Coding for Storage Systems:

Professor Paul H. Siegel
Department of Electrical and Computer Engineering
Center for Magnetic Recording Research
University of California, San Diego

Scope

Abstract:
This course provides a thorough introduction to the theory and application of constrained coding, widely used in digital data storage devices, such as magnetic disk drives found in desktop computers, portable computers, and consumer electronics devices; magnetic tape drives used in data backup and archiving systems; and optical disk drives that record and play CD's and DVD's. Additional topics to be covered include the information-theoretic analysis of noisy digital recording channels and the design of concatenated coding schemes that can, in principle, approach the channel capacity. The course will conclude with an overview of recent results on coding for page-oriented storage devices, such as two-dimensional optical storage and holographic recording.

Introduction:
A constrained code consists of two parts: an encoding algorithm that efficiently transforms (arbitrary) data sequences into sequences that can be more reliably recorded and retrieved from the storage medium; and a decoding algorithm that takes the retrieved sequence – which may contain some errors – and recovers the original data sequence.
There is a deep and rich mathematical theory that serves as a foundation for the analysis, design, and implementation of constrained codes. There is also a substantial engineering literature that addresses practical aspects of constrained coding, including the integration of constrained codes with algebraic and probabilistic error control techniques.

Recording systems are often modeled as noisy finite-state intersymbol interference (ISI) channels. Analysis of the achievable information rates of these channel models provides theoretical limits on recording density and data transfer rate. Techniques for numerically estimating information rates have recently been developed, and progress has been made on the design of capacity-approaching coding schemes that combine constrained codes and error-correction codes.

As traditional track-oriented recording technologies approach physics limits, new page-oriented storage technologies such as two-dimensional optical storage (TwoDOS) and holographic recording offer the possibility of achieving ever higher volumetric storage density and data transfer rates. Techniques for information-theoretic analysis of one-dimensional constrained codes and noisy ISI channels do not extend directly to higher dimensions, and the design of efficient codes for these page-oriented storage systems is an active research area with many challenging open problems.

Topics covered in this course include (time-permitting):

- Recording channel models and coding requirements
- Representation and information-theoretic properties of constrained sequences
- Design of encoders and decoders for constrained systems
- Integration of constrained codes and error control codes
- Information-theoretic analysis of noisy intersymbol-interference channels
- Capacity-approaching concatenated coding schemes
- Coding for page-oriented (two-dimensional) recording channels

Prerequisites

The course requires familiarity with basic concepts in linear algebra, probability theory, and linear systems analysis. Knowledge of elementary information theory, communication theory, and error-correction coding will be helpful, but is not required.

Syllabus

1. **Introduction to constrained systems**
   - Runlength-limited systems and charge-constrained systems
   - Graph presentations
   - Irreducible systems
   - Finite-type and almost-finite-type systems
   - Capacity of constrained systems
2. **Theory of non-negative matrices**
   - Irreducible and primitive matrices
   - Perron-Frobenius Theorem
   - Shannon capacity formula
   - Connections to Markov chain theory
3. **Finite-state encoders for constrained systems**
   - Properties of encoder graphs
   - Approximate eigenvectors and Franaszek algorithm
   - The state-splitting algorithm

- Sliding-block decoders

4. **Other constructions and complexity bounds**
   - Constructions for runlength-limited codes
   - Codes with unconstrained positions
   - Encoder / decoder implementation bounds
5. **Spectral-null constraints**
   - Spectral analysis of constrained systems and codes
   - Codes with simple spectral nulls
   - Codes with multiple spectral nulls
6. **Combined constrained and error control coding**
   - Distance properties of constrained systems
   - Partial response channels
   - Error-event characterization
   - Distance-enhancing constraints (RLL, matched-spectral-null, MTR, time-varying MTR)
   - Constrained parity codes
7. **Capacity of partial response channels**
   - Information rates of noisy intersymbol-interference (ISI) channels
   - Simulation-based estimation of capacity
   - Capacity-approaching coding for partial response channels
8. **Coding for page-oriented storage devices**
   - Capacity of two-dimensional constraints
   - Bit-stuffing bounds
   - Construction of block and finite-state codes
   - Information rates and coding for noisy two-dimensional ISI channels

Reference Books

Primary text:

An Introduction to Coding for Constrained Systems, by B.H. Marcus, R.M. Roth, and P.H. Siegel (draft version downloadable). This is a revised and expanded version of "Constrained systems and coding for recording channels," B.H. Marcus, R. M. Roth, and P.H. Siegel, Chapter 20 in *Handbook of Coding Theory*, V. Pless and W.C. Huffman (editors), Elsevier Press, Amsterdam, 1998, pp. 1635-1764.

Recommended supplemental texts:

An Introduction to Symbolic Dynamics and Coding, by D. Lind and B.H. Marcus, Cambridge University Press, Cambridge, 1995 (second printing, 1999), ISBN 0-521-55900-6 (paperback).

Codes for Mass Data Storage Systems, by K. A. Schouhamer Immink, Shannon Foundation Publishers, The Netherlands, 1999. (See also the earlier edition, *Coding Techniques for Digital Recorders*, Prentice Hall, New York, 1991.)

Other references:

There is a substantial and continually growing literature on the subject of coding for storage devices, and reference will often be made to relevant journal papers, conference papers, technical handbooks, and other textbooks.

Schedule of lectures:

All lectures will be given in Room Buzano of the Department of Mathematics of Politecnico di Torino from 10.30 to 12.30 of the following days: 2,3,4,9,10,11,16,17,18,23,24,25 of May.