including error control and correction techniques, combined with their ease of implementation in *very large-scale integrated* (VLSI) circuits, has led to many successful applications of error control coding in practice.

Error control coding was first applied in deep-space communications where we are confronted with low-power communications channels with virtually unlimited

Trellis and Turbo Coding, by Christian B. Schlegel and Lance C. Pérez
ISBN 0-471-22755-2

bandwidth. On these data links, convolutional codes (Chapter 4) are used with sequential and Viterbi decoding (Chapter 7), and the future will see the application of turbo coding (Chapter 10). The next successful application of error control coding was to storage devices, most notably the compact disk player, which employs powerful Reed–Solomon codes [19] to handle the raw error probability from the optical readout device which is too large for high-fidelity sound reproduction without error correction. Another hurdle taken was the successful application of error control to bandwidth-limited telephone channels, where trellis-coded modulation (Chapter 3) was used to produce impressive improvements and push transmission rates right up toward the theoretical limit of the channel. Nowadays coding is routinely applied to satellite communications [39, 46], teletext broadcasting, computer storage devices, logic circuits, semiconductor memory systems, magnetic recording systems, audio–video systems, and modern mobile communications systems like the pan-European TDMA digital telephony standard GSM [33], IS 95 [44], CDMA2000, and IMT2000, all new digital cellular standards using spread-spectrum techniques.

1.2 THE RISE OF DIGITAL COMMUNICATIONS

Aside from the technological advantages the digital processing has over analog, there are also very good theoretical reasons to limit attention to the processing of digital signals. Modern digital communication theory started in 1928 with Nyquist's seminal work on telegraph transmission theory [34]. The message from Nyquist's theory is that finite bandwidth implies discrete time. That is, a signal whose bandwidth is limited can always be represented by sample values taken at discrete time intervals. The sampling theorem of this theory then asserts that the band-limited signal can always be reconstructed *exactly* from these discrete-time samples.[1] These discrete samples need to be processed by a receiver since they contain all the necessary information of the entire waveform.

The second pillar to establish the supremacy of digital information processing came precisely from Shannon's 1948 theory. Shannon's theory essentially establishes that the discrete-time samples that are used to represent a band-limited signal could be adequately described by a finite number of amplitude samples, the number of which depended on the level of the channel noise. These two theories combined state that a finite number of levels taken at discrete-time intervals were completely sufficient to characterize any band-limited signal in the presence of noise—that is, on any communications system.

[1] Since it is not shown elsewhere in this book, we present Nyquist's sampling theorem here. It is given by the following exact series expansion of the function $s(t)$, which is band-limited to $[-1/2T, 1/2T]$:

$$s(t) = \sum_{i=-\infty}^{\infty} s(iT)\mathrm{sinc}\left(\frac{\pi}{T}(t - iT)\right); \qquad \mathrm{sinc}(x) = \frac{\sin(x)}{x}.$$

With these results, technology has moved toward completely digital communications systems, with error control coding being the key to realize the sufficiency of discrete amplitude levels. We will study Shannon's theorem in more detail in Section 1.5.

1.3 COMMUNICATIONS SYSTEMS

Figure 1.1 shows the basic configuration of a point-to-point digital communications link. The data to be transmitted over this link can come from some analog source, in which case it must first be converted into digital format (digitized), or it can be a digital information source. If this data is a speech signal, for example, the digitizer is a speech codec [20]. Usually the digital data is source-encoded to remove unnecessary redundancy from it; that is, the source data is compressed [12]. This source encoding has the effect that the digital data which enter the encoder has statistics which resemble that of a random symbol source with maximum entropy; that is, all the different digital symbols occur with equal likelihood and are statistically independent. The channel encoder operates on this compressed data, and introduces controlled redundancy for transmission over the

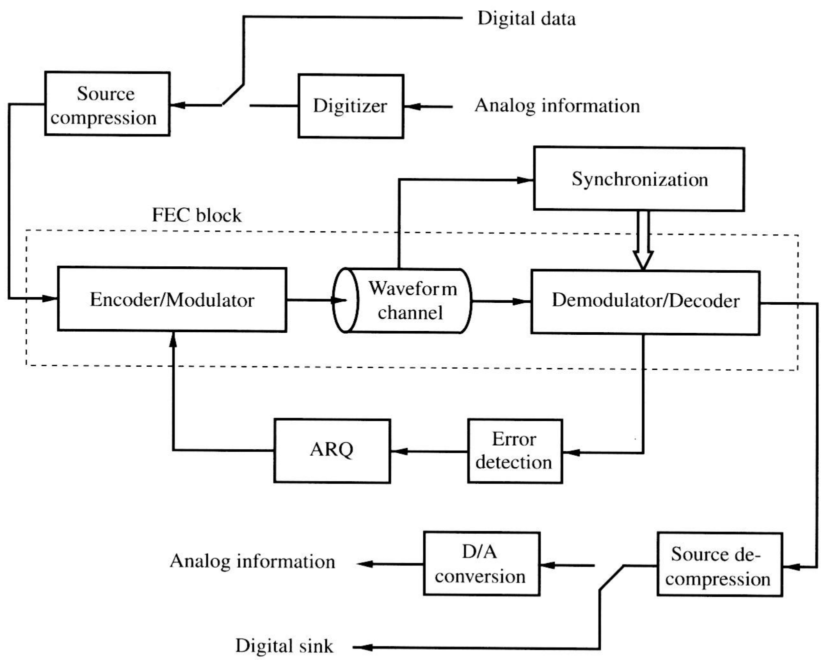


Figure 1.1 System diagram of a complete point-to-point communication system for digital data. The forward error control (FEC) block is the topic of this book.